

Translation-based corpus studies

Translation-based corpus studies
Contrasting English and Portuguese tense
and aspect systems

Diana Santos

To my Mother

Contents

Preface	xi
1 Introduction	1
1.1 Motivation	1
1.2 Language differences are deeper than is generally acknowledged	2
1.3 Bootstrapping of language engineering	3
1.4 The ubiquity of vagueness	4
1.5 Grammatical studies allow for smaller corpora	5
1.6 On the use of literary texts	6
1.7 Conventions: On the “third line”	7
1.8 Aspect for beginners	7
1.9 Summing up	8
2 Parallel corpora and contrastive studies	9
2.1 Parallel corpora	9
2.1.1 Research on very large corpora	10
2.1.2 Research on smaller corpora	12
2.1.3 Studies on one half of a parallel corpus	13
2.2 Contrastive studies	14
2.3 Translation studies and translation theory	16
2.3.1 Machine translation practice – and theory	18
2.3.2 Translationese	20
2.3.3 Baker's definition of “corpus-based translation studies”	22
2.4 Translation data as linguistic data	22
2.4.1 Jakobson and competence vs. performance	23
2.4.2 Dyvik's translational-based semantics	23
2.4.3 Schubert and implicitness in translation	24
2.4.4 Tsujii and the translation of discourse	25
3 Tense, aspect, and semantics	27
3.1 The crosslinguistic analysis of tense and aspect	27
3.1.1 Smith and the perfection of English	28
3.1.2 Dahl (1985): a major work, but one important flaw	29
3.1.3 Slobin and thinking for drawing	31
3.1.4 Gawrońska and article and aspect choice	32
3.2 Vagueness	32
3.2.1 Types of vagueness	34
3.2.2 Vagueness vs. unmarkedness or underspecification	37
3.2.3 Vagueness vs. ambiguity	38
3.2.4 Vagueness and compactness	39
3.3 A working definition of tense and aspect	39
3.3.1 A new definition of aspectual class	41

3.3.2	Vendler's methodology, not Vendler's categories!	41
3.3.3	The Portuguese aspectual classes in a nutshell	43
3.3.4	Comparing the two sets of classes from a Vendlerian angle	50
3.4	The descriptive model	50
3.4.1	The aspectual network of Moens	52
3.4.2	My appropriation of Moens's model	56
3.5	The aspectual network model applied to English and Portuguese	59
3.5.1	An aspectual network for Portuguese	59
3.5.2	An aspectual network for English	66
4	The translation network	69
4.1	Basic issues	69
4.2	Abstract properties of translation	70
4.3	Concrete translation networks	73
4.3.1	Coercion by translation	73
4.3.2	The translation network does not force one analysis	75
4.3.3	Introduction of vagueness or ambiguity by translation	80
4.3.4	Compact categories	83
4.3.5	Lexical issues	88
4.4	Problems and limitations	91
4.4.1	Putting things together	91
4.4.2	"Too free" translations	94
4.4.3	Can one get an overall picture, anyway?	95
4.4.4	Temporal reference and quantification	97
4.4.5	How to evaluate the translation network model	98
4.5	Summing up	98
5	Corpus studies	101
5.1	The corpus	102
5.2	Initial corpus processing	103
5.3	Imperfeito	107
5.4	Perception verbs	113
5.5	The English perfect	119
5.6	Pluperfect and Mais que Perfeito	123
5.7	Studying the I/P distinction and Perfeito	129
5.8	Acquisitions as a relevant class	137
5.9	Obras and Mudanças: a different way of classifying events	142
5.10	Temporal reference	146
5.11	Corpus-based contrastive generalizations	148
6	Language engineering, evaluation and applications	151
6.1	Evaluation of the translation network model	151

Contents	ix
6.2 Direct applications	152
6.3 Advantages of understanding time and translation	155
6.4 In the realm of parallel corpora	156
6.5 A plea for validation	156
6.5.1 Santos & Oksefjell revisited	158
6.5.2 Validation machinery wanted	160
6.6 Concluding remarks	160
References	163

Preface

This book describes research on the tense and aspect systems of English and Portuguese. It is not primarily concerned with tense and aspect, however, but with methodological, theoretical and practical issues that arose and which I tried to answer in the course of the work on my doctoral dissertation.

The present text is in fact a rewriting of substantial parts of my dissertation “Tense and Aspect in English and Portuguese: A contrastive semantical study” (Santos, 1996a), in an attempt to make the general points clear, while leaving out the bulk of the details of the study.

Since the actual writing of the dissertation, in 1995 and the beginning of 1996, I have kept reading and, to some extent, writing on these matters until February 1999. I have updated the material (and, in some cases, the ideas) accordingly. In February 1999 I finished the main writing of this book, however leaving the empirical material for later. That happened only in the last months of 2003, when I also wrote the last chapter, on leave from SINTEF, to which I am grateful for providing the necessary infrastructure.

I hope that this book can be of interest to several kinds of readers, namely corpus linguists, tense and aspect researchers, semanticists, and also translation scholars, given that I used translation corpora (aligned parallel corpora) as the empirical semantic data for my research on tense and aspect. Consequently, much of the present text is devoted to discussing the issues involved in such an approach, and to some extent providing answers to them, in the form of models for handling and organizing contrastive tense and aspect data.

The fact that I have such a wide audience in mind may be taken as a serious drawback of the book, given the current focus on specialization. Nevertheless, I am convinced that it is always fruitful to view things from different perspectives and to use knowledge from different areas. Furthermore, we are dealing with several disciplines applied to the same subject matter – natural language – in the present case. A study is not necessarily deeper if it concentrates exclusively on one aspect of a phenomenon; it should gain in insight what it loses in exhaustiveness.

The reader should also be warned that the aim of this book is not to provide an annotated bibliography. I do not feel the need to mention all and every one who have expressed the same (or the opposite) opinion. I will only cite other people if I do not manage to make a point as well as they do; and will point the readers to other sources mainly to avoid discussing the content of their work in detail.

It should also be clear that I am aware that “my” ideas are not only mine. In some way or other all ideas (with the exception, perhaps, of the specific details of the particular translation network model) have been thought and expressed in a

different form before. It is, however, impossible to trace back every source influencing the following pages.

I will thus stress here the work which (I believe) has influenced me the most, as well as those individuals who have mostly contributed to my scientific development. It is standard practice to say that they cannot be held responsible for their influence on me. This is a moral question I do not want to raise; my hope is, nevertheless, that they will not regret it.

Lauri Carlson is the one I owe most credit to, not only because he accepted to supervise me, but also because he indulged in discussing unclear intuitions concerning Portuguese for years, while proposing several formal models for me to play with in order to say things more precisely.

Since I moved to Oslo, Stig Johansson has been a kind of patron saint for me. He gave me guidance and introduced me to the parallel corpora community and goals. He also helped me with all sorts of practicalities, and I can never thank him enough for allowing me time (in 1997) to prepare the present book.

It is not easy to thank one's family when there is role overlap. Jan Engh – playing the role of traditional linguist and loving husband – has been a significant influence, reviewing critically hundreds of drafts for me. He has pointed out dozens of important references I was unaware of, and has developed my sense for primary sources. He has also tried to make my language more understandable. In a way, this book is an attempt to re-write my dissertation in a form that may please Jan.

Chapter 1

Introduction

Why should one care to do translation-based corpus studies? Is this a new specialized branch of linguistics, or just yet a new name for parallel corpora processing?

I start with a short explanation of what this book offers in Section 1.1, and proceed to discuss in Sections 1.2 to 1.6 some basic assumptions, which were confirmed in the course of the study. They are of quite different sorts. Some will be implicit in the rest of the book, while others will be repeatedly emphasized in different places.

Sections 1.7 and 1.8 are concerned with improving the readability of the present text, respectively explaining the translation conventions employed in the presentation of the translation examples and providing a simple introduction to the notion of aspect. Section 1.9 rounds off the chapter by presenting the overall structure of the book.

1.1 Motivation

The overall goal of this book is to emphasize the importance of linguistics (more specifically, semantics) for translation, and the importance of translation to study natural language. Theoretical as the previous sentence may seem, it actually reflects practical purposes and concerns. And it should be at once stressed that we are concerned here with translation **performance** and linguistic **performance**.

As it will be seen in the remainder of this book, there is nowadays a significant interest in dealing with parallel corpora to deploy translation-based applications. Analogously, the translation studies discipline has also ventured into corpus-based practice. However, it think that neither development is doing justice to language, in the sense that both communities are making use of too shallow analysis tools.

This book is an attempt to go deeper because it deals with semantics, and because it expects that language and translation are complex. Oversimplification can throw the baby with the water, and I believe that a clear understanding of the problems is a necessary prerequisite to be able to solve them. So, this book presents an avowedly complex model to deal with undeniably complex data.

The specific object of study is the meanings associated with the linguistic systems of tense and of lexical aspect, in two languages of the same language family, English and Portuguese. Nevertheless, it is shown that there is hardly a case where total meaning preservation occurs in actual translations.

These linguistic subsystems are related to temporal specification, but they are not only concerned with temporal aspects. Rather, a number of other notions –

related to the general world model informing one language – such as obstacles, attention, habituality, volition, manner and perspective, will be discussed in connection with the analysis of translation performance.

1.2 Language differences are deeper than is generally acknowledged

... more linguistic research should focus on the implications of these language-specific differences before any responsible universal linguistic claims should be made about human or "natural" language
(Tobin, 1994: xviii)

Maybe the most abstract idea contained in my work (cf. Santos, 1996a) is that two languages do not say the same thing (add quantification according to your liking: never say the same thing, sometimes do not say the same thing, may not say the same thing...). This belief arose as a reaction to the overwhelmingly dominant idea, in machine translation and natural language processing alike, that languages are only **formally** different: in other words, the belief that languages simply have different forms to convey the **same** meaning.

In practical terms, this also results in natural language processing being identified by most people as English language processing, together with localization or translation procedures, something I tried to argue against in Santos (1999b).

Brought up with Fernando Pessoa's quotation "A minha pátria é a língua portuguesa" ('My homeland is the Portuguese language'¹), I never doubted that the best contribution I could make to natural language processing would be to the processing of Portuguese, my native language. Culture and behavioural conventions are inextricably mingled with language, in such a way that to master one tongue is much more than being able to produce grammatical and meaningful sentences in it.

Given that tense is one of the most developed grammatical systems in Portuguese (and known to be one of the most difficult to master by foreign speakers as well), it is hardly surprising that it strongly appealed to me. In addition, its (apparent) reference, to time, seemed to be a domain where English and Portuguese could hardly diverge – and, in fact, several transfer-based MT systems suggested an interlingual approach for precisely that domain; cf. Van Eynde (1988) and Hauenschild (1988).²

¹ Bernardo Soares, *O livro do desassossego*.

² For those not familiar with MT jargon, *transfer-based* means that the underlying model of translation is that of changing from one internal structure representing the source language text into an internal structure representing the target language text, while *interlingua* aims at obtaining a common internal representation for two (or more) languages.

If I were able to find significant differences in the organization of the two languages, this would be a stronger argument than if I studied parts of language directly connected with social behaviour, for example. On the other hand, if I could use the same underlying analysis for both languages, I might be able to write a dissertation for a wider audience than Portuguese linguists.

In fact, I started with the second goal in mind. I strove for some time to achieve a unified model that could handle English and Portuguese. It was only when I started to look carefully at the translations that I realized that my old intuition was right and that a unified approach was the wrong way to go. My dissertation can thus be seen as an investigation into how much (or how little) the languages English and Portuguese differ in the restricted domain of reference to time (in linguistic terms, in the fields of tense and aspect).

One of the aims of the present book is to make it clear that, at least methodologically, one cannot perform a contrastive study assuming the same underlying mechanism in two or more languages. Rather, one has to assume that they are different so that it makes sense to compare them.

1.3 Bootstrapping of language engineering

To many laymen, it is incomprehensible that we can build machinery that can convey a man to the moon, but none that can translate even very simple texts into French

(Kay, Gawron & Norvig, 1994: 3)

It is my belief that language engineering³ should proceed by developing tools to help human beings perform difficult tasks, instead of trying to accomplish fully automatically those difficult tasks.

This cannot be said to be a new idea, but my impression is that one constantly sees tools appearing which are meant to be of help in some task, but whose developers had not the least idea of what people actually do in practice. It is very easy to get carried away and provide a lot of ... wrong functionalities!

The present work was inspired by the belief that an important part of research and development in machine translation should be devoted to the study of already existing translations. One of the purposes is to allow people to look at real examples from a more informed perspective than purely quantitative or statistical processing.

As a general trend, NLP goals have moved from full automation to the development of tools to help in a human task. In translation, fully automated machine translation has given way to computer-aided translation as well as tools for the translator. I will claim in this book that, in order to have a useful tool for handling large bodies of parallel texts, one must use semantic and contrastive

³I use interchangeably the terms language engineering and natural language processing (NLP).

knowledge. But the tool in turn becomes a very good testbed for semantic and contrastive analysis: the creation of a tool to perform a human task is one of the best ways to learn how to perform that particular task. This kind of “research bootstrapping” represents, I believe, one of the methodologically sounder conclusions of the research community in natural language processing.

Combining theoretically informed models with useful, but not too ambitious, applications appears to be a good middle ground between the sceptical attitude of those who hold that natural language is too complicated to handle by computer, and that the task is therefore not worth trying (“first the theoretical issues must be settled”), and the naïve attitude of those who attempt to do everything at once, and whose systems, displaying obvious inadequacies, feed the attitude of the sceptic.

1.4 The ubiquity of vagueness

A human language must permit the communication of thoughts in a way that is reasonably efficient relative to the lifespan and cognitive capabilities of human beings. [...] This requirement justifies the claim that human languages are, by nature, imprecise.

Keenan (1978: 162)

Vagueness is a fundamental property of natural language. Even though its existence – and pervasiveness – is generally acknowledged, it is astonishing how little it has actually been taken into account in language description.

I use the term “vagueness” to refer to the situation where a linguistic sign (a word, a grammatical device, a sentence) is not required to have a single classification or, conversely, where the same situation can be described by two or more linguistic signs (see Section 3.2).

This is an extremely common situation if one looks at the competence side of language (there are hardly any lexical items with only one particular meaning); contrary to the view of most researchers, my claim is that it is equally so in performance. In other words, I claim that, in context, many – if not most – utterances do not require the speaker to decide between a lot of possible distinctions that language as a whole would allow them to make.

Given that in translation one tries to match two systems which encompass a myriad of small distinctions which are not shared – and, even when they are, evidence for those distinctions appears in different places of the system⁴ –, vagueness can be seen as one of the most important issues in translation.

⁴The reader is not required to agree with this statement at the outset. I will try to provide enough evidence in the remainder of this book.

Actually, I view translation as a classification of the source language in terms of the target language, therefore (roughly) identifying multiplicity of translation with contrastive vagueness, and (one particular sort of) untranslatability as corresponding to cases where the target language is vague regarding a distinction relevant to the source.

Even though a substantial part of the present book will be devoted to discussing and presenting evidence on the difference between languages focussing especially on vagueness, I consider Keenan (1978) to be unrivalled on the subject, both in the argument and in the wealth of empirical data provided from several distinct languages. Here is what he has to say about the impossibility of meaning equivalence in general:

[I]n considering the set of 'meanings' expressible in two languages we are faced with the following fact, in the general case: first, the basic meanings (i.e., those expressed by the basic sentences of the language) are not exactly the same in the two cases, and second, the ways of forming (expressions of) complex meanings from (those of) simpler ones are not exactly the same. It is surely plausible then that one of the languages would allow us to express some meanings that were not exactly expressible in the other. And it would surely be surprising, and a very strong empirical claim, that different languages using different means to express 'meanings' always arrived at exactly the same end. (Keenan, 1978: 166)

1.5 Grammatical studies allow for smaller corpora

Demographic samples are representative to the extent that they reflect the relative proportions of strata in a population. [...] Language corpora require a *different* notion of representativeness, making proportional sampling inappropriate in this case. [...] Such a corpus would permit summary descriptive statistics for the entire language represented by the corpus. These kinds of generalizations, however, are typically *not of interest for linguistic research*. Rather, researchers require language samples that are representative in the sense that they include the full range of linguistic variation existing in language.

Biber (1993: 247, emphasis mine)

It seems that grammatical studies do not demand corpora of the size required for lexical studies. Or, to put it differently, the notion of corpus can be adapted and still be useful when one does not have access to billions of words.

My main point is that one should not be obsessed with size, or by “representativeness”, in defining a corpus. In addition, a corpus should be defined relative to a goal (and to content). In short, a corpus should be expressed as a **corpus of X to do Y**. For example, gathering a corpus of words to study lexical

distribution according to genre must be quite a different task (and produce quite a different result) than gathering a corpus of sentences to study syntactic patterns. And, likewise, compiling a corpus of translation pairs is different from gathering a corpus of spoken utterances to study dialect, from creating a corpus of handwritten letters to study address modes, and from collecting electronic mail to study the distribution of accented versus non-accented characters in daily communication.

Also, there is a limit to the amount of information explicitly encoded in a corpus (if that information is manually generated). The deeper the annotation the smaller the size. Even though one may always wish to have more data, I should stress that I do not consider an unqualified “it was based on too small a corpus” a valid criticism of the work described in the present book.

In fact, and given that the linguistic phenomena analysed here are required to be present in every sentence, and additionally are not among the conscious choices translators (or authors) make, a corpus considered to be small for other purposes was shown to be rich enough to raise problems and allow hypothesis verification. On the other hand, any study should be validated in larger corpora, and there is always the hope (or the fear) that larger amounts of data will bring, according to Zipf’s law, new cases as yet unaccounted for.

1.6 On the use of literary texts

This work is based on the analysis of literary texts, and therefore on literary translation. This may sound strange to an unprepared reader, who might equate computational concern with technical text.

But there are a bunch of different reasons why the use of published translations makes sense in the context of studying tense and aspect, in addition to practical factors such as their availability, and the larger variety of tenses in a narrative context.

First, as suggested by Doherty (1995), literary translation is bound to be more literal than, say, scientific translation, given the respect due to the author and his language. So, one should expect texts more parallel in this kind of translation.

In the particular works employed in the present study, the style of the authors is not, as far as I can see, marked by any special devices related to tense and aspect. Thus there is no ‘literary bias’ in this respect.

There are several schools of thought that categorize as two extremes literary and other translation. Not denying different concerns and priorities, it should nevertheless be emphasized that, linguistically speaking, many of the same problems appear, although to a different degree. Even the most prosaic and matter-of-fact technical sentence may give rise to interpretation differences and pose complex choices for translation. So, I prefer to look at translation in general.

1.7 Conventions: On the “third line”

The authentic translation examples presented in Chapters 4 and 5 are always accompanied by an English rendering of the Portuguese part, which I refer here as the “third line”. Its primary purpose is to replace the need to understand Portuguese at all, but it should be acknowledged that the rendering I choose serves vastly different secondary goals as well.

Sometimes, it expresses an equally possible translation, sometimes it expresses alternative translations, in other cases it simply yields more literal renderings of the Portuguese text. Whenever different shades of meaning can be conveyed by the Portuguese line, I try to express them in English. In very few cases (always pointed out as such in the text), the third line is offered as a better translation. In most cases, it strives to be rather literal, and may not even be acceptable English.

A reader fully knowledgeable of Portuguese should not need to read that line, except to wonder at alternative choices that might have been done.

As a methodological point, however, it strikes me as desirable to back translate, or doubly translate, the Portuguese text, both to convey the “spirit of the language” and to show, in practice, that a total meaning-preserving rendering is the exception rather than the rule. From my own experience as a reader, I enjoy reading “translation” pairs this way, even in cases where I speak the other language as well.

1.8 Aspect for beginners

Tense, no matter its philosophical puzzles, is a well-know notion to any Romance or English speaker, due to one’s familiarity with it from primary school. However, the term “aspect” is new, and although it does not refer to anything more complex than “tense”, it may sound intimidating to a reader not acquainted with it. My goal with this preliminary section is simply to try to make aspect as familiar and trivial as tense.

Aspect is a verbal category in (among others) Slavic languages, in which a verb can have imperfective or perfective aspect. The notion of aspect is related to matters like completeness, iterativity, internal (temporal) perspective, etc. In other words, these are words and concepts usually employed to explain its meaning.

Since aspect deals with notions related to time, it has become standard practice to talk about a semantic category of “tense and aspect” (or “tense, modality and aspect”).

Later on, the classification of events in a language according to their typical temporal profile became also part and parcel of the field of “tense and aspect”, and it is in this tradition that this book uses the word “aspect”, really most of the time in the expression “tense and aspect”.

1.9 Summing up

In this chapter, I have introduced some fundamental assumptions and standpoints. Let me now give the reader a fair idea of what to expect in the remainder of the book.

Chapters 2 and 3 feature an introduction to matters of parallel corpora and of tense and aspect, respectively. They must consequently also deal with translation and semantics. Chapter 2 first surveys the work on parallel corpora, providing a state of the art description of this emerging field. Then, after brief reviews of the disciplines of contrastive analysis and translation studies, it argues for basing contrastive studies on authentic translations.

Chapter 3, in turn, presents a small subset of the complex problem area of tense and aspect, focussing mainly on semantics, before introducing the model chosen and its application to the two languages under discussion, viz. English and Portuguese.

Chapter 4 presents a representational tool, the translation network, devised to describe translational data with regard to tense and aspect semantics. The translation network is shown to capture some aspects of translation description, and to be sufficiently precise to discuss actual translation pairs. Some generalizations will be illustrated and taken up in detail.

Chapter 5 is a critical summary of the empirical studies that have been conducted. The corpus, its preparation, and preliminary annotation are described, and the quantitative data is displayed. Although a careful and well-designed test methodology was missing, the fact that a number of different — *a priori* unrelated — studies were conducted, and then compared, allowed for several conclusions on tense and aspect contrast. In addition, consideration of many individual examples was essential to any conclusion whatsoever. So, I will not be reporting on general discovery methods to elicit results without explicit analysis of a number of translation pairs. On the contrary, I will do my best to describe my at times painful examination of the data. It is a lengthy chapter, where I attempt both to convey substantive differences between the two languages and discuss questions of corpus-based methodology, trying to give enough examples to illustrate my points without requiring knowledge of Portuguese.

Finally, Chapter 6 discusses the practical implications of the ideas presented for the design of translation tools and natural language processing in general, rounding off the book with a tentative program for corpus-based contrastive studies.

Chapter 2

Parallel corpora and contrastive studies

In this chapter I give a short overview of the field of parallel corpora in Section 2.1, as well as introduce the related disciplines of contrastive studies and translation studies, respectively in Sections 2.2 and 2.3.

It will be shown that (roughly) neither contrastive studies pays attention to translation, nor translation studies focus on (strictly) linguistic aspects of translation.

My main interest in the present chapter, and in the book as a whole, however, is precisely to show how the study of translations can help in the discovery both of the linguistic systems of the languages involved and of their differences. To put it differently, I want to reveal what knowledge actual translation performance can offer. This question is discussed in Section 2.4, paving the way for a more thorough discussion of translation pairs as semantic data in the next chapter, in the more specific domain of tense and aspect.

2.1 Parallel corpora

To compile and study electronic parallel corpora is an activity of the 90's and consequently much of the theoretical underpinning of such activity remains to be provided. There are, nevertheless, a number of examples and projects that helped shape the field, which I will review briefly here.

I use the term *parallel corpora* to describe both the parallel composition of corpora in each language (which are sometimes called *comparable corpora*) and the parallel expression of content in two or more languages (sometimes called *translation corpora*,⁵ or *aligned corpora*), and leave further precision, if needed, to the context. People who insist on the dichotomy between translation corpora and comparable corpora fail to see that a translation corpus (in their sense) may also be seen as a comparable corpus, and that a comparable corpus may also include translated text in both languages, so that – as the English-Norwegian Parallel Corpus (ENPC; see Johansson & Hofland (1994) and also 2.1.2 below) was the first to instantiate – one can have a parallel corpus which is of both sorts.

⁵ This name is inappropriate when there is no true original, as is the case with some EU documents (see Danielsson & Ridings, 1996a), and it may be misleading since it is equally applicable when the parallel corpus only contains several translations of the same text, and not the original, as in the case of the Bible (see Resnik *et al.*, 1997), or when one has texts which are both translations and originals in a translation chain. Moreover, a collection of translated texts without the original, as in Baker (1996), is a more appropriate candidate for being named a translation corpus (even though referred to as a *corpus of translated text*).

This is not only a terminological point, though: On the contrary, the kinds of phenomena studied and the assumptions underlying these studies are more similar than they are different, and this is why I prefer to use the term *parallel corpora* without further discrimination.

There is an upsurge of interest in parallel (and multilingual) corpora these days, even though so far most of the literature concentrates on technical and conceptual aspects of their compilation as opposed to their use. (Incidentally, in some cases illustrating what I mentioned in Section 1.2 as a conceptually problematic overemphasis on tools.)

Here, I am only interested in studies which have actually **used** parallel corpora to make linguistic claims (or to develop NLP systems with goals different than simply processing those corpora).

In this overview, I distinguish between

- a) work using large enough corpora so that statistical techniques can be applied,
- b) work on much smaller corpora, including (detailed) analyses of individual translation pairs.

My belief is that both kinds of studies are important, and have much to offer each other. To study and contrast them should therefore prove relevant for the development of a unified “corpus-based contrastive and translation studies” discipline.

2.1.1 Research on very large corpora

Without trying to survey the field, a fairly representative sample of parallel corpus studies employing very large corpora, usually statistically-oriented, is as follows:

- the IBM group at Yorktown, which uses a statistical model for machine translation fully derived from parameters estimated from an English-French parallel corpus, the Hansards
- the work of Ken Church and his associates at Bell Labs and of Pierre Isabelle’s group at CITI, also mainly on the Hansards, aimed at contrastive lexical research in order to produce or improve NLP applications and/or Hdeveloping translator’s aids
- the work of Ido Dagan and his colleagues using comparable corpora as aids for monolingual analysis, such as lexical disambiguation
- the work of Makoto Nagao and other Japanese researchers on example-based machine translation

Through the use of ingenuous statistical measures, Brown *et al.* (1990) try to model translation from English to French by automatically estimating translation probabilities in the Hansards, a parallel corpus of the Canadian parliamentary debates. Even though the method is computationally very expensive, they want to show that observation of performance alone is enough to translate, or, at least, go a long way in the translation process.

As far as the work on parallel corpora is concerned, Church and his associates have mainly studied two kinds of problems: sentence alignment, and detecting word correspondences automatically (see Gale & Church, 1991; Church & Gale, 1991) as an aid for bilingual lexicographers, and – they claim – as a way of getting sense disambiguated corpora, something which helps in NLP and lexicography in general.⁶ In the Hansards, they were able to obtain a corresponding word for 60% of the source words with a 95% correction.

In the process of developing tools for aiding translators, Isabelle *et al.* (1993) studied the possibility of automatically detecting (a very limited case of) wrong translation in parallel corpora, namely complete deceptive cognates. They conclude positively on the feasibility of such a tool, after looking for 145 pairs extracted from a false-friends dictionary in a tagged version of the Hansards of 1996, and getting 56 real errors (corresponding to 76.7% of the matches).

Sato & Nagao (1990), who have coined the term *example-based machine translation*,⁷ proposed to use a database of previously translated sentences as input to machine translation, instead of making use of hand-coded rules. They suggested ways of partial matching and of choice among several candidates based on estimates of their translatability, though acknowledging that the creation of a thesaurus with both enough translation examples and accurate estimates is a major task. This was independently corroborated by Sadler & Vendelmans (1990),⁸ who create a bilingual knowledge base (BKB) by interactively proposing, to a human user, links between the constituents in the source and target language. Although they mention 20,000-word parallel corpora in the field of computer manuals (*ibid.*: 450), and they mention that the system is self-improving, the only example discussed (an English sentence of eight words, translated by a French sentence of 11) produces 13 translation units. Creation of BKB's along these lines seems thus prohibitive.

No doubt this was the reason why future research followed more modest paths: instead of fully automatic machine translation, Sato (1992) turned to an example-based translation **aid** system, while Utsuro *et al.* (1992) reported on lexical knowledge extraction from bilingual corpora, both for syntactic disambiguation (PP-attachment) and human-assisted verb case-frame compilation. Still in view of full MT, Doi & Muraki (1992) propose to replace translation corpora by the use of non-aligned parallel corpora, using a combination of co-occurrence statistics in the source and target languages to decide between the several options provided by the bilingual dictionary.

⁶ For example, the ability to sort English concordances by sense (making use of the underlying French translation) should be useful even for English lexicographers with absolutely no interest for French. (I disagree with this claim, as I will have ample occasion to show later.)

⁷ In fact, the title of their paper is “memory-based translation”, but is universally described as “example-based translation” afterwards.

⁸ Who speak of “analogy-based” machine translation instead.

Dagan *et al.* (1991) seem to have been the first to propose the use of another language (non-translated) corpus to improve several tasks in MT and in NLP in general, with the slogan “Two languages are more informative than one”. In Dagan & Itai (1994: 576-80), they report the following evaluation of their method: From ten paragraphs picked at random from foreign news sections of Israeli press, they selected 103 ambiguous (in relation to English) Hebrew words. Using a corpus of 25 million American English words (from the Hansards, the Associated Press, and the Washington Post), they were able to choose the right translation in 64 cases – their method did not provide an answer in 33 cases, and gave the wrong result in two cases only.

Although the list above is not exhaustive, it is undeniable that research on very large corpora is centred on the improvement of tools or applications applying automatic data extracted from parallel corpora. The situation is quite different for smaller corpora, to which we now turn.

2.1.2 Research on smaller corpora

Most researchers are forced to use much smaller corpora in their investigation, especially because they must create those corpora from scratch. It is perhaps paradoxical that smaller corpora are often produced with greater care and with a wealth of detail (encoding) impossible for other projects.⁹ Good examples of this are:

- the English-Norwegian Parallel Corpus (ENPC), later expanded into the Oslo Multilingual Corpus (OMC), a corpus of a number of language pairs
- PEDANT, involving Swedish into English and vice versa
- the German-Norwegian parallel corpus, containing original texts in both languages

The three corpora have in common the use of SGML, manual revision, and heavy annotation. In addition, the constitution of each corpus obeys criteria such as balance of genre and number of different authors and translators. In order to appreciate the care put into their preparation, see respectively Johansson *et al.* (1996), Danielsson & Ridings (1996a;b) and Fabricius-Hansen (1991).

Note also the extreme methodological care in the building of the following corpora, comprising parallel original texts like those of

- Sylviane Granger's International Corpus of Learner English (ICLE), discussed e.g. in Granger (1996), encompassing texts written by learners of English as a foreign language (EFL)
- Berman & Slobin's (1994) corpus of elicited oral narratives based on a picture book (which will be discussed further in Section 3.1.3)

⁹ The paradox is that, with the same resources (human and technical), those teams might create much larger corpora, if they did not spend so much effort on encoding and preparing.

Finally, another (very common) case that I will discuss under the label “smaller corpora” (although orthodox corpus scholars might deny them the use of the name *corpus*) are studies based on few (sometimes, only one) sources and their translations. Even though one might refer to them as “studies on parallel texts”, I will use the word *corpus* and leave to normative minds the uninteresting task of drawing the line in a precise way. As an example of the quality that such studies can display, I mention those collected in Maslov’s (1985) book.

The sort of things that researchers aim at with these (three kinds of) corpora, which I have labelled “smaller corpora”, are also usually different in kind from the studies discussed in the previous section. It is not uncommon to use the corpus to do detailed contrastive studies on a particular, specific detail of (one of) the language(s) – hoping, of course, that the same corpus will furnish the raw data for many such studies.

The presentation of data from the corpus is more of the kind “good examples”, or specific interesting cases, or even absolute frequency distribution of translations, but it is rarely the case that explicit hypotheses and methodological premises are stated. One can thus say that, in such corpus studies, the main use of the corpus is the taming of intuition, making more “scientific” the presentation of data, while their authors do not often make methodological points. The two most common patterns are:

- either one tries to see how a corpus performs regarding a particular issue (as is the case of Hasselgård (1995) on word order, and Wikberg (1996) on questions);
- or one uses examples from a corpus to defend/assess a particular theory which was based on previous analyses of such or similar corpora (as is the case of Slobin (1996) on movement verbs or Fabricius-Hansen (1998) on text structuring).

Often, the two ways of proceeding alternate in the researchers’ work: first they probe the corpus for some phenomenon, then they formulate some hypothesis, and consult the corpus to see whether it holds. This was the case of the empirical studies that eventually gave rise to the theories put forward in the present book, which will be presented later on.

There are other cases, though. Barlow (1997), for example, uses parallel material to evaluate a particular hypothesis, put forward independently by Lakoff & Johnson (1980), namely that “good is up”, by checking in parallel texts whether contrastive evidence could confirm this metaphor. What is interesting in this study is that the hypothesis was monolingual, and was not based on parallel (or other) corpus studies.

2.1.3 Studies on one half of a parallel corpus

Even though the title of the present section is not easy to understand, I want to discuss those studies which concentrate on translated text, and therefore only look at one half of a virtual parallel corpus.

- Gellerstam's (1986) discussion on translationese by comparing novels translated into Swedish with original Swedish novels is a hallmark of modern empirical contrastive studies, because he actually measured deviance between the two kinds of Swedish;
- on a larger scale, Baker's compilation of a corpus of translated texts as a tool for contrastive studies, which will be discussed in 2.3.3, cannot be ignored.

This kind of studies will have as their goal the characterization of translated text as somehow deviating from the norm. For the moment it is hard to say very much as far as methodology is concerned, although it seems obvious that they rank with the smaller parallel corpora in both care and aim.

This concludes this short overview, which is clearly outdated in terms of works referred to but, in my view, it still provides a good grid to classify what goes under the label parallel corpus. In section 6.4, I cite more recent books on the subject. I turn now to contrastive studies, which have a longer history, during which there has not been much emphasis on the analysis of authentic translations.

2.2 Contrastive studies

The following passage, from an influential textbook on contrastive analysis, clearly illustrates that little attention has been given to the study of actual translations in the contrastive studies discipline: "There are three branches of interlingual linguistics: *translation theory* [...], *error analysis*; and *contrastive analysis*. [...] I shall not be discussing further translation theory: the reader is referred to [...]" (James, 1980: 4f).

In view of such neglect, James's conclusion in the last chapter is noteworthy:

there are two levels of translation – semantic and pragmatic – and that for CA [Contrastive Analysis] we ought to equate L1 and L2 forms, which, no matter how far they diverge superficially, are semantically and pragmatically equivalent. We conclude that translation equivalence, of this rather rigorously defined sort, is the best available TC [tertium comparationis] for CA. (James, 1980: 178)

In fact, James is locating the foundations of his discipline outside its very scope, thus leaving to translation theorists the task of providing the means to compare or contrast two languages. (The word "available" cannot be taken seriously.)

James also mentions, in passing, one of the most common objections to basing contrastive studies on translation performance, a claim about which I will have much to say in the present book: "Translated texts are an obvious basis for textual CAs. Their main limitation is their potential for translation-distortion, that is, the target-language text can show signs of interference from the source language" (ibid.: 117).

It is interesting to note, as an aside, that while James discusses the use of translations for *textual* studies, he does not even mention their use in *grammatical* contrastive analysis, which is the main subject of my investigation.

Standard contrastive analysis proceeds from a description of types which are comparable, and presumed equivalent. Lipinska-Grzegorek's contrastive study of perception verbs and nouns in English and Polish, following the Contrastive Generative Grammar model, makes this clear: "It is assumed in Krzeszowski's model that sentences and texts of both languages have common semantic representations" (Lipinska-Grzegorek 1977: 3); "Sentences representing each type in English and Polish are supposed to be ideally equivalent (word-for-word grammatical translations)" (ibid.: 5).

The contrastivist thus postulates that two sets of sentence types are equivalent, analyses them in a parallel form in the two languages, and has as the main goal to describe differences of surface form and/or of grammaticality, or cases of similarity between the two languages. As Lipinska-Grzegorek notes,

negative similarity is more interesting than positive similarity. Positive similarity is something that could be expected to occur because the semantic representations were output monitored, i.e., it was a methodological assumption that no semantic entity which was not linguistically relevant would be present in the semantic representations and that only these types of propositional meanings [...] would be discussed which have a linguistic representation for at least one type [of sentence types, DMS] (Lipinska-Grzegorek, 1977: 65).

In other words, the main problems of standard contrastive analysis can be recast into the attempt of explaining the gaps (parallel or not) in an ideal set of sentence types.

If one does not share the view that it is possible to identify a set of sentence types sharing meaning across languages – as I do not – the whole enterprise as described above seems ill-conceived. In fact, Lipinska-Grzegorek's own observations seem to cast doubt on the existence of a *same* set of semantic sentence types from which the actual sentence types of both languages are derived: "Many similarities within this class [Polish verbs of physical perception, DMS] are analogical to the similarities within the class of English verbs of perception, although there is rarely a one-to-one correspondence between these similarities" (ibid.: 9).

In what is called theoretical contrastive analysis, namely, the attempt to produce a methodological basis for the discipline, the following argument is common: "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, 1980: 85). This is something I would like to challenge. My proposal is precisely to use the target language as frame of reference, and use actual translations (as well as intuitive judgements of closeness of meaning) as the

means to link the source language categories to the target languages', thus dispensing with "common", "universal" categories postulated *a priori*.

My point is that, while translating, translators are actually looking at the source language with target language eyes (even though they sometimes look from the opposite direction, i.e., they look at target language with source language eyes and provide us with, sometimes unavoidable, translationese). They are never situated in a universal sphere of common categories. And I believe that the best way to carry out contrastive studies is precisely to study actual translation, and not to postulate abstract categories/rules and provide analyses of data which are flawed from the start by the unreality of such "universal" constructs.

This seemingly extreme position of not granting any common representation language, which is typical of structuralism, has often been criticized on the grounds that linguists should be interested in the commonalities of languages, and that they should strive to provide a level of abstraction high enough to "recognize the picture" in every natural language. Differences and idiosyncracies should be overlooked as uninteresting. I share the opposite view with an increasing number of practitioners of computational linguistics – cf. among others the views expressed by Nagao (1988), Tobin (1993), and Engh (1998) – who see with dismay the work of (some) theoretical linguists, with tools heavily loaded with untestable hypotheses (articles of faith) and data carefully tuned to present only the cases that support their theory.

Furthermore, Van Buren's claim (above) is also in a sense empty, since a large portion of (if not most) contrastive analysis has not been made with reference to common categories, but simply guided by the notion of preferable translation, or rendering (even though, according to James, this is applied – and not theoretical – contrastive analysis).

Let me illustrate this with reference to one contrastive study in particular:

One of the most influential contrastive studies ever (whose main aim, incidentally, was translation teaching) was the *Stylistique Comparée* of Vinay & Darbelnet (1977), who held the view that translation (the passage from one language to another) was in fact a practical application of comparative stylistics. The present book owes a lot to their insights, many of which I "transferred" to the contrast between Portuguese and English. Because of the wealth of examples they present, even if one were to disagree with the way they classify them, one cannot but pause to reflect on the translation regularities that they uncover (or suggest). In fact, a great deal of the contrastive data they present has been taken up later by more specific studies.

They used a fairly complex system in their book, with seven translation processes, three planes of external stylistics (vocabulary, grammar and composition), and an opposition between the plan of reality and that of thought, and introduced a good deal of new concepts and terminology, of which the "chassé-croisé" is a good example:

Chassé-croisé: Process of translation by means of which two signifieds change places and change grammatical category. Ex: "He limped across the

street: Il a traversé la rue en boitant”. The chassé-croisé is an instance of transposition [process of translation by which a signified changes grammatical category]. (Vinay & Darbelnet, 1977: 6, my translation)

One of their main contributions (in addition to the detailed overview) was their emphasis on what they called the spirit of the languages (“le génie de la langue”), and which they maintained (quite rightly, in my opinion) was different for English and French. What they aimed at was to elicit the differences between texts produced by monolingual brains in comparable situations, as they state in the preface.

Vinay and Darbelnet compare different styles, which are actualized, or reflected, in different text organizations, different grammars and different lexicons. By having a holistic view of the language, they succeed in noticing many more differences than if one started from a lower end (e.g., comparing different morphologies would fail to show the differences across levels). Vindicating the structuralist claim that all levels in a language are interlinked, they showed that by considering those levels together they were able to produce a consistent description of an enormous amount of detailed differences.

This has also been noted by Slobin, who, taking up in more detail the description of motion and position in several languages, concluded, that “typologies of grammar have consequences for “typologies of rhetoric”” (Slobin, 1996: 218).¹⁰

I thus hope to have convinced the reader that it is not only logically possible, but practically possible as well, to do contrastive studies without postulating common categories. In the next chapter (Section 3.1), I will come back to the subject of comparing and contrasting languages and will argue that it is in fact the only methodologically sound way to proceed.

I conclude the present section by suggesting that contrastive studies should be based (at least partially) on translation material, be it competence or performance, instead of using an abstract unverifiable *tertium comparationis* in addition to the material in the two languages.

Studies of translation, however, are often far from interested in linguistic differences, as we will see in the next section.

2.3 Translation studies and translation theory

There has been a blossoming of the study of translation, focussing on its interdisciplinary status, after the works of Toury (1995) and Even-Zohar (1990). Matters like reader response, universals of translated literature, the sociological implications of translations, and even the history of translation have inspired many practitioners, as can be seen in the overview given by Øverås (1996).

¹⁰ It is probably irrelevant what comes first, i.e., whether it is the style that leads to different grammatical preferences (as seems to be Vinay and Darbelnet’s view) or whether it is different grammars that condition different styles, as Slobin put it.

But perhaps as a consequence of the need to make it clear that there is much more to translation than simply linguistics, purely linguistic conclusions have been neglected or simply not been focussed on. I consider this an unfortunate result, because, after all, translation is an activity that is essentially linguistic (try to define it without using the word “language”!), which makes it hard to believe that the study of the two individual languages involved has nothing to contribute to the study of translation. One might call their study a pre-requisite, but if it is, then it is because one needs it to formulate hypotheses about translation.

On the other hand, it is also true that translation is not often used in general linguistic theory, as I will show in the next chapter.

The area of translation is so broad that I will only select a few topics related to the purposes of the present book, and direct the reader to Snell-Hornby (1995) for an attempt at a unified description.

I content myself here with presenting some reflections on machine translation practice, discussing the concept of translationese and arguing against the recent statements by Baker (1996) about universals of translation. These three subjects, it should be said, are all somewhat in the periphery of what is fashionable in the translation studies discipline.

2.3.1 Machine translation practice – and theory

The area of machine translation is probably the one with most adherents in the natural language processing community, and the one with more visibility outside research quarters. During its history, it has given birth to many systems based on many different conceptions, trying to tackle a variety of problems.

It is therefore to be expected that the attempt to automate the process of translation performed by so many groups all over the world, its partial success, and its many failures, may bring something of value to the study of translation as a process, or even as a result.

It is in this vein that the following (evidently non-exhaustive) remarks are offered:

- It cannot be denied that there is much in the translation process that is structurally determined, i.e., that in many cases it is possible to derive translation rules that produce an acceptable output in the target language simply by taking into account structural patterns.
- It was painfully discovered that there are cases in which the (structural and lexical) information present in the source text is not enough to render the target language text completely. In some cases this can be unproblematically supplied by information pertaining to the target language only (as is the case with language specific grammatical gender), but in other cases reasoning of unlimited complexity must be performed in order to understand the meaning and therefore arrive at the best (or even, at any acceptable) translation.

- It was also found that the level of the translation unit cannot be determined a priori, and that there may be complex dependencies between different translation units.
- It was early understood that contrastive data could not be derived automatically from “parallel” descriptions of monolingual data (e.g., a dictionary of the source language plus a dictionary of the target language would both make more distinctions than those required by translation and fail to make many others which are crucial for it).

As might be expected, just as in contrastive studies and general linguistics, two “schools of thought” (or kinds of practitioners) formed: one using a universal semantic representation language (or *interlingua*) and the other advocating transfer and putting the emphasis on the particular differences instead of the general similarities. For a long time, these two schools held, in one respect, the same view of machine translation, namely that it should embody translation competence knowledge (to apply to the texts that would be put into it). Recently, however, there has been an emphasis on using translation performance (usually human translation) as part of the translation system, as was briefly described in Section 2.1 above. One could call this “radical transfer”, although strictly speaking there are no longer any transfer rules involved, or one might call it an instance of “direct translation”. In any case, it should be clear that in this last approach no *interlingua* is called into the picture.

One interesting concept that has emerged in the MT literature in the last few years is the notion of “translation mismatch” introduced by Kameyama *et al.* (1991: 194): “TRANSLATION MISMATCHES [...] are found where the grammar of one language does not make a distinction required by the grammar of the other language”.

Let me explain the vagueness of this 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.

In fact, it is interesting to note that the dichotomy competence / performance is essential for the proper definition of *translation mismatch*. A clarification of the meaning of this term will be relevant for the rest of this book. *Translation mismatch* either applies to systematic differences between two languages (a matter of competence) or to a systematization of the actual differences or problems encountered (a matter of performance). For ease of reference, I will henceforth call these two possible definitions *translation competence mismatch* and *translation performance mismatch*. An example of the first is Kameyama *et al.*’s (1991) discussion of definiteness (use of articles) and politeness in the pair Japanese-English, as well as Keenan’s (1976) overview of

semantic differences across languages. Examples of the second are those referred to by the typologies of Nida (1959: 22ff) and Bar-On (1993).

It is important to note that the two concepts are obviously related (as competence and performance must be), and thus a label like “cases where the source language makes a distinction not made by the target language” (and its converse) can be applied to both planes. Either one is talking about the source language *system*, or about a source language *particular utterance*. As Bar-On (1993) notes, most cases of the former mismatches do not really cause problems for translation, because the information is not functionally relevant. (The notion of *functionally relevant* is one of the major contributions of Catford (1967), one of the few books reflecting on translation from a linguistic standpoint.)

I believe that the performance cases are much more interesting, not only because they are much more frequent, but also because they add a dimension (a new classification) to the set of possible cases, namely: *cases where the source language makes a distinction which is not preserved in the target language although it could, strictly grammatically speaking, be preserved* (and the converse case). Furthermore, it must be noted, that this is the rule rather than the exception.

There may be readers who suspect me of inconsistency here with my previous claims; and who might ask: how can I know a distinction is not preserved (or is introduced), if I have no common frame of reference? Let me clarify, then, that I have the reference of “standard translation” of smaller units. In other words:

- I do not give away my bilingual competence (or translator’s competence) as far as the two languages are concerned;
- an empirical way of verifying loss/addition is by providing a standard translation of smaller units and thus compare inside one language.

Most of this book will be devoted to describing actual translation performance mismatches, and show how they can help us understand the differences between the two languages English and Portuguese.

Let me now turn to a specific feature of translations that has been responsible for many linguists giving up the study of language differences.

2.3.2 Translationese

The claim that the language of translated texts sometimes has a “third language” flavour, as mentioned already in the discussion of James (1980) in Section 2.2, was shown clearly in Gellerstam’s (1986) study of Swedish original versus translated text, as well as in Duff’s (1981) interesting survey of translation into

English. This phenomenon, which has been called *translationese* by some scholars,¹¹ has aroused three different kinds of reactions:

1. to consider it a necessary problem, a fact of life, and try to reduce its influence on contrastive studies;
2. to promote it to the status of a universal of translation, and therefore study the translated text in itself (see the next section for a critique of this option);
3. to dismiss the study of translated texts as unreliable, uninteresting, etc.

I want to propose here a fourth and perhaps more optimistic way of looking at this undeniable character of translations: that it also gives us a lot of clues about the two languages; and that translationese is in fact a kind (as any other) of language evolution, which is visible not only in translated text but in the texts of people who are influenced by reading another language or simply by getting most of their reading second hand (that is, translated text proper).¹²

The first view I want to argue against is that translated text should be dismissed: as any detailed study of the translation and of the language of such texts can show, one finds a lot of good language (and fine translation) as well as a lot of language influenced by the source language. Almost every word used by the translator is the result of a choice (not necessarily conscious, of course) and, depending on the level and on the phenomenon one is interested in, the very same utterance can be both a good and a bad example of translation. Considering that one finds a lot of “bad” language in all sorts of other texts written by native speakers, whose explanation may be even more difficult to trace, and that most people do not react to translationese when they read translated texts, we must admit that such a borderline is really more theoretical than real.

The most controversial claim I want to make about translationese is, however, that it is methodologically a good aid in contrastive studies, because it can provide a means to measure what kinds of problems and differences there are between two languages (or, better, appear when one goes from one to the other). In other words, translationese is one of the best types of empirical evidence one has for language differences.

¹¹ In addition to Gellerstam (1986), see also Schmied & Schäffler (1996), Baker (1996), Santos (1995; 1997b).

¹² In fact, it would be interesting to study whether the influence of English on “original” texts in other languages, for example in texts written originally in Portuguese, is stronger among those authors who habitually read in English than among those authors who are used to reading translations from English only. If, on the one hand, the latter seems to be more protected, the influence may turn out to be more insidious because it goes unnoticed (assuming, for the sake of argument, that all translated texts will contain an unavoidable amount of translationese).

2.3.3 Baker's definition of "corpus-based translation studies"

According to Mona Baker (1996: 185), in corpus-based translation studies, "translation scholars are ultimately not interested in the words or syntactic structures themselves. What they are interested in are abstract, global notions such as explicitation and simplification, which are independent of specific languages [...]".

This is, of course, a clear statement of Baker's interests, but I think it is too narrow a view of translation studies. One **can** be interested in studying translation to learn about words and syntactic structures (and semantic structures, incidentally). To claim that this is not translation studies proper, but simply syntax, or semantics, or linguistics, is to deny the interdisciplinary character and the synergetic potential of translation studies themselves, which has been repeatedly emphasized by other scholars; cf. e.g. Snell-Hornby (1995).

My own work in corpus-based translation studies (at least I see it this way) goes against the above statement: I have been interested precisely in semantic, syntactic and lexical features.

Baker's research goal, on the contrary, studying features of translation *per se* independently of context (language pair, type of text, translation school followed by the translator) appears to me absurd: There are no "types of linguistic behaviour which are specific to translated text" which are independent of the languages involved. Furthermore, to say that a translation either simplifies or makes more complex a given text, or that it makes explicit some things while leaving others implicit is almost tautologous.

It seems to me that in any given translation there may be things that get simpler and others more complicated, and different degrees of explicitation and/or implicitation will occur for every feature which does not have an exact counterpart in the other language. This is undeniable, it agrees with common sense: why promote it to a Universal Feature of Translation?

Given that normalization (or sticking to conservative ways of expression) is probably as frequent as translationese (which is a sort of deviation), one again has no reason to say that any given translation (or translated texts in general) will have more of one than of the other.

In sum, whenever there are no precise translation equivalents (which is frequent, and therefore renders translation the exciting activity that requires close study), the process of translation is a tension between the following three pairs: simplification vs. complication; implicitation vs. explicitation; normalization vs. translationese.

To claim that it is possible to decide which features dominate in translation is, in my view, simply to deny the complexity of translation.

2.4 Translation data as linguistic data

As just mentioned, most studies of translation nowadays are not concerned with the linguistic properties that may or may not be involved, or underlie, the

translation process. Conversely, many researchers involved in parallel corpus studies have a much too simplistic view on translation, which may lead either to failure to discover significant facts due to erroneous assumptions, or even to false conclusions.

In my studies, I have tried to achieve a balance between practical feasibility and theoretical appropriateness. This requires that one must not only deal with observable linguistic properties, but also do justice to many subtleties that can be involved in the translation relation, and not assume that it is one of equivalence.

This will be discussed further in Chapter 4, with the presentation of the translation network model, a model designed to describe the contrasts between two languages that are made visible by looking at translation pairs.

Before concluding the present chapter, however, let me discuss some approaches that are related to the issue of translation as data (other approaches related to tense and aspect will be discussed in Section 3.1).

2.4.1 Jakobson and competence vs. performance

A famous article by Jakobson on translation points out that “languages differ essentially in what they **must** convey and not in what they may convey” Jakobson (1959: 236), looking at translation as a bridge between two (language) systems.

However, if we look at translation performance, my conclusion – or my answer to Jakobson’s quote – is that *languages differ in what they **do** convey, even if they **can** convey the same*. It is important to note here the difference between looking at translation in the abstract (a set of categories in one language that correspond to another set in a different language), and studying real translation data produced by real translators.

2.4.2 Dyvik’s translational-based semantics

Dyvik’s (1998) theory of translational-based semantics is apparently similar to my standpoint, but a closer look shows that it diverges crucially in aim and premises. I discuss it here to show that the relationship between translation and meaning is complex and varied, bringing about radically distinct models to join translation and semantics.

Dyvik’s aim is to provide a translational basis for the discipline of semantics, i.e., redefine semantics in terms of the primitive notion of IS_TRANSLATION_OF (or “optimal translation in some not narrowly restricted set of circumstances”, the details are not relevant for my purposes here). This goal I definitely sympathize with, and it is perfectly in line with the standpoints I have just expounded for contrastive studies.

However, Dyvik’s views – and therefore ultimate aims – are the opposite of mine in what concerns universals. As a traditional formal semanticist, he is interested in the *interlingua*, a meaning representation language that underlies every and each possible natural language:

One purpose of general linguistics, and hence of linguistic semantics, is to characterise the limits of variability among possible natural languages. Hence the distinctions we need to draw in our theoretical framework should be closely related to the distinctions that could be drawn in some natural language. [...] The number of distinctions, or “granularity”, we can achieve in semantic representations will [be limited] by the granularity of the world's languages. (Dyvik, 1998: 54)

This seriously compromises his attempt to provide a methodologically sound framework for semantics, if one happens to believe that distinctions made only by one language are equally relevant. It is thus easy to disagree with the following quote:

Even if we intuitively feel able to describe minute meaning differences between the members of a set of signs in a language, this does not legitimise giving them distinct semantic representations unless we can also demonstrate that they have different translational properties with respect to a relevant set of target languages. (Dyvik, 1998: 57)

Another problem with his proposal is that, although using corpora, Dyvik considers only a subset of translation data, namely lexeme to lexeme translation (ibid.: 62). By doing this, he throws away most of the interesting translation data (as I intend to show in Chapter 5). Incidentally, bilingual dictionaries would be more adequate for his purposes.

2.4.3 Schubert and implicitness in translation

Other relevant points related to translation, although expressed long ago, are still generally neglected or unknown.

One I want to revive here is Schubert's (1988) implicitness as a guiding principle to machine translation (that can actually be more generally applied to translation *tout court*). Schubert claimed that much (if not most) of a translation's success was due to a preservation of implicit information which was common in source and target language, and, quite the contrary, the attempt to make explicit the meaning of the source expression before translation (in order to translate), if taken seriously, would produce worse results. No matter certainly against mainstream MT ideals at the time, he had a strong empirical basis (its author had been engaged in large-scale machine translation for years) and it actually follows from what will often be stressed in the remainder of this book:

You lose, or distort, or add, or enhance, every time you translate into another language. If you do it twice (once into the semantic representation language, from there into the target language), you always end up with a less successful or more deviant version.

On the other hand, Schubert's statement also points to the notion of vagueness, which is one of the cornerstones of my analysis of natural language (and translation), and which I will discuss in the next chapter. In a nutshell, he claims that it is wrong to try to make things more precise than natural language itself does.

2.4.4 Tsujii and the translation of discourse

Finally, before ending a chapter devoted to parallel corpora and their analysis, I must raise a subject rarely discussed in connection with the analysis of translation pairs, but which plays an important role in the creation of the final translated text – possibly in a rewriting phase, posterior to the activity of translation itself: the strive for discourse cohesion.

One should never forget that the target “side” of a translation pair is often motivated by previous choices in the discourse, at least as much as it depends on the differences between the two language systems, or on the source content. Failure to take this into consideration may severe the analysis considerably.

In this connection, let me recall Tsujii’s (1986) observation that different languages also have different style conventions and discourse strategies, and that this fact has definite consequences for (machine) translation.

This truth is hardly ever fully appreciated in existing analyses of translation data, but we should keep it in mind.

Chapter 3

Tense, aspect, and semantics

This is not the place to give an overview of the tense and aspect literature or the discipline of semantics. I will simply discuss some of the issues I find most interesting with respect to those issues, and then present the specific model I use in my analyses of tense and aspect semantics in English and Portuguese.

It is the rule rather than the exception to be eclectic in the domain of tense and aspect. Even “schools of thought” like Discourse Representation Theory (Kamp & Reyle, 1993) use a mixture of intuitions and descriptive models coming from obviously disparate sources.

This allows me to move straight to substantive (rather than formal) matters. In Sections 3.1 to 3.3, I take up some methodological and philosophical points which, in a way, any researcher must take a stand on, and then I present in Section 3.4 a model adequately catering for my choices, consisting basically of an adaptation of Moens’s (1987) proposal. Section 3.5 illustrates the models for Portuguese and English.

Some of the questions I will mention are not specifically related to the domain of tense and aspect. They concern a general approach to semantics and linguistics in general. They will be discussed and illustrated exclusively with examples in this area, though.

3.1 The crosslinguistic analysis of tense and aspect

When one has more than one language as one’s object of study, there are basically three ways to proceed:

- One can take a *universalist* approach and look at Language as essentially one. Formal differences become just an irritating detail of no real linguistic interest – and then one is bound up to see clearly the **same** structure, the **same** thing (albeit imperfectly) in all languages.
- One can take the position that languages vary in ways which are non-random, and that classification of languages is possible, desirable, and ultimately the aim of linguistics: the *typologic* approach.
- One can take the position that each language is one original way to encode a culture and a way of looking at the world, and therefore should be studied for its own sake, and neither as a parameterized difference from a group of languages nor as yet another instantiation of the same thing: the *relativistic* approach.¹³

¹³ For a good discussion of linguistic relativity, see Rossi-Landi (1973), who proposes several more precise formulations, investigates some of their consequences, and discusses

Basically, choosing between the three ways is a question of personal world views (as so often happens in science). Work in the three directions (from different premises) is so different that it may be wondered whether it can be compared, or even seriously evaluated by members of the other schools.

Embracing the relativistic view, I will try to show here that it is the methodologically soundest, by criticizing some studies done from other standpoints. I am aware, nevertheless, that to the members of the universalist or typologist schools of thought the relativistic position is uninteresting, reductionist, and even “un-linguistic”.

One thing I hope in any case to make clear is that many studies proceed from assumptions which are strong (and in my view, untenable). For example, most formal work on tense and aspect embraces wholeheartedly the assumption that, at one deep enough level, it all reduces to the same thing, cf. ter Meulen (1995):

Even though the linguistic data in this book are limited to English, and no comparative data have been considered, the interpretive principles and semantic constraints formulated here should prove to find expression in all natural languages. The abstract semantic principles governing tense and aspect constrain the set of possible natural languages considerably (...) (ter Meulen, 1995: x)

Instead of arguing at an abstract level that “No, it does not reduce to the same thing”, I hope to show in the subsequent empirical chapters that it does not (at least, at the level which is interesting and meaningful for linguistic analysis).

I cannot resist, however, showing how primitive such standpoints are when they are applied to actual languages, by reviewing some of Carlota Smith’s (1991) analyses. I will also discuss some problems of typological approaches by reviewing Dahl’s (1985) famous work and Slobin’s studies.

3.1.1 Smith and the perfection of English

Smith (1991) represents the parametric approach to aspect, clearly inspired by Chomskyan linguistics:

Parametric variation occurs across languages in many domains, including the aspectual domain. [...] I will propose formal and substantive principles that underlie the domain of aspect in Universal Grammar. The principles provide the parameters that determine its realization in the grammar of particular languages. [...] the aspectual categories are not language dependent, but are based on human cognitive abilities (Smith, 1991: xviiff).

some philosophical problems around it. In addition, his book shows that this thesis, in one formulation or other, is actually to be found in a wide number of linguists and is not to be associated only to Sapir and Whorf. Examples cited are Benveniste, Hjelmslev, Martinet and Trier.

Her book is concerned with English, French, Russian, Navajo and Mandarin Chinese, and puts forward the concept of situation type, and of the imperfective, perfective and neutral viewpoints as universal concepts. The impression a non-committed reader gets, however, is that what Smith is doing is to devise a system appropriate for English, and analyse the other languages as deviating from it.

Let me present just one example. According to Smith's analysis, there is in English a perfect match between the perfective viewpoint (indicated by the simple past) and situation type. In the other languages she analyses, there is no single viewpoint that matches situation type: the perfective viewpoint is either defective (it does not apply to all situation types) or changes situation type. This makes it almost impossible for us, non-native speakers of English, to believe that the universal system is so strikingly identifiable with English...

3.1.2 Dahl (1985): a major work, but one important flaw

Dahl's book on tense and aspect constitutes a hallmark of (typological) tense and aspect studies. Dahl conducted a large study whose "primary aim [...] was to test the hypothesis that the TMA [tense-mood-aspect] categories that occur in the languages of the world can be reduced to a small set of cross-linguistic category types" (Dahl, 1985: 2). Dahl's study was based on a questionnaire of 250 questions which was translated from English into 63 other languages. The answers to the questionnaire were then reduced to the particular TMA categories of the language. Crosslinguistic categories were elicited by statistical analysis.¹⁴

However, in spite of the considerable care and the statistical apparatus used, it is not guaranteed that Dahl's study was not too dependent on English (or Swedish, his native language), either. In fact, one could hold that what was in fact investigated was the relative crosslinguistic weight of aspectual notions like the English perfect and the English progressive.

First of all, Dahl himself acknowledges that translation is not the best way to elicit meanings. Then, it is symptomatic that the results of the study did not manage to stress any crosslinguistic category which is marginal to English or Swedish. On the contrary, English categories were always ranked high among the universal categories found. Dahl actually notes, too, that the values for the Swedish perfect ranked suspiciously high in the perfect (PFCT) universal category. Third, the fact that the data were elicited through a questionnaire may make the "typical" cases illustrated by the questionnaire language specific, and, therefore, will only illustrate representatively the languages in which those cases

¹⁴ One of Dahl's most original assumptions – irrelevant though for the present point – was that universal categories would correspond to several meanings and uses, i.e., they were prototypical instead of being based on precise meanings; cf. "a set of universal TMA category types have been postulated, from among which language-specific categories are chosen. These universal categories are identified by clusters of semantic and morphosyntactic categories, which characterize their prototypical manifestations" (Dahl, 1984: 3).

are important. (It should be mentioned that care was taken to add distinctions not present in English that were already known beforehand to be of relevance – like witnessed versus unwitnessed assertion – but probably not frequently enough.)

Let me briefly give four specific examples of the bias of Dahl's questionnaire towards English, and consequently, of the difficulty of making sense of the Portuguese answers to it:¹⁵

1. A large number of questionnaire examples (23 cases out of the 67 criterial examples for the perfect) involved the English verb MEET, which belongs to a small class of verbs (acquisitions) which are particularly distinguishable through tense and aspect devices in English (and possibly all Germanic languages), but are typically translated into Portuguese (and possibly all Romance languages) by two lexical verbs of very different aspectual properties and associated patterns. Simply keeping account of the tenses and aspects used in the translations totally misses the point in these examples, cf. *have you met him?* vs. *I meet him often in the university* vs. *have you ever met him on the beach?*, respectively translatable by *Conhece-lo?* (verb CONHECER in Presente) or *Conheceste-o?* (verb CONHECER in Perfeito) if e.g. he is already dead, *Encontro-o muito na faculdade* (verb ENCONTRAR in Presente), and *Alguma vez o encontraste na praia?* (verb ENCONTRAR in Perfeito).¹⁶ In other words, the aspectual distinction made through tenses and aspects in English is lexicalized in Portuguese, and thus, for example, the tense opposition that may be involved in the Portuguese translations is but one part of the rendering of the aspectual distinction. Even more strikingly, the answers to Q30 and Q31 involve in Portuguese two contrasting copular verbs: *ser* and *estar*, but the same formal markers of tense and aspect.
2. One other instance is the verb SIT, which although only occurs once in the questionnaire, illustrates another general problem with this questionnaire-based analysis. The syntactic construction in which it occurs, typical of Swedish and possibly also acceptable in English, has no natural expression in Portuguese with two finite verbs, i.e., *he sat and read* can be rendered by *estava sentado a ler* ('he was sitting reading') or *estava a ler sentado* ('he was reading seated'). Only one of the verbs would be described by a finite tense. Now, an informant could have been led to produce the rather

¹⁵ In this discussion I will not suggest any original analyses of Portuguese categories. Rather, I try to embrace Dahl's assumptions just to show it is still problematic. Although readers are welcome to skip them, I include question numbers so that my claims can be checked.

¹⁶ The utterance *Já o conheces?* (for *Have you met him?*) would be more felicitous if the person referred by 'him' were present. Likewise, *Já o encontraste na praia?* would be equally acceptable for *Have you ever met him on the beach?* I tried, however, to present examples without *já*. (See point 3 below.)

awkward sentences *ele estava sentado e lia* or *ele estava sentado e estava a ler*, which, if this was the case, would imply that Dahl got translationese and not Portuguese.

3. Thirdly, the absence of the category *já* + Perfeito from Dahl's results, although warranted by my intuitions and by my own answers to his questionnaire,¹⁷ led me to the following disturbing conclusion: the Portuguese informant(s) simply did not understand that it was the perfect interpretation of the English bare verb which was at stake, because perfect in Portuguese is a covert category. In fact, barring grammatical "laws" (of the kind "use it with the adverbs *already*, *never*, *always*, *ever*, *lately* or with *since* adverbials"), the (present) perfect is one of the most difficult distinctions for a Portuguese speaker to master in English, given that both present perfect and past simple are translated by Perfeito. I additionally assume that, if the words *already* or *ever*, for example, had appeared in the questionnaire, then, even though the Portuguese informant would use *já* + Perfeito, the questionnaire researchers would take *já* as corresponding to *already*. This, although in a way true, would conceal the fact that *já* must "go together" with Perfeito.
4. Finally, Dahl classified the Portuguese Imperfeito as PASTi, i.e., a category indicating pastness in imperfective contexts. However, the questionnaire sentences dedicated to habitual past (HABPAST) are a proper subset of those of PASTi, and Imperfeito is the answer to the first and not to those that correspond to the set difference of the two. These latter ones would have to be rendered by Imperfeito + Progressiva in Portuguese. This is arguably due to the encoding of the complex category Imperfeito + Progressiva as both Imperfeito and Progressiva, which is what Dahl refers to as 'double accounting' (Dahl, 1985: 67). While this is not an *a priori* wrong assumption (it has a methodological justification), it resulted in the missed generalization of (bare) Imperfeito as HABPAST, while obscuring, in addition, the contrastive influence of Progressiva versus Imperfeito.

I believe that these examples serve to show that the linguistic bias is strictly inescapable, no matter the intent to make the investigation crosslinguistically neutral.

3.1.3 Slobin and thinking for drawing

In fact, one might go even further and try to avoid the use of a previous / primary language at all, as Slobin does, eliciting children's oral narratives from a picture book without text, which are then claimed to be original narratives in each language. I have, however, also doubts about the language-independence of this

¹⁷ In my opinion, *já* would be actually obligatory in many prototypical sentences of perfect, e.g. Dahl's Q53, Q56, Q64, Q153, and preferential in e.g. Q42, Q136, Q37.

method. Slobin acknowledges the existence of what he calls “thinking for speaking”

[Speakers seem] to be guided in how they choose to talk about experience by the most available grammatical means provided by their native language (Slobin, 1987: 443)

but fails to acknowledge that the creation of the story also required “thinking for drawing” (and therefore telling the story), and that, consequently, he is studying the differences of the languages towards English (or whatever Germanic language of the writer/drawer), and not original texts in themselves. (This is, in addition, emphasized by his apparent requirement of (at least) one sentence per drawing. From his examples, most notably those of Slobin (1987: 439-440), I believe it is clear that in some languages this does not lead to a “natural” narrative.)

I therefore conclude that, no matter the quality of the studies done in the typological paradigm, there is an inescapable bias towards one language, which will flaw the results.

3.1.4 Gawrońska and article and aspect choice

A particularly inspiring work is Gawrońska’s (1993) study of translation between Polish and Russian, on the one hand, and English and Swedish, on the other. She argues convincingly for the use of language specific categories as points of departure, even if the languages involved are typologically similar.

I cite here her description of how “universal” categories come about:

The way of reasoning to be found in many studies on ‘universal aspect’ may be characterized as follows: assumptions about the nature of the ‘typical’ aspectual distinctions give rise to a search for similar distinctions in languages which lack a corresponding morphological aspect category. Since several features [...] can evidently be expressed by other means in other languages, the category of ‘aspect’ is claimed to be universal. Then, further investigations on the nature of the semantic distinctions between different predications in an ‘untypical aspect language’ (e.g. English) lead to reformulations of the ‘universal’ notion of aspect. This reformulated notion is in turn assumed to give an explanation of the ‘typical’, morphologically marked verbal aspect. (Gawrońska, 1993: 75f)

Gawrońska’s work, although not corpus-based, is a noteworthy example of how a fresh look at problems can bring new and valuable insights. Her focus is the specification of rules for machine translation; however she makes use of both theoretical linguistics and translator’s intuitions.

3.2 Vagueness

Vagueness has been identified as a **problem**, by formal semanticists as well as lexicographers, by AI researchers as well as philosophers. In my view, however, it is one of the fundamental properties of natural language, a strength instead of a

weakness, in sum: a relevant property that should be investigated, exploited, and even mimicked.

Not many people share this view, but I can for example invoke Neustupný's (1978) opening words of a chapter dedicated to "Vagueness in language":

If the development of applied and mathematical linguistics has led to some decrease of interest in that property of language phenomena which I call vagueness, then the further road of linguistics will most likely lead to emphasizing its importance and to its recognition as a methodological principle of the first rank without which truly realistic linguistics cannot advance. (Neustupný, 1978: 41)

In fact, I believe that most problems in contemporary formal approaches, as well as in traditional linguistics, stem from the lack of consideration of this pervasive property of natural language.

As a matter of fact, it is astonishing that so much work and consideration have been put into ambiguity, which is a minor "sister" of vagueness, but whose description appears in every linguistic textbook. This may stem from the fact that it is easier to model (and deal with) in formal grammars, compilers etc., which focus on artificial languages.¹⁸

It is true that there are schools primarily concerned with vagueness: in artificial intelligence, the fuzzy logic paradigm, see Zadeh (1975); in linguistics, the cognitive linguistics paradigm, see Taylor (1989); and in mathematical logic and philosophy, where vagueness is a central issue, see Keefe & Smith (1996). This happens even in some formal (natural language) semantics quarters, as the books by Kempson (1975) and Pinkal (1995) show.

Nevertheless, consistent consideration of vagueness in a computational perspective has so far been almost completely lacking, even though one can identify it almost everywhere. Let me present several different studies which have, in some way or other, pointed to the relevance or inevitability of vagueness.

An interesting example is prepositional phrase attachment: depending on how to count their data, from 8.7% up to 21.7% of Hindle & Rooth's prepositional phrase examples were vague between attaching to the object noun or to the verb, given that "an attachment is semantically indeterminate if situations that verify the meaning associated with one attachment also make the meaning associated with the other attachment true" (Hindle & Rooth, 1993: 113).

Likewise, concerning anaphoric reference, Sampson (1987) argues that in a significant number of cases it is not trivial to decide the antecedent of the pronoun *it* – again due to the vagueness of the text.

Also, the task of corpus annotation was claimed to be difficult because tokens are often vague between several categories; cf. Leech (1993: 280):

¹⁸ It is also true that many phenomena that I consider as displaying vagueness have been (erroneously, in my view) described as ambiguity. This may account for its "success" in the literature.

“Experience with corpora suggests that uncertainties of category assignment are quite frequent [...] because of the prototypical, or fuzzy, nature of most linguistic categories”.

3.2.1 Types of vagueness

As most general terms in linguistics, “vagueness” is vaguely understood and used differently by almost everyone who deals with it.¹⁹

Vagueness can be thought of as the opposite of precision or well-delimited definition, but as such it can encompass at least underspecification, imprecision, and ambiguity, and perhaps also redundancy.²⁰

The vagueness I am interested in (or, alternatively, the way I want to define vagueness as a concept for **linguistic** research) does not include all kinds of underspecification (only the relevant one) or imprecision (only the systematic one); and it contrasts with ambiguity (see below). Furthermore, as far as redundancy is concerned, no matter whether it hinders a precise definition or not, it is, in my view, not related to vagueness at all.

Vagueness is an attribute of a classification (system). An object (a linguistic object, for instance) can be vague regarding as many classifications as it is subject to. A first definition of “vague” is thus: When an object can be classified as more than one thing in one particular classification scheme, then it is vague with regard to the distinction presupposed by the two (or more) values. Such an object can thus be used as a member of either category, and it can be taken to represent both. In addition, when a vague linguistic object is employed, “a speaker can remain indifferent between alternatives *a* and *b* and still meaningfully [use it]”, as Keenan (1978: 173) put it.

One can distinguish between several different kinds of vagueness (though here the classification is also individual, see Kempson (1975) and Neustupný (1978) for other classifications).

REFERENTIAL VAGUENESS applies to objects of the world, when they are to be classified by words. It is quite often the case that the words themselves are clear and precise enough to be understood and used by speakers without miscommunication, but there is not a well-delimited definition for these concepts in terms of the real world. For example, when classification is made in terms of a physical quantity (as is the case in the pairs *bald/not bald*, French *fleuve/rivière*), many would-be referents are vague between either word (either classification). If one wants to classify a “mid-sized” river in French or a person with little hair, one has — in some cases, at least — the possibility to classify the river as both *fleuve*

¹⁹ Neustupný observes that linguistic terminology has varieties (just like languages do) and he deplores the state of affairs in which “too many linguists [...] are complete monolinguals in the variety in which they happened to be trained” (Neustupný, 1978: 36).

²⁰ Interestingly, one can attribute this wide range of possible interpretations to the vagueness of the “opposite” concept itself — see e.g. Lyons (1968: 10.4.1; 1977: 9.1-2) on the many (linguistically relevant) phenomena concerning opposition and contrast.

and *rivière* and the man as both *bald* or *not bald*. A specific case of referential vagueness is *gradable opposites* – first discussed in Sapir (1944) –, in which definition a comparison is implicit, such as *good/bad*, *big/small*, *friendly/unfriendly*. Many subjects of the classification may be vague between e.g. *big* and *not big*.

If this were all there was to it, however, vagueness would not be particularly interesting from a linguistic point of view.²¹ Referential vagueness in language is a must: as Keenan (1978: 160) puts it, “human language must be imprecise in order to permit efficient communication”.

What is interesting is that language itself, when analysed as the object to be studied by whatever metalanguage (e.g. grammar, another natural language, logic) shows the same behaviour the “world” described by language displays. That is, words, phonemes, sentences will display the same categorization properties as birds, rivers and sounds. This leads to the conclusion that one of the constitutive properties of language is precisely vagueness.

One could thus say that there are such phenomena as grammatical vagueness, logical vagueness, contrastive vagueness, depending on which metalanguage (or level of analysis) one is concerned with.

Clear examples of GRAMMATICAL VAGUENESS, i.e., words or phrases which can equally well be analysed or interpreted in two different ways – syntactically or semantically speaking – are:

- Is *amigo* Adjectivo or Nome²² in *Ele é amigo do Pedro* (‘he is a friend of Pedro’s’)? Note that *amigo* can be modified by *muito* (‘very’), as well as inflected for the superlative (*amicíssimo*) as is the rule with Portuguese adjectives; it is, on the other hand, analogous to *ele é tio do Pedro* (‘he is Pedro’s uncle’), where *tio* is a Nome, and it can be used in the (nominal) construction *um dos amigos do Pedro* (‘one of Pedro’s friends’).
- Is *pintado* an Adjectivo, or a Particípio passado in *O muro pintado de branco tem de ser deitado abaixo* (‘the wall painted in white has to be destroyed’)? Are we describing a wall that is white (an active interpretation) or a wall that has just been painted white (a passive one)?
- Is *se* passive or ergative²³ in *O barco afundou-se*? (respectively ‘we drowned the boat’, colloquial, and ‘the boat sank’).
- Is *se* reciprocal or reflexive in *Eles zangaram-se* (‘they became angry’)? Angry with each other or just each one on his own?

²¹ But note that this is the vagueness referred to in philosophy.

²² I use the words *nome* and *adjectivo* to name the Portuguese part-of-speech categories that roughly correspond to noun and adjective in English. During the discussion in the present book I will try to use each language’s terminology, hoping to convince the reader that contrastive grammar is as necessary as contrastive lexicography/lexicology.

²³ The terminology is translated from Mateus *et al.* (1989), who discuss *se passivo* and *se ergativo* on pages 225 and 213 respectively.

- Is *conhecer* inceptive or stative in *Conhecer pessoas como ele é uma aventura* ('To know/meet people like him is fascinating')?
- Temporal or causal Gerúndio in *Uma vez, passando por casa dela, lembrei-me do irmão* ('I recalled her brother once, passing by her home')? At that moment, or because of passing?
- Consequential or temporal Gerúndio in *Fechei a porta da cozinha, deixando-a nos seus domínios* ('I closed the kitchen door, leaving her in her realm')? Consequently, or at the same time?

LOGICAL VAGUENESS, concerns sentences which can be given equally plausibly different logical translations. Some clear quantification puzzles are:

- *Alguns homens cantaram uma canção* ('some men sang a song'): all at the same time, the same song, or each one sang the same song, one after the other, or each man sang his own song, different and at a different time?
- *Todos os professores leram três livros* ('all teachers read three books'): Did they read the same three books?
- *Não gosto dos livros da Maria* ('I don't like Maria's books'): In addition to the well known vagueness of *de* (I don't like the books that she owns, the books that she publishes, the books that she writes?), this sentence is vague regarding the plural entity: I don't like her collection of books? I don't like any of her books, or I don't like most of her books (in general, though there may be exceptions)?

Let me at once emphasize that monolingual vagueness is solved in normal communicative situations in either of two ways: it simply does not matter to the understanding, or the context leaves it clear which option is meant (in which case the whole expression works the same way as if it were ambiguous).

What I will be mainly interested in here, however, is CONTRASTIVE VAGUENESS. Most contrastive and translation studies (though not acknowledging it) perform a classification of source language words and expressions in terms of another language (which can be viewed as a classification scheme).

In fact, translation out of context (e.g. as displayed in a bilingual dictionary) is a classification where there is often more than one candidate – one-to-one entries are very infrequent for most language pairs. The same is obviously also true for grammar, as pointed out by Nida (1959), and brought to the fore in machine translation quarters by Kameyama *et al.* (1991) under the name of *translation mismatches*, as discussed in 2.3.1 above.

In itself, this multiplicity of classifications is not necessarily a problem if in every context it is evident for the user of the dictionary (or translator) which one of the classifications (translations) to use. The degree of overlapping **in context** of the possible candidates is here the touchstone. (And, in fact, as I will discuss in 3.2.3 below, if they do not overlap at all, the source word is not vague, but simply ambiguous). The further apart they are, the easier it is to decide – it is for example difficult to conceive of hesitation between *terrorize* and *land* as translations

(classifications) of the Portuguese verb *aterrar*. In Santos (1997a) I presented a set of translations from Portuguese to English and German where the difficulty to choose between translations was grounded in some commonality that could justify the use of the word in the source text without requiring a conscious choice.

Note that this is a problem different in kind (and in practical relevance, as I will try to demonstrate) from the fact that the terms themselves may be referentially vague in each language. In fact, in the absence of the referent to which the source language word corresponds, if the target language happens to have imprecision cutting the pie in different places, this may complicate the process of translation because the translator must make a guess. There are, however, arguably few texts where precision as regards physical identification criteria is essential for accurate communication; such cases can, furthermore, be straightforwardly “solved” by translators’ footnotes or comments.

So, while what is small (or cold) may differ from culture to culture, as well as what can be good (or clever), the vague mapping of two vague terms is not so important as the mapping that translation itself imposes, producing several alternative translations for which there often is no reason to decide in the source language text.

Anticipating what I intend to show in the remainder of the book, I believe vagueness is relevant for translation in two ways: Sometimes, vagueness in the source language causes problems for the translator, because the vagueness cannot be maintained in the target language²⁴. Often, however, it is the translation process that induces the vagueness, a vagueness that is not felt (is not relevant) in the source language.

3.2.2 Vagueness vs. unmarkedness or underspecification

One might conclude from the paragraphs above that vagueness as I present it is nothing else than structuralist unmarkedness. Even though it is true that unmarkedness entails vagueness – if an expression is unmarked relative to, say, gender, we can say it is vague between all possible genders –, unmarkedness cannot be equated with vagueness; only subsumed by it.

In other words, there are vague expressions which are not unmarked: *x* can be vague between denoting A or B without neither A nor B being expressible on a markedness scale. In more technical terms, A and B are neither required to be in a privative nor in an equipollent opposition in order for an expression to be vague between them.

Also, not all underspecification can be counted as vagueness: Only pieces of information that are related to the classification should be taken into consideration. That is, a word can be vague regarding gender in a classification of

²⁴ Most distinctions relevant in (a particular) language (and which have therefore been categorized by linguists) can be neutralized in some contexts – and are often actually neutralized in language use. Consequently, monolingual vagueness is far from exotic, as will be amply shown in the following chapters.

gender in a language where expressions can be marked for gender, but one word cannot be vague regarding whether it is snowing or not, concerning gender, even though it is obviously underspecified about the weather.

3.2.3 Vagueness vs. ambiguity

Both vagueness and ambiguity can be loosely characterized as describing “one linguistic expression that has more than one interpretation or classification”. However, the two (or more) interpretations are mutually exclusive in ambiguity, not in vagueness. Moreover, I want to argue that vagueness is present in natural language both in competence and in performance, while ambiguity is a property only of the latter.

Vagueness is related to the language system, that is, it is systematic and, furthermore, a frequent property of language. On the other hand, I see ambiguity as an unsystematic and accidental property (though also pervasive).²⁵

In fact, for an expression to be vague between A and B, there must be a non-null intersection, a shared content between A and B; so that it is easy – and economical – to have the same linguistic object doing double duty. This discards the cases of “either A or not A” as instances of vagueness. Since vagueness is a property of the linguistic system, the knowledge to use it is part and parcel of a native speaker’s competence – and so, it is also automatically reflected in his or her performance. In comparison, most ambiguities produced by speakers go unnoticed and are irrelevant from a communicative point of view, therefore causing no problems for a translator.²⁶

Finally, a very important difference between ambiguity and vagueness is that choosing one of the alternatives at hand is information preserving in ambiguity (if one chooses the right one) – with the exception of the cases discussed in the previous footnote – while choice in vagueness always implies a loss, as noted by Kaplan *et al.* (1989).

None of these remarks can, however, provide a clear criterion to distinguish between ambiguity or vagueness in every case, for cases remain where we would need to be inside the author’s mind to know whether s/he was indifferent or, rather, meant one of two possibly opposing interpretations (and, in the latter case, also whether s/he was conscious of the ambiguity). In this respect,

²⁵ One argument for the difference of systematicity is the following: ambiguity at a particular level tends to be reduced (disambiguated) at a higher level (with more context); while vagueness can be preserved or enforced at all levels.

²⁶ Barring, of course, the cases where the ambiguity is intentional, or perceived (see Poesio, 1995) – such as puns, or, according to Rydning (1998), some political discourse. In those cases, it is functionally relevant in the sense of Catford (1967), and one may well be confronted with an untranslatable language exchange. This does not concern me here. In fact, although statistically marginal, this sort of problem is much more discussed in the translation literature than contrastive vagueness.

one should keep in mind that tests and definitions are different issues, as Gillon (1990) stresses.

3.2.4 Vagueness and compactness

The mirror image of (contrastive) vagueness is COMPACTNESS, i.e., a situation where a particular device (word, grammatical marker or construction, quantifier, etc.) conveys information that is perceived by another language as more than one piece of different information.

Monolingual compactness is at the heart of a linguistic system (bunches of features are generally assigned to each grammatical device) and its recognition has become one paradigm in linguistics (cognitive linguistics). In addition, it has had some impact in philosophy and artificial intelligence. Accounting for different shades of meaning is of interest to stylistics (to explain why some constructions may be preferred) or diachronic studies (to trace how the language has evolved).

CONTRASTIVE COMPACTNESS is often a problem for the translator: it is either difficult to specify all pieces of information involved, or the result is simply awkward. If the translator opts for conveying only some pieces, then it is not always straightforward what to choose.

In fact, an interesting observation I made in my corpus was that translators, whenever possible, when faced with a contrastively compact expression tended to use a vague translation, precisely in order not to make a choice.

3.3 A working definition of tense and aspect

TENSE is information having to do with order, position in a time line (or structure). Deictic tense is based on the concept of “now” (the present moment) and defines past and future as prior to or following “now”. Relative tense is concerned with the ordering relations between two times. ASPECT is concerned with the temporal shape of an event or situation, how it distributes in time. Some authors restrict the term to “temporal aspect”, cf. Carlson (1981: 32), defined as “properties of sets of periods which essentially involve the concept of (initial, internal, final) subperiod”. I take it more broadly as the temporal profile associated with an expression. Here, too, one might separate internal aspect (independent of “external” time) and external aspect, defined with respect to the time line.

Even though one may discuss whether tense and aspect exist in the “real” world, for the analysis of natural language it is only relevant to understand that these are categories essential to the way Indo-European languages (at least English and Portuguese) work, and are therefore to be considered as properties of the linguistic system.

There is an on-going discussion of which elements can meaningfully be attributed tense and aspect properties, which materializes in the following questions:

- Is there a distinction between aspectual class (Aktionsart) and aspect?

- What are the bearers of aspect? Verbs, verb classes or sentences?
- Should one use aspectual classes or aspectual features?

My answer to each of these points will be brief, hoping that the results achieved by the choices will argue in their favour.

Some researchers separate lexical aspect (or aspectual class) from other aspectual properties only definable with respect to the time line proper, like iterativity, habituality, inchoativity, progressiveness. The two main arguments are 1) that the devices that carry the two kinds of information are different (aspectualizers, morphologic derivation vs. lexical items) and that 2) a type/token distinction is involved: lexical aspect or aspectual class concerns the description of the types, while aspect proper or grammatical aspect concerns their existence as tokens, on a particular time line. But, are these differences enough to justify the stipulation of two distinct categories? In my view, these are all manifestations of the same category, concerning temporal organization. In fact, in the model I use, the interactions between lexicon and grammar can be satisfactorily handled under the assumption that they deal with the same categories.

This is also the answer I give to the next question. All (verbal) components carry some aspectual information. That is actually why one can classify verbs, verb phrases and sentences (and, even though sharing the main verb, not necessarily in the same category), as well as being able to specify the changes in aspect (aspectual class) that are produced from verb to verb phrase or to sentence. This aspectual calculus, first suggested by Verkuyl (1972), shows that it is not the case that anything goes; rather, there are principled ways to arrive at a final aspectual class starting from the properties of its components.

Finally, there are researchers who are not happy with the idea of class.²⁷ They suggest that features describing a class are more important and more fundamental than the classes themselves. In my view, this is simply wishful thinking. Classes are fundamental to cognition, features are a logician's artifact to try to understand the classes. Vendler was right in pointing out the typical classes in English. Others have since then pointed out that some logically definable classes were missing, as if that were a valid criticism.²⁸ In the present book, I will

²⁷ For example, Verkuyl (1989, 1993) has claimed that “aspectual class” is simply a distracting concept. Rather, its explanatory power is embodied in the features that are used to analyse it. That is, in Verkuyl’s terms, aspectual classes are no more than an idiosyncratic set of features, which, furthermore, are different according to different scholars. Sandström also makes the point that features are more basic than aspectual class; cf. “rules should make explicit reference to parts of a nucleus that are required to be present, or that are added, by a temporal expression, rather than to aspectual class” (Sandström, 1990: 66).

²⁸ In fact, it is the other way around: A theory that departs from aspectual classes can simply postulate that such and such aspectual classes exist in one language (and is welcome to provide a description of them in semantic terms), but a theory that departs

show that different languages, e.g. Portuguese and English, operate with different aspectual classes; neither language is claimed to be logically more complete or better.

3.3.1 A new definition of aspectual class

I will thus work with a notion of aspectual class as a primitive: an aspectual class is a kind of expression in a language which is relevant to the overall organization of that language. Both because the language displays a set of members sharing enough properties; and because it includes grammatical markers that create such kind of expressions and/or are sensitive to them.

How, then, should one go about finding aspectual classes in a language? One good starting point would be to look at the tenses and other verbal markers (like auxiliaries, aspectualizers) to try to define the distinctions they are sensitive to, and the distinctions they produce.

In the case of languages where grammatical literature is available, one could try to test the claims / features that have been published on those grammatical devices, as was the case in Santos (1994b), where I tried to assign to every occurrence of Imperfeito in a small corpus a set of features that had been identified as relevant, and ended up with clusters of features which I was unable – and later unwilling – to untie.

Another – in my experience more fruitful – course of action is to look at contrastive data including a well-studied language such as English, and see (unexpected) differences emerging. This gives a lot of data on the differences, but may fail to thoroughly investigate similarities, which incidentally may only be apparent.

One ingredient of aspectual classes that I found extremely important to take into consideration is the fact that categories vague between different (non-vague) aspectual classes is a widespread feature of languages.²⁹ If one pauses to think, one should not really be surprised, since vagueness is associated with efficiency of human communication.

3.3.2 Vendler's methodology, not Vendler's categories!

I begin to explain why one should look for Portuguese characteristics instead of using or adapting Vendler's classes. Most people merrily adapt Vendler's partition of English to other languages, taking its universality for granted. However, I believe this is to misread Vendler. In my opinion, one should follow Vendler's methodology instead, and look for clear grammatical contrasts that distinguish predicates in a language.

from (independent) semantic primitives is at odds to explain the lack of logically possible combinations.

²⁹ The previous sentence does not assert the existence of more general classes (for example having less discriminating features). As will be shown below, most vague classes are not amenable to a more general specification than the classes they “correspond” to.

I am not suggesting that the categories he elicited are specific for English. Since so many people imported the distinctions into their own language, it is beyond doubt that it is possible to identify the Vendlerian categories in several other languages. What I claim, instead, is that the **importance** that the English language as a system confers to those categories is **specific for English**. Other languages, in turn, may give more importance to other distinctions and even neglect those that English favours. Vendler, in fact, clearly states that his aim is “to describe the most common time schemata implied by the use of **English verbs**” (Vendler, 1967: 98f, emphasis added).

It seems uncontroversial that systematic distinctions in one language appear at all linguistic levels, and, therefore, instead of trying to organize the lexicon by grouping together individual items, one should look at the grammatical system and see what distinctions it allows the speaker to make. From there, one should be able to look for these very same features in the lexicon, and arrive at a motivated partitioning of it.

This was exactly what Vendler did for English. The following is a very short description of his tests: possessing continuous tenses or not; if yes, if someone is V-ing and stops, it may not be true that he V-ed, or it must be true. The first cases (accomplishments) accept the question *how long did it take*, the second (activities) accept *for how long*. For those which do not accept continuous tenses (the progressive), some (states) accept *for how long*, others (achievements) only *at what time*.

Summing up, the classes singled out by Vendler depend crucially on the meanings and distinctions effected by the English progressive and by the English prepositions *for*, *in*, and *at*. Now, everyone knows that prepositional meaning is one of the most language specific features there is. Likewise, the meaning of a particular aspect in one language is also never exactly mirrored in others (and this is particularly true as far as the progressive is concerned). Thus, it should be obvious that the categories Vendler uncovers are primarily relevant for English.³⁰

Now, one could argue against the claims above by saying that it was clear what Vendler meant with his tests, and that therefore those tests could be translated/rendered in any other language. This is at least approximately true. In fact, Vendler himself, arguing for the use of linguistic facts in philosophy, makes

³⁰ This is also obvious in the way Vendler distinguishes between generic states and specific states, depending on the (lexical) existence of a corresponding activity: a *grocer*, a *ruler* or an *educator* is never ‘grocing’, ‘ruling’ or ‘educating’, while a *smoker*, a *dogcatcher* or a *painter* are sometimes smoking, catching dogs or painting. Any attempt to transfer this distinction into Portuguese will fail, not only because there is no derivational similarity between the corresponding words (*mercador/merceiro*, *governante*, *educador*, *fumador*, *?apanhador de cães*, *pintor*), but also because the conclusions regarding the same meanings do not hold, either. In fact, a *mercador* (an old term) is sometimes “a *mercar*”, the word *merceiro* (‘grocer’) is not derived from a verb, just like *sapateiro* (‘shoe-maker’) is derived from the product (shoe), while *lavadeira* (‘washing woman’) is derived from the verb (to wash).

the following claims: (i) “philosophical arguments should explore the necessary truths embedded in some actual language or other” (ibid.: 26); (ii) “A statement like *One cannot know something erroneously* is true in all languages provided it is well translated” (ibid.), i.e., provided the translator “succeeded in reconstructing in his own language a conceptual model sufficiently similar to the linguistic environment of the English word” (ibid.: 25). But, if, for example, in Portuguese, *for* is translated in two different ways, and if *at what time* has no unique translation except a sort of *when*, one may doubt whether the same distinctions are being made in the other language.

This was precisely what led me to look for independent language-internal characteristics on which to base a classification of linguistic expressions. Here again I am supported by Vendler: “Philosophical statements mirroring some idiosyncratic aspect of a particular language are no less true than the ones corresponding to some common feature. The difference is that assertions of the former kind will be more difficult to translate than assertions of the latter kind” (ibid.: 28).

Note that the reason for this discussion is not a mere philosophical or methodological one. Its practical consequences should be overwhelming. By using Portuguese-internal distinctions, I expect to arrive at a model that explains a fairly wide range of phenomena whose explication would so far have been hindered by attempts to use English-specific categories.

To show how much it can be hindered, let me try out the Vendlerian categories on Portuguese, comparing it with the model arrived at for the Portuguese language that will be described later.

3.3.3 The Portuguese aspectual classes in a nutshell

Since this book is not intended primarily to those scholars whose main interest (and competence) is in the Portuguese language, I will be brief in the description of the aspectual system of Portuguese the way I see it. It is almost impossible to describe the (mental) process I went through to arrive at the model – even though it was to a large extent corpus based. I therefore content myself with describing the result here, providing in later chapters some corpus based (sort of) confirmation.

Qualidades, Estados and events

There are three major categories in Portuguese as regards their behaviour with respect to time: *Qualidades*, *Estados*, and events (*Qualidades* will also be referred to as properties or permanent states, while *Estados* will also be referred to as temporary states). Portuguese conceptualizes reality separating these three cases.

Events (as opposed to *Qualidades* and *Estados*) occur in time, i.e., event tokens come with a unique and definite temporal location (and spatial, as well). Events have thus temporal identification criteria, the same participants in the same place at a different time instantiate a different event (albeit possibly of the same event type).

The distinction between Portuguese temporary states and permanent states, or properties, has turned out to be exceedingly difficult to explain, even though I believe to have amassed considerable evidence for it being operational as a grammatical category.

Of course, I could say that this is a genuinely Portuguese conceptualization, and that there are simply no tests in English to help English speakers pick out the same properties Portuguese speakers do. Even though this may be true, it should not be impossible to explain the distinction in English.³¹ I will therefore try to explain where the boundaries are, in the following.

Note that this distinction is, in most cases, not done in content words (as opposed, for example, to Greg Carlson's (1977) distinction between stage-level and individual level predicates). Rather, it seems very much grammaticalized, especially if one considers *ser* and *estar* as grammatical markers of *Qualidade* and *Estado*, respectively, and not fully as main verbs. This means that questions of the following type are misguided for Portuguese: Are "young", "successful", "stationary", temporary states? Is "round" a property? From the point of view of Portuguese, these words (or their corresponding ones) are vague, and can be conceptualized as **either**; see examples (1)-(2):

- (1) *As malas são redondas, mas estão um bocado amachucadas agora.*
The bags are (SER) round, but they are (ESTAR) a little out of shape now
- (2) *Olha, a tua sombra está redonda.*
Look, your shadow is (ESTAR) round

On the other hand, words like *novo* ('new' or 'young'), even though by definition hold in temporary periods in time (in the everyday sense of the word temporary), are most commonly conceptualized as permanent states, i.e., as essential properties. This is probably so because they are inescapable for living things, as illustrated in example (3):

- (3) *O João é novo*
João is (SER) young

By contrast, when age is not so much measured by (inescapable) time but by use, one can describe it either as an essential or a contingent property:

- (4) *Este vestido está novo*
This dress is (ESTAR) new

Example (4) can be used to describe a 50 year-old dress which was found intact (and hardly used) in some cellar. Example (5) points this out even more clearly:

³¹ Incidentally, it seems that it can even be exported into English: John Harris, in a lecture at the University of Oslo in March 1996, noted that both Irish English and Caribbean English distinguish between a habitual or punctual property (his terms), the examples being: *I do be sick* vs. *I am sick*. This seems an exact match for the permanent / temporary distinction expressed in Portuguese with *ser doente* / *estar doente*.

- (5) *Comprámos estas calças ao mesmo tempo, e as minhas **estão** velhas e as tuas novas!*

We bought these trousers at the same time, and mine are (ESTAR) old and yours new!

There are obviously many other examples of the distinction encoded in *ser* vs. *estar* in contexts other than adjectives. I present locative descriptions in example (6) and passives in example (7):

- (6) (Ao novo jardineiro:) *As estátuas **são** no jardim, mas **estão** aqui por causa das obras.*

To the new gardener: The statues are (SER) in the garden, but they are (ESTAR) here (now) because of the construction work

- (7) *O aspirador **é** arrumado na despensa mas **está** arrumado aqui hoje porque estamos a pintar a despensa.*

The vacuum-cleaner is (SER) kept in the storage room but is (ESTAR) kept here today because we are painting the storage room

Surprisingly, the best description of the difference between properties and states I was able to find, and with which I intend to mirror the natural use of property-inducing grammatical devices by any native speaker of Portuguese, is Hempel's description of a scientific law (Qualidade) vs. an accidental generalization (Estado):

A scientific law cannot be adequately defined as a true statement of universal form: this characterization expresses a necessary, but not sufficient, condition for laws of the kind here under discussion. What distinguishes genuine laws from accidental generalizations? [...] a law can, whereas an accidental generalization cannot, serve to support counterfactual conditionals, i.e., statements of the form 'If A were (had been) the case, then B would be (would have been) the case', where in fact A is not (has not been) the case. [...] Similarly, a law, in contrast to an accidentally true generalization, can support subjunctive conditionals, i.e., sentences of the type 'If A should come to pass, then so would B', where it is left open whether or not A will in fact come to pass. [...] a law can, whereas an accidental generalization cannot, serve as a basis for an explanation. (Hempel, 1966: 55f)

Let me just mention that the same "objective" situation³² can be conceptualized differently by different speakers, as a contingent or an essential property. Consider the alternative formulations: *Ela está gorda* vs. *Ela é gorda* ('She is fat'). However referring to the same weight and external appearance, the two sentences (and associated beliefs) may differ strikingly as far as their psychological consequences are concerned.

³² Strictly speaking, situations would only be objective if there were language without speakers, i.e., perceivers of the world.

In order to show that the trio events-Qualidades-Estados are actually observable at all levels of Portuguese grammar, I argued in Santos (1996a;b) that:

1. The three past tenses Perfeito, Imperfeito and Pretérito Perfeito Composto (PPC) correspond to these categories, and produce different interpretations for arguments of each kind. Table 3.1 displays a simplified representation, where I try to show with lexical predicates roughly corresponding to be good, be sick and fall down that different tenses produce different interpretations (the slash means disjunction, and “prop. t” stands for a property of the particular time *t*).

Table 3.1: Aspectual profile of three Portuguese past tenses

	Qualidade Imperfeito	Estado PPC	Acontecimento Perfeito
Qualidade <i>ser bom</i>	<i>Ele era bom</i> Qualidade	<i>Ele tem sido bom</i> Acont./Estado	<i>Ele foi bom</i> Acont./Estado
Estado <i>estar doente</i>	<i>Ele estava doente</i> Estado (prop. t)	<i>Ele tem estado doente</i> Estado	<i>Ele esteve doente</i> Estado limitado
Acontecimento <i>cair</i>	<i>Ele caía</i> Qualidade/Acont.	<i>Ele tem caído</i> Estado	<i>Ele caiu</i> Acontecimento

2. Portuguese aspectualizers also select (or create) different categories. There are thus different lexical items for conveying a Qualidade or a temporary state, as in the case of example (8). Likewise, different aspectualizers select (or prefer) a different interpretation of the situation they apply to, as in examples (9) and (10).

- (8) *Ele parou de correr*
He stopped running; he is not running now
Ele deixou de correr
He quit running; his running is over
- (9) *Ele tornou a fumar*
He smoked again; he took another cigarette
Ele voltou a fumar
He smoked again; he went back to smoking
- (10) *Ele tornou a ler o jornal*
He read the newspaper again
Ele voltou a ler o jornal
He went back to his habit of reading the newspaper

3. There is a sharp distinction between a temporary or permanent goal encoded by the prepositions *a* and *para* in European Portuguese, as examples (11) to (14) make clear.

- (11) *Vou à universidade (e volto já).*
I go by the university (and will be back soon)
Vou para a universidade.

I go to the university and will stay there for some time

Note that *a* does not imply that I go and come back, even though this is often the case. It only implies that I drop by but do not linger there, as in example (12).

- (12) *Vou ao merceiro e depois para casa*
I go by the grocer's and then home

Other examples are (13) and (14):³³

- (13) *O Luís foi para o Brasil em 1974.*
Luís moved to Brazil in 1974
Nós fomos ao Brasil nas férias de Natal.
We've been to Brazil this Christmas

- (14) *A Rosa foi à polícia.*
Rosa went to the police station, and should be back soon
A Rosa foi para a polícia.
Rosa enrolled in the police corps; Rosa went to the police station to remain there, to be interrogated, or to do some work there

4. A similar distinction between a *Qualidade* and an event (a general property of a temporal location, or an accidental temporal location) can also be formally marked by the preposition *em* or *a* in temporal adverbials; cf. example (15), where *na* and *à* are contracted forms of *em* and *a*. There are also some temporal adverbials which can only have a *Qualidade* interpretation (introduced by the preposition *de*) as in example (16).

- (15) *Na quinta-feira vou/fui ao cinema*
Thursday I'll go/I went to the movies
À quinta-feira vou/ia ao cinema
On Thursdays I go/I used to go to the movies

- (16) *De dia quase não há polícias aqui*
By daylight there are almost no police here
Ele só consegue estudar de noite.
He can only study at night

5. Temporal connectives are also clearly sensitive to the threefold difference between *Qualidades*, *Estados* and events: *Quando* ('when') can only be felicitously applied to events or *Estados*; and *desde* (temporal 'since') maps differently events, *Estados* and *Qualidades*: the whole event corresponds to the beginning of the extended period, cf. example (17); while the whole *Estado* or *Qualidade* is co-temporal with the extended period described by the *desde*-clause, as seen in examples (18)-(19).

- (17) *Desde que o João tomou banho no rio, está terrivelmente constipado.*

³³ This is a feature that only makes sense with change of location verbs, namely *ir*, *vir*, *levar*, *trazer*, *subir*, *descer*, and a few others; nevertheless, it is used all the time.

Since John bathed in the river, he has a terrible cold

- (18) *Desde que estou a tomar banho já tocaram duas vezes à porta*
Since I've been taking a bath the doorbell has already rung twice
- (19) *Desde que dou aulas à noite ando muito cansada*
Since I am an evening teacher I feel very tired

Obras and Mudanças

When we look closer at the class of events, it is also possible to find two main kinds which have no exact correspondence to any of the three English classes of events (activities, accomplishments and achievements): *Obras* and *Mudanças* are respectively defined as those events that take time, and those which bring about a change.

Again, this distinction was motivated by a large number of observations concerning different linguistic behaviour (as well as by the translation regularities which will be described in later chapters). The reader should nevertheless be aware that some of the statements below are simplified.

1. The Progressiva (*estar a* + Infinitivo) means in progress with *Obras*, see examples (20)-(21), and on the verge of with *Mudanças*, cf. examples (22)-(23):

- (20) *Estou a escrever um livro.*
I'm writing a book
- (21) *Estou a tomar banho.*
I'm taking a bath
- (22) *O comboio deve estar a chegar.*
The train must be arriving
- (23) *O vaso está quase a cair.*
The pot is nearly falling

2. The aspectualizer *acabar de* means “to finish” only with *Obras*, cf. examples (24)-(25). With *Mudanças*, it can only mean “just happened”, see examples (26)-(27):

- (24) *Acabei de pôr a mesa às 2.*
I finished laying the table at two
- (25) *Acabei de ler o livro, podes levá-lo.*
I finished reading the book, you can take it
- (26) *Ele acabou de sair.*
He has just left
- (27) *Ele acabou de voltar da loja.*
He has just returned from the shop

3. The aspectualizer *ir* followed by Gerúndio means “almost” with *Mudanças*, cf. examples (28)-(30a); and continued activity with *Obras*, cf. examples (30b)-(32):

- (28) *eu ia caindo*
I almost fell down
- (29) *eu ia-me esquecendo*
I almost forgot
- (30) *eu ia ganhando*
a) I almost won / b) I kept earning/winning
- (31) *eu ia trabalhando*
I worked from time to time
- (32) *eu ia lendo*
I kept reading

4. *Já* with *Mudanças* can be paraphrased with *já* with the result state, cf. examples (33)-(34); *já* with *Obras* is mainly existentially interpreted, cf. examples (35)-(36):

- (33) *Já cheguei = já cá estou*
I've arrived = I'm already here
- (34) *Já entrei na floresta = já estou na floresta*
I've entered the forest = I'm already inside the forest
- (35) *Já comi um bolo ≠ já há um bolo no meu estômago.*
I've eaten a cake ≠ there is already a cake in my stomach
- (36) *Já escrevi um livro ≠ já há um livro escrito por mim.*
I've written a book ≠ there is already a book written by me

5. *De ... a ...* can only be used with *Obras*, where it is grammatical, cf. examples (37)-(38).

- (37) *Ele construiu a casa de Janeiro a Março.*
He built the house from January until March
- (38) *Ele correu das duas às quatro.*
He ran from two to four

6. In contexts where habituality is not possible, Imperfeito can be applied to *Obras*, see example (39), expressing perspective from inside (perspectival aspect), but not to *Mudanças*.

- (39) *Quando ele se banhava no rio, apareceu uma criança.*
When he bathed in the river, a child appeared

This ends an introduction to the way situations are conceptualized in Portuguese. Many of these points will be taken up again later on. Let me now contrast this Portuguese-centred classification with a classification based on the original Vendlerian tests.

3.3.4 Comparing the two sets of classes from a Vendlerian angle

On the semantic side, Vendler's tests distinguish between resultative and non-resultative events, and then, for the resultative, between extended events and punctual events (as seen from an English perspective).

Now, a possible way to describe the situation in Portuguese is by somehow reversing the order of the tests, and to notice that punctual events are identified with change (not result in itself, but this is possibly a too subtle question to be of interest here). Figure 3.1 provides a graphical illustration of the contrast between Portuguese and English events.

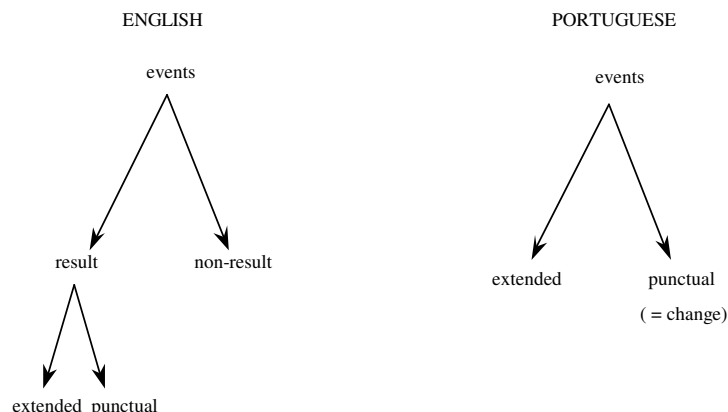


Figure 3.1: Comparing event structure in the two languages

On the linguistic side, Vendler's tests are based on the perfect – the right thing to do if it selects the result. But what is the analogue of the English perfect in Portuguese (in its use of selecting the result)? There is none, precisely because the notion of result is not grammatically relevant in Portuguese.

The other major test suggested by Vendler, namely, cooccurrence with the progressive, intends to look for extendedness as an inherently relevant property. And since this is a relevant category in Portuguese, it is a more Portuguese-oriented test, and thus, as was shown above, I also used it (or its analogue) to distinguish between *Mudanças* and *Obras*.

However, there is a difference between the two languages. The progressive is not grammatically relevant in Portuguese to the same extent as it is in English – and, in fact, many other criteria were just presented to distinguish between *Obras* and *Mudanças*. More to the point, there is a somewhat similar (or competing) test: *Imperfeito* denoting perspectival aspect or habituality (the first for the extended type, the second for the non-extended type). This test may be more adequate as a “translation” of Vendler's progressive test, given the relative frequencies of the *Progressivo* and *Imperfeito* in Portuguese, cf. Chapter 5.

The other tests Vendler provided to show result or durativity, namely cooccurrence tests with *in-* and *for-*adverbials, can be subject to analogous

remarks. However, it is important to note that these tests do not look for the same properties: *in* indicates boundedness (not resultativity)³⁴ and *for* extendedness. It is a property of English that in most cases there is a conflict between the two tests, i.e., inherent bounded and inherent extended seem to conflict, though this is neither natural nor logically necessary: it is equally acceptable, in my view, to say (though conveying different meanings) *I painted the house in two hours* or *I painted the house for two hours* (or with *read the book* substituted for *paint the house*). But, in most cases, in English you can either only use *in-* or *for-*adverbials with a particular verb, as Vendler pointed out.

Now, as one would expect, analogue tests in Portuguese will fail to arrive at a relevant distinction: *Obras* will be felicitous with both *em* and *durante* (roughly corresponding to *in* and *for*), cf. examples (40)-(41);³⁵ and *Mudanças* with neither. *Mudanças* would be felicitous with the adverbial *dali a*, which simply indicates relative location, not duration, cf. example (42).

(40) *Comi em dez minutos, estava com tanta pressa!*

I ate in ten minutes, I was in a hurry!

(41) *Comi durante uma hora, já não posso mais!*

I ate for one hour, I can't eat anymore!

(42) *Ele entrou dali a uma hora.*

He entered one hour later / in one hour

This concludes my point on Vendlerian categories for Portuguese. What Vendler did was remarkably appropriate for English, but he pin-pointed properties of the English language, which are not by any means conceptually or logically necessary. He did not aim, nor (accidentally) manage, to identify distinctions which are universally valid and relevant for the grammar of all languages of the world. Or, to remain within the present subject, for Portuguese.

3.4 The descriptive model

I will start by describing Moens's (1987) model of the interaction between several pieces of tense and aspect information in section 3.4.1. This model is appealing

³⁴ See e.g. Dahl (1981) for a discussion that shows that inherent boundedness (telicity) is often mixed with the notion of completion (or external boundedness). Even though he ends the paper by claiming that one of these categories is not relevant, I believe that his conclusion is misguided, and derives from his attempt to arrive at a universal definition drawing from examples of several languages which are incompatible in this respect.

Egg (1995), on the contrary, explicitly argues for the existence of two independent features "telic" and "bounded", and classifies bounded non-telic events as a new aspectual class that he calls "intergressive".

³⁵ Admittedly, there are *Obras* which are only natural with one of these adverbials as in English, cf. *construir a capela em dois meses* ('build the chapel in two months') and *trabalhar na capela durante dois meses* ('work in the chapel for two months'), but in my view this is not enough to claim yet another main distinction inside *Obras*.

because of two characteristics: it allows one to handle vagueness and express markedness, and it provides a systematic and encompassing description of English to start with.

In section 3.4.2, I will describe the changes to the model necessary to make it conform to the ideas expressed above, when they were not present in the original Moensian model. Some examples of the model's usefulness for both English and Portuguese will be provided in section 3.5.

3.4.1 The aspectual network of Moens

Since Moens's dissertation in 1987 (Moens, 1987; Moens & Steedman, 1988), the concept of aspectual network has significantly influenced the computational linguistics community, especially in tense and aspect studies. See e.g. Caenepeel (1989), Sandström (1993), Kent (1993), Lascarides (1988) and Santos (1993) for proposals using and extending Moens's insights.

Here I summarize Moens's main contributions, the way I see them:

1. First, Moens provided a formalization of aspectual class in terms of the concept of a nucleus, i.e., "a consequentially related sequence of preparatory process, goal event, and consequent state" (Moens, 1987: 3), glued together by a notion of **contingency**, rather than simple temporal order. Different aspectual classes correspond to linguistic expressions depicting different parts of a nucleus.
2. Second, Moens provided a formalization of the aspect calculus (i.e., the way several linguistic ingredients combine to give the final sentence its aspect) by geometrically modelling it as a path in a finite network. In addition to a purely compositional framework, where the existence of some devices would bring expected new aspect values, he also tackled the cases where apparently incompatible devices add up to perfectly understandable English sentences: For this latter case, he introduced the notion of coercion, i.e. the reinterpretation of (prior) aspect brought about by the application of an aspect operator.

For example, in *he played the sonata for ten hours*, the *for*-phrase forces its argument (typically a culminated process³⁶) to be read as a process, and so *he played the sonata* is coerced into a process (by iteration, in this particular case, but note that a process could be obtained by selecting a part of the sonata, which would be the reasonable case if *ten minutes* had been used). The end result of adding the *for*-expression is then a new culminated process.

³⁶ With his proposal, Moens introduces a new terminology for events, namely processes, culminated processes, culminations and points. The first three correspond roughly to Vendler's (1967) activities, accomplishments and achievements. Points correspond to punctual non-resultative events such as *tap* or *knock*. I will use Moens's terminology when describing his work.

This concept of coercion models what Sandström calls the “contextual dependency of aspectual class”, i.e., the context causes the reinterpretation of the aspectual class a given sentence (or part thereof) belongs to.

3. Finally, Moens suggested a uniform solution to several puzzles of English tense and aspect with the machinery above:
 - a) He related the perfect and the progressive aspect operators to parts of the nucleus, and modelled marked cases of the use of such operators by coercion in the aspectual network;
 - b) He related narrative progression and temporal connectives also to operations that selected, or added, parts of the nucleus.

In sum, by giving more structure to the semantics of tense and aspect, Moens was able to provide more adequate descriptions of the behaviour of several tense and aspect mechanisms.

Of the various proposals made by Moens, I will concentrate on the computational model, the aspectual network proper, whose rationale is to model the import of the several bearers of aspect in text, through transitions among the nodes in that network. Here are Moens's own words of motivation:

It would seem rather strange to say that *run* or a sentence containing the verb *run* is ambiguous between all these categories. It seems like a better idea of a Vendlerian taxonomy not as a fixed classification, but as a dynamic one where an element like *run* [...] can move from one category to the other, provided the context is such as to support the assumptions associated with the transition in question (Moens, 1987: 44).

Throughout his dissertation, Moens is concerned with English, and he purports to describe all and only possible aspect transitions of English with the network. Let me describe his model in some detail.

The ontology. Each node represents an aspectual class. Aspectual class is equated with event type. Aspectual class is defined in terms of a nucleus, which maximally comprises a preparatory phase, a culmination point and a consequent state. Basically, changes in aspectual class add part to or delete parts from events.

The nodes. Moens describes English as containing four nodes of the event kind (culminations only, culminations with resulting state, culminations with preparatory process and resulting state, and preparatory process only), and four nodes of the state kind (comprising consequent states, and three additional kinds of states: progressive, habitual, and lexical states).

Every node has the same status in Moens's network, that of classifying a possible English proposition. To start with, basic propositions in English are defined as those denoted by sentences in the simple past, with subjects syntactically and semantically singular. However, Moens also mentions that “the network can be used to classify all these linguistic units [smaller units than

sentences, such as verbs or verb phrases]” (Moens, 1987: 94), by classifying sentences whose remaining constituents do not bring any aspectual changes.

The transitions. The network specifies the transitions among aspectual classes that are possible in English. Figure 3.2 is my own drawing of Moens's original network.

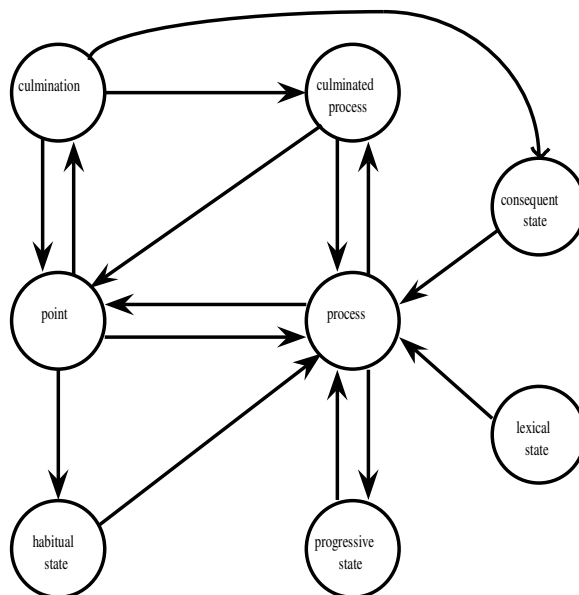


Figure 3.2: Redrawing Moens (1987) original aspectual network

According to the motivation (and evidence) for the transitions suggested by Moens, there are three kinds of transitions contained in his aspectual network:

1. Transitions corresponding to aspectual operators, which are brought about by some morphosyntactic device, such as the perfect, the progressive, or a *for*-adverbial. Moens describes specific linguistic operators in a functional way:

OPERATOR: aspectual class 1 ---> aspectual class 2.

2. Transitions brought about by **coercion**, i.e., caused by the application of an aspectual operator whose definition requires a different kind of input.
3. Transitions unmarked “all the way down”, i.e., transitions made without an explicit clue, in which an expression is interpreted as belonging to a different aspectual class without any objective operator to license it (this is how Moens handles habituality in English, see below).

These distinctions were inspired by the discussion in Kent (1993). The terms *marked* and *unmarked* were introduced by Kent to denote, respectively, “transitions which correspond directly to syntactic constructions”, and those “which do not correspond to any syntactic constructions” (Kent, 1993: 90). The question is whether this definition is understood in terms of the system, or in terms of a particular instance. For “without an exhaustive study of syntactic constructions in English one cannot tell whether all so-called unmarked transitions actually are associated with a corresponding syntactic construction” (ibid.: 91). Kent suggests the terms *explicit* and *implicit* marking for what I have labelled transitions of the first and second kinds.

Examples. I will use Moens’s original aspectual network to illustrate the transitions. To make it possible to follow the examples easily (annotated with the page number in Moens’s dissertation), I present the definitions given by Moens of the operators FOR, IN, UNTIL and the progressive:

for: PROCESS ---> CULMINATED PROCESS
 in: CULMINATED PROCESS ---> CULMINATED PROCESS
 until: PROCESS ---> CULMINATED PROCESS
 progressive: PROCESS ---> PROGRESSIVE STATE

So, the sentence *John ran for two hours* (p.46) is analysed as *John ran*, which is a process, plus a *for*-adverbial, yielding a final culminated process. Schematically, underlining the path label:

PROCESS (John ran) for (two hours) CULMINATED PROCESS

This simple example only features one transition of the first kind.

Now, in *he read a novel for a few minutes* (p.45), the expression *he read a novel* is a culminated process, so a transition from culminated process to process is effected through coercion caused by the *for*-adverbial, and then the addition of *for* brings about a new culminated process. I use the symbol “@>” to denote coerced transitions. Two different transitions were thus involved in the analysis of this sentence.

CULMINATED PROCESS (he read a novel) @> PROCESS for (a few minutes)
 CULMINATED PROCESS

A similar situation occurs in *The mountaineers reached the top in less than 5 hours* (p.47), where *The mountaineers reached the top*, which is basically a culmination, is first coerced into a culminated process, and then a new culminated process is obtained by the addition of the *in*-phrase:

CULMINATION (the mountaineers reached the top) @> CULMINATED PROCESS
in (less than five hours) CULMINATED PROCESS.

And similarly, the analysis of the sentence *My program ran in less than four minutes this morning* (p.47) proceeds by adding a culmination to a process due to the presence of the *in*-adverbial:

PROCESS (My program ran) @> CULMINATED PROCESS in (less than four minutes) CULMINATED PROCESS

Even though Moens states that “for other transitions, such as the one resulting in an iterated process or an habitual statement English has no explicit markers and they can be made freely” (Moens, 1987: 51), he does not present an explicit example of habituality only, such as for example the sentence *John worked for IBM*. I will, however, present his example *I was walking to work last winter* (p.62), which, according to Moens, involves a transition into habitual state:

CULMINATED PROCESS (I walked to work) @> POINT @> HABITUAL STATE
@> PROCESS progressive PROGRESSIVE STATE

This case is not the only one where coercion involves more than one transition. More than one unmarked transition appears in several analyses, as I will show. A case in point is the progressive of culminations: *John is reaching the top* (p. 59) is analysed by Moens in the following way:

CULMINATION (John reached the top) @> CULMINATED PROCESS @>
PROCESS progressive PROGRESSIVE STATE

Yet another example is *John arrived late until his boss told him off* (p.52), where one has the point *John arrived late* coerced into a process by iteration:

CULMINATION (John arrived late) @> POINT @> PROCESS until (his boss told him off) CULMINATED PROCESS

Finally, for *John was solving the puzzle in 5 minutes* (p.54), Moens offers two readings, respectively

1. CULMINATED PROCESS (John solved the puzzle) in (5 minutes) CULMINATED
PROCESS @> PROCESS progressive PROGRESSIVE STATE
2. CULMINATED PROCESS (J. solved the puzzle) @> PROCESS progressive
PROGRESSIVE STATE @> PROCESS @> POINT @> CULMINATION @>
CULMINATED PROCESS in (5 minutes) CULMINATED PROCESS

This illustrates the possibility of having more than one path in the aspectual network, corresponding to different interpretations.

3.4.2 My appropriation of Moens's model

Though using most of the machinery proposed by Moens for the description of English and Portuguese, I made the following changes to the interpretation of aspectual networks in general:

1. I have dropped the identification of aspectual classes with features representing missing or present parts of an abstract entity called the nucleus (containing a preparatory phase, a culmination, and a consequent state). This was both due to my consideration of class as more basic than features (see Section 3.3 above) and due to specific properties of Moens's own network that showed the fragility of such identification: First, most (English) states could not be accommodated in such a model anyway. Second, and though it would be logically conceivable that an aspect class

should be formed by a preparatory state plus a culmination (without consequent state), it does not appear in Moens's network (probably because it is not relevant for the description of English).

2. I have added the possibility of having nodes representing syntactic entities lower than sentences/propositions, so that the aspectual network might be used for the contribution of arguments and not only of adjuncts. In fact, I consider Moens's model highly appropriate to render the changes in aspectual class (and properties) brought about by plural arguments to "singular" events and by singular arguments to "plural" events, such as *build bridges* and *run a mile*, respectively.
3. I have also extended the use of the aspectual network to tenses. It is understandable that, given the few tenses in English, Moens did not have to deal with the matter. Still, his stipulation of basic propositions in English (sentences in the simple past with subjects syntactically and semantically singular) is very suspect and, from my point of view, unnecessary. The problem is that Moens defends the use of the past tense by claiming that "tense fulfills the function of disambiguating the aspectual character of sentences which would otherwise be aspectually ambiguous" (Moens, 1987: 95). But then, how can one analyse sentences other than those in the simple past? As Kent (1993: 82) puts it, "But then how is the past element of meaning, which must be part of this proposition, removed?" By removing the requirement that every node must represent a tensed proposition, I am able to express, for example, the aspectual import of the English simple present as:

simple present: POINT ---> HABITUAL STATE

4. Yet another relevant criticism made by Kent (1993) is whether Moens's unmarked transitions are to be taken as unmarked transition types (in English) or as unmarked transitions in a particular utterance. This question can be re-stated as follows: can one define a one-to-one mapping between transitions and arcs, or can a given transition be performed by travelling different arcs, some of them even unlabelled? Though Moens seems to have had the first interpretation in mind, I will opt for the second one, that is, there may be two nodes in an aspectual network linked by several different arcs.³⁷
5. In addition, I have restricted Moens's model in the ability to perform unmarked transitions, as will be shown presently. The motivation runs as follows: paths involving several unmarked transitions which are not objectively motivated (by some grammatical operator) introduce an element of indeterminism which prevents evaluation ("anything goes").

³⁷ This will, among other things, help explain language evolution, when eventually one arc may be preferred.

Contrary to the marked transitions and to the initial and final nodes of a coercion path, any intermediate nodes are not directly “observable” and may look suspiciously artificial. In the very few cases where such a path seems to have to be taken, I impose a strong requirement of perceived markedness, i.e., only marked utterances can be analysed by such paths, not ordinary ones.

6. Finally, I have not used Moens’s model for modelling narrative progression, basically because I did not have the time to look at the phenomenon in my own studies. It is nevertheless a very interesting issue that has been taken further by Caenepeel (1989) and by Sandström (1993), to which I will only make some sporadic remarks in this book. I have, however, in line with Sandström’s (1990) remarks that *when-* and *while-clauses* have only coercion capability, suggested an alternative for dealing adequately with subordinate clauses, separating the statement of the aspect of the subordinated clause from that of the main clause.

Let me summarize the redesigned version of the aspectual network, for further reference:

The aspectual network is simultaneously meant as a description of the grammatical system (a static model), and a dynamic model for the description of any particular sentence as far as aspect is concerned. Thus, an aspectual network is a representational device formed by nodes and labelled directional arcs that link the nodes, to which an interpretation is provided in terms of world situations as seen by the language. And the aspectual network can also be seen as a computational device for getting the reference to situations of the right type.

Vagueness is modelled in the network by the introduction of nodes linked to (at least two) other nodes by unlabelled arcs. Discourse or grammatical context alone will be responsible for those transitions.

Nodes correspond to kinds of relations between linguistic expressions and situations according to their relationship with time (cf. my definition of aspectual class above). More generally, they correspond to (sets of) semantically defined situation types. The number of nodes, as well as their names, is motivated solely by the language they purport to describe.

Arcs correspond to changes in kind of linguistic expression. These changes always add information to the kind of situation described, even though the information itself may be, for example, that something is missing or that it did not occur. When arcs are labelled by grammatical devices, transitions along those arcs correspond, syntactically, to the use of that device.

In Chapter 4, I will suggest that a particular extra-linguistic context, that of translating into a different language, may be responsible for unlabelled transitions as well.

Paths describe the interpretation of a particular sentence. A path in the network is a sequence of connected arcs. A path starts with a lexical node and ends with a

sentential one. At most one unlabelled arc can precede a labelled arc in a non-marked utterance. Unlabelled arcs model coercion, and are thus licensed by the grammatical operator that labels the following arc in the path. Labelled arcs can only be followed (in other words, belong to a path) if the corresponding grammatical device appears in the sentence. Conversely, all grammatical devices appearing in the sentence must have contributed for the traversal of at least one arc.

The network is used for computing the final aspect of a sentence, starting from a classification of verbs. Arguments change aspect: for example, goal arguments of activity verbs make them accomplishments, as in *run to the store*.³⁸ In principle, all sorts of linguistic clues coming from the noun phrase domain can have import in the aspectual network as well. Explicit quantification, however, may be strictly outside the scope of the aspectual network.

Recursiveness: when clauses function as arguments or temporal modifiers to other clauses, the aspectual network must be used recursively. I therefore suggest arc labels which themselves identify nodes. For example, the label “temporal adverbial”, which might apply to any node representing a (main) clause, is both the name of a node in the aspectual network and the final node of a path coming from a clausal node by an arc labelled e.g. *when*.

3.5 The aspectual network model applied to English and Portuguese

I will not discuss the whole system of any of the languages here, since a significant part will be presented in the remainder of the book anyway. An analysis of English or Portuguese grammar is better left for a book on each language respectively; I expect the reader will be more interested in finding out how the particular analyses were arrived at with the help of corpus studies (the subject of Chapter 5) than in the analyses themselves.

Still, I will try to give some examples of the (redesigned) model’s power and identify some of its strengths and weaknesses. I will thus present one full aspectual network for each language and discuss some particular cases.

3.5.1 An aspectual network for Portuguese

There are a lot of grammatical devices that are related to tense and aspect in Portuguese, resulting in quite a complex graphical presentation of an aspectual network. Nevertheless, I present one in Figure 3.3, and then go on to show how some selected properties of the Portuguese language are rendered in the aspectual model, redrawing subparts of it for readability.

³⁸ It is conceivable that, as more complex expressions containing a verb are built, aspect values can occur which are not available at the lexical level. Examples of these non-lexical aspect values in English may well be habituality and plurality (brought about by plural arguments or tenses). A distinction between lexical and non-lexical nodes, however, did not seem to be an essential property of the model.

Let me explain the different values of the Progressiva (*estar a*) by showing increasingly complex paths and the way they add to the meaning of the other components. The Progressiva is modelled thus:

estar a: OBRA ---> ESTADO TEMPORÁRIO

estar a: SÉRIE ---> ESTADO TEMPORÁRIO

In other words, the Progressiva creates a Estado out of an Obra. The semantic operation relative to the Progressiva is the description of a temporary state that holds during the time “occupied by an event.” Additionally, the Progressiva does the same to a sequence of events (Série).

Progressiva is marked with Mudanças. It can either add the connotation of graduality, or the imminence interpretation (on the verge of). This is modelled as coercion from Mudança to Obra, prior to the import of the Progressiva proper. A sentence like *Ele está a ficar velho* (‘He is getting old’) is thus modelled by the following path:

MUDANÇA (ficar velho) @> OBRA estar a ESTADO TEMPORÁRIO

The original Mudança was coerced to a process that takes time (an Obra), and the sentence states that ‘he’ is in the middle of such process.

Figure 3.4 shows the relevant arcs and nodes in this connection:

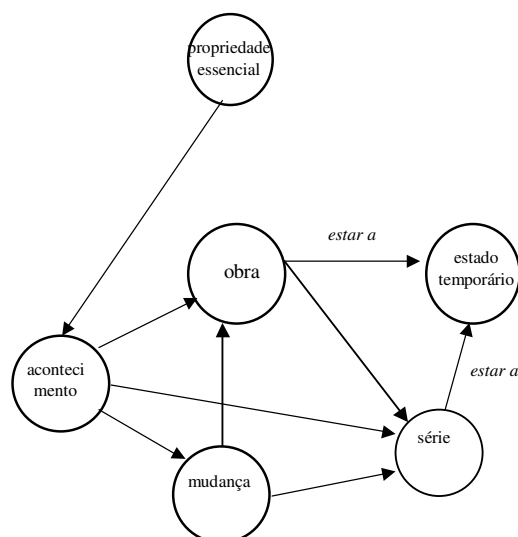


Figure 3.4: Zoom into the Progressiva in the Portuguese aspectual network

Likewise, in *As visitas estão a chegar e o almoço ainda está tão atrasado!* (‘The guests are coming and the lunch is still not ready!’), the Mudança relative to the arrival is expressed as an Obra which is already in progress (the guests are already

in the middle of a process which will make them arrive), i.e. we also have a coercion from *Mudança* to *Obra*.

MUDANÇA (chegar) $\xrightarrow{e>}$ OBRA $\xrightarrow{\text{estar } a}$ ESTADO TEMPORÁRIO

Alternatively, one could interpret the sentence as a *Série* (a series of arrivals), and have instead:

MUDANÇA (chegar) $\xrightarrow{e>}$ SÉRIE $\xrightarrow{\text{estar } a}$ ESTADO TEMPORÁRIO

Finally, one of the most marked uses of the *Progressiva*, applying to *Propriedades essenciais* (a subgroup of *Qualidades*), as in cases like *Ele está a ser simpático* ('He's being nice'), can be modelled by two coercions in sequence, from *Propriedade essencial* to event and then either to *Obra* or to *Série*:

PROPRIEDADE ESSENCIAL (ser simpático) $\xrightarrow{e>}$ ACONTECIMENTO $\xrightarrow{e>}$ OBRA $\xrightarrow{\text{estar } a}$ ESTADO TEMPORÁRIO

PROPRIEDADE ESSENCIAL (ser simpático) $\xrightarrow{e>}$ ACONTECIMENTO $\xrightarrow{e>}$ SÉRIE $\xrightarrow{\text{estar } a}$ ESTADO TEMPORÁRIO

The semantic import of the *Progressiva* in this case is the replacement of the property by an event which is itself endowed with such a property. This event is then either coerced in turn into an *Obra* (i.e., is viewed as taking time) in order for the *Progressiva* to be successfully interpreted, or one has to consider a sequence of events of being nice; the *Progressiva* places us then into the middle of such a sequence.

The fact that two unlabelled transitions have to be followed shows how marked this use of the *Progressiva* is.

Let me now show the way I have modelled the import of the three past tenses mentioned in Table 3.1, and follow some example transitions.³⁹ The word *ACONTECIMENTO* ('event') is used as a shorthand for both *Obras* and *Mudanças*, meaning that the operator would not change the aspectual type in such cases.

Perfeito: ACONTECIMENTO ---> ACONTECIMENTO PASSADO

Perfeito: ESTADO ---> ESTADO PASSADO

Imperfeito: ESTADO ---> ESTADO PASSADO

Imperfeito: SÉRIE ---> QUALIDADE PASSADA

Imperfeito: QUALIDADE ---> QUALIDADE PASSADA

PPC: ESTADO ---> ESTADO EM XN

PPC: SÉRIE ---> ESTADO EM XN

These examples show that the same operator can have different basic import for different input categories. In other words, there is no requirement of one single arc with each label. This means that it is equally basic (non marked) to apply *Imperfeito* to *Estados* than it is to apply it to *Qualidades*. What is interesting

³⁹ The words *passado/a* ('past') or *em XN* ('in the extended now') are just meant to make explicit that tenses have more import than simply aspectual: they have a temporal localization feature. In the aspectual network I will not add the corresponding tensed categories, nor will I do so in the examples that follow.

is to find out the result of the operators when they apply to input types that do not belong to their basic transitions. For example, if Imperfeito is applied to a *Mudança*, as in example (43), what will happen? The answer is simple, and parallel to the case of PPC applied to a *Mudança* as well, cf. example (44): the *Mudança* is coerced into a *Série*, which will then turn into a *Qualidade* by application of Imperfeito, or into a *Estado* by using PPC.

(43) *Ele arranjava relógios*

‘He fixed clocks’; he was a repairman of clocks

(44) *Ele tem lido*

Literally, ‘he has read’; conveying ‘he has been reading’

The paths involved in the aspectual network are (43p) and (44p):

(43p) OBRA (arranjar relógios) $\xrightarrow{\text{Imperfeito}}$ SÉRIE $\xrightarrow{\text{Imperfeito}}$ QUALIDADE

(44p) OBRA (ler) $\xrightarrow{\text{PPC}}$ SÉRIE $\xrightarrow{\text{PPC}}$ ESTADO

On the other hand, if *Perfeito* is applied to a *Qualidade*, it will have to be coerced into an event or series of events in order to be felicitous, as in example (45); or into a temporary state by other means of reasoning, as in example (46). Figure 3.5 shows the relevant arcs and nodes in this connection.

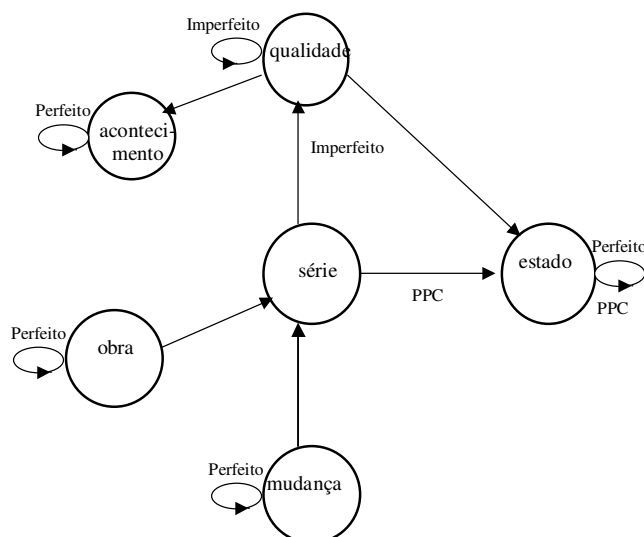


Figure 3.5: Zoom into the past tenses in the Portuguese aspectual network

(45) *Ele foi generoso*

He was generous; he acted generously

(45p) QUALIDADE (ser generoso) $\xrightarrow{\text{Perfeito}}$ OBRA $\xrightarrow{\text{Perfeito}}$ OBRA

- (46) *Ele foi professor na Universidade da Beira Interior*
He was a teacher at UBI; he has taught there

(46p) QUALIDADE (ser professor) $\xrightarrow{à}$ ESTADO Perfeito ESTADO

Note that the acceptance of these sentences requires a more complicated processing by the hearer (and therefore also more information gets to be transmitted), and this is mirrored by greater complexity in the network model.

Finally, temporal adverbials, which, following Sandström, I take to denote a time (a part of the time line), are handled by allowing them both coercion capabilities (by requiring to be applied to a particular kind of “object”/sentence) and transforming capabilities (by their presence transforming the final output into something else).

à 5ª feira: ACONTECIMENTO ---> QUALIDADE

durante: OBRA ---> OBRA

durante: SÉRIE ---> SÉRIE

durante: ESTADO ---> ESTADO

An interesting case is the result of applying *durante* to a *Mudança*: if felicitous, it measures the duration of the result state, coercing thus the *Mudança* into a *Estado*, cf. (47) and (47p).

- (47) *Ele saiu durante uma hora*
He left for an hour; he was away for an hour

(47p) MUDANÇA (sair) $\xrightarrow{à}$ ESTADO durante ESTADO Perfeito ESTADO

Figure 3.6 shows the relevant arcs and nodes in this connection:

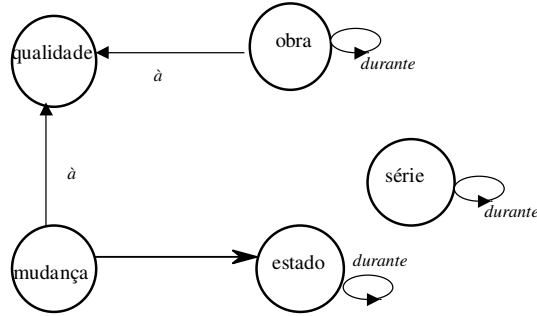


Figure 3.6: Zoom into *durante* and *à* in the Portuguese aspectual network

Also following Sandström in clearly distinguishing between the import of temporal connectives to the subordinate and to the main clause, I distinguish the import of *desde* in the *desde*-clause, and the import of the *desde*-clause (*oração-desde*) to the main clause:

desde: MUDANÇA ---> ESTADO
 desde: ESTADO ---> ESTADO
 oração-desde: ESTADO ---> ESTADO

Whether the subordinated clause contains a *Mudança* (in which case the beginning of the *Estado* is marked by it) or already an *Estado* (in that case co-temporal with the *Estado* described by the *desde*-clause), the result of the application of the *desde*-clause to another (main?⁴⁰) clause forces this one to be interpreted as an *Estado* as well.

- (48) *Desde que ele caiu (que) anda de muletas*
 Ever since he fell down he walks on crutches

(48p) MUDANÇA (cair) Perfeito MUDANÇA desde ESTADO
 QUALIDADE (anda de muletas) @> ESTADO oração-desde ("desde que ele
 caiu") ESTADO

This is an example of using recursiveness: the *oração-desde* functions as operator in the second line of (48p), but it was “composed” also in the aspectual network in the first line of (48p). Figure 3.7 hopefully helps visualizing this.

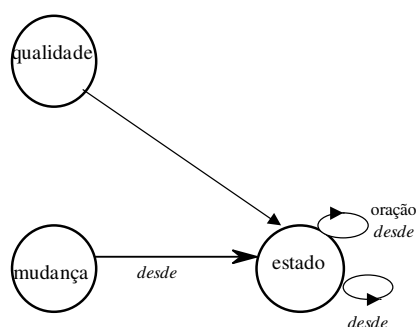


Figure 3.7: Zoom into *desde* in the Portuguese aspectual network

This ends a first presentation of the aspectual network for Portuguese. The reader will still be exposed to more – and authentic – examples and my attempt to make sense of them contrastively. There are a few details (as well as further classes and kinds of operators) which I have not discussed here. They should not hinder the understanding of the points I want to make in this book, though.

⁴⁰ The fact that, in most cases – see (48), a *que* must be inserted, with corresponding changes in clitic order, etc. goes against an analysis of main and subordinated clause for *desde*. As far as I know, this is the only temporal connective in Portuguese in which this phenomenon occurs.

3.5.2 An aspectual network for English

For English it is easier to present the whole system, not only because people is in general quite familiar with Vendler's and Moens's work but because it is not necessary to explain or translate literally the examples.

I will start by providing the full picture, and then turn to the places where I deviate from the norm in analysing English. I am mainly interested in explaining how the vague classes work in the model (I ignored this subject in the Portuguese section for the sake of simplicity).

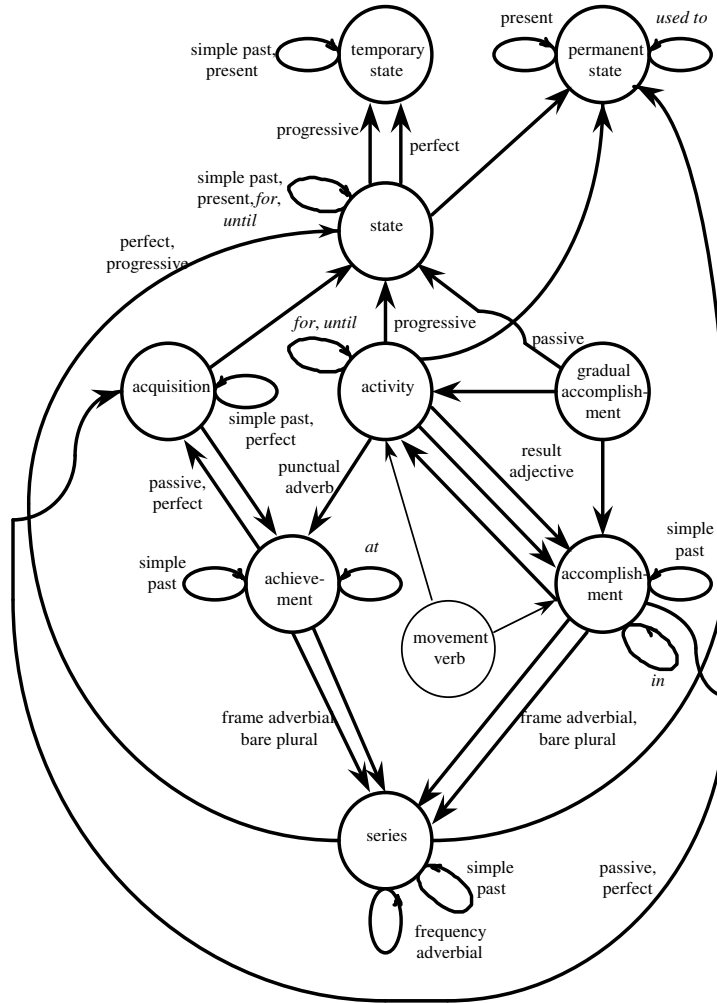


Figure 3.8: The English aspectual network

I use the same label (the same kind of state), temporary states, for the result of the progressive and of the perfect in English (even though they correspond to different parts of an event, they have never, as far as I know, shown to have different aspectual properties). Even though this is not essential, I keep a node permanent states in English as well.

Then I introduce an aspectual class which I call series, and which I have borrowed from Freed's (1979) description of English aspectualizers, cf.: "A series is an aspectual verb-type which has the temporal characteristics of either an activity, an accomplishment, or an achievement (or some states), which is considered to occur repeatedly or habitually" (Freed, 1979: 53). In my view, a series can mean both an individual change and a set of them, therefore merging Moens's points, Kent's repeated accomplishments, and Freed's series.

This inherently vague category – which, in my opinion, is unrelated to quantification – is necessary to describe the output category of aspectualizers like *keep*, and one use of the progressive; cf. *He kept bothering us* and *Judith is eating sandwiches*, as well as the input category for aspectualizers like *quit*, cf. *he quit betting on horses*.

I then use two other vague categories, which can be identified by having unlabelled arcs into the two possible non-vague alternatives, namely a) acquisitions – that have unlabelled arcs to both states and achievements; b) gradual accomplishments – that have unlabelled arcs to both accomplishments and activities. A precedent to gradual accomplishments in the literature are Dillon's (1977) "degree-inchoatives" and "gradual achievements". As far as acquisitions are concerned, namely expressions vague between a change of state and the state itself, they will be copiously described and used in the rest of this book. To my knowledge they were first proposed as a class in Santos (1991), although recognized as having special properties by Dowty (1979) and Heinämäki (1984).

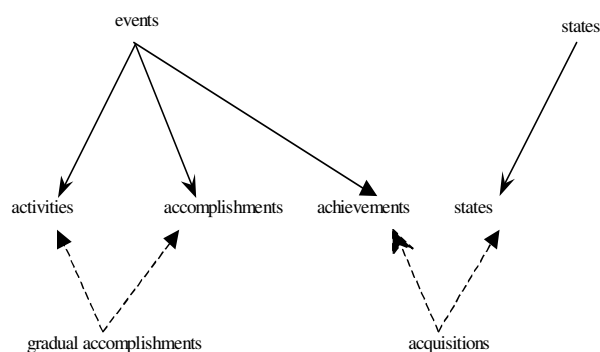


Figure 3.9: The English aspectual classes

Figure 3.9 provides a structural description of the classes in the English system, so that it may be easier to grasp them. Dotted lines mean “vague between”, solid ones correspond to the usual set inclusion.

To end the presentation of the English aspectual network, and given that a large amount of examples in the original Moens’s network has already been given in Section 3.2.1, I will use one single example, namely, the case of *for* adverbials. It has been chosen since I use a different analysis from that of Moens: Moens claimed that a *for*-phrase transforms the output into an accomplishment, while Sandström (1990: 66) noted that sentences with *for* do not combine with *when* (as “real” accomplishments do).

for: ACTIVITY ---> ACTIVITY
 for: SERIES ---> SERIES
 for: STATE ---> STATE

In fact, the difference between *walk* and *walk for two hours* is smaller than the difference between *walk* and *build a house*, for example, and aspectually *walk* and *walk for two hours* seem to have the same properties: accepting the progressive, not accepting *in*-adverbials, sounding better with the perfect progressive than with the simple perfect. In order for another aspectual class to be felicitous with *for*, it has to be constrained to either a series, an activity or a state, as illustrated respectively by examples (51), (50) and (49):

(49) *He disappeared for a long time.*

(49p) ACHIEVEMENT (disappear) @> STATE for STATE

(50) *He built the house for a while*

(50p) ACCOMPLISHMENT (build) @> ACTIVITY for ACTIVITY

(51) *He gave that particular opening speech for years*

(51p) ACCOMPLISHMENT (give a speech) @> SERIES for SERIES

This closes my presentation of the tense and aspect model that I will use, and on which I have based a new model to do contrastive studies. This new model will be presented in the next chapter.

Chapter 4

The translation network

In this chapter, I will present a model for the description of translation and translations, termed the “translation network”. I will begin by suggesting a set of basic requirements for such a model. Then, I will present a specific instantiation in the domain of tense and aspect, which will be used to cast the contrastive analyses of the rest of the book. I will illustrate its general properties in some detail by referring to several examples.

4.1 Basic issues

As discussed in the previous chapters, I believe it is necessary to describe the two languages on their own terms – and thus not resort to common categories a priori; even though the possibility that there are shared categories/meanings is not ruled out. Rather, it becomes an empirical question.

Translation is seen as establishing a mapping from the categories specific to the source language into the categories of the target language, subject to the specific constraints of the latter. Metaphorically speaking, translation is like viewing a source text with target language eyes.

What I will try to argue with the help of this model – in the present chapter and in the book as a whole – is that, compared to what a source language native speaker sees,

- a translator can see more
- a translator can see less

In other words, the mapping in general is not functional.

Anticipating a little, some cases of seeing more will be called “coercion brought about by translation”, borrowing from Moens (1987). I believe that the need for translation (or for understanding the original text when one “thinks” in the target language; cf. Slobins’s (1987) “thinking for speaking”) coerces the meaning of the original text to something which might not be there in the first place.

In other cases, a different situation holds: the potentialities of the target language lead the translator to express information that is commonly expressed, just to sound natural or produce a clearer formulation. In this case, one would not say that the translator interpreted the text in that way in the first place, but rather that he embellished it for the target language’s sake.

Of course, it is necessary to be cautious about the question of “what is there” really means. Though for practical purposes we contrast two sentences or two sequences of sentences (both while analysing the translation **and** while

translating) we must not forget that the translator may draw on information he has read in virtually any part of the book to be translated, so that addition of information may be just local, not global. With this in mind, one may still find it useful, when contrasting two languages, to observe the cases where local addition or removal of information has taken place.

Another even more fundamental question concerns the objectivity of what is there in the first place. There is no reality without experiencer, and there is no text without an interpreter; cf. Steiner (1992). What “really” is there is thus not objectively measurable: one’s analyses necessarily involve a certain amount of subjectivity. Still, the detailed comparison of two persons’ subjectivity – the translator’s and mine, as an approximation to language intersubjectivity, may be worth while to pursue and formalize.

4.2 Abstract properties of translation

For the translation network, I assume that an appropriate language model can be defined as a set of categories which are linked, working as the input and output of a set of operators. This is a reasonable model for tense and aspect, as the work by Moens (1987) and his followers shows, cf. Chapter 3.

A model for the translation between two languages will consequently include two such language models, one for the source language and one for the target language. Additionally, translation from the source to the target language must induce some mapping between the categories (and possibly between the operators as well⁴¹).

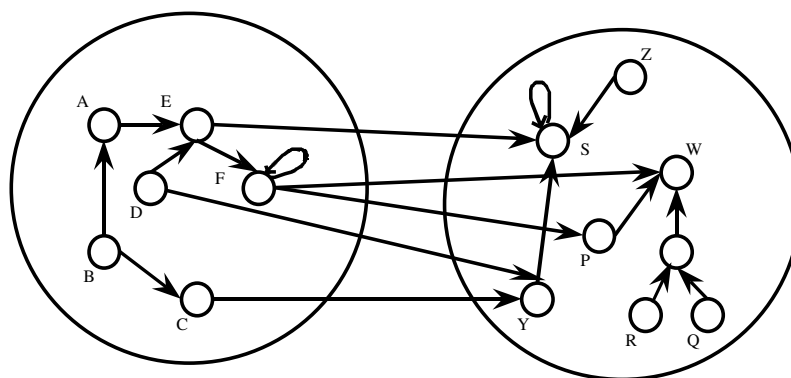


Figure 4.1: An abstract Translation Network

⁴¹ It is up to the analyst to describe the categories in the language as including all meaning (operators just being transitions, marked or unmarked, between the categories) or leave some meaning to be expressed by the transition. For the present discussion I assume the first alternative, but I will come back to this subject later on.

A simple example is depicted in Figure 4.1. (Small circles represent “nodes”, and stand for categories, arrows depict “arcs” and stand for transitions. All nodes are labelled; arcs may be labelled, but not necessarily. The two larger circles enclose source and target language, arcs from the former to the latter are called “translational arcs”).

Some of the points illustrated in the figure are worth noting:

- node names are different (there is no common metalanguage)
- there are source nodes which have no correspondence in the target network (namely A and B)
- there are target nodes which do not correspond to any source node (such as Z, R and Q)
- the same source node may correspond to more than one target node (cf. F)
- the same target node can correspond to more than one source node (e.g. Y)

Now, in order to make use of such a model, one needs to be able to analyse a source text as a path in the source translation network, a target text as a path in the target translation network, and decide which correspondences hold (i.e., which translational arcs are involved).

In this connection, note that, in this model of translation, there is no requirement for path isomorphism of any kind. The paths in the two languages can have different lengths, and visit the same or different nodes (i.e., have loops or not). Furthermore, not all nodes need be linked by translational arcs.

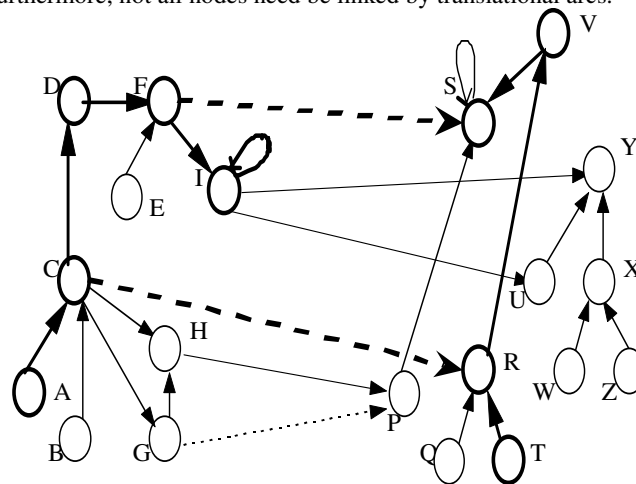


Figure 4.2: The description of a translation pair (in bold) in a translation network

Figure 4.2 illustrates these points with another fictitious translation network, where the translation pair is marked as bold in the two monolingual networks.

While the source path (A C D E I I) includes six nodes (five different ones) the target path only contains four (T R V S). Also, only the second and the fourth nodes of the source path are translationally related to some node in the target language model. Finally, note that, **in this particular translation pair**, the last node in the source path has no correspondence in the target language.

An interesting consideration is thus in order here: Even though there may be (translationally) related categories in the two languages, they may not be (monolingually) connected in the same way, so that it is not necessarily possible that “corresponding” lexical items (path beginnings) can arrive at “corresponding” final nodes. So, in the example of Figure 4.2, even though nodes I and Y are translationally related in the system, there is no path from V or S to Y in the target language network.

Another question is what Catford (1967) has termed “level shift” (namely that a particular shade of meaning may be conveyed by lexical or grammatical means in different languages). This can be expressed in the translation network model simply by an extra initial path, as illustrated in Figure 4.3 (cf. my remarks about lexicon vs. grammar in Sections 4.3.5 and 5.11, though).

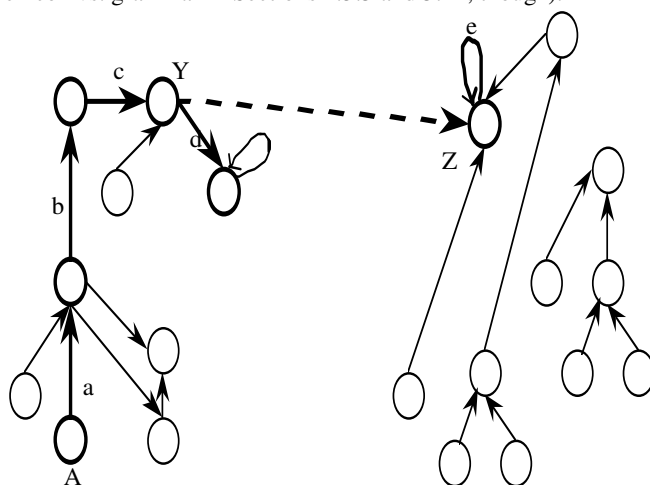


Figure 4.3: Modelling lexical vs. grammatical import in a specific translation in a generic translation network

While Y and Z are translationally related, Z is lexically expressed in the target language, but was grammatically built in the source language (starting with lexical category A, and transformed by operators a, b and c).

Finally, an important feature of these language models, which will also play a key role in translation, is vagueness. As described in Chapter 3, vagueness is modelled by the existence of unlabelled arcs (i.e., no overt grammatical operators) from a node to two or more nodes, and it is the context which makes

the choice (in case such choice is required). We say that the expression is **coerced** into that category. In the translation model I suggest, if there are no corresponding vague categories in the target language and no choice is required by the source text, translation implies a choice which always corresponds to a loss of information.

The notation to describe translation pairs is a sequence of node names intertwined with arc names. When an arc has no label, @> denotes coercion and T> coercion by translation.

4.3 Concrete translation networks

4.3.1 Coercion by translation

To illustrate coercion by translation, let us turn to the concrete example of English states, which, in a contrastive perspective, are vague between denoting an essential property and an accidental state. There is no way (most of the time) to preserve such vagueness in Portuguese, so translation will have to decide which interpretation to render. This is a case of contrastive vagueness, and so coercion (the need to take an unlabelled arc) is brought about by translation, and that is what the T's stand for in Figure 4.4.

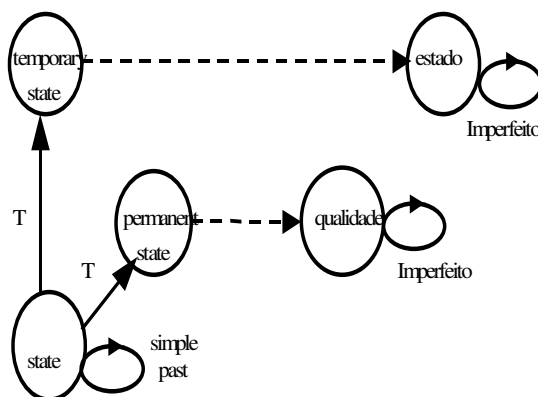


Figure 4.4: Illustrating coercion by translation, in the translation of English states into Portuguese

Let me note, first, that what we see here are **only** the relevant part(s) of a complete translation network that would describe the translation from English to Portuguese, and that the monolingual paths accepted by the two monolingual networks are absolutely minimal, simply taking a lexical item belonging to the STATE class (or to the QUALIDADE or ESTADO classes) and leaving it in the same class after the application of the simple past tense (or of the Imperfeito tense).

Still, this simple case is enough to show the difficulties posed by contrastive vagueness and to illustrate the way the model works. Consider translation pair (1) and its alternative translation (1’):

- (1) *Beside him on a table **was** a small Oriental gong and a bowl of cigarettes.*
*ao lado, na banca de cabeceira, **havia** um pequeno tantã oriental e um*
maço de cigarros.

beside, on the bedtable, **there was** a small oriental gong and a cigarette pack

- (1’) *ao lado, na banca de cabeceira, **estava** um pequeno tantã oriental e um*
maço de cigarros.

The different choices can be modelled by different pairs of paths (the subscript ‘p’ stands for ‘pair of paths’), namely:

- (1p) STATE (be on) simple past STATE T> PERMANENT STATE
 QUALIDADE (have) Imperfeito QUALIDADE
- (1’p) STATE (be on) simple past STATE T> TEMPORARY STATE
 ESTADO (estar) Imperfeito ESTADO

Consider now example (2), which is more involved (though not from the point of view of English):⁴²

- (2) *And he drank a little pulque and that was breakfast*
Bebeu um pouco de pulque. E foi o seu pequeno almoço.

He drank a little of pulque. And it was his breakfast.

- (2’) *bebeu um pouco de pulque e isto era o seu pequeno almoço*

Its alternative translation (2’) follows, in fact, the same path as (1) above, namely

- (2’p) STATE (it be breakfast) simple past STATE T> PERMANENT STATE
 QUALIDADE (ser) Imperfeito QUALIDADE

The relevance of these examples is that both alternatives are “correct” translations but presuppose a vagueness resolution that is outside what is conveyed by the English text proper.

In fact, *that was breakfast* in the English original could just as well refer to that particular instance of breakfast, or to the habitual sequence of events that constituted breakfast for ‘him’, or to **both**. Note, furthermore, that if ‘he’ had had a typical instance of breakfast at the time of the narrative, the two interpretations could not even be objectively distinguished.

The same is true about the position of the Oriental gong in example (1), although not leading to such striking meaning differences. Does the English sentence refer to the gong’s position, immutable since long ago, or to its being there then, because e.g. the maid or the doctor had taken it there? One may well

⁴² The curious reader may like to know that its analysis is like that of (45) in Chapter 3.

say that the English text does not lead us to pose this kind of questions, but each of the translations convey one of these meanings.

As a reader, you may say, what does it matter? In other words, it is undeniable that many details are not functionally relevant (using Catford's (1967) terminology). But the translator must decide, in each case, which formulation to use, and he must also decide whether the "details" are relevant or not.

Returning to example (2), given that 'his' eating breakfast is not relevant to the rest of the story, one could point out that the author was describing the character's habitual actions to show how poor he was, and that the particular eating instance in itself was not part of the main sequence of events. One could thus claim that, given a target language where the difference between description and action is linguistically marked, the second (the alternative) translation would be preferable. However, a defender of the first (the published) translation is entitled to argue that replacing pure description with explanation actually changes the narrator's position and is therefore unacceptable.⁴³

The above discussion gives, I hope, a flavour of the many subtleties that can be involved in the translation of instances which may be trivial from a source language point of view, but which require a choice involving many factors when translating, and which may lead to far-reaching consequences.

4.3.2 The translation network does not force one analysis

Let me show more examples of (partial) translation networks where other features of this descriptive model can be appreciated. For this purpose, I extend the previous domain by including the English passive, as depicted in Figure 4.5.

I want to show that this particular partial translation network allows us to account for many *a priori* unrelated and relatively intricate cases of translation, such as the ones in examples (3) to (6).

(3) *and the canoes were drawn up in front of this area.*

e era em frente dessa zona que os barcos se alinhavam.

and it was in front of that zone that the boats stood aligned

(3p) EVENT (draw up the canoes) passive ACQUISITION @> STATE simple past
STATE T> PERMANENT STATE
OBRA (os barcos alinharem-se) @> SÉRIE Imperfeito QUALIDADE

My analysis of example (3), (3p), interprets the English sentence as describing a state of position, and the Portuguese sentence as describing a habitual situation (location).

⁴³ In a more radical form, it would be like adding an explicit moral to a parable.

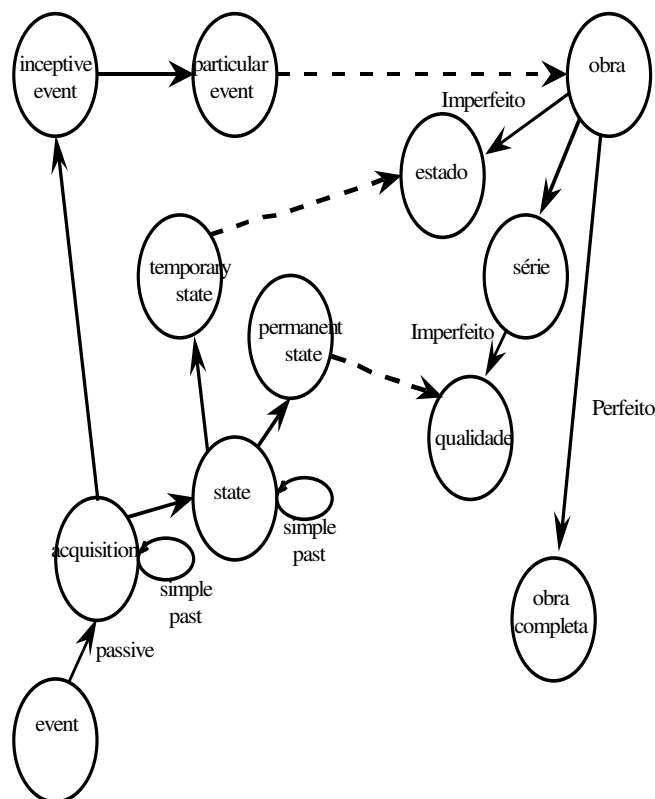


Figure 4.5: Translation model for the English passive

In order for the reader to be able to decide for himself which interpretation(s) is appropriate, I provide a larger context:

(...) Part of the far shore disappeared into a shimmer that looked like water. There was no certainty in seeing, no proof that what you saw was there or was not there. And the people of the Gulf expected all places were that way, and it was not strange to them. A copper haze hung over the water, and the hot morning sun beat on it and made it vibrate blindingly.

Os confins da praia desapareciam numa luz vacilante que parecia água. Tudo se esbatia e ninguém poderia dizer se o que via existia ou não. E a gente do Golfo olhava para tudo, deste modo. Havia uma névoa de cobre sobre a água, na qual o quente sol da manhã incidia, dando-lhe um brilho que cegava.

The brush houses of the fishing people were back from the beach on the right-hand side of the town, **and the canoes were drawn up in front of this area.**

Kino and Juana came slowly down to the beach and to Kino's canoe, which was the one thing of value he owned in the world. (19-20)

As cabanas dos pescadores ficavam por detrás da praia, à direita da cidade, **e era em frente dessa zona que os barcos se alinhavam.**

Kino e Joana desceram vagarosamente até à praia e aproximaram-se do barco de Kino, que era a única coisa de valor que ele tinha no mundo. (24-25)

It is important to note, however, that the particular translation network presented does not **force** the interpretation upon us. On the contrary, I will show in the following that it allows several alternative paths corresponding to precisely the same wording:

1. In the English original, one may claim that there are no grounds to choose a state interpretation, and the English monolingual analysis stops at the acquisition node (the rest is coerced by the translation), corresponding thus to:

(3p₁) EVENT (draw up the canoes) passive ACQUISITION simple past
 ACQUISITION $\overline{T}$ STATE $\overline{T}$ PERMANENT STATE
 OBRA (os barcos alinharem-se) $\overline{e}$ SÉRIE Imperfeito QUALIDADE

2. Yet another interpretation, which I find inappropriate for extra-linguistic reasons, is the one in which the English sentence describes an action done in the past by some unspecified character (they had drawn the canoes there). This would correspond to the following paths (note the different categorization of *alinhar*):

(3p₂) EVENT (draw up the canoes) passive ACQUISITION simple past
 ACQUISITION $\overline{T}$ INCEPTIVE EVENT $\overline{T}$ PARTICULAR EVENT
 OBRA (alinhar os barcos) $\overline{e}$ SÉRIE Imperfeito QUALIDADE

Such a situation accounts well for the analysis of the English original in e.g. examples (4) and (5), though:

- (4) *for the hundreds of years of subjugation **were cut** deep in him.*
*Porque centenas de anos de escravidão **tinham cavado** fundo nele.*
 Because hundreds of years of slavery had dug deep in him.
- (5) *The songs remained; Kino knew them, but no new songs **were added***
*as canções tinham ficado. Kino sabia-as, mas não lhes **tinha juntado***
nenhuma

the songs had remained. Kino knew them, but he had not added any

The subtle change in meaning produced by the voice reversal – incidentally, quite a frequent translation regularity – is left to the reader's appreciation. I present here the analysis of (4) only:

(4p) EVENT (cut years of subj. deep in him) passive ACQUISITION $\overline{T}$
 INCEPTIVE EVENT
 OBRA (cavar fundo nele) Mais que Perfeito OBRA COMPLETA

3. As for the interpretation of the Portuguese translation in (3), there is also another possibility, corresponding to coercion to another node in the English side, namely:

(3p₃) EVENT (draw up the canoes) passive ACQUISITION @> STATE T>
 TEMPORARY STATE
 OBRA (os barcos alinharem-se) Imperfeito ESTADO

This translation can be paraphrased (and therefore could have been expressed) by *que os barcos estavam alinhados* (Imperfeito progressivo), and this is the reason why I do not prefer this interpretation. Note, though, that this is a case where, theoretically, the vagueness is preservable, since the two paths corresponding to choice in the source language, (3p₁) and (3p₃), can be translated by the same device in the target language (the Imperfeito). However, Portuguese speakers tend to select a more specific meaning, due to their habits of “thinking for speaking”.⁴⁴

4. Example (3) is, in fact, particularly interesting because it can be analysed in yet another way, this time not covered by the network of Figure 4.5, namely: What if *alinhar-se* was an Aquisição (instead of an Obra)?⁴⁵ Then, we would have a relatively parallel path in the two languages, displayed in Figure 4.6, and no decision about kind of state in English would be necessary.

The new path would then be:

(3p₄) EVENT (draw up the canoes) passive ACQUISITION simple past
 ACQUISITION @> STATE
 AQUISIÇÃO (os barcos alinharem-se) @> ESTADO Imperfeito ESTADO

Thus we have here an example of translation introducing vagueness (or ambiguity – this question will be discussed below), because this translation network (and the translation network model in general) allows more than one alternative path in the target network for the same source path. That is, a non-vague source text may be translated by a vague (or ambiguous) target text. Furthermore, this example shows that vagueness introduction does not occur only in cases where a vague category of the target language is at stake: one may simply have alternative paths that arise from lexical vagueness (or ambiguity).

⁴⁴ At least it would be fairly common to interpret the Portuguese rendering of (3) as meaning “in general”, and not at that particular moment only.

⁴⁵ The uncertainty (vagueness?) between Aquisição and Obra in this case is closely related to the interpretation of the plural subject NP, i.e., the distributive and group readings go hand in hand with either aspectual classification. The alternatives are:

- each canoe stood in the line (distributive NP, Obra)
- the set of canoes formed a line (group NP, Aquisição)

Again, most probably the author (and in this case also the translator) meant **both**. Note furthermore that there are also several alternatives available for *se*.

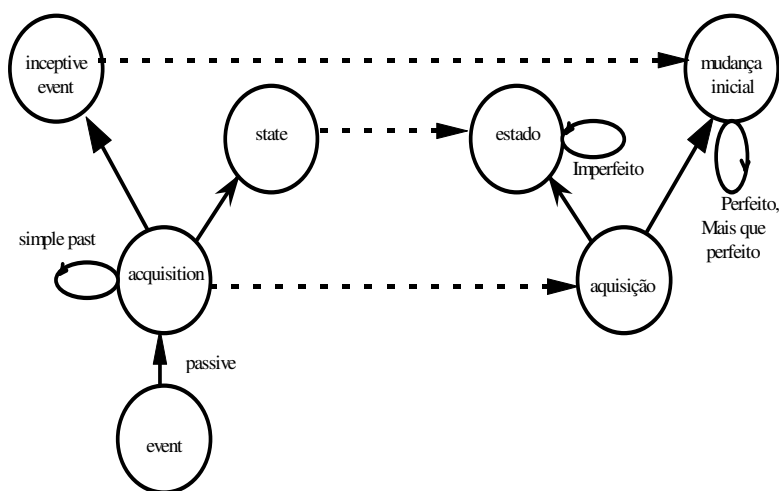


Figure 4.6: Translation network dealing with acquisitions and Aquisições

An impatient reader might say that the translation network in Figure 4.5 was not needed after all, given that the one in Figure 4.6 might be enough to deal with example (3). I must stress that the translation network in Figure 4.5 was created after observation of very many different translation pairs, and not only – or especially – to cater for example (3). I will show some examples – (6) and (7) – where passive is translated by Imperfeito with a habitual interpretation, providing, for purposes of comparison, alternative translations (6')-(7') with a particular event reading:

- (6) *The gathering procession was solemn, for they sensed the importance of this day, and any children who showed a tendency to scuffle, to scream, to cry out, to steal hats and rumple hair, **were hissed** to silence by the elders. Era um cortejo compacto e solene. Todos sentiam a importância do dia. Se qualquer criança tentava brigar, gritar, chorar, roubar chapéus ou puxar os cabelos, logo os irmãos mais velhos a **obrigavam** a estar quieta.*
if some child tried to fight, ... at once the elder brothers **forced** it to be still
- (6') (...) e todas as crianças agitadas (...) **foram silenciadas** pelos mais velhos com assobios.
and all children who showed excitement were kept silent by the hisses of the elders
- (7) *Thus Kino's future was real, but having set it up, other forces **were set up** to destroy it, and this he knew, so that he had to prepare to meet the attack. Assim, o futuro de Kino era real; mas mal o construía, logo outras forças*

se formavam para o destruir. Kino sabia-o bem.

but as soon as he built it, at once other forces **formed** to destroy it

- (7') (...) *mas, tendo-o construído, outras forças **apareceram** para o destruir*
(...)

but having built it, other forces **turned** up to destroy it

What is important with examples (6) and (7), whose analyses involve paths similar to (3p₁), is that neither *obrigar* nor *formar-se* can be interpreted as Aquisições in the particular contexts at stake, and therefore an option like (3p₄) is not available.

4.3.3 Introduction of vagueness or ambiguity by translation

Let us consider example (8) to have another look at vagueness or ambiguity in connection with (the interpretation of) translation:

- (8) *And the rhythm of the family song was the grinding stone where Juana **worked** the corn for the morning cakes*
*o ritmo dessa música familiar vinha da mó com que Joana **preparava** o milho para a refeição da manhã.*

the rhythm of that family song came from the grinding stone with which Juana **prepared** the corn for the morning meal

Here, we can identify the past form *worked* as describing either a particular occurrence or a general property of the grinding stone.⁴⁶ The first meaning would be uniquely conveyed by the progressive *was working*, but the simple form is vague between the two (which actually can be true simultaneously).

In order to analyse this case, we have, again, two paths induced by translation – and no possibility to avoid coercion – in the network of Figure 4.7:

- (8p₁) EVENT (Juana work the corn) simple past EVENT T> HABITUAL STATE
OBRA (Joana preparar o milho) @> SÉRIE Imperfeito QUALIDADE
- (8p₂) EVENT (Juana work the corn) simple past EVENT T> PARTICULAR EVENT
OBRA (Joana preparar o milho) Imperfeito ESTADO⁴⁷

Although the translation network of Figure 4.7 represents the general import of the English simple past which, if applied to acquisitions, is three times vague regarding Portuguese (i.e. there are three different nodes that can be chosen for coercion by translation), *work* is no acquisition. Therefore, we have in

⁴⁶ The latter interpretation would also mean that the *where*-relative clause was necessarily restrictive, while the former interpretation allows both kinds of relative clauses.

⁴⁷ This second alternative would be mandatory in case the Portuguese text had the Progressiva:

- (8') *o ritmo dessa música familiar vinha da mó com que Joana **estava a preparar** o milho para a refeição da manhã.*

example (8) only a twofold vagueness, modelled by the paths from EVENT to HABITUAL STATE OR PARTICULAR EVENT.

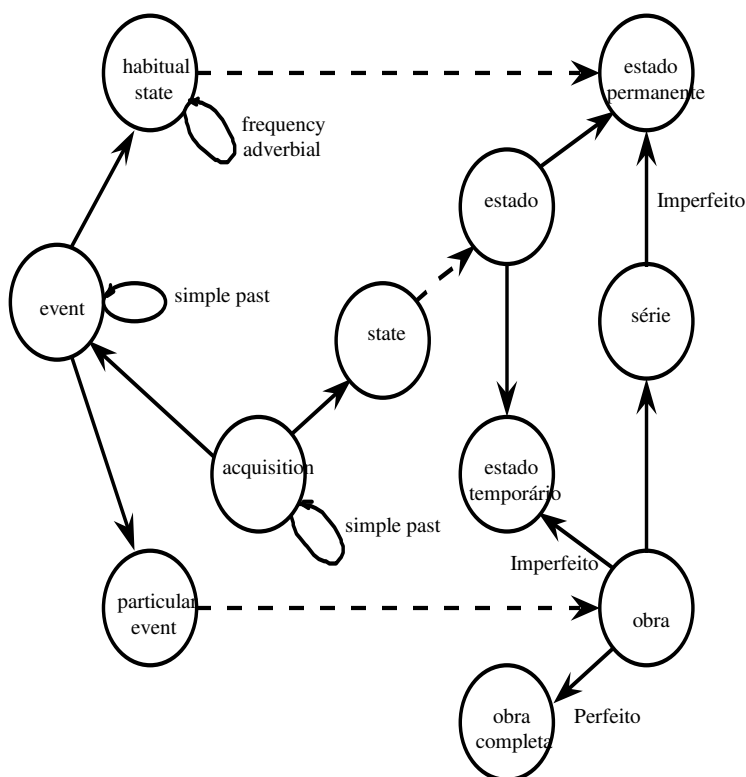


Figure 4.7: A translation network depicting simple past

In this case, we may have a case of vagueness preserving: lexical (contrastive) vagueness (of *work*) preserved by the grammatical vagueness of Imperfeito, or we could have a target language sentence which is ambiguous *tout court*.

Let us pursue the first hypothesis first: Imperfeito can be considered vague between the two meanings, habitual and particular event, and therefore is similar to simple past, which is also vague. However, simple past and Imperfeito are very different in terms of markedness: while it can be said of *worked* that it is by default interpreted at the current time, one can say the opposite about *preparava*, which conveys a property-like character by default.

This can be shown in the difference in strangeness between the two (grammatical) sentences in the two languages which select only the habitual interpretation, in (8[^]) below. The English sentence is rather marked, constituting a sort of temporal garden path, while the Portuguese one is perfectly acceptable:

- (8[^]) And the rhythm of the family song was the grinding stone where Juana **worked** the corn for the morning cakes, which Louisa **was using**.
 o ritmo dessa música familiar vinha da mó com que Joana **preparava** o milho para a refeição da manhã e que a Luísa **estava a usar**.

There are other instances, however, which suggest the (second) hypothesis of ambiguity of Imperfeito as a better analysis. Let us have a look at them:

1. The first example is an ambiguous source sentence in Imperfeito which is mistranslated. I give the alternative translation of example (9) in the gloss:

- (9) *onde entrara em rapazinho quando as paredes da casa ainda se **erguiam**.
 which he had entered as a young lad when the walls of the buildings **were**
 still **being raised**.*

where he entered as a small boy when the house walls **were** still **standing**

Now, the time when you build the walls of a monastery and the time when they (still) stand up are disjoint, so the Portuguese sentence can only mean one of them (namely the second). It is, however, instructive to notice that the English translator was misled precisely because the default in his own language is the eventive reading.

2. The second example, on the other hand, can be described as introducing ambiguity through the use of Imperfeito:

- (10) *They were high and graceful canoes with curving bow and stern and a braced section midships where a mast **could** be stepped to carry a small lateen sail*

*Eram barcos esguios e graciosos, com a proa e a popa curvas e um barrote no meio, onde se **cravava** o mastro para fixar a pequena vela latina.*

They were slim, gracious boats, with curving prow and stern and a beam in the middle, where the mast **was stuck** in in order to fasten the small Latin sail

In a way similar to example (3), the two interpretations of *cravava* correspond to either a stative positional description (stemming from the interpretation of *cravar* as an Aquisição) or to a habitual action (probably done by the boat owners, described by the impersonal *se*). Leaving aside the interpretation of the modal in the source sentence, the two paths in the target language can be described as follows:

(10p₁) MUDANÇA (cravar o mastro) @> SÉRIE Imperfeito QUALIDADE

(10p₂) AQUISIÇÃO (o mastro cravar-se) @> ESTADO Imperfeito ESTADO

The relevant difference between the target parts of examples (10) and (3) is the difference between an *Obra* and a *Mudança*. While (3p) and (10p₁), and (3p₄) and (10p₂), are structurally similar, there is no counterpart to (3p₂) or (3p₃), because the action of *cravar* does not take time, and so one cannot be “inside” it.

Another (striking) difference is, of course, the form of the source English sentence that corresponds to the Portuguese translations. For, while in the case of example (3) we were happy to consider the source language as being most vague between the two interpretations corresponding to the different paths in the target network, regarding example (10) we cannot but agree that at least interpretation (10p₂) – and possibly (10p₁) as well – goes far beyond what is actually stated in the English text: To say that a mast can be stepped does neither mean that it is there all the time, nor that boat owners regularly do it.

While this particular question is irrelevant in the context of the book in question (the way canoes are operated is not relevant to *The Pearl*'s plot), it points to a regularity often found in translation between the two languages (and therefore to a language difference), which can be cast as follows: “While English is very often concerned with possibility, Portuguese concentrates on assigning properties resulting from habitual behaviour”. This will be looked into more closely in Chapter 5, where the translation into Imperfeito, and from Imperfeito, in the corpus will be analysed.

4.3.4 Compact categories

We have so far been concerned with addition of interpretations, that is, addition of information (which may result in a loss). Let us look at the opposite case, where information is left out in the translation.

This happens when one has compact categories in the source language with no equally compact counterpart in the target language: An alternative to creating syntactically more complex sentences with information not functionally relevant is simply choosing the most natural “bits” of information to convey.

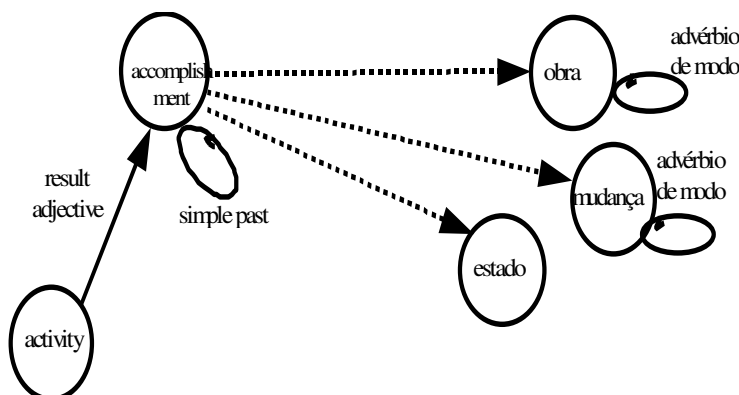


Figure 4.8: Translation network for accomplishments

This is clearly the translation strategy used to render English accomplishments (which express simultaneously an activity and a change of state). Given that there is no corresponding compact class in Portuguese, if the translator does not want to

introduce two clauses instead of one, he has to choose among the three alternatives displayed in the translation network of Figure 4.8: (1) an *Obra* formulation, describing something that takes time, and may or may not have a result, (2) a *Mudança* formulation, concentrating on the change, or even (3) a state formulation, focussing on the end result.

Let me present four examples of these different ways of translating accomplishments. Example (11) shows the possibility to express more than one part of the meaning of the compact expression, contrasting with example (14), which omits the meaning. While the Portuguese manner adverbial in example (11) compensates for what is expressed by the English activity verb, cases like example (14), where the activity of brushing is totally lost, seem to be rather more frequent. (One could, in this case, also wonder whether a translator should repeat a metaphor.) Examples (12) and (13) show translation of accomplishments respectively by a *Estado* and a *Mudança*.

- (11) *and he **pried open** the baby's mouth and poured it down.*

abriu à força a boca da criança e fê-la beber.

he **opened** with force the mouth of the child and made it drink

- (11p) ACTIVITY (pry) adjective ACCOMPLISHM simple past ACCOMPLISHM
MUDANÇA (abrir) advérbio de modo MUDANÇA Perfeito MUDANÇA

- (12) *And his face **grew crafty**.*

*O rosto **tinha** agora uma expressão de astúcia.*

The face **had** now an expression of astuteness.

- (12p) ACTIVITY (grow) adjective ACCOMPL simple past ACCOMPL
ESTADO (ter) Imperfeito ESTADO

- (13) *Kino's face **grew dark and dangerous**.*

*A expressão de Kino **tornou-se** sombria e ameaçadora.*

Kino's expression **became** somber and threatening

- (13p) ACTIVITY (grow) adjective ACCOMPL simple past ACCOMPL
MUDANÇA (tornar-se) Perfeito MUDANÇA

- (14) *the sky was **brushed clean** by the wind*

*O vento **limpara** o céu*

the wind **had cleaned** the sky

- (14p) ACTIVITY (brush) adjective ACCOMPL passive ACQUISITION
OBRA (limpar) Mais que Perfeito OBRA COMPLETA⁴⁸

⁴⁸ In order to make sense of these paths, the reader must also check Figure 4.6 (with accomplishment substituted for event), and accept for the moment that the correct description of the import of *Mais que Perfeito* applied to an *Obra* is that it turns it into an *Obra completa*. The first part of the path, shared with the other examples, is the most

An interesting subcase of this question, incidentally one of the most discussed matters in contrastive translation studies, is the one concerning verbs of movement and position. A lot has been written about the subject, from Talmy's (1985) typological claim that languages are either frame- or path-centered to Slobin's (1996, 1997a, 1997b) and Aske's (1989) detailed studies involving English and Spanish.

From the point of view of English, the cases of movement are aspectually very similar to the examples discussed above, as can be appreciated in the translation network of Figure 4.9.

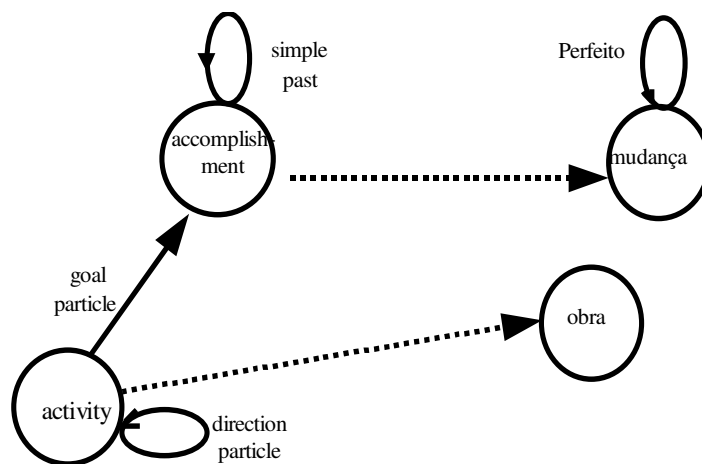


Figure 4.9: Translation network dealing with particles

The goal particle conveys the result of the movement, and the verb the manner of movement. In Portuguese, however, movement is conceptualized as a simple change (of position) and thus one has a *Mudança*, as examples (15) and (16) illustrate. The way that change is effected is of (almost) no interest for Portuguese speakers.⁴⁹

- (15) *Her hand crept out from under her shawl*
O braço dela saiu do xale
 Her arm **left** the shawl

relevant to the discussion here, but it is important to note that all these phenomena are connected and make sense in a larger context.

⁴⁹ However, the **sort** of change may be emphasized, like in example (a):

- (a) *And the last light went out of the embers in the fire hole*
E, dentro em pouco, até a claridade da última brasa da fogueira se extinguiu.
 And, in a while, even the clarity of the last ember of the fire **had extinguished**

- (15p) ACTIVITY (creep) goal particle ACCOMPL simple past ACCOMPL
 MUDANÇA (sair) Perfeito MUDANÇA
- (16) *He slipped off his sandals*
Descalçou as sandálias
 He **took off** the sandals
- Very often, though, English particles are vague regarding goal or direction (and thus the resultant expression verb + particle is vague between activity and accomplishment). Whenever a translator interprets the particle as directional, there is a clear tendency to omit the directional information altogether,⁵⁰ as illustrated by examples (17) and (18); while, if the particle is interpreted as goal, a Mudança results, as shown in examples (19) or (20).
- (17) *She looked up at him, her eyes as cold as the eyes of a lioness.*
ela fixou-o com os olhos gelados de uma leoa.
 she stared at him with the frozen eyes of a lioness
- (17p) ACTIVITY(look) directional particle ACTIVITY simple past ACTIVITY
 OBRA (fixar) Perfeito OBRA COMPLETA
- (18) *The third man scrabbled away like a crab*
O terceiro homem arrastava-se como um caranguejo
 The third man **dragged** like a crab
- (18p) ACTIVITY (scrabble) directional particle ACTIVI simple past ACTIVITY
 OBRA (arrastar) Imperfeito OBRA EM PROGRESSO
- (19) *Kino looked down to cover his eyes from the glare.*
Kino baixou as pálpebras para fugir à dureza da luz.
 Kino **lowered** his eyelids to escape the hardness of the light
- (19p) ACTIVITY (look) goal particle ACCOMPL simple past ACCOMPL
 MUDANÇA (baixar) Perfeito MUDANÇA
- (20) *The news stirred up something infinitely black and evil in the town*
A novidade erguera na cidade qualquer coisa profundamente negra e diabólica
 The news **had raised** in the town something deeply black and evil
- (20p) ACTIVITY (stir) goal particle ACCOMPL simple past ACCOMPL

⁵⁰ There is a tendency to leave out directional information in general – not only in the case of particles – as example (b) shows:

- (b) *Now Kino and Juana slid the canoe down the beach to the water*
Kino e Juana empurraram o barco para a água
 Kino and Juana pushed the boat into the water

MUDANÇA (erguer) Mais que Perfeito AQUISIÇÃO

Portuguese prefers to describe change of place rather than motion, as the pure replacement of a (movement) activity by a Mudança in example (21) shows.

- (21) *the neighbors were **tumbling** from their houses now*
*Os vizinhos tinham **acorrído***

The neighbours had (quickly) **come**

Not content with completely removing the information about the way the neighbours were leaving their houses, the translator adds the information that they were all coming towards Kino and Juana, something which is only implicit in the original wording. To model this with a translation network, one must add, to the English side, path information on the goal (in bold below), and then use the usual rules:

- (21p) ACTIVITY (tumble) past progressive SERIES goal NP ACCOMPL
 MUDANÇA (acorrer) Mais que Perfeito AQUISIÇÃO

For the sake of completeness, let me note that yet another compact category involving activities exists in English, this time with locative prepositional phrases but, contrary to the two previous cases, it generally gets translated in a structurally similar way in Portuguese (cf. examples (15) and (b) in note 50). Even in this case, however, whenever they can get rid of the manner of movement, translators do it, as can be appreciated in example (22).

- (22) *Kino and Juana **walked through** the city as though it were not there.*
*Kino e Joana **atravessaram** a cidade como se ela não existisse.*

Kino and Juana crossed the city as if it did not exist.

- (22p) ACTIVITY (walk) path-PP ACCOMPL simple past ACCOMPL
 OBRA (atravessar) Perfeito OBRA COMPLETA

Before leaving the subject of compact categories, I want to point out that this phenomenon is not restricted to syntactically complex expressions. In fact, in addition to the obvious cases of lexical compactness (like lexical accomplishments), I have used compactness to describe grammatical operators such as the Imperfeito and the perfect.

The next – and last – example of Portuguese “disrespect for manner” involves both an accomplishment and the perfect. The English elaborate formulation in example (23), referring to a concrete process having led to a result which is still holding, is translated into Portuguese by an abstract Obra em progresso (in progress), which makes no reference to the concrete way in which the result was achieved (or to the metaphor employed).

- (23) *the day **had drawn** only a pale wash of light in the lower sky*
*o dia mal **espalhava** uma pálida claridade no céu*

the day hardly **spread** a pale clarity in the sky

- (23p) ACCOMPL (draw) perfect ACQUISITION simple past ACQUISITION

4.3.5 Lexical issues

As hinted in the previous section, the relationship of phenomena like compactness, vagueness, to the lexicon deserves closer consideration. In many cases, it is fair to say that there is a lexical counterpart of the situations described above in terms of general categories.

But another interesting case is that, lexically, general classes in the source language may correspond to several less general classes in the target language, with different aspectual and grammatical properties altogether. Consider the case of Portuguese *Mudanças*.

Even though, aspectually, in Portuguese there is no difference between changes of (state of) position and changes of other states, the way these two domains are (differently) conceptualized in English results in different translation patterns. On the one hand, “normal” changes (from the point of view of English) are rendered as achievements, as in example (24) or accomplishments as in example (25), since Portuguese is vague regarding duration of *Mudanças*. Actually, it is often the case that the English translator adds graduality, if not manner altogether (translating a *Mudança* by an accomplishment).

- (24) *se ninguém lembra já o que mais de quarenta anos de Império tornou vulgar,*
*if no one remembers any longer what after more than forty years of Empire has **become** common knowledge*
 if nobody remembers any longer what more than forty years of Empire **made** common

- (25) *Assustou-se, ficou mesmo estranhamente assustado.*
*He **grew** frightened, he actually felt a strange fear*
 He **got** frightened, he got really frightened, strange as it were

Neither translation is fully successful: in example (24) because the cause of the (change of) state is not stated in English; and in example (25) it looks as an over-“Englification” to interpret *assustar-se* as a gradual process, when it is much more probable that the Portuguese sentence described a sudden fright.

Conversely, in example (26) the translator uses an achievement when one might imagine that ‘he’ gets confused over time:

- (26) *Ou ele [...] **ficara** um pouco confuso.*
*Or else he [...] **had become** a little confused.*
 Or his mind eventually **showed** signs of confusion

As simple sentences are not enough to let the reader have an opinion, here is a larger context:

E foi isso mesmo que o distraiu e o trouxe de volta à sua inactividade quotidiana, fazendo-o demorar-se mais longamente, dia a dia, nos seus postos de observação.

Porque era como se a descida do deus Kama tivesse estabelecido, na transmigração das almas, a maior das confusões. Ou ele, que nunca em sua humilde vida de génio vira um deus, **ficara** um pouco confuso. (PE, p. 128)

And that was what distracted him and brought him back to his daily inactivity, making him delay longer, day by day, in his observation posts.

For it was as if the descent of the god Kama had greatly confused the transmigration of souls. Or else he, who in his humble genie's life had never seen a god, **had become** a little confused. (PE, p. 93)

The verb *ficar*, in its unambiguous *Mudança* interpretation, is difficult to translate, as shown in the many different paths found in the authentic translations. In (27), *ficar* was rendered by an activity in English, and its result turned into an apposition.

- (27) *Ele **ficara** horrorizado, sem saber que fazer.*
*The genie **looked on**, horrified, not knowing what to do.*
 He **got** horrified, without knowing what to do

Changes of position, the other subclass of *Mudanças*, tend to get a rendering of movement that conforms more to English, often expressed by activities, or at least with the help of an activity main verb. Examples (28) and (29) show addition of manner (and thus translation of a change of location by an accomplishment):

- (28) *Depois, **foi** a uma mesa próxima*
*He then **walked** to a nearby table*
 Then, he **went** to a nearby table
- (29) *O escravo, tão suavemente como entrara, **saiu**.*
*The slave **slipped out** as quietly as he had entered.*
 The slave, as softly as he had entered, exited.

It is much more common to add manner when translating changes of position than when translating other *Mudanças*, even though this is not impossible, as example (27) above has shown.

Generalizing this situation, this brings consequences in terms of the translation network model. In order to adequately describe translation, one may need to specify the source aspectual network further, introducing unlabelled arcs pointing to categories which are more specific than the source language cares to differentiate grammatically, as in Figure 4.10. This is necessary in order to state that only a specific subset of the elements of a class enter a particular translational

relation, as is the case with the two kinds of Mudanças when translated into English.⁵¹

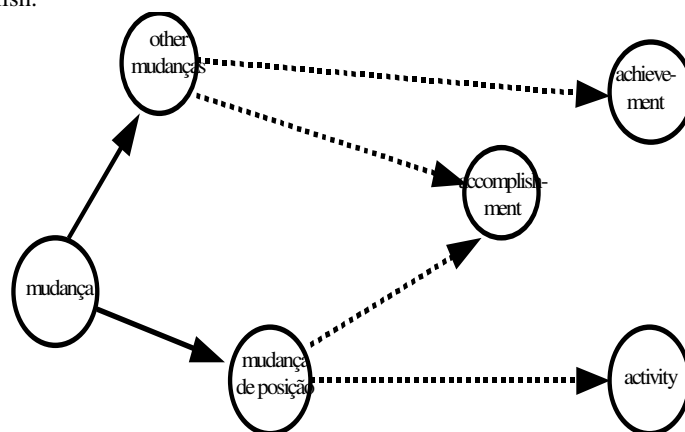


Figure 4.10: Translation network for Mudanças

Another possibility would be to label the translational arcs with specific semantic classes – i.e. from the unique node of Mudanças, have two arcs labelled "change of position" and "change of state" –, but this might not suffice in case also grammatical operators were different (corresponding thus to different paths), as will be discussed in Section 4.4.1 below.

Note that this is different from coercion by translation. In coercion by translation, the source language is vague, i.e. does not make the distinction required in the target language; here, there is no vagueness between changes of position and changes of state (no native speaker would fail to make the distinction). But the difference is signalled by the lexicon only, not by the grammar. In translation, this is a very common situation: a given (non-vague) lexical item must often be translated into either of two (or several) more specific words; e.g. Portuguese *toalha*, can be rendered in English as *towel* or *tablecloth*. Lyons (1977: 237f) gives an interesting presentation of the lexical difficulties in translating *The cat is on the mat* into French.

⁵¹ In fact, change of position verbs are also somewhat special in Portuguese, because they can be modified by manner adverbials and become Obras, which is something that other Mudanças cannot. (c) shows an example of this:

- (c) *Ia de olhos baixos, temendo encontrar alguém conhecido*
He walked with his eyes down, afraid of meeting anyone he knew
He went with eyes in the ground, fearing to meet anyone known

The parallel with the lexicon was also noted when describing vagueness above. It could therefore be claimed that the same translation problems, strategies and regularities are common to all levels of language: lexical, syntactic, semantic, ... even pragmatic. Note in this respect Neustúpný's (1978: 160) considerations: “language means the totality of rules necessary for human communication; in other words, [...] language means not simply grammatical but **communicative competence**”. One should thus not refrain from applying translation-theoretic concepts to all levels of linguistic communication when translation is involved.

The translation network model actually makes the implicit claim that the same categories are operative at every level.

4.4 Problems and limitations

I would not like to give the impression that the translation network I present here is the ultimate and complete solution to the problem of describing real translations and the systematic relations occurring between two languages. Even though I hope it constitutes a significant step in that direction, it still has many weaknesses, some of which will be pointed out below.

4.4.1 Putting things together

So far, the examples of translation relations have all been node-centered – depending simply on the node(s) arrived at in the source (and correspondingly in the target) aspectual network(s). However, things are not always so simple. In fact, previous choices may prevent translation relationships, i.e. the translation of a category may also depend on the path taken so far, as is the case of the translation of *Obras completas*. (“*Obras completas*” are *Obras* to which a *Perfeito* or *Mais que Perfeito* has been applied, consequently expressing completion.)

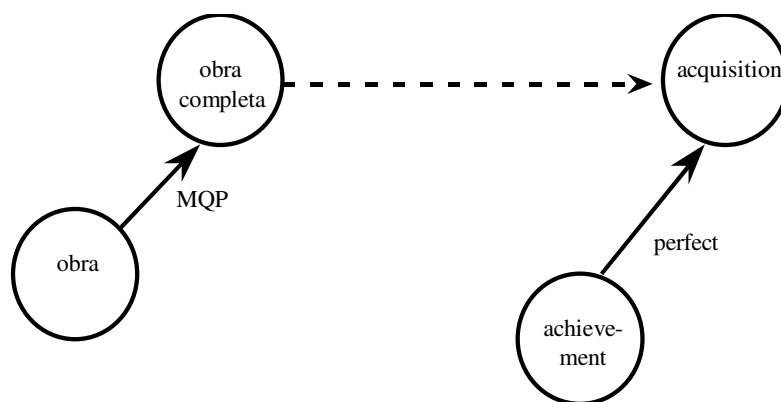


Figure 4.11: Translation of a completed *Obra* by an acquisition

Examples (30) and (31) correspond to different translation choices – and different translational arcs – which are displayed respectively in Figures 4.11 and 4.12.

- (30) *que se interessara demasiado pela homenagem,*
 who **had become** too interested in the commemoration
 who **had been** too interested in the homage
- (31) *O papá discursou?*
 Did you **give** a speech, Papa?
Did Papa talk?

(30p) OBRA (interessar) ^{MQP} OBRA COMPLETA
 ACHIEVEMENT (become) ^{perfect} ACQUISITION

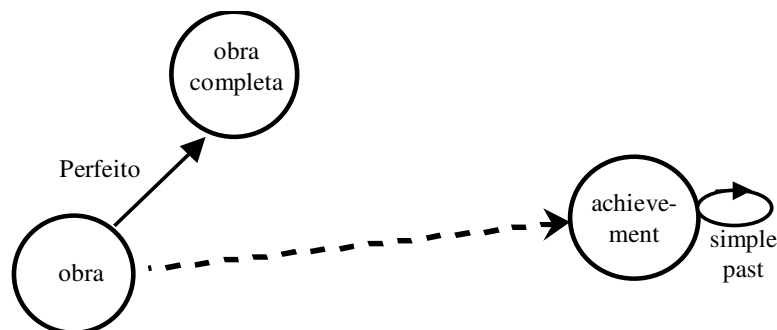


Figure 4.12: Ordinary translation of an Obra by an achievement

(31p) OBRA (discursar) ^{Perfeito} OBRA COMPLETA
 ACHIEVEMENT (give) ^{simple past} ACHIEVEMENT

In order to describe the two possibilities, one cannot simply merge the two networks, like in Figure 4.13.

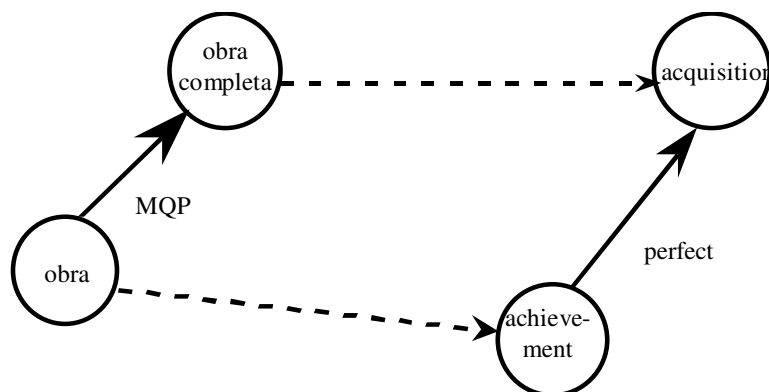


Figure 4.13: A **wrong** way to combine the two cases

Rather, one needs to unfold the two situations as in Figure 4.14, to which I have added the node `ACCOMPLISHMENT` for completeness's sake:

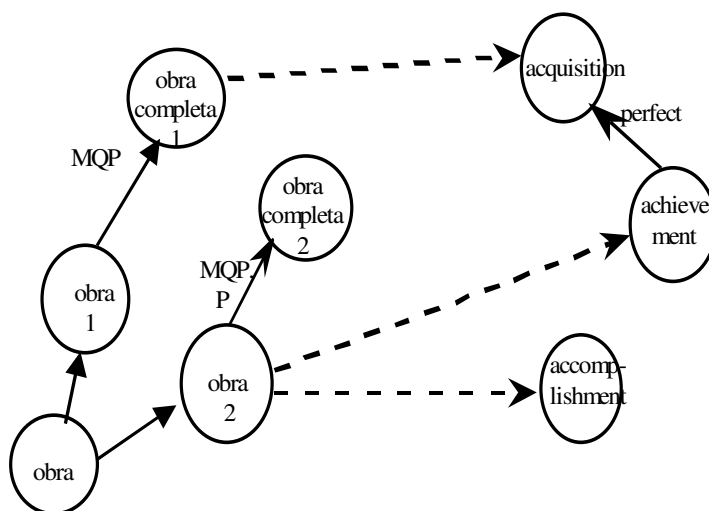


Figure 4.14: A possible rendering of the two translations in a common network

This partial translation network expresses that if the path `OBRA #> OBRA-1 MQP OBRA COMPLETA-1` is taken, then there is no direct translation relationship (no translational arc) between the achievement to which the perfect is applied and *Obra 1*. (I use the symbol `#>` to express “further specification”, and not coercion – represented by `@>` –, which is required language-internally.)

The inescapable conclusion is that a full translation network, covering a whole language translated into another language, cannot be easily obtained by uncritically adding up all partial translation networks that were found to hold. Rather, a lot more work is required to model the translation regularities between two languages consistently.

This may reduce the interest in the translation network model, since it does not allow an easy confrontation of the systems of two languages. It is not just a question of pairing the aspectual networks of two languages, adding coercions induced by translation. As shown in the two last sections, nodes and arcs may have to be unfolded into more complex patterns.

On the other hand, as Lab (1990: 52) puts it, having a linguistic description of two languages is not the same as having a linguistic description of the translation among them. Describing translation between two languages is more difficult and complex than describing either of them.

Kamp seems to acknowledge this when he notes that:

There have been few systematic attempts to explain the differences between the distinct ways in which different European languages use their

tenses and temporal adverbs to express what appears to be the same information – at least that have made use of a precise formal framework. [...] explanations of such interlingual distinctions – explanations, in other words, of what might be considered idiosyncrasies of the tense-and-aspect systems of individual languages – require a **more refined machinery than** one may have seemed to need in **the (non-comparative) study of tense-and-aspect systems of single languages**. (Kamp, 1991: 64, my emphasis)

However, believing that the same information is ultimately present in different languages,⁵² Kamp goes on to argue against such a conclusion in the following sentence: “The impression that analyses of individual systems would be able to make do with substantially simpler machinery, however, must, I think, be illusory” (ibid.). In the absence of a satisfactory model for contrastive and monolingual data, it is impossible to prove or falsify such statements, but I would say that the sections above lead us to the opposite conclusion.

4.4.2 “Too free” translations

Other problems remain when the translations are not literal enough – literal enough in the sense of displaying a clause to clause translation, not necessarily in the same order.

Example (32) – which, in principle, is as simple as examples (11), (14) or (18), displaying reduction of manner – is complicated because, syntactically, two clauses were translated by one.

- (32) *The wind **blew** fierce and strong, and it **pelted** them with bits of sticks, sand, and little rocks.*

*O vento forte e furioso **fustigava-os** com ramos partidos, areia e cascalho.*

The strong, furious wind **pelted** them with broken branches, sand, and rubble

There is, at the moment, no way to express in the translation network something like *blow* and *pelt* getting one clause with main verb *fustigar*, even though linguistically speaking this is undoubtedly what happens. In fact, we have here an instance analogous to causatives in English (like *blow something open*), which was considered a compact device when corresponding to one clause in the translation; see Figure 4.8 above.

In this particular example, however, English requires two clauses (even two sentences), and I have not considered, in the English aspectual network, neither the device “coordinative conjunction” nor the operation “coordination”. A tentative way to describe this particular translation pair would be as in (32p), but obviously a whole field of English grammar (and of Portuguese grammar) is to be studied before adding such general elements to the model.

⁵² Kamp claims that his aim is to “compare the tense and aspect systems of French and English within the context set by the account of the French system that is proposed in [Kamp & Rohrer’s book on French]” (Kamp, 1991: 41). In other words, to describe English in a theory empirically motivated by French.

(32p) ACTIVITY (blow) and ACCOMPLISH (pelt) simple past ACCOMPLISH
 OBRA (fustigar) Imperfeito OBRA EM PROGRESSO

Even more obvious cases not covered by the model are simple omissions of translations, which may be represented by a null node in the target aspectual network. Far from considering this an irrelevant technical detail, I note that to know which categories were omitted might be significant. This would add yet another level of indeterminacy to the model, however, because one would have to decide between e.g. a 2-1 translation or a 1-1 plus a 1-0 translation. (Incidentally, this also applies to example (32).)

4.4.3 Can one get an overall picture, anyway?

Figures 4.1 to 4.3 may have given us the hope that it is possible to get at the contrast between English and Portuguese at a glance, by drawing the complete translation networks. Although the sections above have shown that this was not feasible using only the nodes relevant to each language (i.e. each language's own aspectual network), one can still draw a rough picture by adding up all translational arcs found.

This is what I have done in Figures 4.15 and 4.16, which should only be read as approximate models. They are also incomplete, since I have only drawn the translational arcs for which I have found some evidence from my limited corpus studies, not all possible arcs.

Two things can be learned from this exercise, though. The first is that the translation networks are far from symmetric, which should not surprise us: Only in the cases where the translational arc denoted a fully equivalent rendering should it be expected to be found inverted in the opposite direction, and there were hardly any such cases.

I also realized that, by adding up all possible cases, the two systems appeared much more similar than the translations show in practice, i.e. the resulting translation model displayed an illusory similarity. For, it is true that I found it relevant that a node “acquisitions” and a node “Aquisições” existed, but most of the time there was no translation relation among them, i.e. English acquisitions were not translated by Portuguese Aquisições or vice-versa (cf. section 5.8 below). The same with “Série” and “series”, which, though analogous, were invoked in different “places” of each language's grammar.

It turns out, therefore, that bigger translation networks, trying to encompass at the same time several different and unrelated cases of translation may turn out to be less informative than partial, or specific, translation networks. These must, obviously, handle all grammar required to deal with the particular translation pair (in the limiting case containing the whole source aspectual network, and the whole target aspectual network), but no other translational arcs than the ones potentially involved should be present.

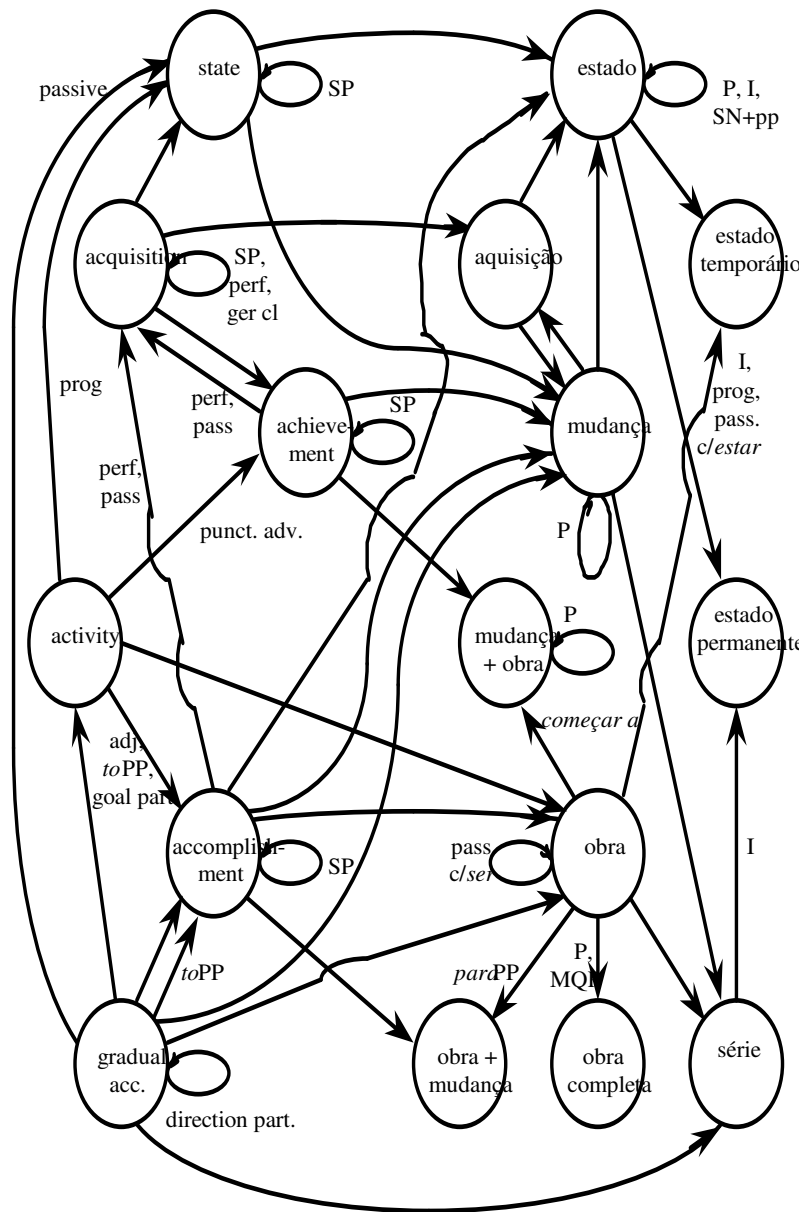


Figure 4.15: An approximate translation network from English to Portuguese

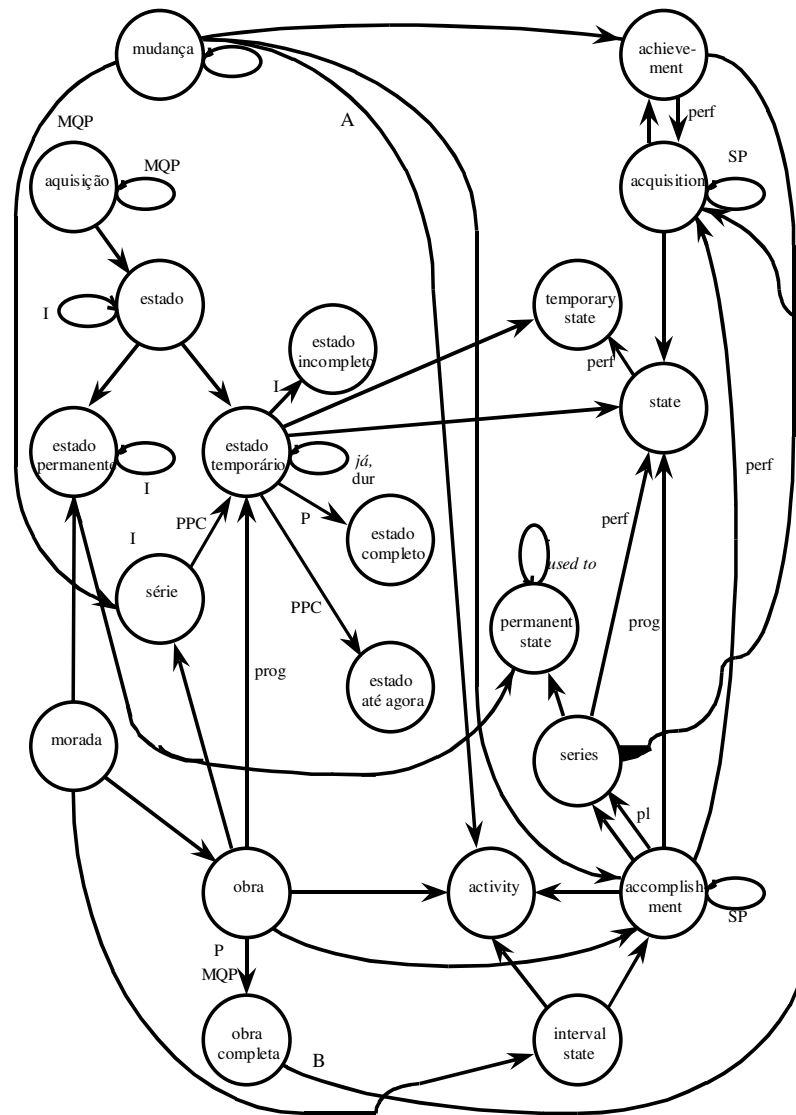


Figure 4.16: An approximate translation network from Portuguese to English

4.4.4 Temporal reference and quantification

Recalling that tense and aspect involve not only aspectual categories, but also reference to periods of time, we should also bear in mind that the translation model (and the aspectual network models) does not provide a solution to the

treatment of translation pairs involving (solely) such matters. The general questions involved, however, appear to be exactly parallel – and even crucially related to each language's categories, as will be discussed in Section 5.10.

In any case, and getting back to the main purpose of the present section, the translation network model should be extended to include also temporal translation differences.

4.4.5 How to evaluate the translation network model

One of the most important criticisms that can be made regarding the translation network is actually the issue of verifiability. An aspectual network analysis can be – at least partially – verified by the sequences of linguistic operators that are included in the sentence, given that formal, observable, devices label most arcs. But translational arcs have no labels ... and, in fact, there are no formal traces of the translation process in a translation.

In chapter 6, I will sketch possible ways to engineer a validation of the translation network.

A positive word on the application of the translation network model to my own contrastive studies should be stated, though. To have some diagram about general categories to which I had to commit myself (instead of working only with analyses of individual translation pairs) gave rise to numerous revisions of the diagrams, reanalyses of many translation pairs and rethinking of many problems. This would not have been possible if I had not worked with the associated general “language” of aspectual categories, coercions, and translational arcs.

In addition, the use of the translation network made me discover and try to solve (at least) the three kinds of problems that a simple mapping of two aspectual networks would not solve, namely:

- coercion brought about by translation
- splitting of a category required by translation
- specification of history-based categories required by translation

I have also internalized, although not yet made explicit, and thus formalized, rules about comparison of alternative paths; it is this competence that allowed me to discuss the translations presented in the present chapter.

It is an empirical question whether using this model makes it possible to efficiently communicate with readers who potentially will want to make use of it, too. One has to wait for alternative analyses of the same phenomena (or of the same translation pairs) to see if there is an objective way to compare them and settle on one.

4.5 Summing up

In this chapter, I have outlined a model for the description of translation, the **translation network**, which relies heavily on the assumption that translation does not need to be X-preserving, whatever the X, and that considerations relevant to

the overall economy of the target language condition the form and content of the result to a larger extent than is generally acknowledged.

Using relatively simple (real) examples and few different (partial) networks from English into Portuguese, I showed cases of

- coercion brought about by translation
- addition of interpretations triggered by translation
- creation of vagueness by translation
- preservation of vagueness in translation
- choice of the wrong alternative in translation
- choice of part of a compact meaning by translation

I also noted several limitations, and unresolved problems, pertaining to the translation network model, in order to provide a fair assessment of the model.

Before closing the chapter, let me present, in a nutshell, a list of general assumptions that motivate the model. In Santos (2000), I claimed that, in real translation:

- Sentence structure and overall text organization are not always preserved
- Translators do not favour literal translation
- Literal translation can be problematic
- Translation cannot be handled without significant processing of both texts (i.e., every translation decision can have far-reaching consequences, paradigmatic relations must be taken into account in both languages)
- One sentence does not convey a single meaning

And, not only a model of real translation must account for these points, but

- A model of translation must accommodate “noise”, i.e., explain both mistranslation and translationese

In the next chapter, I will be concerned with the content, that is, the actual differences I was able to spot between the two languages. If I am able to make significant generalizations and explain troublesome translation pairs by using the translation network as a descriptive tool, however, this will constitute additional *a posteriori* evidence for the adequacy of the model.

One should nevertheless remark that the translation network proposal has two levels: the general formalism to describe translation, on the one hand, and the actual categories, differences, and translation strategies that have been encoded in the model – regarding Portuguese to English and Portuguese to English translation – on the other. It is only to the latter that the corpus studies described in the next chapter can provide evidence. To provide evidence to the former, a more indirect confirmation has to be sought, namely that if an English to Portuguese translation network allows one to describe and formalize interesting contrasts, then the general translation network model may be an adequate descriptive device for translation practice and language contrast.

Chapter 5

Corpus studies

How can corpus studies be brought in support of the translation (and monolingual) models presented so far? I presume anyone accepts the claim that my work was “corpus-inspired”, since I present authentic examples of translation. But how can corpus studies produce evidence for (the particular instantiation of) the translation model suggested, if not as a testbed for the application of the model (i.e., to validate it)?

In fact, this last task was not even systematically attempted, as stated in the previous chapter. This would not make the model or theory “corpus-based”, anyway – it would only make them “corpus-validated”. (And, incidentally, not particularly well validated, if one shares the concern that test and training corpora should be kept separate.)

My aim in this chapter is to explain the process by which, with the aid of a small corpus in the two directions, I was able to arrive at several of the conclusions presented in this book.

I have to acknowledge that I have not produced a set of well-designed experiments to discover the kind of findings I have arrived at. Rather, I did some basic annotation in order to map the translation and, in view of the results, I investigated the tentative rules and/or exceptions that emerged. In some cases, this led to the refining of hypotheses and further studies. In other cases, the results concerned only methodological issues, or the studies were just inconclusive.

One thing remains crystal clear, however: I had to analyse a large number of translation pairs before my ideas on the contrast of the two languages, and therefore on the “style” of each language, eventually took shape. Since this process spanned over more than three years, there is no point in trying to document it in detail. So, this chapter begins by presenting a sketch of the preliminary studies, reporting in turn several deeper studies. The emphasis is on showing how these studies helped shape the model of translation and how they uncovered the differences between the two languages – stressing the collaboration among different pieces of evidence.

The focus of the present chapter is more on the process than on the contrastive results themselves. An alternative organization would be to give, for each claim, a description of the supporting evidence. This would better suit an audience whose main interest was Portuguese grammar and its contrast with English, and not particularly – or so much – corpus methodology.

5.1 The corpus

The corpus consists of two books and their translation into the other language. These will be referred to as the EP corpus (i) and the PE corpus (ii) respectively. Statistics concerning the two corpora are summed up in Tables 5.1 and 5.2:

(i) one American English novel, *The Pearl*, by John Steinbeck, Bantam Books, 1975 (1st edition, 1945). (Each chapter has the reference EPn, *n* ranging from 1 to 6.)

(i') and its translation into European Portuguese by Mário Dionísio: John Steinbeck, *A pérola*, Publicações Europa-América, 1977.

Table 5.1: Quantitative data on the English to Portuguese part (EP corpus)

	Words in English	Words in Portuguese	Sentences (Eng)	Sentences (Port)	Transl. pairs	Tensed transl
EP1	3416	3051	211	224	207	535
EP2	2233	2017	115	127	113	248
EP3	5815	5172	349	410	340	782
EP4	5051	4554	347	393	337	701
EP5	2827	2472	194	235	191	396
EP6	6718	5996	412	472	412	1082
Total	26060	23262	1628	1861	1602	3744

(ii) a selection of Portuguese short stories from Jorge de Sena, *Antigas e Novas Andanças do Demónio*, Edições 70, 5^a edição, 1984 (1st edition, 1978).

(ii') and their translation into (American) English: Jorge de Sena, *By the rivers of Babylon and other stories*, edited and with a Preface by Daphne Patai, Rutgers University Press, 1989.

Table 5.2: Quantitative data on the Portuguese to English part (PE corpus)

	Words in Portuguese	Words in English	Sentences (Port)	Sentences (Eng)	Transl. pairs	Tensed transl
PE10	4410	4898	323	324	322	596
PE11	1501	1719	70	71	70	217
PE12	4944	5425	179	181	179	728
PE3	3447	3695	210	211	210	456
PE6	4460	5258	352	355	351	595
PE8	2019	2279	107	107	106	239
PE9	4393	4698	208	210	208	487
Total	25174	27972	1449	1459	1446	3318

The short stories will have the referenc PEn [numbers correspond to chapter numbers in the Portuguese edition]. Each is presented by title, together with the date of its writing, translated title and translator.

PE10: *A noite que fora de Natal* (1961) – A Night of Nativity (Frederick G. Williams);

PE11: *O grande segredo* (1961) – The Great Secret (Daphne Patai);
 PE12: *Super flumina Babylonis* (1964) – By the rivers of Babylon (Daphne Patai);
 PE3: *Mar de pedras* (1960) – Sea of Stone (Christopher C. Lund);
 PE6: *A comemoração* (1946) – The Commemoration (Edward V. Caughlin);
 PE8: *A campanha da Rússia* (1946-1960) – The Russian Campaign (Daphne Patai);
 PE9: *Kama e o génio* (1964) – Kama and the Genie (K.D. Jackson).

5.2 Initial corpus processing

The preprocessing of the texts included the following tasks: (i) scanning and subsequent proof-reading, as well as general corrections (undoing hyphenation because of line breaks, removing page numbers and special formatting, etc.), (ii) sentence separation, and (iii) sentence alignment.

Some of the figures in Tables 5.1 and 5.2 actually depend on the choices taken during the alignment process (assuming that neither proof corrections nor sentence separation were controversial). “Translation pairs” are the result of alignment, therefore counting each case of 1-1, 1-0, 2-1, 1-2, etc. aligned sentences as **one** translation pair. Alignment here simply means a manual (computer-aided) reassignment of line numbers, from an initial state of one sentence per line to a situation where line numbers count (sets of) sentences which correspond in translation.

Then, for each translation pair, I annotated in a separate file all tense translations, *from the point of view of the source language*. This means that the label “tensed translations” basically counts tensed clauses in the source language (the translation itself may be tensed or not).

This annotation could have been done more carefully, since I did not annotate cases of tensed clause not translated by a clause (they were just omitted). In addition, I did not record cases of “tense creation” or even “clause creation” in the target language. Furthermore, while I did record the cases of a source tense being translated by an aspectualizer for which there was no equivalent in the source, I did not record any other cases of aspectualizers (either their translation or disappearance). However, I did annotate separately the main verbs *be* and *ser*, *estar*, and *haver* in the source text, since I expected statives to require special consideration.

Let me give some concrete examples of the annotation to illustrate some problems, showing that the annotation often fails to render the complexity of the pairs. To give the reader an idea of what was actually involved, I will display a selection of sentences, which obviously did not occur in sequence, followed by the annotation that was added.

1. Only his eyes searched in the darkness, and in the pale light of the moon that crept through the holes in the brush house Kino saw Juana arise silently from beside him. He saw her move toward the fireplace.

1. Só os olhos vogaram na escuridão. E, à vaga luz da Lua que se infiltrava pelas fendas da cabana, Kino viu, ao seu lado, Joana levantar-se sem ruído e aproximar-se da fomalha.

- | | |
|--|---|
| <p>2. "I go to inform myself," and he closed the gate and slid the bolt home.</p> <p>3. She put her hand on Coyotito's covered head. "Open it," she said softly.</p> <p>4. He turned away from her and walked up the beach and through the brush line. His senses were dulled by his emotion.</p> <p>5. Her back was bent with pain and her head was low.</p> <p>6. His senses were burningly alive, but his mind went back to the deep participation with all things, the gift he had from his people.</p> <p>7. Há sempre quem suponha, na sua paixão, que destruir Roma, a devassa Roma, a pecadora Roma, é dar testemunho dos desígnios de Deus.</p> <p>8. Não era uma tentação que repelia assim; mas era, como bem sabia, um esforço para que o céu se contentasse com as relações espirituais de uma oração.</p> <p>9. – Que foi que ele revelou antes de desfalecer?</p> | <p>2. – Vou saber. Fechou a porta e correu o trinco.</p> <p>3. Pôs a mão na cabeça coberta de Coyotito e disse baixinho: – Abre-a.</p> <p>4. Voltou as costas à mulher, subiu pelo areal, atravessou as sebes, com os sentidos embotados pela emoção.</p> <p>5. Levava as costas curvadas pela dor e a cabeça caída.</p> <p>6. Os seus sentidos ardiam, vivos, mas, com aquele dom que os antepassados lhe tinham transmitido, regressava à íntima comunhão com todas as coisas.</p> <p>7. There are always those who, in their zeal, believe that by destroying Rome, that debauched Rome, that sinful Rome, they are carrying out the designs of God.</p> <p>8. It was not a temptation that she repelled in this way; but it was, as she well knew, an effort to satisfy the heavens with the spiritual offering of a prayer.</p> <p>9. "What did he reveal before he lost consciousness?"</p> |
|--|---|

The annotations of tense translation taking place in the sentences above were as follows (in parentheses I present some remarks as to alternative annotations or regarding complexity not apparent in the annotation) [PS – simple past; I – Imperfeito; P – Perfeito, etc.]:

- 1 PS – P; PS – I; PS – P (should one add PS – 0 as well?)
- 2 pres; 2 x PS – P (should state pres – ir futuro instead?)
- 3 PS – P; imper; PS – P (different source text order)
- 4 2 x PS – P; pass – pp (one PS gets two P)
- 5 pass – I; be PS – I (or simply 2 x be PS – I?)
- 6 be PS – I; PS – I; PS – MQP (different source text order)
- 7 haver pres; PC – pres (should one add ser – prog or ser-0 ?)
- 8 ser I – PS; I – PS; ser I – PS; I – PS; IC – inf
- 9 P-PS (question emphasis ignored; tense addition disregarded)

Notwithstanding all the limitations of this preliminary work, I believe it was wise to start by such a simple preprocessing, given that, in every deep study that followed, the numbers and data recorded here were never more than a preliminary indication that some particular aspects should be given careful attention. Since one cannot know, at the outset, what one will eventually investigate, I hold the view that the simpler initial data collection and gathering, the better.

After performing the tense annotation, I wrote some scripts to count tense translation, identifying at the same time which translation pairs illustrated which type. The kind of results are illustrated for the simple past, displayed in Tables 5.3 and 5.4.

Table 5.3: The translation of simple past in the EP corpus

	EP1	EP2	EP3	EP4	EP5	EP6	Total
SIMPLE PAST	323	185	499	412	242	644	2305
Imperfeito	125	85	175	179	89	260	913
Perfeito	161	75	266	192	123	318	1135
Infinitivo	5	4	10	9	9	20	57
Gerúndio	5	2	10	3	6	9	35
Imperfeito conjuntivo	3		3	10		10	26
Mais que Perfeito	6	5	11	10	4	3	39
Condicional	1		4	2	4	3	14
Presente		5	5			4	14
Presente conjuntivo		1			1	3	5
Part. pass.	4	3	3	5		3	18
<i>ir</i> + Gerúndio	2	2	4	1	2	1	12
<i>ficou</i> + Part. p./adj			1	1		2	4
<i>ficou a</i> + Infinitivo			2			2	4
<i>pôs-se a</i> + Infinitivo	2	2		1		5	10
<i>estar</i> + Part. pass.	2		1	1	1	1	6

Table 5.4: Portuguese tenses which were translated into simple past

	PE10	PE11	PE12	PE3	PE6	PE8	PE9	Total
SIMPLE PAST	317	126	281	234	256	137	284	1635
Perfeito	205	33	118	104	108	70	92	730
Imperfeito	102	91	152	118	136	64	155	818
Mais que Perfeito	3	1	3	4	5		24	40
Presente	1		2		1			4
Imperf. Conj.	5	1	5	4	4	2	9	30
Condicional				1			2	3
<i>ir</i> + Gerúndio Imp.			1	2			1	4
<i>ir</i> + Gerúndio Perf.	1				1			2
<i>vir</i> + Gerúndio Imp.				1				1
<i>vir</i> + Gerúndio Perf.					1		1	2
Progressiva						1		1

Note that, due to the deficient annotation policy followed, the numbers concerning tense distribution in the translated text were not indicative of the full distribution of tenses in the target texts (for example, simple past clauses in the English

translation that corresponded to infinitive clauses have not been taken into account in Table 5.4). Anyway, this sort of output produced a rough overview of the translation patterns. A more condensed display of the tense distribution in the whole corpus is shown in Tables 5.5 and 5.6.

Table 5.5: Portuguese tenses in original and translated text

Tense	Original Portuguese (3318 tensed verbs)		Translated from English	
	No.	%	No.	%
Imperfeito	1102	33%	1041	33%
Perfeito	770	23%	1221	39%
Presente	433	13%	278	9.9%
Mais que Perfeito	353	11%	159	5.1%
PPC	3	0.09%	2	0.06%

To arrive at an estimated total number of tenses in translated text, in order to compute the percentages shown, I subtracted from the total number of source tenses the cases of translation into non-tensed or non-clauses which had been recorded (139 for PE and 190 for EP). Any numbers relative to translated texts are, anyway, always smaller than the exact ones, as explained above (since tense creation was not recorded).

Table 5.6: English tenses in original and translated text

Tense	Original English (3744 tensed verbs)		Translated from Portuguese	
	No.	%	No.	%
Simple past	2305	62%	1635	45%
Past progressive	58	1.6%	102	2.8%
Present	260	6.9%	439	12%
Pluperfect	142	3.8%	340	9.4%
Present perfect	48	1.3%	52	1.4%

Doing this work taught me that even “simple” tasks may be difficult when one deals with real text, a realization that is hardly new to anyone who has performed corpus studies. A more original discovery was that clause alignment was definitely more complicated than sentence alignment: While most of the sentences were found to have a 1-1 correspondence and did not change order in the translation, clause addition or omission was quite frequent, and considerable reordering of clauses inside and across sentences occurred. Also, the matter of finite versus non-finite clauses seemed to vary a lot with translation; my annotation scheme, however, focussing on tensed (and thus finite) clauses, did not allow me to quantify this. Some numbers and examples about clause and sentence alignment were published in Santos (1994a).

The most important realization of this preliminary study was that no relevant conclusions could be made on the basis of a quantitative wrap up alone. Therefore, it did not take me long to engage in more specific studies.

Given that the full studies are publically available, I will not dwell on their particulars and will rather extract only their most important features. It should, in any case, be made clear that most studies were exploratory, in the sense that I first looked at the distribution of a particular feature, tried to find relevant explanations, and only then devised tests to confirm or reject them.

The studies will be preceded by some remarks on why a particular phenomenon was thought worthwhile to investigate, followed by a description of the studies and findings, and how they produced guidance for further studies.

5.3 Imperfeito

I started⁵³ by looking at the Imperfeito past tense for many reasons: it was very frequent (the most frequent tense in the PE corpus, the second most frequent in EP); I was sure it had an habitual meaning as its most important feature, but had no idea how to explain this; it is one of the most difficult tenses for foreign learners to grasp.

In the first study conducted I looked at every occurrence of Imperfeito in both original and translated text in EP1 and PE11, classified the reason(s) for its use and attempted to check whether there were significant differences in use. I concluded that Imperfeito conveyed a bundle of meanings, which were related and available to the reader (such as habituality, graduality, extendedness, marking of perspective, stativeness, plurality, etc). The translator would in most cases have to choose (not necessarily a conscious choice) the most salient one(s), or the easiest one to convey.

Table 5.7. Translation of Imperfeito into English

	PE10	PE11	PE12	PE3	PE6	PE8	PE9	Total
IMPERFEITO	127	109	214	155	198	99	200	1102
simple past	101	91	147	120	136	64	156	623
past progressive	11	5	14	8	17	17	10	82
gerund	5	3	2	2	5	3	6	26
<i>could</i>	3	2	8	4	4	4	7	32
conditional	1	1	22	8	15	1	4	53
<i>used to</i>	2		9	1		2		13
passive	1		2	1	5	4	4	16
pluperfect			3	1	5	3	4	16
infinitive			1	2				3

⁵³ This section is an abstract of Santos (1994a; 1998a; 1996a: Chapter 10).

The raw contrastive facts, shown in Tables 5.7 and 5.8, revealed several interesting features, namely a larger number of translations of Imperfeito into English pluperfect than expected, a smaller number of translations into past progressive, and a high frequency of insertion of the modal *could* in the translation into English. Conversely, *could* frequently was rendered into bare Imperfeito when translating from English.

In Table 5.8, two kinds of Progressiva are recorded, separated by “+”, Progressiva with *ir* Gerúndio and with *estar a* Infinitivo.

Table 5.8. English origins of Imperfeito in translated text

	EP1	EP2	EP3	EP4	EP5	EP6	Total
IMPERFEITO	145	96	211	194	100	295	1041
simple past	126	85	172	179	88	260	910
progressive	4	4	12		5	8	33
<i>could</i>	6	3	6	7	3	14	39
pluperfect	2		5	7	2	1	17
conditional	1					5	6
present	1	2	7		1	4	14
passive		2	8	9	1	2	22
<i>might</i>						1	1
<i>should, must</i>			1	1			2
subjunctive				1			1
+ PROGRESSIVA	3+2	2+1	1+2	0+3	1+3	1+4	8+15
progressive	2+1	0+1	1+2	0+3	0+2	0+4	3+13
simple past	1+1	2+0			1+0	1+0	5+1
<i>can</i>					0+1		0+1

The unexpected translation by the past perfect was one of several clues to the fact that translators have to choose one from many meanings of Imperfeito. Pluperfect was one strategy for conveying several of these meaning units, namely: a habitual situation turned into an indefinite plurality of situations⁵⁴, as in examples (1)-(2); a property assigned to a vaguely defined time turned into the result of an action causing that property, as in examples (3)-(4); and a past state translated into the action which gave rise to it, as in examples (5)-(6).

- (1) *Ele, que via almas de brâmanes passarem a cães,*
He, who had seen Brahmins' souls become dogs,
He, who saw Brahmin's souls turn to dogs
- (2) *Era por todos risonhamente admirada aquela partida feita ao Rebelo, um coitado que era casado e chefe de posto fronteiriço.*
They had all gleefully admired that trick played on Rebelo, a poor fellow

⁵⁴ Both when the events are repeated, or when the actors are a plural indefinite number.

who was married and head of a border post.

It **was** by all cheerfully **admired** that practical joke made to Rebelo, a poor guy who was married and head of a border post.

- (3) *Depois, em verdade, a aldeia não crescera tanto quanto, naquela época, prometia*

*Afterwards, in fact, the village had not grown as much as it **had promised** at that time*

Afterwards, true, the village had not grown so much as, at that time, it **promised**

- (4) *Lera, uma vez, uma história qualquer que até se **passava** ali, no Alto de S. João.*

*Once he had read some story or other that **had** even **taken** place there in Alto de São João.*

He had read, once, some story which even **happened** there, in Alto de SJ

- (5) *Porque **era** afinal, e agora, um ladrão temível que o povo procurara.*

*For it **had** finally **become**, now, a dreadful thief whom the people had been hunting.*

Because he **was** after all, and now, a dreadful thief that people had looked for.

- (6) *onde **estavam** instalados os escritórios e salões da Sociedade Espírita in which the offices and meeting rooms of the Spiritist Society **had been** installed*

where **were** located the offices and rooms of the Spiritist Society

This last translation strategy, namely coupling both a tense shift and a lexical shift (using a verb of another aspectual class) has been found to be extremely common, and will be taken up in a later section. Let me just add that the reverse was equally common, namely, to translate the result of an event in the past perfect by a state:

- (7) *I **had not heard** of it*

*Não **sabia** de nada...*

I **did** not know anything...

- (8) *and it was all right; she **had accepted** it,*

*e achava-a justa. **Aceitava** a ideia.*

and she found it fair. She accepted the idea.

To explain the fact that most Imperfeito occurrences were rendered by the simple past, I hypothesize that there is a price to pay when translating an unmarked device into a marked one (as would be the case of translating Imperfeito by the English past progressive), so this is a translation strategy generally avoided by translators.

In fact, many translations of Imperfeito by past progressive actually seemed to involve an additional English rationale: the progressive in examples (9)-(10) indicates cooccurrence (and/or selects the “in progress” meaning of the English activity *run*), something which is left vague in Portuguese:

- (9) *eu não sabia o que **fazia***
*I didn't know what I **was doing***
 I did not know what I **did**
- (10) *E, seguindo Átis que **corria***
*And, following after Athis, who **was running***
 And, following Athis who **ran**

Finally, another interesting translation pattern detected was Imperfeito to conditional. As examples (11)-(12) illustrate, something conveyed as an habitual action in Portuguese can be rendered as a hypothesis, rule or generalization in English:

- (11) *Apenas **saía** para pregar nas aldeias próximas*
*He **would** leave only to preach in the nearby villages*
 He only **left** to preach in the near villages
- (12) *Era agora um gato ajuizadíssimo, que não **caçava** mesmo um mísero maina que lhe pousasse ao pé*
*It now became an immensely wise cat, which **wouldn't** even hunt a small mynah that landed at its foot,*
 It was now an extremely wise cat, which **did** not even hunt a miserable mynah that landed near

Let me stress the point of a tense conveying several meanings. When the target language does not feature a similar tense with the same constellation of meaning units as that of the source language, the choice of one interpretation is required. Translation errors caused by choosing a wrong lexical unit constitute clear evidence for this choice, as in examples (13)-(14), as well as cases such as example (15) when another interpretation (here provided in the gloss) is equally possible:

- (13) *onde entrara em rapazinho quando as paredes da casa ainda se **erguiam**.*
*which he had entered as a young lad when the walls of the buildings **were still being raised**.*
 where he entered as a small boy when the house walls were still up
- (14) *Agora, assim alquebrado e impotente, tudo o que **pensava**, se o escrevesse, lhe parecia que [...].*
*Now, thus decrepit and impotent, everything that he **used to think**, if he were to write it, seemed to him to [...].*
 Now, thus decrepit and impotent, everything he **was thinking**, if he wrote it, seemed...

- (15) *Umas vezes escrevera na verdade para saber o que **pensava***
*At times he had indeed written in order to know what he **was thinking***
 Some times he had written really to know his own opinions

Also, additional (spurious) meanings may appear when a tense conveying a specific meaning is used as the translation of one without that same meaning. In fact, it was easy to find translations employing Imperfeito where habituality had been added without being mentioned in the source text, as in examples (16)-(17):

- (16) *The beach was cut and scarred and padded by the feet of animals that **had come** to drink and to hunt.*
*A praia estava cortada, marcada, patinhada pelos pés dos animais que ali **vinham** beber e caçar.*

The beach was cut, marked, waded by the feet of the animals which **came/used to come** there to drink and hunt

- (17) *and the canoes **were drawn** up in front of this area.*
*e era em frente dessa zona que os barcos se **alinhavam**.*

and it was facing that zone that the boats **were aligned/used to align**

After I had looked carefully at every non-Imperfeito to past simple translation (and, in the opposite direction, at every non-past simple to Imperfeito translation), it became clear that several grammatical peculiarities (such as subordinated clause status, modals or indirect speech) also interact with tense choice, either blocking, or licensing, a different tense. Thus, *since* requires the perfect in example (18), while example (19) has an idiomatic translation:

- (18) *D. Conceição **sabia** disso, desde que "princiara a comemoração"*
*Dona Conceição **had known** about it, ever since "the commemoration had begun,"*

D. C. **knew** about that, since "the commemoration had begun"

- (19) *como **era** músico*
what a musician

how he **was** musical

This process led me to look at aspectualizers and their translation, the progressive in particular. I started by investigating whether there was a direct mapping of progressives in the two languages, but the data displayed in Table 5.9 showed quite a different picture, indicating that progressive vs. past simple and Imperfeito vs. Imperfeito Progressiva must cover different parts of the “meaning sphere”.

The main observations were: Progressive addition in the English text is conspicuous when the information “in the middle of a gradual process” needs to be conveyed, as in example (20); while the need to emphasize “at that particular moment” seems to be the explanation for example (21) (as well as for example (10) above).

Table 5.9: The translation of the English past progressive

	EP1	EP2	EP3	EP4	EP5	EP6	Total	%
PAST PROG	8	5	19	4	8	14	58	
Imperfeito	3	4	12		5	8	32	55
Imperfeito prog.	1	1	2	3	2	4	13	22
Gerúndio			2			1	3	0.5
Mais que Perfeito	2				1		3	
adjectivo						1	1	
Presente				1			1	
<i>ir</i> + Gerúndio	2		1				3	
<i>estar</i> + Part.pass.			1				1	
Infinitivo			1				1	

- (20) *fitou [...] a bacia onde o sangue **coalhava** espesso*
*He looked [...] at the basin where the blood **was** thickly **coagulating***
 he stared at the bowl where the blood **coagulated**, thick

- (21) *Marco Semprônio **seguia** os movimentos do imperador*
*Marcus Sempronius **was following** the emperor's movements*
 Marcus Sempronius **followed** the movement of the emperor

Progressiva removal in translation into English is considerably rarer, and is often connected with other markers. In example (22), the adverb *constantly* does the job:

- (22) *Então um livro desses [...] precisa que tu **estejas** sempre **a explicar** o que é isto [...] ?*
*Does a book like this [...] need you constantly to **explain** what this is [...]?*
 So such a book [...] needs that you **are** always **explaining** what this is [...]?

Progressiva addition in the Portuguese translation, on the other hand, was often connected with the aspectualizer *ir* (as seen in Table 5.3), which conveys graduality, as seen in example (23).

- (23) *the hurrying newcomers **were being** softly informed*
*os que se haviam agregado nos arredores **iam sendo** informados*
precipitadamente e em voz baixa
 those who had joined in the suburbs **were being** informed hastily and in a soft voice

Progressive omission in the Portuguese translation seems to have several distinct explanations: in example (24), due to lack of concern with manner, the Portuguese text states the result instead; in example (25), the English text was interpreted as conveying perspective, which had to be rendered in Portuguese by Imperfeito.

- (24) *The neighbors **were tumbling** from their houses now, and they watched the falling sparks and stamped them out to save their own houses.*
*Os vizinhos **tinham acorrido** e estavam atentos às faúlhas que voavam e que eles pisavam logo para defenderem as suas próprias casas.*

The neighbours **had come** in haste and were attentive to the sparks that flared and that they stamped at once to defend their own houses

- (25) *Juana **was whispering**, "Coyotito." She **was begging** him to be quiet.*
*Joana **ciciava**: "Coyotito." **Pedia-lhe** que estivesse quieto.*

Juana **murmured**: "Coyotito." She **asked** him to be quiet.

But by far the most frequent explanation for translating the progressive by non-Progressiva examples seemed to be the inexistence, in Portuguese, of an aspectual class like English activities. In other words, the function of the progressive with accomplishments, selecting the activity part, is simply missing, given that the closer aspectual class, Obras, has only "activity" (no result), and thus it is enough to render it with a single tense, as examples (26) to (28) show. Using the Progressiva with *estar* in Portuguese would make these occurrences marked, going against good translation practice.

- (26) *The weak moon **was losing** its light and the clouds thickened and curdled to the southward.*

*A pálida Lua **esmaecia** mais, as nuvens engrossavam e acumulavam-se para os lados do sul*

The pale moon **paled** more, the clouds thickened and gathered on the South side

- (27) *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

- (28) *And to meet the attack, Kino **was** already **making** a hard skin for himself against the world.*

*Preparava-se para o ataque. **Urdia** já uma forte couraça para se defender do mundo.*

He prepared for the attack. He **built** already a hard armour to defend himself from the world

5.4 Perception verbs

Following a pure grammatical feature, I looked at perception verbs next. Reasons for this choice were their high frequency, the patterns of *could* with Imperfeito alluded to above, and the identification of a typical clause rearrangement involving such verbs.

Again, the specific studies are available elsewhere⁵⁵. A quantitative description of perception verbs (in the past) in the corpus is provided in Tables 5.10 and 5.11.

Table 5.10: Translation of the 76 past perception verbs in the PE corpus

<i>ver</i>	37	20 Perfeito	14 simple past, 3 <i>could</i> , 2 present perfect, 1 past participle
		17 Imperfeito	9 simple past, 5 <i>could</i> , 1 <i>in sight</i> , 1 <i>visible</i> , 1 pluperfect
<i>ouvir</i>	19	14 Perfeito	12 simple past, 1 <i>could</i> , 1 present perfect
		5 Imperfeito	3 simple past, 1 <i>could</i> , 1 past progressive
<i>sentir</i>	20	10 Perfeito	9 simple past, 1 present perfect
		10 Imperfeito	9 simple past, 1 <i>could</i>

Two general conclusions were revealed. First, apparently similar words behave differently in the two languages, and the lexicon unveils, to some extent, the same differences that grammatical studies do. Second, the modal system is intricately intertwined with the tense system, at least when one translates between a “mainly tense” language (Portuguese) and a “mainly modal” language (English).

Table 5.11: Translation of the 161 past perception verbs in the EP corpus

<i>see</i>	89	31 <i>could</i>	11 Perfeito, 11 Imperfeito, 3 pôde, 2 Condicional, 1 pudesse, 3 other
		58 <i>saw</i>	40 Perfeito, 10 Imperfeito, 3 Infinitivo, 1 Mais que Perfeito, 4 other
<i>hear</i>	43	11 <i>could</i>	5 Imperfeito, 2 Perfeito, 2 Gerúndio, 1 Imperfeito do Conjuntivo, 1 <i>ficou a ger</i>
		27 <i>heard</i>	12 Imperfeito, 13 Perfeito, 1 <i>começou a ger</i> , 1 Mais que Perfeito
		5 present perfect	5 Perfeito
<i>feel</i>	24	7 <i>could</i>	3 Perfeito, 3 Imperfeito, 1 other
		17 <i>felt</i>	10 Perfeito, 5 Imperfeito, 2 other
<i>smell</i>	5	3 <i>could</i>	2 Perf, 1 Imp
		2 <i>smelled</i>	1 Perf, 1 Imp

Given that the translation pattern uncovered a preference for *could* with perception verbs, three contexts could be identified: while unattended (but unavoidable) perception in English was expressed with *could* as in example (29), this auxiliary verb was most often employed to describe physical ability to perceive – and thus obstacles are mentioned as part of the co-text in example (30). Another common use of *could* was inception of perception, making *could* a typical case of acquisition, as in example (31).

⁵⁵ This section is an abstract of Santos (1996a: Chapter 11; 1998b).

- (29) *And he **could smell** the sharp odor of exposed kelp from the receding tide.*
Sentia o cheiro penetrante das algas que a maré vazia deixava na areia.
 He **felt** the sharp smell of seaweed which the low water left on the sand
- (30) *He **could see** the outline of the watcher, but the sleeping men were below his vision.*
Via o perfil da sentinela, mas não os dois homens adormecidos.
 He **saw** the sentinel's profile, but not the two men asleep.
- (31) *until the water cleared and he **could see***
por fim, a água tornou-se clara e pôde ver
 finally, the water became clear and he **was able to see**

From the point of view of translation, there is an important distinction here. While the first group is by default translated by Imperfeito, and the third group often employs the Portuguese modal *poder*, the translation of the second, never employing *poder*, seems to depend on various other features. For example, narrative context may be important in cases like example (32):

- (32) *In the moonlight he **could see** the frantic frightened eyes*
Ao luar, viu ainda uns olhos inquietos e assustados.
 In the moonlight, he **saw** still some anxious, frightened eyes

The most important evidence for the fact that (just like Imperfeito in the previous section) *could* has a bundle of meanings, not necessarily mutually exclusive (among which there is no need to choose), and which can all be conveyed simultaneously (the ideal use), is the possibility to supply four “interpretations” for example (33).

- (33) *And, looking down, she **could see** the cigarette of the man on watch.*
Por baixo da gruta, Juana viu o cigarro da sentinela.
 Below the cave, Juana **saw** the cigarette of the sentinel

Does *could* in example (33) express possibility, ability, inception of ability after some changes, or difficulties to perception? Does it express one of these, some of these, or even all of these? In addition to the plain fact of perception, chosen in the actual translation, one could have understood and rephrased it in three more ways (corresponding translation in square brackets):

- Perception at last: Juana looked down and could see the cigarette [*viu* or *pôde ver*]
- Simple possibility, or background conditions: If she looked down, she would see the cigarette [*via* or *podia ver*]
- Difficult perception: Looking down, she was able to see the cigarette [*via* or *viu*]

In many cases, a translator must choose without an objective clue from either source or target language, just his own interpretation or personal choice.

Considering now the opposite direction, the translation of Portuguese perception verbs into English illustrates new concerns and problems: First, habitual Imperfeitos, as in example (34), were never translated by *could*.

- (34) *todos os escriturários se abstinham de o interromper, quando o **viam** curvado sobre a secretária*
*all the clerks avoided interrupting him when they **saw** him bent over his desk.*

all clerks avoided interrupting him, when they **saw** him bent over the desk

The use of Imperfeito vs. Perfeito in Portuguese seemed to be clearly related to whether a (perception) situation was taking place or whether a perception event advanced the narrative. Since English does not normally allow the progressive with perception verbs, most translations into simple past fail to make these fine distinctions, present in the source text. Thus, in example (35), the English text fails to convey the fact that “he saw and continued to see all along”, while in example (36) the hearing event advances unambiguously the narrative in Portuguese, but not so in English.

- (35) *lançou-se para o portão, que **via** lá ao fundo*
*raced for the gate which he **saw** at the back*

he aimed at the gate, which he **could see** there far away

- (36) *um silêncio se demorou na biblioteca, durante o qual se **ouviram**, abafados, no crepúsculo que punha sombras pelos cantos, os ruídos domésticos da 'vila'.*
*in the library there lingered a silence through which **could be heard** [...] the muffled sounds of domestic life in the villa.*

and a silence lasted in the library, during which **were heard**, muffled, in the crepuscule which put shadows in all corners, the domestic sounds of the villa

Interestingly, negation of perception is less marked with *couldn't* than with *didn't* (which presupposes that there was something to be perceived). So, in example (37) the man is probably not to be seen, while in example (38) the most probable interpretation is that M.S. left. (Note that this information, however, is not conveyed by the Portuguese texts.)

- (37) *Não viu o sujeito*
*He **couldn't see** the man.*

He **did not see** the fellow

- (38) *Marco Semprônio não o viu sair.*
*Marcus Sempronius **did not see** him leave.*

Marco Semprônio **did not see** him leave

An intriguing observation was also made: even though both languages have “similar” syntactic structures concerning the perception of events (namely

non-finite clauses as grammatical objects of a perception verb), the two languages seem to use them differently. In English, one has infinitive or gerund clauses⁵⁶; in Portuguese, there is still another syntactic alternative⁵⁷: infinitive, *a*+infinitive (or gerund), and past participle.

First of all, there were just twelve cases in Portuguese original text against 43 in English; second, only 25 of those were rendered by this kind of structure in translated Portuguese (see Table 5.12).

Table 5.12: Translation of clausal objects of perception verbs in the EP corpus.

VPinf stands for infinitive, VPainf *a*+ infinitive, VPger gerund

<i>could</i>	VPinf 2	1 Perfeito VPinf, 1 gerúndio VPainf
	VPger 10	3 Imperfeito, 1 Imperfeito VPger, 1 Imperfeito VPainf, 1 Imperfeito VPpastpart, 1 Perfeito VPainf, 1 Perfeito VPinf, 1 gerúndio, 1 <i>pôde ver</i>
simple past	VPinf 16	7 Perfeito VPinf, 6 Perfeito, 2 Imperfeito, 1 Imperfeito VPinf
	VPger 12	3 Perfeito VPger, 3 Perfeito VPpastpart, 3 Perfeito, 1 Perfeito VPainf, 1 Imperfeito VPainf, 1 Mais que Perfeito
pres perf	VPinf 3	1 Perfeito VPinf, 1 Perfeito VPainf, 1 Perfeito

The most straightforward correspondence was the perception of positions, expressed in English with gerund and in Portuguese with past participle, as in example (39).

- (39) *The doctor looked past his aged patient and **saw** himself **sitting** in a restaurant in Paris*

*O médico desviou o olhar da doente idosa que tinha na frente e **viu-se sentado** num restaurante de Paris*

The doctor deviated the look from the aged patient he had in front and **saw** himself **seated** in a restaurant in Paris

It is interesting to note in this connection that, while this seems to be a rule followed by Portuguese translators, it does not seem to be known by their English-speaking colleagues, as the poor translation in example (40) shows.

- (40) *No prédio fronteiro, **viu** o calafate **sentado** à mesa, que o observava amigavelmente por cima da escudela fumegante.*

*In the building opposite, he **saw** the caulker **seated** at the table, observing him amicably above his steaming wooden bowl.*

In the house in front, he **saw** the caulker **sitting** at the table, who was

⁵⁶ According to Palmer (1965), this opposition expresses progressiveness.

⁵⁷ Actually, there is a myriad of other possibilities as well, considering use of personal vs. impersonal infinitive, subject inversion and placement of clitics, etc., all matters which I will disregard here.

watching him in a friendly mood over the steaming plate.

It should also be noted that only two of the 12 original Portuguese cases had *a*+infinitive or gerund – and one, example (34), the past participle); nevertheless, the English translators rendered two of the Portuguese infinitives as gerund, examples (41)-(42), clearly because of the English activity verb employed.

- (41) *Sentia-se descer lentamente, num poço sombrio e húmido, sem fundo.*
She felt herself slowly descending into a dark and humid bottomless pool.
 She **felt** herself **go** slowly **down**, in a dark humid well, without bottom
- (42) *ele, empoleirado num dos ramos frondosos, viu vir do santuário uma luminosidade.*
he, perched on one of the leafy branches, saw a glow coming from the sanctuary.
 he, up in one of the leafy branches, **saw** a light **come** from the sanctuary

In the English to Portuguese direction, a quite frequent situation was the creative rephrasing of the whole sentence to avoid this construction, exemplified here by examples (43) and (44). Translations often demonstrate that in Portuguese less attention to manner and detail is paid.

- (43) *through the knife he could feel the muscle tighten hard*
sentiu na faca a resistência do manto.
 he **felt** in the knife the resistance of the muscle
- (44) *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 eyes, irresistibly fixed at the ground near the post of the hut
- Finally, the study of perception verbs was also triggered by a perceived freer addition and omission of those verbs by translation than in the case of other verbs, a phenomenon which was often connected with the expression of perspective⁵⁸, cf. examples (45)-(46).
- (45) *Kino held the great pearl in his hand, and it was warm and alive in his hand.*
Kino apertou a pérola na mão e sentiu-a quente e viva.
 Kino clasped the pearl in the hand and **felt** it warm and alive
- (46) *But now it was gone, and there was no retrieving it.*
Mas agora via que tudo estava acabado, sem remédio possível.

⁵⁸ Once again, something which is closely connected with tense and aspect in both languages, albeit in a different way.

But now she **saw** that everything was finished, without possible remedy.

Explicitation and ambiguation, two well-known translation strategies, were both found, cf. examples (47) and (48) respectively.

(47) *Haveis-me entendido, haveis gostado?*

*Have you understood me and have you liked what you **heard**?*

Have you understood me, have you liked?

(48) *They **heard** from the secret gardens the singing of caged birds and **heard** the splash of cooling water on hot flagstones.*

*Desses secretos jardins **vinha** o canto de pássaros engaiolados e o cair da água fresca nas lajes escaldantes.*

From those secret gardens **came** the singing of caged birds and the falling of cold water on the hot flagstones

5.5 The English perfect

Although one of the most fascinating subjects in the English tense and aspect literature, the English perfect has been less studied in my work for several reasons: Lack of enough data in the corpus⁵⁹, the fact that in American English the perfect is less used and its meaning constellation is less clear, and, last but not least, my belief that present perfect as a grammatical category has no equivalent in Portuguese. More specifically, the crucial notion of result is absent from Portuguese grammar, and the form of dividing the temporal line in the two languages is also radically different.

Table 5.13: Translations of the 48 present perfects in the EP corpus

	EP1	EP2	EP3	EP4	EP5	EP6	Total
PRESENT PERFECT	3	0	18	14	10	3	48
Perfeito	2		13	13	7	1	26
PPC				1			1
Presente			3		1		4
passiva	1						1
Mais que Perfeito			1				1
Futuro conjuntivo			1			1	2
Presente conjuntivo					1	1	2
Infinitivo					1		1

But let me report the particular studies first.⁶⁰ In a first overview⁶¹ of present perfect meaning units, namely resultative, experiential, hot news and universal

⁵⁹ Narrative text is not in general a good place to find instances of present (perfect).

⁶⁰ Reported in Santos (1995a: section 3.2; 1996a: Chapter 12).

⁶¹ Based mainly on Sandström (1993), but see other sources in Santos (1996a).

perfect, I was struck by the similarity to monolingual descriptions of the adverb *já*. (Its use with *just* also seemed to correspond exactly to the aspectualizer *acabar de*.⁶²) Also, the much studied Pretérito Perfeito Composto (PPC), an intriguing but rare Portuguese tense, seemed clearly to bear a family resemblance to the present perfect progressive. Finally, the present perfect was often illustrated with quantifiers like *never* or *always*, intricately connected with its meaning.

Table 5.13 provides an overview of the translations of the present perfect in the EP corpus, while Table 5.14 shows its origins in the PE corpus.

Table 5.14: Translations into 52 present perfects in the PE corpus

	PE10	PE11	PE12	PE3	PE6	PE8	PE9	Total
PRESENT PERFECT	23		15	5	5	2	2	52
Perfeito	15		8	3	5	2	2	35
Presente	6		3					9
PPC	1		2					3
<i>haver</i> + Part.pass.				2				2
Imperfeito			1					1
Presente conj.			1					1
Mais que Perfeito	1							1

No matter, the similarity of the theoretical descriptions of *já* and the perfect, translation performance shows, again, that these markers are used in different contexts, and possibly also with different motivations. The translation pattern showed only two instances of *já*, the first in example (49) marking inevitability (no way to go back), the second in example (50), in my opinion a mistranslation (I would use *Uma vez ouvi...*⁶³).

- (49) *Kino has said so*
Kino já disse
 Kino **has already said**

- (50) *I've heard a coyote pup cry like a baby.*
Já ouvi o choro de uma lobazita que era tal qual o duma criancinha.
 I've **already heard** the cry of a small wolf which was exactly like the one of a small child

The other translations simply could not have used *já*, because it would mean that the event had been planned, and in no case was this meant. Rather, the most frequent use of the perfect in the English text was to focus on the result. So,

⁶² Unfortunately, there was only one example (incidentally confirming the similarity), in the PE corpus, of both *acabar de* and *just*.

⁶³ Again, this translation “mistake” just makes it clear that the present perfect has many meanings, and that mistranslation was caused by a wrong choice.

the main messages of examples (51) and (52) are “evil is here” and “I’m here”, respectively.

- (51) *It **has brought** evil*
*Ela **trouxe** o mal com ela*
 It **brought** evil with it

- (52) *I **have come** to see the baby*
***vim** ver o pequeno*
 I **came** to see the baby

Quantification is often present and almost always so in negative sentences. In these cases, Portuguese can provide a much closer translation: Example (53) shows that a negated event extends to the present, while example (54) uses PPC to convey repetition up to the present.

- (53) *It is a pearl such as one **has never seen***
*– É uma pérola como nunca se **viu**.*
 It is a pearl as **was never seen** (before)

- (54) *"If that is so, then all of us **have been cheated** all of our lives."*
*– Se isso é possível, quer dizer que **temos sido roubados** toda a vida.*
 If that is possible, it means that we **have been stolen** the whole life

Relative past time reference is conveyed by Futuro do conjuntivo, and it is conspicuous that the result part is again left out in the translation. The difference in meaning is that English expresses that the transaction should be completed in example (55) and that the night should already be there in example (56), while Portuguese simply mentions the two events (of selling and coming) as temporal signposts.

- (55) *When I **have sold** my pearl I will pay you," Kino said.*
*– Quando **vender** a minha pérola – respondeu Kino.*
 "When I **sell** my pearl", answered Kino.

- (56) *"Only until a day **has passed** and the new night **has come**."*
*– É só por um dia, só até a noite **chegar**.*
 "It's only for a day, only until the night **comes**."

The specification of an extended temporal period encompassing both past and present (the “extended now”) provides clear examples of a need to choose in the tense and aspect domain. However, curiously, none of the examples found in the corpus occurred in the present perfect. Example (57) may be related if we consider the past perfect as the past of the present perfect; example (58) was in the passive.

- (57) *Kino's people had sung of everything that **happened** or **existed***
*A gente de Kino cantara tudo que **acontecera** ou **existira**.*

The people of Kino had sung everything that **had happened** or **had been**

- (58) *He was trapped as his people **were** always **trapped***
*Estava peado, como todos os da sua raça sempre **tinham estado***
 He was hindered, as every one of his people **had** always **been**

It should be noted in passing the close connection between present perfect and passive – for some verbs the two forms even merge, as in example (59). So, while example (58) pairs neatly with (57) as an example of the need to choose between one of the two time periods in the translation, example (60) patterns the same translation strategy as example (53), in negative sentences.

- (59) *"The pearl **is gone**."*
*A pérola **desapareceu**.*
 The pearl **disappeared**.

- (60) *And he **was** never **heard** of again and the pearls were lost.*
*Mas nunca mais **ouviram falar** nele e as pérolas desapareceram.*
 But they no longer **heard** about him and the pearls disappeared.

Let us now look at the opposite direction, present perfect as translation in the PE corpus. The most interesting case is the translation of *há* with *Presente*, rendered by two present perfects, as in example (61):

- (61) *Mas também **há** quase setenta anos que me **aturo***
*But then, it's **been** almost seventy years I've **had** to put up with me*
 But also there **are** almost seventy years since I **suffer** myself

Otherwise, predictably, quantified sentences are duly rendered with the corresponding quantifiers and present perfect. Likewise, in direct speech including actions with present consequences the Portuguese tenses are often rendered by the perfect, cf. examples (62)-(63).

- (62) *E a comadre Joaquina deu-me este pastel que aqui **trago***
*And your godmother Joaquina gave me this pie I've **brought***
 And your godmother Joaquina has given me this pie I **bring** here
- (63) *Se [...] não mostrasses como és um homem arrependido da má vida que **teve***
*if you hadn't [...] demonstrated that you are a man repentant of the bad life he's **led***
 If [...] you did not show how you are a man sorry for the bad life he **had**

Já + *Perfeito*, in his context, seems to correspond nicely to, and be correctly rendered by, English present perfect. Out of the six cases, five are thus rendered, four without *already*. Actually, the two cases employing *already*, examples (64)-(65), seem to me not the best of translations (I suggest my own renderings as better):

- (64) – *Os meus amigos velhos **já morreram** todos*
*"All my old friends **have already died**"*

My old friends **are** all **dead**

- (65) *Eu **já** ceei em casa dela*
*I **already** had my supper at her house*
 I've **had** supper at her place

The same pattern can be shown in the 8 cases of *já* + Mais que Perfeito (arguably here the past of *já* + Perfeito), where the most idiomatic translations, examples (66)-(67), are the ones which do not include *already*.

- (66) *meio encostado à parede, cuja cal **já** se **esvaíra** havia muito*
*half leaning against the wall whose whitewash **had** long since **faded***
 half leaning against the wall, whose lime **had already faded** long ago
- (67) *e ele disse que tu não acabavas nunca, porque tu eras um grande poeta,*
*um dos maiores que **já tinha havido** no mundo*
he said that that you would never wear out, because you were a great
*poet, one of the greatest that **had ever existed** in the world*
 and he said that you never ended, because you were a great poet, one of the greatest who **had already existed** in the world

In fact, my impression is that all instances of *já* + Mais que Perfeito translated into *already* + pluperfect are cases of grammatical translationese.

This study, although quantitatively little representative, showed the need to look at yet other parts of the grammatical system, and repeated the observation that negated contexts behave differently in translation.

5.6 Pluperfect and Mais que Perfeito

It is clearly arbitrary to split the perfect into present perfect and past perfect, as was seen in the previous section. But one has to make such choices, and look at a subset of the phenomena at different times.

One of the reasons for this division of the perfect was obviously my Portuguese perspective. Pluperfect and Mais que Perfeito, just like perception verbs, seemed to mirror a similar distinction in the two languages. However, a large number of miscorrespondences could immediately be spotted in the rough translation patterns displayed in Tables 5.15 and 5.16.

Table 5.15: Translations of the pluperfect in the EP corpus

Number of pluperfects: 142	pluperfect -> MQP	97	Number of MQPs: 159
	pluperfect -> Imperfeito	17	
	pluperfect -> Perfeito	13	
	pluperfect -> Imperf. conj.	7	
	pluperfect -> Passiva	2	
	simple past -> MQP	39	
	passive -> MQP	18	

Table 5.16: Translations of Mais que Perfeito in the PE corpus

Number of MQPs: 353	MQP -> pluperfect	297	Number of pluperfects: 340
	MQP -> simple past	41	
	MQP -> passive	7	
	Imperfeito -> plup.	17	
	Perfeito -> pluperfect	8	
	Passiva -> pluperfect	2	

In the English to Portuguese direction, “pluperfectness” is preserved in only 68% of the cases (97 out of 142), while it is added (see Table 5.17) in 39% of the cases (62 out of 159).

Table 5.17: Origins of pluperfect in the PE corpus

	EP1	EP2	EP3	EP4	EP5	EP6	Total
MAIS QUE PERFEITO	29	24	38	26	9	33	159
pluperfect	17	17	19	13	3	28	97
simple past	6	5	11	10	4	3	39
passive	4	2	6	3	1	2	18
progressive	2				1		3
present perfect			1				1
<i>could</i>			1				1

In the Portuguese to English direction, Tables 5.16 and 5.18 indicate preservation in 84% of the cases (297 out of 353) and addition in 13% (43 out of 340).

Table 5.18: Origins of pluperfect in the PE corpus [MQP – Mais que Perfeito]

	PE10	PE11	PE12	PE3	PE6	PE8	PE9	Total
PLUPERFECT	20	24	52	57	106	13	68	340
Mais que Perfeito	18	23	43	51	93	6	62	296
Imperfeito		1	3	1	5	3	4	17
Perfeito			2	1	3	3		9
Imperfeito conj.			3	1	2		1	7
MQP conjuntivo			2	2	1		1	6
MQP condicional	2							2
Condicional					1			1
Passiva (<i>ser</i>)					1			1
MQP Progressivo						1		1

The studies performed⁶⁴ yielded the following results: A change of aspectual class in the translation (with the consequent temporal change) was confirmed to be very common. Sometimes, the translation would subtly change the meaning, as

⁶⁴ Reported in Santos (1996a: Chapter 13; 1999a).

in example (68), but it generally yielded a more idiomatic translation, as in examples (69)-(70).

- (68) *But now, by saying what his future was going to be like, he **had created** it.*
*Mas agora, planeando o futuro, era ele que o **criava**.*
 But now, planning the future, it was him who **was creating** it
- (69) *They **had been** suspicious of it from the first.*
***Desconfiavam** dela desde o princípio.*
 They **suspected** it from the start
- (70) *For a moment his body was black in the cave entrance, crouched and silent, and then he **was gone**.*
*Por um instante, o corpo recortou-se-lhe, negro, na entrada da gruta, atento e silencioso. Depois – **saiu**.*
 For a moment, his body was silhouetted, back, at the cave entrance, attentive and silent. Then – he **went** out.

Another regular pattern uncovered was connected to *when*-clauses, where English needs a pluperfect to convey temporal sequence (simple past would convey simultaneity), while Portuguese uses Perfeito in the *quando*-clause; cf. examples (71) and (72).

- (71) *And when it **had passed** out of sight, he went back to the roadway*
*Quando o carro **deixou** de se ver, Kino voltou à estrada*
 When the car **stopped** being seen, Kino returned to the road
- (72) *When the doctor **had gone** and all the neighbors **had** reluctantly **returned** to their houses, Kino squatted beside [...]*
*Quando o médico **partiu** e todos os vizinhos, com pouca vontade, **regressaram** a casa, Kino agachou-se ao lado das [...]*
 When the doctor **left** and all neighbours, reluctantly, **returned** home, Kino squatted beside [...]

A similar case is involved in one (incidentally, quite complex) instance of sequence of tense, see example (73), where pluperfect in Portuguese could not be used:

- (73) *"Do you think they would take you back alive to say they **had stolen** it?"*
*– **Pensas que te levarão vivo para dizerem que ta **roubaram**?***
 Do you think they will take you alive to say that they **stole** it from you?

Looking now at the opposite translation direction, we see a kind of converse regularity involving *quando*-clauses emerge. To express simultaneity in Portuguese between a main clause in Mais que Perfeito and a *quando*-clause, the latter must also have Mais que Perfeito, while in English only the simple past is idiomatic, see examples (74)-(75).

- (74) *...meditou em porque esquecera a bacia, a não vira quando se **sentara**...*

... he wondered why he had overlooked the basin, had not seen it when he **sat** down...

he considered why he had forgotten the bowl, had not seen it when he **had sat** down

- (75) *Quando, enfim, **entrara** nele, também como agora se encostara à porta*
*When finally she **entered** it, she had leaned, even as now, against the door...*

When, at last, she **had entered** it, also like now she had leaned against the door

Another common reason for having Mais que Perfeito in Portuguese main clauses stems from a wish to describe relevant traits of characters by narrating events in a vague, indefinite, past. This literary tradition seems to be alien to English, and thus translators simply use the simple past with corresponding adverbs, as in examples (76)-(77).

- (76) ... ***chegara**, de uma vez, a arriscar que ...*
 ... and even **ventured**, on one occasion, the opinion that ...

he **had gone** so far as, once, to venture that

- (77) ... *ele **acabara** conhecendo todas as almas de que dispunha aquela pequena parcela da humanidade.*
 ... he **ended** up knowing all the souls contained in that small parcel of humanity.

he **had ended** up knowing every soul which that small portion of mankind had at its disposal

Some translations, however, bring a distortion of the meaning conveyed, since they (may) describe a past until the present situation, while the Portuguese originals clearly only refer to a past delimited time. Examples (78)-(79) have their possible implications in square brackets.

- (78) *Mas **havam concordado** ambos em que ...*
*But they both **agreed** that ...*

But they **had agreed** both that [they may no longer agree]

- (79) *O amor para ele **fora** carne e espírito, ...*
*Love **was**, for him, flesh and spirit, ...*

Love to him **had been** flesh and soul [he has no love now]

Example (80) features a mistranslation, which can be readily understood since it involves both a complex Portuguese aspectual class (Aquisição) and já:

- (80) *os habitantes da aldeia (que ele já **conhecera** como antepassados deles mesmos ou de animais que lhes circulavam ao pé) ...*
*the villagers (whom he **knew** as their own ancestors or as animals that gathered around them) ...*

which he **had** once **met** (OR which he **had** already **known** before) as their own ancestors or as ancestors of animals gathering near

Looking now into the addition of pluperfectness by the English translator, we find the converse cases: use of the event that produces the result described in the original; indefinite time marking; conveying a series of events when the original is in Imperfeito.

The only new pattern I was able to observe concerned English gradual achievements in the passive, see example (81). Given that Portuguese has no corresponding aspectual class (a gradual achievement renders both a result and an associated gradual process), the translations employed Mais que Perfeito:

- (81) *The little hole **was** slightly **enlarged** and its edges **whitened** from the sucking, but the red swelling extended farther around it ...*

*A pequenina mancha **aumentara** um pouco, o contorno **desinflamara-se** com a sucção, mas toda a bolha vermelha se alargara...*

the little spot **had increased** a little, the contour **had disinfamed** with the sucking, but the whole red bubble had enlarged

This shows the need to consider the English passive as a device focusing on the result, while the four Portuguese passives seem to place more attention on the way the state is felt (and created) and still leave room for Mais que Perfeito to be employed when the result is in question. In this context, it is interesting to have a closer look at example (82), where the original text hints at a succession of events involving payment to a greater extent than the English rendering.

- (82) *... vindos não se sabia de onde, pois todos os servos **estavam** bem **pagos**.
... not knowing where they were coming from, for all his servants **had been** well **paid off**.*

come from God knows where, since all servants **were being** well **paid**

However, the most unexpected finding was that an extended period up to the present (the hallmark of present perfect, see section above) is also sometimes conveyed in Portuguese, with Imperfeito. Some of the sentences translated by pluperfect fail then to express this temporal meaning, as examples (83)-(84) illustrate.

- (83) *teria finalmente a ressurreição da sua vida que o peso de uma pedra imensa, que era o seu destino, não **permitia** que surgisse e caminhasse.*

*would at last experience the resurrection of her life that the weight of an immense stone, which was her destiny, **had not allowed** to rise and go forth*

she would have at last the resurrection of her life which the weight of a huge stone, her fate, **was not allowing** to appear and walk

- (84) *Não fôra uma serpente vagabunda que, por esse tempo, se instalara entre as raízes da árvore, e lhe **fazia** o favor de comer aquilo tudo,*

*If it had not been for a loafing snake who during this time **had** settled in*

*among the roots of the tree and **done** him the favor of eating it all up,*

Had it not been a homeless snake who, by then, had moved into the tree roots, and **was doing** him the favour of eating all that

Note that only in the fourth study this “new” feature of Imperfeito was discovered, after focussing on temporal periods and uncovering partial correspondences for other tenses. While this particular property suggests yet another reason why pluperfect might be chosen as the translation of Imperfeito – namely, by being the past of the perfect and therefore conveying a period until then –, it also brings to the fore other hints of addition and loss in otherwise tolerable translation strategies. In fact, although past perfect may convey repetition, it fails in example (85) to convey its probably most important message: until now. And that, in example (86), the practical joke is still admired at the time of the narrative.

- (85) *E disse-lhe que a tua tença estava atrasada e que não a **pagavam***
*And I told him that your pension was late and **had not been** paid*
 And I told her your pension was late and they **kept not paying** it

- (86) ***Era** por todos risonhamente **admirada** aquela partida feita ao Rebelo, um coitado que era casado e chefe de posto fronteiriço.*
*They **had** all gleefully **admired** that trick played on Rebelo, a poor fellow who was married and head of a border post.*

It **was** by all cheerfully **admired** that practical joke made to Rebelo, a poor guy who was married and chief of a border post.

Surprisingly, the same loss and the same temporal implications could be seen in translation from Perfeito to pluperfect. It is true that example (87) is vague regarding when “the years” finish, but the main point is that the whole sentence is relevant up till now.

- (87) *Eu, que **leve**i anos a curar-me dos sustos, fiquei estarecido*
*I, who **had spent** years recovering from the fears, I was appalled*
 I, who **have taken** years to heal from the frights, I got appalled

In example (88), on the other hand, the translator seems to be too strictly obeying the indirect speech sequence of tense rules, when the Portuguese writer on purpose has violated them. (Another case of similar conventions in two languages, but no parallel use.) The Portuguese sentence, in addition to reporting speech, is also asserting the fact that, until now, the work has not been finished, while the English one merely reports the speech event as if the “I” character had no beliefs about the truth of what she had just said.

- (88) *E eu disse que tu **ainda não** acabaste e que logo acabas*
*And I said that you **had not yet finished** but would soon do so*
 And I said that you **haven't finished** yet and that will finish soon

A last interesting example, also a translation of Perfeito into pluperfect, concerns rendering of perspective. Example (89) describes the musings of the father arriving home, wondering whether his daughter will be there or not. In this case pluperfect in English may be employed to mark subjectiveness, mirroring free indirect speech.

(89) *Claro que **ainda não** veio.*

*Of course, she still **hadn't come**.*

It is clear that she **hasn't yet come**.

In sum, we see again that theoretically similar devices display an array of subtle, but enlightening, differences on how either language modulates meaning, and that translation practice uniquely allows us to uncover them.

5.7 Studying the I/P distinction and Perfeito

My last empirical study⁶⁵, before starting to look for models to accommodate the data, revisited the Imperfeito/Perfeito distinction and looked more carefully at Perfeito.

Table 5.19: *Ser* and *estar* in Imperfeito or Perfeito in the PE corpus

<i>ser</i>	442	Imperfeito	219	simple past	205
		Perfeito	36	simple past	33
<i>estar</i>	77	Imperfeito	41	simple past	37
		Perfeito	2	simple past	2

The first investigation focused on the *ser/estar* dichotomy and its relationship with the simple past tenses. Table 5.19 shows their distribution in the PE corpus. Perfeito is obviously the marked choice.

Perfeito with *ser* is prototypically connected with a property which has ended (and which held during a definite period). There does not seem to be a construction in English with similar properties, cf. examples (90)-(91).

(90) *a minha condição de cidadão romano, que meu Mestre não **foi***
*my Roman citizenship, which my Master did not **have***

my condition of Roman citizen, which my Master **has not had**

(91) *quando **foste** o pretor de Antióquia*
*when you **were** praetor of Antioch*

in the time you **worked** as praetor of Antioch

Next, Perfeito is also used when an event is assigned a property. An event is classified as an error in example (92), another event is qualified as beautiful in example (93). In example (94), referring to an event cotermporal with the night hours, the translator felt the need to employ a different verb.

⁶⁵ Reported in Santos (1996a: Chapter 14).

- (92) *Foi um erro monstruoso que é preciso corrigir*
It's all a monstrous mistake that must be set right
 It **was** a terrible mistake that it is necessary to correct
- (93) *Foi bonito?*
Was it pretty?
 Did you find it beautiful?
- (94) *as horas da noite foram [...] uma distração contínua*
the night hours provided a continuous distraction
 the night hours **were** a neverending fun
 Thirdly, *Perfeito* with *ser* is also used to temporally locate events, as illustrated by examples (95)-(96).
- (95) *Dizes que isso foi há quarenta anos*
You say this happened forty years ago.
 You say that **was** forty years ago
- (96) *Não foi, portanto, nessa hora que o meu divino Mestre nasceu de novo [...]?*
Wouldn't that be, therefore, the very hour when my divine Master was born again [...] ?
Wasn't thus on that time that my divine Master was born anew ?
 This last example borders with simple clefting, of which there were thirteen occurrences in the corpus. Interestingly, clefts display different tense strategies in the two languages. While in Portuguese the cleft must agree in tense with the main clause, this is not the case in English, cf. examples (97) and (92) above.
- (97) *mas foi o que eles ouviram*
but that's what they heard
 but that **was** what they heard
Perfeito with *estar* occurred only twice, and in both cases it described a temporally bound event in the past, advancing the narrative. The first translation, in example (98), also illustrates a recurring situation: manner of position is by default added in English. On the other hand, a main locative clause in English simple past seems to convey the same message as *Perfeito* with statives, as in example (99). Or, the choice may have been conditioned by the fact that the alternatives (such as present perfect) would additionally convey an erroneous “up to the present” implication.
- (98) *assim esteve um longo tempo*
thus he stood for a long time
 thus he **was** for a long time
- (99) *Esteve hoje cá o Padre Manuel*

*Padre Manuel **was** here today*

Padre Manuel **has been** here today

The Perfeito of *ser* in translation from English to Portuguese confirms the findings above. Especially interesting is example (100), where Portuguese expresses a property of an event (of speaking), and by all means not of the priest.

- (100) *and the Father **made** that very clear. The loss of the pearl was a punishment visited on those who tried to leave their station. And the Father **made** it clear that [...]*

*O padre **foi** bem claro a esse respeito: o desaparecimento das pérolas era um castigo para aqueles que tentavam abandonar o seu posto. O Padre **foi** bem claro: [...]*

The priest **was** crystal clear about it: the disappearance of the pearls was a punishment for those who tried to quit their post. The priest **was** crystal clear: [...]

In addition, a regularity was found in translations of *be* using Perfeito, which became one of the most important pieces of evidence on the differences in aspectual character of the verbs in the two languages. In fact, *be* was very often used inchoatively in English original text and, given that it could not be translated by *ser* or *estar*, other verbs in Perfeito had to be employed in examples (101)-(103).

- (101) *Now uncertainty **was** in Kino*
*A dúvida **apoderou-se** de Kino.*

The doubt **took** over Kino.

- (102) *Instantly Juana **was** on her feet and her hand lay on his arm.*
*De um salto, Joana **pôs-se** de pé e pousou a mão no braço de Kino.*

With a jump, Juana **stood** up and put the hand on Kino's arm.

- (103) *And she **was** silent, for his voice was command.*
*Ela **calou-se**, porque a voz dele era uma ordem.*

She **shut** up, because his voice was an order.

Several other strategies were used to solve the incoherence of lexical and aspectual values in *be* and *ser*. A common one was addition of perception, already illustrated in example (45). Another was addition of intellectual subjectivity, as in examples (46) above and (104) below. In those cases, the translation expresses an (added) thought event, but misses stating the original situation.

- (104) *He looked then for weakness in her face, for fear or irresolution, and there **was** none.*

*Ele procurou então no rosto da mulher fraqueza, medo ou irresolução. Nada disso **achou**.*

He searched then in his wife's face weakness, fear or irresolution. He **found** nothing of the sort.

Non optimal translations are, again, additional proof that several meanings or narrative strategies are involved at the same time in the use of *be*. In example (105), the English sentence describes a background condition, while the translator read *be* inchoatively and expressed restlessness as the result of the wind.⁶⁶

- (105) *The wind blew freshly into the estuary, a nervous, restless wind with the smell of storm on its breath, and there **was** change and uneasiness in the air.*

*Um vento frio, irrequieto, arrastando consigo o cheiro da tempestade, soprou no estuário. A instabilidade e a inquietação **ficaram** no ar.*

A cold and restless wind, dragging the smell of tempest, blew in the estuary. Instability and unquiteness **remained** in the air.

Finally, the most inchoative of Portuguese verbs, *ficar*, was translated in example (87) by English *be*.

I did not do a study of the addition of *be* in translation (since I had not annotated the main verb of the translation). Still, example (106) is a striking demonstration that Imperfeito with events sometimes means property assignment.

- (106) *Gostavam pela certa, **cantava** bem.*
*Certainly they liked her, she **was** a good singer.*

For sure they would like, she **sang** well.

Looking at the Imperfeito/Perfeito opposition in general (and not only with stative verbs), I was unable to consider every instance due to the huge number of cases involved. Thus, I first looked at the non-past simple translations of Perfeito. Three cases of Perfeito were rendered by the gerund, showing that gerund clauses in English can also advance the narrative (therefore working as acquisitions too), see example (107).

- (107) *Apurando a vista, examinou-o minuciosamente, e **deteve** os olhos no pequeno golpe no pescoço,*
*Concentrating his gaze, he examined him minutely, **detaining** his eyes on the small cut on his neck,*

Straining his eyes, he examined him minutely, and **detained** the eyes on the small cut in the neck,

Another interesting case was related to the negated verb *imaginar*, which seems to be construed as a stative predicate up to the present in Portuguese, but requires the conditional and the perfect in English; cf. example (108).

- (108) *Não **imaginei**, porém, que fosse preciso vir da Síria conspirar em Roma, Saulo...*
*I **wouldn't** have imagined, however, that it was necessary to come from*

⁶⁶ My suggested translation is *O vento soprava frio no estuário, um vento nervoso e agitado com cheiro de tempestade, e **havia** mudança e inquietação no ar* ('... and there was change and unquietness in the air').

Syria to conspire in Rome, Saul...

I **could not imagine**, however, that it was necessary to come from Syria to conspire in Rome, Saul...

Then, concerning translations into the simple past, I selected 60 random cases of Imperfeito and as many of Perfeito in the PE corpus, and sampled the EP corpus for 50 examples of simple past into Perfeito and 50 cases of simple past into Imperfeito. The main point of this sampling was to identify cases not yet investigated, as well as to get a rough idea of the distribution of the meanings involved in these tense pairings.

As regards Imperfeito to simple past, one fourth were considered habituais, one third stative, one fourth were considered to indicate simultaneity or temporal inclusion, and one sixth perspective. Two new situations were, in any case, identified: comparative clauses and continuous, homogeneous, time-taking processes (a kind of Obras generally translated by a gradual achievement in English, see examples (109)-(111)).

- (109) *O frio, a água, a treva, **penetravam**-no.*

*The cold, the water, the darkness, **penetrated** him.*

Cold, water, darkness, **were penetrating** him.

- (110) *embora seguro do caminho, senão da noite e da solidão que se **confundiam** cada vez mais e a perder de vista.*

*sure of his way although less so of the night that **mingled** with the solitude, fast becoming one.*

although he was certain of the way, if not of the night and solitude too which increasingly **mixed** and as far as the eye could reach

- (111) *Mas, das profundezas da árvore, **subia** um delicioso perfume.*

*But from the depths of the tree there **arose** a delicious perfume.*

But, from the tree depths, a wonderful smell **was rising**

In example (111), Imperfeito conveys not only the gradual development of the event, but also the possibility of its being in the background, and that it is the perspective of the main character which is being reported. The English translator may have had to choose and has thus lost some of the possibilities.

The converse translation relation, simple past into Imperfeito, was often employed for activities, or for actions which the translator considered as simultaneous with the previous one. In the four cases chosen to illustrate this situation, displayed in examples (112)-(115), world knowledge and subjective interpretation have to be involved for a proper translation. The cases are ordered from my total agreement with the translator, in example (112), to radical disagreement, in example (115).

- (112) *Greedy fingers went through his clothes, frantic fingers **searched** him*

*Dedos ávidos percorriam-lhe o fato, dedos frenéticos **rebuscaram**-no.*

Greedy fingers were running through the clothes, frantic fingers **ransacked**

him/them.

- (113) *"We must go west," he said, and his eyes **searched** the stone shoulder behind the cleft.*

*– Temos de ir para oeste – disse ele, enquanto **fixava** a proeminência da rocha do outro lado da garganta.*

*"We must go west", said he, while he **stared** at the rock shoulder on the other side of the gorge.*

- (114) *The hot sun beat on the earth so that Kino and Juana moved into the lacy shade of the brush, and small gray birds **scampered** on the ground in the shade.*

*O sol ardente causticava tanto a terra que Kino e Joana tiveram de procurar a sombra rendilhada dos arbustos, donde pássaros **fugiam** apressados.*

The burning sun cauterized so much the earth that Kino and Juana had to look for the bushes' lacy shadow, from where birds **were escaping** hastily.

- (115) *The third man **scrabbled** away like a crab, slipped into the pool, and then he began to climb frantically*

*O terceiro homem **arrastava-se** como um caranguejo, escorregara até à lagoa e começou a trepar freneticamente*

The third man **was dragging** himself like a crab, he had slipped to the lagoon and he started to climb up frantically

As regards *Perfeito* into simple past, also some interesting cases were uncovered, namely those in which *Mudanças* (verbs denoting a change) in *Perfeito* get translated by an activity where its endpoint (or startpoint) is not stated; see example (116).

- (116) *Marco Semprônio **veio** até junto do vulto agachado. Sob o manto, sem desagasalhar-se, tirou da bolsa moedas que tilintaram no lajedo.*

*Marcus Sempronius **approached** the crouching figure. Beneath his mantle, without loosening it, he took coins from his pocket that tinkled on the flagstone.*

Marcus Sempronius **came** until near the crouching figure. [...]

In the case of *Obras*, *Perfeito* makes them bounded, while simple past fails to convey inception, see examples (117) and (118):

- (117) *nela dois velhos imemoriais, emaciados, de repas gordurosas, **fitaram-me**. from it two long-forgotten and emaciated old men, with thin greasy hair, **stared** at me.*

on it two ancient old men, emaciated, with greasy hair, **set** their eyes on me

- (118) *O homem **olhou-o** com o mesmo olhar de polícia, e depois, com espanto, fixou as flores.*

*The man **looked** at him with the look of a policeman and then, surprised,*

stared at the flowers.

The man **gave** him the same policeman's look, and then, with astonishment, gazed at the flowers.

In fact, event completion is only rarely expressed in English through the use of particles, as in example (119), even though *sat down* may be interpreted (in other contexts) as stative.

- (119) *o velho veio como uma seta, **sentou-se** ao meu lado, e perguntou-me flautadamente:*
*the old man approached me like an arrow, **sat down**, and asked me sprightly:*
 the old man came like an arrow, **sat** beside me, and asked me in a careless way

Finally, I looked at frequency adverbials and words ending in *-ever* with the simple past, amounting to 20 cases, further specified in Table 5.20. The reason for this particular topic was the vague quantification associated with some Portuguese tenses (Imperfeito, PPC), as well as the wish to see whether temporal reference values added by this kind of adverbials were kept or changed in translation.

Table 5.20: Rendering of simple past with frequency adverbials in the EP corpus
 (“corr.” stands for corresponding adverbial)

adverbial	frequency	translation
<i>-ever</i>	3	Imperfeito + corr.
<i>sometimes</i>	2	Imperfeito + corr.
<i>now and then</i>	2	Imperfeito + corr.
<i>nearly always</i>	2	Imperfeito + corr; <i>a maior parte</i>
<i>as always</i>	3	Imperfeito + corr; Presente + corr; MQP
<i>as always when</i>	1	Imperfeito + corr.
<i>always</i>	1	<i>normal</i>
<i>never</i>	3	Imperfeito; Perfeito <i>nunca mais</i> ; <i>sem</i>
<i>never again</i>	1	Perfeito + corr.
<i>every time</i>	1	Imperfeito + corr.
<i>in the times when</i>	1	<i>na época das</i>

As expected, the vast majority of translations employed Imperfeito. The most interesting feature was the rendering of *always* as an assertion of normalcy in example (120) and as an atemporal rule – and thus stated in Presente – in example (121), again a translation choice:

- (120) *And he didn't know whether or not it was **always** a little blue.*
*Não sabia se era **normal**, ou não, aquele azul.*

He did not know whether it was **normal**, or not, that blue.

- (121) *and on the fear came rage, **as** it **always** did.*

Mas, como sempre acontece, o medo trouxe a raiva.

But, as **always** happens, fear brought rage.

Let me also revisit the case of negative contexts, where Perfeito covers a definite period – see example (122); while Imperfeito describes a property of a past time, or of a person in a past time; cf. example (123):

- (122) *The doctor shrugged, and his wet eyes **never** left Kino's eyes.*

*O médico encolheu os ombros, e os seus húmidos olhos **nunca mais** se afastaram de Kino.*

The doctor shrugged, and his wet eyes **no longer** moved away from Kino.

- (123) *The doctor **never** came to the cluster of brush houses.*

*o médico não vinha **nunca** àquele lugarejo de cabanas.*

The doctor would **never** come to that hamlet of huts.

Another observation was that English seemed to have a kind of “locative” way of expressing repetition that, in Portuguese, would more often than not be expressed with frequency adverbials and Imperfeito, as examples (124) and (125) show.

- (124) *Juana walked in darkness **for a moment** and in light **the next**.*

*E Joana caminhava, **ora** às escuras, **ora** iluminada.*

And Juana walked, **now** in the dark, **now** in light.

- (125) ***Here**, one would pause and look at the earth, while the other joined him.*

***De vez em quando** um parava a inspeccionar o chão e o outro aproximava-se.*

Once in a while one stopped to inspect the ground and the other approached him.

Finally, I also looked at the 41 cases of simple past translated into Imperfeito + *já*. Given that no case of *already* + simple past was found in the corresponding source text, and only one case of *no longer* (*já não*) was identified, further investigation was needed to uncover the reasons for this combination in translation.

Several different (not mutually exclusive) explanations to adding *já* emerged, namely, the translation of English contrastive *now*, the rendering of temporal clauses implying immediacy, the signalling of predictability, and the need to deal correctly with acquisitions. I will illustrate the two last cases here.

Predictability is added in the translation. Here again, we have cases where arguably another interpretation could be chosen. Example (126) implies that it was predictable that they would get confused, while example (127) assumes that the trackers would come near.

- (126) *and they discussed the great theme of the morning. They **did** not know, it seemed a fine pearl to them,*

*discutindo o grande acontecimento da manhã. **Já** não **sabiam** que pensar.*

Parecia-lhes que a pérola era de fina qualidade,

discussing the great event of the morning. They **no longer knew** what to think. It seemed to them that the pearl was of high quality,

- (127) *When the trackers **came** near, Kino could see only their legs and only the legs of the horse from under the fallen branch.*

*Quando os batedores **já vinham** perto, Kino, debaixo do ramo quebrado, **já** só lhes **via** as pernas e as pernas do cavalo.*

When the trackers **were already coming** near, Kino, under the fallen branch, **no longer saw** but their legs and the horse's legs.

The other relevant observation was that *já* was often employed with Aquisições, to locate in time an (otherwise) atemporal state, holding before another time or event. In examples (128)-(129), *já* also stresses that the starting event precedes the time mentioned by a short time. That is, the learning event occurred shortly before dawn in example (128), and the getting of the baby occurred shortly before 'now' in example (129).

- (128) *Thus, in La Paz, it **was known** in the early morning through the whole town that Kino was going to sell his pearl that day.*

*Assim, mal a manhã rompera, **já** em toda a cidade de La Paz se **sabia** que Kino iria vender a pérola naquele dia.*

So, hardly had the morning broke, **already** in the whole city of La Paz people **knew** that Kino would be selling the pearl that day

- (129) *But Juana **had** the baby in her arms now.*

*mas Joana **já tinha** o filho nos braços.*

but Juana **had already** the son in her arms

In example (130), the core meaning of *já* is the same, namely indicating that the state held before the just mentioned event, but there is no implication that there is any closeness in time between learning and the other event. Probably, in cases where an event cannot be repeated, *já* is intended to stress that the result (still) holds. The original text arguably does not refer to a time before hearing, but the translation is undeniably more idiomatic with *já*.

- (130) *And she nodded and smiled a little at him, for she **knew** these things.*

*Ela concordou com um sorriso, porque **já sabia** tudo aquilo.*

*She agreed with a smile, for she **already knew** all that.*

I have now exhausted the description of the individual studies and will present, in retrospect, evidence for two of my most original findings.

5.8 Acquisitions as a relevant class

It is not always possible to pin down the main ideas to specific corpus studies. For example, the concept of Aquisições occurred at a relatively early stage – as evidenced in Santos (1991). However, only much later did vague classes become

a central issue in my work, after constantly finding instances that could only be adequately described in that way.

Depending on the context, sometimes a given verb seemed to belong to class 1, other times to class 2. First, I pursued the ambiguity hypothesis, in which grammatical context selected, or demanded, a different interpretation. But systematic ambiguity is intuitively unsatisfying, and the contexts did not seem to give a consistent answer – they were only “ambiguous” with those verbs! Which was a clear indication that the “problem” was with the verbs themselves.

Only later the idea that those verbs could belong to both classes, or to a class vague between the two, occurred to me. I then tried to look for contexts where both meanings could be rendered together. By looking at the particular instances of the verbs themselves, I also ended up amassing a lot of evidence for these classes.⁶⁷

Although I have illustrated acquisitions and Aquisições in the previous sections, let me now reconstruct the history of their discovery.

The first class of puzzling verbs was position verbs. Most of the time they seemed to convey both position and the process of its coming about, therefore causing problems for the translator. In example (131), Juana had squatted before, but was still squatting, while who can tell in example (132) when the boats started lying lined? Incidentally the published translation is very good, eluding the question.

- (131) *who **squatted** on Kino's right hand*
*que se **agachara** à direita de Kino*
 who **had squatted** to the right of Kino

- (132) *And the morning of this day the canoes **lay** lined up on the beach;*
*Na manhã daquele dia os barcos **ficaram** na areia.*

On the morning of that day the boats **remained** ashore.

Quite often it was possible to find mistranslations (or choices of a less generally preferred interpretation) involving these verbs:

- (133) *After a while they **lay** down together on the sleeping mat*
*Momentos depois, **estavam estendidos**, lado a lado, na esteira.*

Moments after, they **were lying**, side by side, on the mat.

- (134) *and when the day had come, in the offices of the pearl buyers, each man **sat** alone with his little black velvet tray*
*Mal rompeu a manhã, cada um dos compradores **sentou-se** sozinho na sua loja*

As soon as the morning broke, each one of the buyers **sat down** alone in his little shop

⁶⁷ A corpus-based description of acquisitions was first presented in Santos (1997a).

- (135) *And when Juana **had seated** herself and had settled to nurse the baby, Kino went back to the road.*

*Quando Joana se **sentou** e deu o peito à criança, Kino voltou atrás.*

When Juana **sat** and gave the breast to the child, Kino returned.

One can see that the close temporal sequence of the two clauses in examples (134)-(135) is not preserved in the translations.

Verbs of movement were also found as sources of vagueness and translation inaccuracy, although, interestingly, not to the same extent. In fact, there were other translation difficulties connected with them (such as lack of manner information and of locative detail in translation), but an analogous situation to the vagueness of acquisitions persisted: The English text provides no clue as to whether start of movement or development (or both) is involved. Thus one might interpret the *when*-clause in example (136) as “when the boat started to float”, and that the rising of the shoulders in example (137) followed (and not accompanied) the utterance.

- (136) *and when the bow **floated**, Juana climbed in*
*e a proa já **flutuava** quando Joana saltou para dentro dele*
 and the bow **was** already **floating** when Juana jumped into it

- (137) *"I am sorry, my friend," he said, and his shoulders **rose** a little to indicate that the misfortune was no fault of his.*

*– Lamento muito, meu amigo – disse ele. E **erguia** um pouco os ombros para mostrar bem que não tinha culpa nenhuma daquela contrariedade.*

"I am very sorry, my friend" said he. And he **was raising** a little the shoulders to show well he had no fault whatsoever in that

Example (138) is an interesting (and a typical) case of conveying movement with particles, only possible to express in Portuguese with two clauses describing the position before and after. Incidentally, the translator imagined Juana bent, while I thought she was squatting.

- (138) *As he came through the door, Juana **stood up from** the glowing fire pit*
*Joana, que **estava inclinada** para o lume, **endireitou-se***

Juana, who **was bent** towards the fire, **raised**

Likewise, in example (115) above, the translation only specifies manner of movement, and not its goal.

English activities in general (of which movement verbs are the most conspicuous subclass) were actually found to display this very same behaviour very often. In example (139) the translator again has to choose, while in example (140) he employs *Mais que Perfeito* as a vagueness preserving device.

- (139) *It stopped, and its tail rose up over its back in little jerks and the curved thorn on the tail's end **glistened**.*

*imobilizou-se. Em pequenas sacudidelas, foi erguendo a cauda no ar, e, no extremo dela, o ferrão curvo **brilhou**.*

it stopped. In little jerks, it gradually raised the tail in the air, and, in its extreme, the curved thorn **started to glisten**

- (140) *only a few small barnacles **adhered** to the shell*
*só algumas pequenas lapas se lhe **tinham agarrado***
 only some few limpets **had stuck** to it

As noted in previous sections, the verb *be* was overwhelmingly interpreted as inception, and the same occurred with perception verbs. This was also the case with *have* (supposedly yet another prototypical stative verb), as example (141) shows.

- (141) *Then, snarling, Kino **had** it, **had** it in his fingers, rubbing it to a paste in his hands.*
*Então Kino soltou um grito e **agarrou-o**, **agarrou-o** com os dedos, esborrachou-o nas mãos.*

Then Kino gave a shout and **caught** it, **caught** it with his fingers, rubbed it in the hands

This led me to consider also the passive, the pluperfect, and even gerund clauses, as displaying acquisitional behaviour in English.

Now, it is very interesting to stress that some Portuguese verbs also show a similar behaviour (but, even more interestingly, in general not the translationally related ones). Actually, the discovery process occurred in tandem in the two directions of translation.

The whole process started by acknowledging that there were some verbs which denoted both a state and an event of change of state (a *Mudança*). They were dubbed *Aquisições*, and they were first found in an attempt to formalize the Imperfeito / Perfeito dichotomy. Even though I give some English evidence in Santos (1991), it is clear that it was the Portuguese examples that gave origin to my puzzlement.⁶⁸

The first contrastive fact I had to deal with was the translation of one lexical item into different ones depending on the way to conceive it, often signalled by the tense. Is *conhecer* better rendered by *meet* or *know*? Is *saber* closer to *know* or *learn*? In the cases where the vagueness can be preserved in Portuguese (such as is the case in *Mais que Perfeito*), the translator has to choose. Examples (142)-(143) illustrate that the difference is rather subtle... Nothing is said as to the degree of intimacy between the genius and the old man in example (142), but for all we know he could have known him very well, as in example (143).

- (142) *deitado virtuosamente à porta da casa de um seu neto (neto do velho que*

⁶⁸ It should be noted that the term “acquisitions” in Santos (1991) does not refer exactly to what *Aquisições* and acquisitions came to denote later. In fact, they were still “language-independent” classes and, instead of considering them vague, I gave them a complex temporal profile.

*ele **conhecera** ao instalar-se na árvore)*
lying virtuously at the door of the house belonging to one of its grandsons
*(grandson of the old man the genie **had met** when he moved into the tree)*
lying virtuously at the door of the house of a grandson of his (grandson of
*the old man he **had known** when he settled there)*

- (143) *Marco Semprônio recordou os imperadores que a todos **tinha conhecido** de perto*
*Marcus Sempronius reflected on the emperors he **had known** intimately*
 M.S. remembered all emperors he **had known** closely

Position verbs in Portuguese are also prone to this double interpretation, especially when appearing in subordinate clauses. In principle, English should not have a problem here, because position verbs are also acquisitions in English, but examples (144) and (145) show different translation patterns (perhaps distinguishing between permanent vs. temporary location).

- (144) *o manto se pegava nos espinhos da roseira que a **envolia**.*
*caught his mantle on some thorns of the rosebush that **encircled** it.*
 the mantle was adhering to the thorns of the rosebush **surrounding** it (a statue)

- (145) *Recolheu sobre o seio a mão que **pendia** para fora do catre*
*She drew to her bosom the hand that **was hanging** over the cot*
 She put over her bosom the hand which **hang** out of the cot

Another interesting case of Aquisições is verbs describing intellectual states, such as *apetecer* in example (146). In *Perfeito*, they may denote the inception of the state, or a delimited time period where the state held, while in *Imperfeito* the mere state is asserted to hold. The English translation accommodates the two *Perfeito* meanings, but it also allows a third interpretation (impossible in the original), namely that the wishing state overlaps with previous and subsequent events.

- (146) ***Apeteceu-lhe** estar no café a encarar o Pinheiro Couto*
*He **wished** he were in the café to confront Pinheiro Couto*
 He **felt** like being in the café facing Pinheiro Couto

Finally, the most typical case of Aquisição is *ficar*, a verb which can be used as aspectualizer or main verb and, rather surprisingly for a non-native speaker, expresses a change into a state and the state of remaining in this state, respectively illustrated in examples (147)-(148).

- (147) *Ou ele [...] **ficara** um pouco confuso.*
*Or else he [...] **had become** a little confused.*
 Or he [...] **had gotten** a little confused
- (148) *O frade, com boas palavras, despediu-os, e **ficou** na estrada a vê-los*

afastarem-se

*The friar, with good words, bade them farewell and **stood** in the road watching them leave*

The friar, with good words, took leave from them, and **remained** in the road seeing them go away

The difficulties of dealing with *ficar* are clear in the next three examples. While the translation of example (149) simply yields the wrong interpretation, example (150) fails to express the clear change in position that is the focus of the original, rendering only the final situation. Finally, example (151) is an obvious instance of translationese.

- (149) *Marco Semprônio **ficou** pensativo, e depois fitou o rosto moreno, Marcus Sempronius **remained** pensive, and then his eyes rested on the visitor's dark face*

M.S. **became** thoughtful, and then stared at the dark face

- (150) *Tropeçou e caiu. O caminho era um lamaçal em que **ficou** sentado, dentro de um charco.*

*He tripped and fell. The path was a swamp in which he **sat** mired.*

He stumbled and fell. The path was a mire into which he ended up sitting inside a puddle.

- (151) *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 lay **on**, savoring a discomfort that was exhausted repose

Summing up, let me stress that apparently similar phenomena in two languages do not necessarily mean that translation between them occurs without problems. In fact, the verbs discussed in this section may even produce more translation problems than those belonging to the other classes. After all, the classes were identified on the basis of repeatedly finding problematic translations involving these verbs.

5.9 Obras and Mudanças: a different way of classifying events

Let me now turn to another puzzle which bothered me for a long time and which eventually resulted in the proposal of a different organization of the event realm in Portuguese.

It was perhaps due to the need to explain the absence of the present perfect and the notion of result that I got the idea that Portuguese aspectual classes simply do not distinguish between activity and accomplishment. But this idea only imposed itself after having found that the translation of activities consistently deviated more from the original than the translation of other aspectual classes. On

the other hand, accomplishments were often translated by *Mais que Perfeito*, again a sign that additional devices seemed to be required in that case.

Of course, it was well known, as testified in Santos (1991), that, on the one hand, most distinctions between present perfect and the simple past were neutralized in Portuguese, both being translated into the *Perfeito*. But, on the other hand, as far as statives were concerned, Portuguese would split the present perfect “space” in two, rendered either by *Presente* or *Perfeito*. This showed, first, that the temporal realm was conceptualized differently in the two languages, but later on I realized that also the aspectual categories involved were different.

The picture that eventually emerged is this: Portuguese has two kinds of event types. The first kind, *Obras*, develops in time; the second kind, *Mudanças*, involves a change. It is grammatically irrelevant whether (events linguistically realized by) *Obras* have a result or not; and it is equally grammatically irrelevant whether (events linguistically realized by) *Mudanças* take time or not.

The first contrastive evidence I found to support the category of *Mudanças* was movement verbs in Portuguese (or better, change of place verbs), which do not show any concern for the manner of movement. It is enough to express a change (of place). In example (152), for instance, it is not even idiomatic to employ the verb *andar*!

- (152) *Depois, foi a uma mesa próxima*
He then walked to a nearby table

Then, he **went** to a nearby table

The second observation that called my attention was that *Mudanças* tended to be translated by situations that stretched over time, cf. examples (153)–(154), as if English had to describe reality as it is conceptualized by English, no matter the lack of specification in the original text.

- (153) *e a voz tornou-se-lhe amarga para acrescentar: –*
and his voice grew bitter as he added, "
and his voice turned bitter to add:"

- (154) *Assustou-se, ficou mesmo estranhamente assustado.*
He grew frightened, he actually felt a strange fear
 He **got** frightened, he got even strangely frightened

Mudanças are rather like English achievements. What is interesting is that, lexically, there are significant mismatches between what is conceptualized as an achievement and a *Mudança* in the two languages.

In turn, *Obras* show that result is simply not an issue in Portuguese. Events occur, but their occurrence is enough. It is more important to state causes, regularities, reasons, properties, than it is to specify actual facts (and their results).

So, there is no device to select the result, or the activity part, (contrary to English, which has perfect and progressive, respectively), since grammatically an *Obra* has neither. Still, one can locate something in the middle of an event that takes time (i.e., an *Obra*) – and thus the *Progressiva* shares this meaning with its

English counterpart – and one can insist on a vague succession of similar events up to the present, and thus use the PPC. Both Progressiva and PPC are more temporal in kind than their English “counterparts”.

This contrast between English and Portuguese is clearly shown in the fact that Obras are difficult to render in English, as their common (and poor) translation into (weakly related) achievements reveals, cf. examples (155)-(157).

- (155) *O caso **dera** que falar,*
*The episode soon **became** famous*
 the case **had given** origin to much talk
- (156) *que se **interessara** demasiado pela homenagem,*
*who had **become** too **interested** in the commemoration*
 who **had been** too **interested** in the homage
- (157) *onde agora, como não onde o frade se **distráira**, as pedras eram tão numerosas*
*where now, unlike when the friar **had become** distracted, stones were as numerous*
 where now, not like where the friar **had not paid attention**, the stones were so many

Another strategy for their translation is to add manner and render them as English activities instead, cf. examples (158)-(159). In both cases, however, the translator has to make assumptions not necessarily supported by the original text. In example (158) the narrator might be running, and in example (159) the character may not have really worked, but could have given speeches, or even opposed the project actively.

- (158) ***Passei** entre as mesas empilhadas.*
*I **walked** among the stacks of tables.*
 I **went by** between the stacked tables
- (159) *o Gustavo Dorés **revelou-se** incansável.*
*Gustavo Dorés **worked** untiringly on the project.*
 G.D. **revealed** himself tireless

Yet another translation possibility is to convey the lack of result of an Obra by the progressive of a corresponding accomplishment, as is the case in examples (160) and (20). Very often, however, an Obra in Imperfeito is rendered by pluperfect, as described in the beginning of this chapter, cf. examples (2)-(4).

- (160) *Porque era afinal, e agora, um ladrão temível que o povo **procurara**.*
*For it had finally become, now, a dreadful thief whom the people **had been hunting**.*
 Since, after all, he was now a terrible thief people **had chased**

Interestingly, negation again makes translation more acceptable. In example (161), a negated *Obra* is tolerably similar to a negated accomplishment.

- (161) *os olhos [...] não perscrutavam, apenas alongavam por ela dentro imagens*
his eyes [...] sought out nothing, merely stretched along the distance the images
 the eyes did not **peer**, they only stretched into it images

In example (162), however, when negation co-occurs with the (habitual) Imperfeito, representing quantification and expressing a property (of the mentioned *Obras*), English has to resort to the negation of an ability, which is stronger than the original.

- (162) *Agora, não mais lia, já não escrevia.*
Now he could no longer read, he could no longer write.
 Now, he **did** no longer read, no longer write

Conversely, it is often the case that *Obras* are employed to render accomplishments, and then additional devices must be found by the Portuguese translator to convey result. In example (163), contrastive *agora* ('now') was inserted (implying that, before, the situation was different), while in example (164) a degree adverbial was added.

- (163) *All manner of people grew interested in Kino – people with things to sell and people with favours to ask.*
Toda a gente agora se interessava por Kino. Toda a gente tinha coisas para lhe vender ou favores a pedir-lhe.
 Everyone **was now interested** in Kino. Everyone had things to sell him or favours to ask him.

- (164) *The way began to rise, and as it did the rocks grew larger.*
O caminho começou a tornar-se íngreme e as pedras eram cada vez maiores.

The path began getting steep and the stones **were each time** larger

Finally, there were some cases of *Obras* as translation of activities, generally when the English verb had no result, and thus was closer in meaning to an *Obra*; see examples (165)-(167).

- (165) *Kino stood in the door, filling it, and hatred raged and flamed in back of his eyes*
Kino, de pé, barrava a entrada, com o ódio e a ira chispando no fundo dos olhos.
 Kino, standing, **obstructed** the entrance, with the hate and the rage flaming in the back of the eyes

- (166) *In his house Kino squatted on his sleeping mat, brooding.*
E Kino, em casa, acorado na esteira, meditava longamente.

And Kino at home, squatted on the mat, **meditated** for a long time

- (167) *Kino lay on the ground, **struggling** to rise, and there was no one near him.*
*Kino jazia por terra, **fazendo** por erguer-se, e não havia ninguém junto dele.*

Kino lay on the ground, **working** to rise, and nobody was near him

While this different conceptualization of events seemed to explain most translation patterns, it should be emphasized that it was the result of several distinct – and in principle, unconnected – pieces of evidence that were collected while performing different studies concentrating on different research questions.

That several studies help piece together evidence seems to be one of the most important conclusions regarding corpus-based work, to which I will turn in the last section. Before that, however, I will have a quick look at temporal reference.

5.10 Temporal reference

I have commented on temporal reference in many places in the present and previous chapters; so here I want to have a more systematic look at this, so far, neglected part of the tense and aspect system.

In fact, most of the observations made in connection with temporal periods subordinated them to aspectual or even perspectival considerations, so it is time to look at temporal specification in its own right.

It has already been noted, in connection with examples (57)-(58), that a translator may have to choose between two parts of a larger temporal period specified in the source text. The same happens when the source language is vague regarding the location of the period in question: example (168) is vague about whether the centuries-long period leads up to now (as translated), or whether it is about a period wholly in the past.

- (168) *For centuries men **had dived** down*
*Há séculos que os homens **mergulham***

It's centuries since the men **dive**

Example (169) presents an example of the converse situation: the source language employs a compact way to refer to what, from the point of view of the target language, are distinct periods, namely, a period which includes both the past and the present (of the narrative) – or at least it can be interpreted like that. Translation into Portuguese, however, must choose one of two (or employ two clauses).

- (169) *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

It is also clear that the way some constructs (like *há*, *for*, and particular tenses) work is inextricably related to the aspectual categories relevant to each language.

It has been mentioned above that English uses *for* both to express duration and distance from now (therefore producing vague utterances from a Portuguese point of view), and that this goes hand in hand with the meaning of the perfect (that expresses relevance irrespective of exactly when something took place). Portuguese, on the other hand, has no grammatical marker of relevance or result, and therefore duration of such a vague character cannot be measured, even though it can be expressed with the PPC.

Rather, Portuguese uses *há* plus Presente (holding until now) or *há* plus Perfeito (happened *x* time from now) to express precise localization in time, always measured from a present point of view (now). (Incidentally, *há* is ungrammatical with PPC, which describes an imprecise period up to now, an extended now.) The form *havia* plus Mais que Perfeito is also used in “sequence of tense” rules, as example (170) shows (although featuring an awkward translation, in my opinion).

- (170) *E ele disse que o amo estava muito arreliado contigo, que **havia** mais que muitos meses que tinha feito a encomenda*

*And he said that his master is greatly vexed with you, for it **has been** more than many months since he commissioned them*

*And he said that the master was very annoyed with you, for he had done the commission many months **ago***

Another interesting example is (171), which consists of an idiomatic expression conventionally used in Portuguese when meeting someone you have not seen for a long time, and where the understood second clause is omitted. This means that such utterances leave room for convenient continuation in each language. I would continue the Portuguese sentence as “*não nos vemos*” (Presente), and the English as “since we last met”.

- (171) – ***Há** quantos anos!*

*"It's **been** so many years!*

There are so many years since!

Interestingly, *ago*, which is a “similar” device in English, can only be used to unambiguously locate something completely in the past, and must therefore be used with the past simple (or pluperfect) and never with the present perfect.

Another difference, also connected to these duration markers, is this: Portuguese distinguishes sharply between the precise localization of an event in time (*há*), and the measurement of its duration (*durante*). *Durante*, however, is also used to specify inclusion in larger temporal periods; see examples (172)-(173). These two different uses may be unified by saying that *durante* is vague between specifying surrounding or exact temporal periods.

- (172) *e um silêncio se demorou na biblioteca, **durante** o qual se ouviram,*

abafados, no crepúsculo que punha sombras pelos cantos, os ruídos domésticos da 'vila'.

*and in the library there lingered a silence **through** which could be heard, in the twilight that brought shadows to all corners, the muffled sounds of domestic life in the villa.*

and a silence lasted in the library, **during** which were heard, muffled, in the crepuscule which put shadows in all corners, the domestic sounds of the villa

- (173) *durante o dia, contudo, batia sempre
nevertheless, **during** the day he always knocked
during the day, however, he would always knock*

Now, *durante* is only felicitous with Obras and Estados when used to measure duration (incidentally, a puzzling observation already made in Santos (1991), and now possible to explain with reference to different aspectual classes), but neutral when qualifying a period temporally surrounding the event. In English, however, while *for* applies to activities and states (with the perfect it applies to the result state) to measure duration, *during* has to be used to identify a period instead of simply measuring it.

The obvious conclusion concerning temporal reference is that again we face only an illusory similarity between the two languages. The two presents (how the present time is conceptualized) differ, as well as the two languages “extended nows”: Even though both languages have devices to express an extended now, they are used in different contexts. English uses perfect with *for* mainly in connection with the result of a single action, while Portuguese employs the PPC for conveying a repetition that threatens to become a habit and can undergo no precise measurement.

Finally, temporal clauses have also been said to differ as to what they express and which information they co-occur with. I end the many examples in this chapter with the first sentence of *The Pearl* in example (174):

- (174) *Kino awakened in the near dark.
Quando Kino acordou era de noite.
When Kino awoke, it was night.*

5.11 Corpus-based contrastive generalizations

In this section I will try to sum up the more general conclusions these corpus studies have led to and what I have learned from them.

A relatively common model of how tense and aspect work is to define the final temporal-aspectual value as a combination of imports from tense, aspect, lexis, discourse, syntax, etc, $G=f(A,B,C,D,E)$, G is a function of $A...E$, which for expository purposes I will represent as “+”:

$$G=A+B+C+D+E$$

What my contrastive corpus studies showed was that this model was too simplified, because each component in turn had a bundle of different imports as well. Therefore, one should, instead, write

$$G=A_1+A_2+A_3+A_4+B_1+B_2+B_3+C_1+C_2+C_3+C_4+ \dots^{69}$$

The translator's job is then to find a sequence that renders most of these features in another language.

$$\Gamma=\Pi_1+\Pi_2+\Pi_3+\epsilon_1+\epsilon_2+\epsilon_3+\epsilon_4+\epsilon_5+X_1+X_2+X_3+X_4+ \dots$$

Here, we see that the corresponding translation often uses different blocks and therefore has both less and more, at the same time, something which I hope has already been demonstrated by the many examples commented on above. Let me go back to Keenan's (1978: 166) quotation cited in Chapter 1: "it would surely be surprising, and a very strong empirical claim, that different languages using different means to express 'meanings' always arrived at *exactly* the same end". Putting it into the equation mode above, it would be surprising that $G=\Gamma$. In fact, in most cases I looked into, G was different from Γ .

Of course, not all features (or parts) have equal importance in the different contexts. In fact, Catford's functional relevance is, again, a guideline for the translator's choice. Another guideline is typicality (or the surprise factor): One cannot disregard markedness, therefore frequency (and idiomatycy) play an important role.

One can formalize some of the translation problems that face a translator in terms of the units of meaning that are involved in the transformations represented above. So, in addition to the well-known problem of lexical gaps, one can talk about aspectual gaps, temporal gaps, sequence gaps, modal gaps, quantification gaps, negation gaps, syntactic gaps... And one still has to make room for the cases where more than one kind of gap makes itself apparent in a particular translation pair, which is quite common as exemplified in the preceding sections.

However, the word "gap" is misleading, because it presupposes that the *same* structure exists in the two languages, while we see that the systems, monolingually, have no obvious gaps – they present a coherent meaning structure to the native speaker. The point is that the structures themselves are also different. And this led to the idea of the translation network presented in the previous chapter where the translation relation connects two systems without gaps.

To round off the present chapter, let me offer some remarks on the methodology of corpus studies: I learned that it was extremely important to reassess the data several times, as more and more complicated hypotheses were being suggested. Very often, particular examples exemplify two, three, four complex differences between the languages (with consequent translator's

⁶⁹ Note that, if one considers all associations given by a lexical item or a rythmical pattern in a particular language, it may well be that this is an infinite sequence. I will restrict my analysis to the grammatical features.

decisions), which cannot be expected to be identified in just one pass over the data. To give an idea of what is involved, I tried in this chapter to illustrate how the data took me by surprise and how this led me to identify new features.

It has also been shown, in the present chapter, that mistranslations often give a valuable clue to what the translator has to choose between, i.e., mistranslations are not random. They make us focus our attention on genuine differences. Note that by “mistranslation” no criticism is implied. The word has been used here basically meaning “difference of opinion between the translator and me”. It should also be emphasized that the term has often been used in the discussion of details that in most cases would be completely irrelevant to a reader who is not a tense and aspect researcher.

In my thesis, I started to present a classification of translation pairs according to the semantic relation between them (Santos, 1995b; 1996a: 37-41), but what I eventually came to consider as the most important result, was to be able to pin down, through the translation network, why these different semantic mappings emerged.

Another proposal, put forward in Santos (1995a), concerning grammatical translationese, suggested that the detection of significant distribution differences in translated texts vs. original ones would pinpoint the influence of the source language. Still, in the overwhelming majority of the cases, so many factors were related that the distribution of features was just the top of the iceberg, and I no longer believe that any relevant conclusion can be obtained only by comparing two particular features. Every time I started in this way, I found a myriad of interesting factors affecting the distribution of that particular feature.

The conclusion seems to be that such studies (comparing one or two features) are useful as a probe, but cannot alone yield results (concerning translationese or language differences) without a detailed study of the whole system (or at least significant parts of it).

One of the most satisfying properties of a corpus based investigation is that one can always return to the data and ask other questions, come back to see things in different ways. Looking at one single aspect of language – doing one measurement only – is generally never enough. But doing subsequent apparently unrelated measurements (i.e., taking other parts of language into account) will provide evidence for (or disprove) what one may have caught a glimpse of in one's first steps of the research.

On the other hand, it has to be acknowledged that, without a translation network model to help and guide me, I would have been lost in too much data. Although the translation network model evolved during the course of the studies, it helped me concentrate on general categories and operators, as well as select – and test – related phenomena. This is the reason why the translation network was presented prior to the actual corpus studies, in order to allow the reader to use a similar conceptual formalism when presented with the examples.

Chapter 6

Language engineering, evaluation and applications

In the last chapter of this book, I will suggest how the sort of work described before can be of interest to language engineering applications. After all, my wish was to bridge the gap between theoretical models and empirical studies for sciences like linguistics and translation studies, on the one hand, and the need to computationally process large amounts of text in several real world applications, on the other.

I will start by describing how to evaluate what was presented in previous chapters with the aid of language engineering, and proceed to discuss possible applications. Very briefly, I will review some application areas more or less tightly connected to both tense and aspect and translation, suggesting ways of making use of the knowledge gained. This will be rather brief because it reflects ideas for further development rather than actual work.

While the first part (sections 6.1-6.3) is primarily addressing the language engineering research community and its concerns, I try to address the corpus linguistics research community in the second part (sections 6.4-6.5), discussing the state of the art of parallel corpora and the problem of validation of (contrastive) corpus studies. Finally, I will end the book with a summary of what I would like the general reader to have learned from this book.

6.1 Evaluation of the translation network model

The first and most straightforward evaluation scheme of the translation network model would be to run large numbers of sentence pairs against the two aspectual networks and register the possible correspondences. This would require a parser extracting the tense and aspect categories and operators used, together with an exhaustive verbal classification of the two languages. Both things are still impossible requirements with current capabilities and resources.

But even with such tools at our disposal, a long iterative process would still be required, involving a lot of human intervention, since refinements to the aspectual networks of the source languages would be needed, as described in Sections 4.3.5 and 4.4.1. Also, as a refinement, one might annotate the translational arcs with translation quality information, too. That would eventually lead to a thorough, consistent and fairly complete model of the translation between the two languages, hopefully catering for automatic analysis of new (aligned) corpus data.

Now, it should be obvious that I have not undertaken this job. The corpus analyses described in the previous chapter were done in order to arrive at the

model itself, not to evaluate it. Furthermore, in view of the enormous task of comparing the two languages (although restricted to tense and aspect), one cannot even hope to have done enough corpus studies for such a purpose.

I will therefore discuss other more indirect ways to evaluate the model(s), namely by embedding it into language engineering applications and evaluate these.

6.2 Direct applications

This need for evaluation agrees well with the attempt to answer the following (highly relevant) questions: Are there any direct applications of the work described above? Can it be used for language engineering in any way?

One might evaluate the correctness of the model by using it as a (simple) generator: Employing standard translations for lexical items, would one get a sensible set of translation proposals? If not robust enough to perform machine translation, this use of a particular translation network might be a good testbed for the correctness of the (knowledge embedded in the) system. The system could also be employed as a component of pedagogical applications aimed at teaching foreign languages and the translation between them. Again, this has not been tested, although I hope some innovation-minded teachers may jump on the idea.

Another even less ambitious suggestion would be to translate each aspectual network (and the translational arcs associated) into discovery rules for aspectual classes and test these rules in larger corpora, in a way similar to Brent (1991) or Klavans & Chodorow (1992). In other words, the translation network information would constitute a seed for building a temporal ontology for Portuguese, semi-automatically. The sort of rules I have in mind – which at the moment are pure speculation – are illustrated in Figure 6.1.

```
if a verb is often in Perfeito co-occurring with já &
it is also frequently used in the pluperfect &
it never occurs with ficar a --> it is a Mudança
```

Figure 6.1: An example of a rule for automatic verb classification

A confidence figure might also be computed from the corpora (as the previously cited works do), but this is not relevant for the point made here.

```
if a verb is often in Perfeito co-occurring with já &
it is frequently translated by the present perfect &
it is translated by Presente as well -->
its translation into past simple may be wrong
```

Figure 6.2: An example of a rule in a translation checker

Furthermore, if these discovery rules included translation correspondences, one might also come up with a (partial) method of translation checking similar to the one described by des Tombe & Armstrong-Warwick (1993), but now relative to tense and aspect domain (see Figure 6.2 for another tentative rule). As is the case

with function words, tense and aspect information is very frequent both in original and translated text in parallel corpora.

Evaluating such a translation checker would give some indication of the appropriateness of the underlying translation network model.

However, last but not least, the primary application of the work described here is the obvious enhancement of translation browsers, especially when they are used to study (a vicious circle?) the translation of tense and aspect.

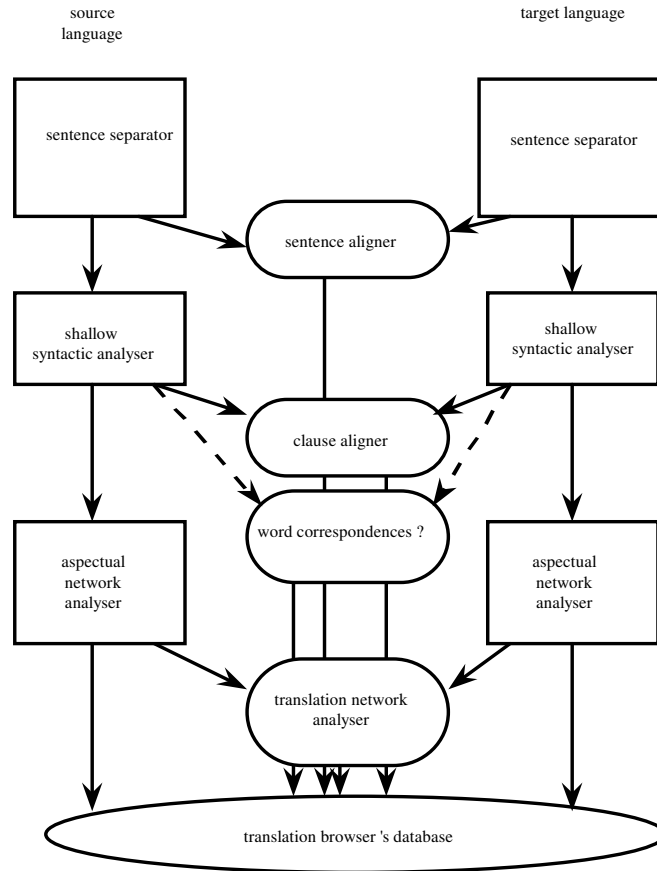


Figure 6.3: An enhanced translation browser

Part of what I learned in the parallel corpora studies reported in this book has been put to use in COMPARA (Frankenberg-Garcia & Santos, 2003). In fact, DISPARA (Santos, 2002), COMPARA's translation browser, strictly separates the issues of reordering text in translation and putting sentences into correspondence. It also allows for search in terms of sentence alignment. Unfortunately, both

copyright issues and lack of resources have so far prevented us from having COMPARA's texts grammatically annotated in either language, which is a precondition for the kinds of studies here envisaged. In any case, I'm not aware of any parallel corpus that is aligned at the clause level (let alone annotated with tense correspondences) such as the one on which I based my studies. This is one reason why repeating a few points from section 8.1 of my dissertation may be worthwhile. There, I looked into the translation browser application (see Figure 6.3) as a basis for several other more complex activities, and claimed that the more complex the application using the browser, the more important it would be to improve the translation browser's capabilities.

In other words, while a human user, browsing through translation material, might be able to skip problems just by relying on his or her judgement, automatic applications would need better precision in order not to compromise the output of the system. For example, if a translation browser were a component of a translation tutor, a translation evaluator, a filter for translation memories, an example-based machine translation system, etc., the matching of translation correspondences and the assessment of their quality would become crucial for the overall performance of the larger applications.

The competence in translation browsing, no matter the complexity and success of the other modules involved in the systems just mentioned, restricts the system's overall potential, so that the improvement of the browser deserves serious attention.

On the other hand, the advantage of trying to improve an application for human users – as most translation browsers that have been developed are – should not be underrated. Not only users with knowledge of translation may help develop the system, pointing to its weaknesses, asking for new functionalities and, in the process, amassing useful data. Also, early feedback increases the usefulness of this module, and contact with users provides satisfaction to its developer.

Contrary to most user-invisible tasks – i.e., tasks which are not understood by the user, just by the language engineer (Gaizauskas, 1998) –, this is a case of a system whose users are both larger systems and human beings, with the corresponding increase in usefulness, testability and visibility.

Let me thus illustrate some of the capabilities of a translation browser that is informed with tense and aspect knowledge. One could get:

- all examples of *fazer* in *Perfeito* translated by the English perfect
- the distribution of tenses that translate clauses in the progressive
- examples of *Mais que Perfeito* translated by the English passive
- accomplishments translated by *Estados*
- the distribution of the translations of the word *evitar*, by tense
- simple past translated by *Imperfeito*, without (radical) change of aspectual class
- whatever is translated by *again* and does not include *outra vez*
- addition of aspectualizers

In any case, it should be noted that I suggest as counterpart to Isabelle's (1992: 83) famous quote that “existing translations contain more solutions to more translation problems than any other existing resource” the converse **existing translations contain more translation problems than any other resource**. Therefore, a translation browser is in itself a very complex system and there is no way to avoid it.

6.3 Advantages of understanding time and translation

Leaving the topic of direct applications and concrete uses, let me also touch upon the interest a study in contrastive semantics of tense and aspect systems may have for language engineering in general.

A bold statement (which is obviously wrong) is that “any natural language processing application needs to understand time”. But hedging it as “it might be useful for any NLP application to understand time”, may be more acceptable. Time itself is a very broad concept, and there are different kinds of temporal understanding which are needed by different applications:

- exact temporal specification of events could be interesting for topic detection or news sifting
- relationship between different sentences or descriptions mentioned in the same text would be helpful for understanding complex events
- duration and kind of different activities should be interesting to understand concepts and reason about them

Arguably, different pieces of the information studied in this book could be of help in these different contexts. However, I will not look into this sort of applications and their information needs in detail, but will instead adduce some arguments for understanding translation and language differences. Note that I am not interested solely in the applications that perform translation (or have it as a building block). Rather, I want to stress that there is a larger group requiring knowledge of translation whether they have to do it or not.

This is generally relevant because the Web constantly enters our daily life, and more and more applications are related to the Web, in subtle ways. Since the Web is multilingual, one needs to know about translation to be comfortable on the Web and to be able to make sense of the information one gets, especially if it is mediated by an automatic broker.

Also, since more and more innovative applications are made for English and then “ported” to other languages later on, knowledge of what and how a given language differs from (or resembles) English might come in handy in this process. This, although apparently a sensible suggestion, has as far as I know not (yet) been backed by any evidence.

6.4 In the realm of parallel corpora

So far this chapter has been written to an audience of language engineers, trying to suggest different ways to consider, appraise or criticize the work. Another audience should be contemplated, namely those people that have defined themselves as parallel corpora researchers, and have published, compiled, and used parallel corpora.

Early researchers in bilingual corpora were eager to point out that obvious applications of parallel corpora would be “lexicography, language teaching and translation” (Aijmer & Altenberg, 1996: 12). This is typical of early times in any discipline. However, the methodology of creating corpus-based bilingual dictionaries does not come into being all by itself, the pedagogical concerns and programmes of language teachers do not suit perfectly the aims of corpus builders, and the daily life of translators does not magically change...

In fact, in the books devoted to language teaching, bilingual lexicography and parallel corpora, I see no proof of bilingual corpora being used for these aims. (People are, of course, working towards those goals, and this is a young research area, still...)

Let us start with teaching:

- First, *language* teaching. Jager *et al.* (1995), ten years after Holland *et al.* (1985), is another book devoted to language technology and computer aided language learning (CALL); however, it is striking to see how little role corpus-based approaches actually play.
- Rosner *et al.* (1999), a workshop devoted to “Computer and Internet Supported Education in Language and Speech Technology”, i.e., *NLP* teaching, has not shown a significantly different picture: use of (monolingual) corpora is just mentioned as an interesting technology, not in connection with any reports of real use.
- Maia *et al.* (2002), devoted to *translation* teaching, features only one example of using a parallel corpus in order to teach translation (although two other corpora allegedly created for that purpose are also introduced). Schmied (2002) creates a graphical representation of the meanings of prepositions and their translations based on bilingual dictionaries, and then checks and improves the result in real translation (corpora). He also suggests (semi-automatic?) “summaries of usage” created from the corpus.

Let us turn now to books devoted to parallel corpora.

- Véronis (2000) presents the field from the language engineer's standpoint. It includes the description of applications such as bilingual terminology extraction, improvement of translation memories (with clause, phrase and skeleton sentences), and bilingual lexicon construction (for cross-language information retrieval (CLIR) and for specific genres). Not surprisingly, parallel corpora were mainly seen as a tool to get more specific resources

for higher-level applications. The only article devoted to CALL simply presented the user with bilingual concordances in a small corpus.

- Borin (2002), surveying the parallel corpus reality from the corpus linguist's standpoint, does not even include a section on applications, but only on “linguistic applications” (studying minority languages, reversing a bilingual dictionary). The only real application reported in the book is the derivation of a bilingual dictionary from the Web (as a comparable corpus).

Finally, I was not able to find a single book devoted to bilingual lexicography that made reference to use of parallel corpora. Altenberg & Granger (2002), although featuring a four papers' section entitled “Corpus-based Bilingual Lexicography”, fail to include a single paper reporting the use of parallel (aligned) corpora in lexicography. Teubert (2002) presents one example and adds, carefully:

Since there are, to date, no bilingual dictionaries of general language based on parallel corpora, we still do not know to what extent they can complement, improve and validate existing dictionaries (Teubert, 2002: 204)

While Roberts & Cormier (2000), the only paper in a book about bilingual lexicography making reference to a parallel corpus, state that the particular parallel corpus in question:

is not used to start investigating equivalence, since, even if the texts are well translated, it is always a translation corpus, therefore should be used with care. Transbase is used rather as a memory aid. If the lexicographer finds there an interesting equivalent he had not thought about, he verifies its use searching in TEXTUM [monolingual corpus]. (Roberts & Cormier, 2000: 236, my translation)

In any case, this is not supposed to be a comprehensive, but simply indicative, picture of where matters stand as regards actual use and application of parallel corpora, and its simple goal was to motivate the following claim:

It is time to stress that corpus-based methodology requires research, especially **parallel**-corpus-based methodology, which cannot be a simple adaptation of monolingual corpus-based methodology, because of the additional variables of language differences and translation.

6.5 A plea for validation

Instead of devoting this chapter to what could be done, I prefer to devote it to what has already been accomplished, suggesting that, instead of building new resources or doing new individual studies, students should consider the problems of validating previous work and suggesting methodologies for practical applications.

In Santos & Oksefjell (1999), we faced two major problems when trying to validate other researchers' claims:

- too few specified claims involving too many variables (not made explicit)
- a scarcity of available bilingual reference corpora, where other variables might be tested on the same data and without having to redo previous studies from scratch

We suggested that, rather than comparing corpora, it would be more interesting and more useful to compare claims, especially because two corpora could differ in some ways and be similar in others. Incidentally, our choice of claims to validate was not ideal, as will be seen below. In fact, sometimes we looked at observable translation regularities instead of explanatory claims.

6.5.1 Santos & Oksefjell revisited

Looking first at Slobin's (1996; 1997b) description of journeys in English translated into a (typologically different, Romance) language, we tried to convert his study into some predictions testable in a corpus, using the English-Portuguese part of OMC/ENPC (Johansson *et al.*, 1999) for it.

An interesting process then took place: how to translate Slobin's description of a journey – “In describing real-world or fictional events [...] a narrator may present a series of linked paths or a path with waystations. I will call a complex path a *journey*” (Slobin, 1996: 202) – into a searchable item in a corpus? How to look for movement (or movement verbs) in a non-semantically tagged corpus?

We selected the particles *into* and *out of* as plausible indicators of complex (boundary crossing) movement (therefore losing instances of journeys or movements without them), and selected those cooccurring with verbs that could be analysed as MOTION + MANNER verbs (the prototypical case). This was a decision that probably took away the most interesting contrastive cases, but was done with the intention of replicating as well as possible what Slobin might have done.

More specifically, we ignored cases of non-motion verbs modified by those particles, like *slam*, *whisper*, *roar*, or *struggle*, which were employed to describe a movement, as well as verbs of position like *rise*, *dangle*, *head* or *cram*, describing the background instead of a concrete movement event. We also excluded from our consideration all verbs that move the direct object and not the subject, such as *put*, *throw*, or *push*, because Slobin did not mention them.

Even after all these (controversial) decisions, a lot still remained for the analyst to subjectively decide: what is a MOTION + MANNER verb? In which cases is it being used that way? In other words, when is a given clause or sentence describing a motion event and the manner of moving? Would our intuitions agree with Slobin's if he were to examine the sentences we classified?

Another problem we faced was how to interpret the results: if the translation into Portuguese did not display the expected decrease of manner

description, or the predicted increase in number of clauses in boundary crossing (incidentally, alternative and opposite ways to deal with the same phenomenon), who should be blamed? Could it not be the translators' fault? For, if translation does not behave as Slobin (and contrastive linguistics) says, it does not automatically follow that languages do not differ in that way. Rather, it may simply mean that the unaware translators are producing translationese, or, from a translator-friendly perspective, that some concessions must be made in translation. After all, if the original text keeps talking about movement, there is not much a translator can do but talk about movement in the translation, no matter whether he would have omitted it all, if he had written it from scratch.

This prompts the following remark: methodologically, it is more reliable to look first at translation practice and see what it can unveil about language differences, as was done in the present book, than to start with language differences and try to foresee translation practice, as Slobin does.

The other validation study was in a way less interesting, because we were looking at a study already based on corpus material (Slobin selected journeys looking at random pages in a printed book), but it raised several kinds of relevant issues anyway. We tried to validate Santos (1998b), a paper on perception verbs, only by considering the published article.

First of all, the presentation of the data – it goes without saying that people do not present all corpus evidence in full in an article – was found to be insufficient to follow other alleys of research: If the data on perception verbs had been separated according to verb and not all clumped together, one could have tested whether their distribution was similar to the one in the new corpus. Likewise, given that the new corpus contained only the beginning sections of books, the influence of final vs. initial chapter in a book could have been further investigated if the data had been classified according to chapter.

However, the most important problem with the validation process in this case was the actual choice of “claims” to look at. We set out to confirm (among other things) whether the translation of VP objects of perception verbs, with or without *could*, “does not show any reliable translation regularity” (ibid.: 124)! In fact, looking now with some detachment, it is evident that such a negative and useless “claim” would not give any satisfactory results, “confirmed” or not. What one should try to confirm is positive claims having an underlying explanatory hypothesis, such as e.g.

- English activities are responsible for most translations of gerund into infinitive; or
- *could* with perception verbs is omitted in the translation more often than with other verbs.

Nevertheless, an interesting exercise in connection with this case study was to play around with different subcorpora to (informally) see whether different combinations of texts would display strikingly different results.

6.5.2 Validation machinery wanted

In any case, a notion of quantitative validation was painfully missed in both studies. How can one compare different numbers and have a measure of reliability? What is random behaviour, common translation behaviour and significantly interesting translation behaviour? How can these kinds of behaviour be indexed regarding size and corpus homogeneity, etc.? All this is work still to be done, before we can, with some confidence, compare numbers dealing with different phenomena and corpora.

This is, in my view, the most important key to understand quantitative parallel corpora studies. Without it, only qualitative effects like surprise and oracle function can be employed. They are important in themselves, as I hope this book shows, but one also needs familiarity with numbers reflecting contrastive phenomena in order to be able to interpret quantitative data.

Ideally, one would have a repository of measures and numbers so that one could and should be able to replicate studies analysing the same data in order to somehow improve the reliability of those numbers and studies.

In a way, by presenting numbers without a frame of reference, isolated researchers are building such a knowledge base the hard way. But, if one could build it on top of already analysed data, using shared corpora, the studies would be both more effective and more reliable, since there are always plenty of hidden hypotheses that people implicitly make and which are embedded in any system, tool or resource, from tokenization to sentence separation and classification counts.

In this connection, the remarks in Johansson (1975) on the need to evaluate thoroughly empirical claims in the realm of applied linguistics remain strikingly modern, and urgent.

6.6 Concluding remarks

In retrospect, a dissertation is as unpredictable as one's life. I learned a lot from working on it, but the paths taken and the results I came to appreciate the most were very different from my expectations.

On the one hand, it is instructive to look at the extremely ambitious hopes I had before starting the PhD work. In the application to enroll as a PhD student, back in 1989, I suggested to describe the differences between the two languages and to implement a temporal reasoning system that could be used for machine translation!

I have come a long way down the path of having more realistic expectations. Interestingly, my definition of what can be useful has changed considerably, too. Toy systems that solve one or two "problems" no longer deserve my attention. Neither do systems that are based on simplistic and untested hypotheses about languages.

Also, the value of NLP as an aid to humans instead of a replacement of humans became clear, especially by realizing that there is always a ceiling effect

in all human technologies. Over a given threshold, consensus is never possible. It is the task of the language engineer (or AI engineer) to allow for this while still deploying informed and interesting systems.

In these last pages, I want to summarize what I believe are the most relevant contributions of this book:

- The insistence on defining language (aspectual) categories using Vendler's method, not Vendler's classes
- The use of real translations as a initial step to arrive at those categories, and their constant validation on the basis of translation material
- The highlighting of contrastive vagueness as a key concept to understand translation and language differences
- The set of aspectual classes arrived at for Portuguese
- The model arrived to describe translation performance (the translation network) as well as translation as a human activity
- The set of translation regularities found between Portuguese and English
- The suggestion of the translation browser as a key application.

Summing up my general learning process, time and again I learned that everything in language is related, and that refinement of analysis always brings other features into the picture.

Using a corpus is like indulging in a conversation which may never end – but we can learn more and more by asking more questions.

I also learned that vagueness is the most relevant property of a natural language, and only with its help can we understand language.

I hope I have been able to convey adequately these messages in the book. Let me conclude by wishing that I some day will be able to create a publically available parallel corpus with a translation browser that allows the readers to check or challenge my claims.

References

- Aijmer, Karin & Bengt Altenberg. 1996. "Introduction". In Karin Aijmer, Bengt Altenberg & Mats Johansson (eds.), *Languages in Contrast: Papers from a Symposium on Text-based Cross-linguistic Studies (Lund, 4-5 March 1994)*. Lund: Lund University Press, pp. 11-16.
- Altenberg, Bengt & Sylviane Granger. 2002. *Lexis in Contrast: Corpus-based approaches*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Aske, Jon. 1989. "Path predicates in English and Spanish: A closer look". In *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society* (February 18-20, 1989). Berkeley, CA: BLS, pp. 1-14.
- Baker, Mona. 1996. "Corpus-based translation studies: The challenges that lie ahead". In Harold Somers (ed.), *Terminology, LSP and Translation: Studies in language engineering in honour of Juan C. Sager*. Amsterdam & Philadelphia: John Benjamins Publishing Company, pp. 175-86.
- Bar-On, Dorit. 1993. "Indeterminacy of Translation: Theory and Practice". *Philosophy and Phenomenological Research* **LIII**, Number 4, December 1993, pp. 781-810.
- Barlow, Michael. 1997. Paper given at the *Fourth Nordic Symposium on Text-based Contrastive Studies*, April 1997, Centre for Advanced Study, Norwegian Academy of Science and Letters, Oslo.
- Berman, Ruth A. & Dan Isaac Slobin. 1994. *Relating Events in Narrative: A Crosslinguistic Development Study*. Hillsdale, NJ: Lawrence Erlbaum.
- Biber, Douglas. 1993. "Representativeness in Corpus Design". *Literary and Linguistic Computing* **8**, Number 4, 1993, pp. 243-57.
- Borin, Lars (ed.). 2002. *Parallel Corpora, Parallel Worlds*. Amsterdam & New York: Rodopi.
- Brent, Michael R. 1991. "Automatic Semantic Classification of Verbs from their Syntactic Contexts: An Implemented Classifier for Stativity". In *Proceedings of the 5th Conference of the European Chapter of the ACL* (Berlin, 9-11 April 1991), pp. 222-226.
- Brown, P., J. Cocke, S. Della Pietra, V.J. Della Pietra, F. Jelinek, J.D. Lafferty, R.L. Mercer & P.S. Roossin. 1990. "A Statistical Approach to Machine Translation". *Computational Linguistics* **16**, Number 2, June 1990, pp. 79-85.
- van Buren, Paul. 1980. "Contrastive analysis". In Jasek Fisiak (ed.), *Theoretical Issues in Contrastive Linguistics*. Amsterdam: John Benjamins, pp. 83-117.
- Caenepeel, Mimo. 1989. "Aspect, Temporal Ordering and Perspective in Narrative Fiction". PhD dissertation, University of Edinburgh.
- Carlson, Gregory Norman. 1977. "Reference to Kinds in English", PhD thesis,

- Univ. of California, Irvine, 1977; published as *Reference to Kinds in English*. New York: Garland Publishing, 1980.
- Carlson, Lauri. 1981. "Aspect and Quantification". In Philip Tedeschi & Annie Zaenen (eds.), *Syntax and Semantics, Volume 14: Tense and Aspect*. New York: Academic Press, pp. 31-64.
- Catford, J.C. 1967. *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford: Oxford University Press.
- Church, Kenneth W. & William A. Gale. 1991. "Concordances for Parallel Text". In *Using Corpora: Proceedings of the Eight Annual Conference of the UW Centre for the New OED and Text Research* (Oxford, September 29 - October 1, 1991), pp. 40-62.
- Dagan, Ido, Alon Itai & Ulrike Schwall. 1991. "Two languages are more informative than one". In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics* (Berkeley, 18-21 June 1991), pp. 130-137.
- Dagan, Ido & Alon Itai. 1994. "Word Sense Disambiguation Using a Second Language Monolingual Corpus". *Computational Linguistics* 20, Number 4, December 1994, pp. 563-596.
- Dahl, Östen. 1981. "On the definition of the telic-atelic (bounded-non-bounded) distinction". In Philip Tedeschi & Annie Zaenen (eds.), *Syntax and Semantics, Volume 14: Tense and Aspect*. New York: Academic Press, pp. 79-90.
- Dahl, Östen. 1984. "Perfectivity in Slavonic and other languages". In Casper de Groot & Hannu Tammola (eds.), *Aspect bound: a voyage into the realm of Germanic, Slavonic and Finno-Ugrian Aspectology*. Dordrecht: Foris, pp. 3-22.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Basil Blackwell.
- Danielsson, Pernilla & Daniel Ridings. 1996a. "PEDANT: Parallel texts in Göteborg". Research report GU-ISS-96-2, Department of Swedish, Göteborg University. [accessed 11 December 2003] <http://spraakbanken.gu.se/lb/pedant/parabank/parabank.html>
- Danielsson, Pernilla & Daniel Ridings. 1996b. "Annotating Parallel Texts with the NSL library". Research report from the Department of Swedish, Göteborg University, GU-ISS-96-7.
- Dillon, George L. 1977. *Introduction to Contemporary Linguistic Semantics*. Englewood Cliffs, NJ: Prentice-Hall.
- Doherty, Monika. 1997. "Übersetzen im Spannungsfeld zwischen Grammatik und Pragmatik". In Rudi Keller (ed.), *Linguistik und Literaturübersetzen*. Tübingen: Narr, pp. 79-102.
- Doi, Shinichi & Kazunori Muraki. 1992. "Translation Ambiguity resolution based on text corpora of source and target languages". *Proceedings of COLING'92* (Nantes, 23-28/7/98), Vol. II, pp. 525-529.

- Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: D. Reidel.
- Duff, Alan. 1981. *The third language: Recurrent problems of translation into English*. Oxford: Pergamon Press.
- Dyvik, Helge. 1998. "A translational basis for semantics". In Stig Johansson & Signe Oksefjell (eds.), *Corpora and Crosslinguistic Research: Theory, Method, and Case Studies*. Amsterdam & Atlanta, GA: Rodopi, pp. 51-86.
- Egg, Markus. 1995. "The Intergressive as a New Category of Verbal Aktionsart". *Journal of Semantics* 12, 1995, pp. 311-56.
- Engh, Jan. 1998. "Normer, grammatikk og databehandling". In Ruth Vatvedt Fjeld & Boye Wangsten (eds.), *Normer og Regler: Festskrift til Dag Gundersen, 15. januar 1998*. Oslo: Universitetsforlaget, pp. 344-359.
- Even-Zohar, Itamar. 1990. *Polysystem Studies*. A special issue of *Poetics Today* 11, No. 1, 1990. Durham: Duke University Press.
- Van Eynde, Frank. 1988. "The analysis of tense and aspect in Eurotra". In *Proceedings of COLING'88* (Budapest, 22-27 August 1988), pp. 699-704.
- Fabricius-Hansen, Cathrine. 1991. "Contrastive stylistics: Outline of a research project on German and Norwegian non-fictional prose". In Karen M. Lauridsen & Ole Lauridsen (eds.), *Contrastive linguistics. Papers from the CL Symposium at the Aarhus School of Business (28-30 August 1989)*. Aarhus: Aarhus School of Business, pp. 51-76.
- Fabricius-Hansen, Cathrine. 1998. "Information density and translation, with special reference to German-Norwegian-English". In Stig Johansson & Signe Oksefjell (eds.), *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Amsterdam & Atlanta, GA: Rodopi, pp. 197-234.
- Frankenberg-Garcia, Ana & Diana Santos. 2003. "Introducing COMPARA, the Portuguese-English parallel translation corpus". In Federico Zanettin, Silvia Bernardini and Dominic Stewart (eds.), *Corpora in Translation Education*. Manchester: St. Jerome Publishing, pp. 71-87.
- Freed, Alice F. 1979. *The Semantics of English Aspectual Complementation*. Dordrecht: D. Reidel.
- Gaizauskas, Robert. 1998. "Evaluation in language and speech technology". *Computer Speech and Language* 12 (4), pp. 249-262.
- Gale, William A. & Kenneth W. Church. 1991. "Identifying Word Correspondences in Parallel Texts". In *Proceedings of 4th DARPA Workshop on Speech and Natural Language*, pp. 152-157.
- Gawrońska, Barbara. 1993. *An MT Oriented Model of Aspect and Article Semantics*. Lund: Lund University Press.
- Gellerstam, Martin. 1986. "Translationese in Swedish novels translated from English". In Lars Wollin & Hans Lindquist (eds.), *Translation studies in Scandinavia*. Lund: CWK Gleerup, pp. 88-95.

- Gillon, Brendan S. 1990. "Ambiguity, generality, and indeterminacy: tests and definitions". *Synthese* **85** (1990), pp. 391-416.
- Granger, Sylviane. 1996. "From CA to CIA and back: An integrated approach to computerized bilingual learner corpora". In Karin Aijmer, Bengt Altenberg & Mats Johansson (eds.), *Languages in Contrast: Papers from a Symposium on Text-based Cross-linguistic Studies (Lund, 4-5 March 1994)*. Lund: Lund University Press, pp. 37-51.
- Hasselgård, Hilde. 1996. "Some methodological issues in a contrastive study of word order in English and Norwegian". In Karin Aijmer, Bengt Altenberg & Mats Johansson (eds.), *Languages in Contrast: Papers from a Symposium on Text-based Cross-linguistic Studies (Lund, 4-5 March 1994)*. Lund: Lund University Press, pp. 113-126.
- Hauenschild, Christa. 1988. "Discourse Structure - Some Implications for Machine Translation". In Dan Maxwell, Klaus Schubert & A.P.M. Witkam (eds.), *New Directions in Machine Translation, Conference Proceedings, Budapest 18-19/8/88*. Dordrecht & Providence: Foris, pp. 145-156.
- Heinämäki, Orvokki. 1984. "Aspect in Finnish". In Casper de Groot & Hannu Tammola (eds.), *Aspect bound: a voyage into the realm of Germanic, Slavonic and Finno-Ugrian Aspectology*. Dordrecht: Foris, pp. 153-177.
- Hempel, Carl G. 1966. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall.
- Hindle, Donald & Mats Rooth. 1993. "Structural Ambiguity and Lexical Relations", *Computational Linguistics* **19**, Number 1, March 1993, pp. 103-120.
- Holland, V. Melissa, Jonathan D. Kaplan & Michelle R. Sams (eds.). 1985. *Intelligent language tutors: Theory shaping technology*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Isabelle, Pierre. 1992. "Bi-Textual Aids for Translators". In *Screening Words: User Interfaces for Text, Proceedings of the Eight Annual Conference of the UW Centre for the New OED and Text Research* (Waterloo, October 18-20, 1992), pp. 76-89.
- Isabelle, Pierre, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliot Macklovitch, François Perrault, Xiaobo Ren & Michel Simard. 1993. "Translation Analysis and Translation Automation". In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation, TMI'93* (Kyoto, July 14-16, 1993), pp. 201-217.
- Jager, Sake, John A. Nerbonne & A.J. van Essen (eds.). 1995. *Language teaching and language technology*. Lisse: Swets & Zeitlinger.
- Jakobson, Roman. 1959. "On Linguistic Aspects of Translation". In Reuben A. Brower (ed.), *On Translation*. Cambridge, Mass: Harvard University Press, pp. 232-239.
- James, Carl. 1980. *Contrastive analysis*. Harlow, Essex: Longman.
- Johansson, Stig. 1975. *Papers in Contrastive Linguistics and Language Testing*. Lund: CWK Gleerup.

- Johansson, Stig. 1998. "On the role of corpora in cross-linguistic research". In Stig Johansson and Signe Oksefjell (eds.), *Corpora and cross-linguistic research: theory, method, and case studies*. Amsterdam & Atlanta, GA: Rodopi, pp. 3-24.
- Johansson, Stig, Jarle Ebeling & Knut Hofland. 1996. "Coding and aligning the English-Norwegian Parallel Corpus". In Karin Aijmer, Bengt Altenberg & Mats Johansson (eds.), *Languages in Contrast: Papers from a Symposium on Text-based Cross-linguistic Studies (Lund, 4-5 March 1994)*. Lund: Lund University Press, pp. 87-112.
- Johansson, Stig, Jarle Ebeling & Signe Oksefjell. 1999. "English-Norwegian Parallel Corpus: Manual". Oslo: Department of British and American Studies, University of Oslo, [accessed 8 July 1999], <http://www.hf.uio.no/iba/prosjekt/ENPCmanual.html>.
- Johansson, Stig & Knut Hofland. 1994. "Towards an English-Norwegian parallel corpus". In U. Fries, G. Tottie & P. Schneider (eds.), *Creating and using English Language corpora, Papers from the fourteenth International Conference on English Language Research on Computerized Corpora, Zürich 1993*. Amsterdam & Atlanta: Rodopi, pp. 25-37.
- Kameyama, Megumi, Ryo Ochitani & Stanley Peters. 1991. "Resolving Translation Mismatches With Information Flow". In *Proceedings of the 29th Annual Meeting of the ACL* (Berkeley, 18-21 June 1991), pp. 193-200.
- Kamp, Hans. 1991. "The Perfect and other Tenses in French and English". In Hans Kamp (ed.), *Tense and Aspect in English and French*, DYANA Deliverable R2.3.B, January 1991, pp. 41-64.
- Kamp, Hans & Uwe Reyle. 1993. *From Discourse to Logic: an Introduction to Modeltheoretic Semantics of Natural language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer Academic Press, 1993.
- Kaplan, Ronald M., Klaus Netter, Jürgen Wedekind & Annie Zaenen. 1989. "Translation by Structural Correspondences". In *Proceedings of the 4th Conference of the European Chapter of the ACL* (Manchester, 10-12 April 1989), pp. 272-281.
- Kay, Martin, Jean Mark Gawron & Peter Norvig. 1994. *VerbMobil: A Translation System for Face-to-Face Dialog*. CSLI Lecture Notes, No. 33, Stanford, Calif.: Center for the Study of Language and Information.
- Keefe, Rosanna & Peter Smith (eds.). 1996. *Vagueness: A Reader*. Cambridge, MA: The MIT Press.
- Keenan, Edward L. 1976. "The Logical Diversity of Natural Languages". In H. B. Steklis, S. R. Harnad & J. Lancaster (eds.), *Origins and Evolution of Language and Speech*. New York: Annals of the New York Academy of Sciences 280, pp. 73-91.
- Keenan, Edward L. 1978. "Some Logical Problems in Translation". In F. Guenther & M. Guenther-Reutter (eds.), *Meaning and Translation:*

- Philosophical and Linguistic Approaches*. London: Duckworth, pp. 157-89.
- Kempson, Ruth M. 1975. *Presupposition and the delimitation of semantics*. Cambridge: Cambridge University Press.
- Kent, Stuart John Harding. 1993. "Modelling Events from Natural Language". PhD thesis, Department of Computing, Imperial College of Science, Technology and Medicine, July 1993.
- Klavans, Judith L. & Martin Chodorow. 1992. "Degrees of Stativity: The Lexical Representation of Verb Aspect". In *Proceedings of COLING'92* (Nantes, 23-28/7/92), Vol. IV, pp. 1126-1131.
- Lab, Frédérique. 1990. "Le Temps de la Linguistique". *T.A. Information* **31**, N.1, 1990, pp.49-55.
- Lakoff, George & Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lascarides, Alexandra. 1988. "A Formal Semantic Analysis of the Progressive". PhD thesis, University of Edinburgh.
- Leech, Geoffrey. 1993. "Corpus Annotation Schemes". *Literary and Linguistic Computing* **8**, 1993, pp. 275-281.
- Lipinska-Grzegorek, Maria. 1977. *Some Problems of Contrastive Analysis: Sentences with Nouns and Verbs of Sensual Perception in English and Polish*. Edmonton: Linguistic Research Inc.
- Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Maia, Belinda, Johann Haller & Margherita Ulrych (eds.). 2002. *Training the language services provider for the New Millenium, Proceedings of the III Encontros de Tradução de Astra-FLUP*, Porto: FLUP.
- Maslov, Yuriy S. (ed.) 1985 [1978]. *Contrastive Studies in Verbal Aspect*. Heidelberg: Julius Groos Verlag, 1985, translated and annotated by James Forsyth, from the original 1978 Russian version.
- Mateus, Maria Helena Mira, Ana Maria Brito, Inês Duarte & Isabel Hub Faria. 1989 [1971]. *Gramática da Língua Portuguesa*. Lisboa: Caminho, 3ª edição.
- ter Meulen, Alice G.B. 1995. *Representing Time in Natural Language: The Dynamic Interpretation of Tense and Aspect*. Cambridge, MA: The MIT Press.
- Moens, Marc. 1987. "Tense, Aspect and Temporal Reference". PhD thesis, University of Edinburgh.
- Moens, Marc & Mark Steedman. 1988. "Temporal Ontology and Temporal Reference". *Computational Linguistics* **14**, Number 2, June 1988, pp. 15-28.
- Nagao, Makoto. 1988. "Language Engineering: the Real Bottle Neck of Natural Language Processing". In *Proceedings of COLING'88* (Budapest, 22-27 August 1988), pp. 448-453.

- Neustupný, J.V. 1978. *Post-structural Approaches to Language: Language Theory in a Japanese Context*. Tokyo: University of Tokyo Press.
- Nida, Eugene A. 1959. "Principles of translation as exemplified by Bible translating". In Reuben A. Brower (ed.), *On Translation*. Cambridge, Mass: Harvard University Press, pp. 11-31.
- Palmer, F.R. 1965. *A Linguistic Study of the English Verb*. London: Longman.
- Pinkal, Manfred. 1995 [1985]. *Logic and lexicon : the semantics of the indefinite*. Dordrecht: Kluwer, 1995, translation by Geoffrey Simmons of *Logik und Lexikon: die Semantik des Unbestimmten*, Berlin: de Gruyter, 1985.
- Poesio, Massimo. 1995. "Semantic Ambiguity and Perceived Ambiguity". Research Report HCRC/RP-68, Human Communication Research Centre, University of Edinburgh, May 1995.
- Resnik, Philip, Mari Broman Olsen & Mona Diab. 1997. "Creating a Parallel Corpus from the "Book of 2000 Tongues"". In *Text Encoding Initiative 10th Anniversary User Conference (TEI-10)*, Providence, November 1997. Revised and expanded version in *Computers and the Humanities* **33**, 1999, pp. 129-153.
- Roberts, Roda P. & Monique C. Cormier. 2000. "L'analyse des corpus pour l'élaboration du Dictionnaire canadien bilingue", In Szende, Tamás (ed.), *Approches contrastives en lexicographie bilingue*. Paris: Honoré Champion, 2000, pp. 223-240.
- Rosner et al. (eds.) 1999. *Computer and Internet Supported Education in Language and Speech Technology*, EACL99, Proceedings of a Workshop Sponsored by ELSNET and the Association of Computational Linguistics, 12 June 1999, University of Bergen, Norway.
- Rossi-Landi, Ferruccio. 1973 [1968]. *Ideologies of Linguistic Relativity*, The Hague: Mouton. First published in Italian in *Ideologie* **4** (1968), pp. 3-69.
- Rydning, Antin Fougner. 1998. "La notion d'ambigüité en traduction". *TradTerm* **5**, no. 1, 1998, Centro interdepartamental de tradução e terminologia, FFLCH - Universidade de São Paulo, pp. 11-40.
- Sadler, Victor & Ronald Vendelmans. 1990. "Pilot implementation of a bilingual knowledge bank". In Hans Karlgren (ed.), *Proceedings of COLING'90*, Vol. 3, pp. 449-451.
- Sampson, Geoffrey. 1987. "MT: A nonconformist's view of the state of the art". In Margaret King (ed.), *Machine Translation Today: The State of the Art (Proceedings of the Third Lugano Tutorial, 2-7 April 1984)*. Edinburgh: Edinburgh University Press, pp. 91-108.
- Sandström, Görel. 1990. "Temporal Adverbial Clauses in Narrative Discourse". Report nr. 31, Department of General Linguistics, University of Umeå, December 1990.
- Sandström, Görel. 1993. "When-clauses and the temporal interpretation of narrative discourse". PhD dissertation, Report nr. 34, DGL-UUM-R-34, Department of General Linguistics, University of Umeå, May 1993.
- Santos, Diana. 1991. "Contrastive Tense and Aspect Data", INESC Report no.

- RT/57-91, October 1991; also available at http://www.linguatca.pt/Diana/download/cont_data91.ps.
- Santos, Diana. 1993. "Integrating tense, aspect and genericity". In *Actas do IX Encontro da Associação Portuguesa de Linguística* (Coimbra, 29 de Setembro - 1 de Outubro de 1993), pp. 391-405.
- Santos, Diana. 1994a. "Bilingual alignment and tense". In *Proceedings of the Second Annual Workshop on Very Large Corpora* (Kyoto, 4 August 1994), pp. 129-141. Extended version as INESC Report no. AR/10-94.
- Santos, Diana. 1994b. "Imperfeito: a broad-coverage study". In *Actas do X Encontro da Associação Portuguesa de Linguística* (Évora, 6-8 de Outubro de 1994), pp. 525-539.
- Santos, Diana. 1995a. "On grammatical translationese". In *Short papers presented at the Tenth Scandinavian Conference on Computational Linguistics* (Helsinki, 29-30th May 1995), compiled by Kimmo Koskeniemi, pp. 59-66.
- Santos, Diana. 1995b. "On the use of parallel texts in the comparison of languages". In *Actas do XI Encontro da Associação Portuguesa de Linguística* (Lisboa, 2-4 de Outubro de 1995), pp. 217-239.
- Santos, Diana Maria de Sousa Marques Pinto dos. 1996a. "Tense and aspect in English and Portuguese: a contrastive semantical study". PhD thesis, Instituto Superior Técnico, June 1996.
- Santos, Diana. 1996b. "Para uma classificação aspectual portuguesa do português". In Ivo Castro (ed.), *Actas do XII Encontro da Associação Portuguesa de Linguística* (Braga-Guimarães, 1-3 de Outubro de 1996), pp. 299-315.
- Santos, Diana. 1997a. "The importance of vagueness in translation: Examples from English to Portuguese". *Romansk Forum* 5, June 1997, University of Oslo, pp. 43-69. Revised version: "The relevance of vagueness for translation: Examples from English to Portuguese", *TradTerm* 5, 1, 1998, Universidade de São Paulo, pp. 71-98.
- Santos, Diana. 1997b. "O tradutês na literatura infantil traduzida em Portugal". In *Actas do XII Encontro da Associação Portuguesa de Linguística* (Lisboa, 1-3 de Outubro de 1997), pp. 259-274.
- Santos, Diana. 1998a. "L'Imperfeito português: étude systématique de ses fonctions et de comment en rendre compte en traduisant vers l'anglais". In *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza* (Palermo, 18-24 settembre 1995). Tübingen: Max Niemeyer Verlag, pp. 777-787.
- Santos, Diana. 1998b. "Perception verbs in English and Portuguese", in Johansson, Stig and Signe Oksefjell (eds.), *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Amsterdam & Atlanta, GA: Rodopi, pp. 319-342.
- Santos, Diana. 1999a. "The Pluperfect in English and Portuguese: What Translations Patterns Show". In Hilde Hasselgård & Signe Oksefjell (eds.),

- Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam & Atlanta, GA: Rodopi, pp. 283-299.
- Santos, Diana. 1999b. "Toward Language-specific Applications". *Machine Translation* **14**, n. 2, June 1999, pp. 83-112.
- Santos, Diana. 2000. "The translation network: A model for the fine-grained description of translations". In Jean Véronis (ed.), *Parallel Text Processing*. Dordrecht: Kluwer Academic Publishers, pp. 169-186.
- Santos, Diana. 2002. "DISPARA, a system for distributing parallel corpora on the Web". In Elisabete Ranchhod & Nuno J. Mamede (eds.), *Advances in Natural Language Processing (Third International Conference, PorTAL 2002, Faro, Portugal, June 2002, Proceedings)*, LNAI 2389. Berlin & Heidelberg: Springer, pp. 209-218.
- Santos, Diana & Signe Oksefjell. 1999. "Using a Parallel Corpus to Validate Independent Claims". *Languages in contrast* **2**, 1999, pp. 117-132.
- Sapir, Edward. 1944. "On grading: a study in semantics". *Philosophy of Science* **2**, 1944, pp. 93-116.
- Sato, Satoshi & Makoto Nagao. 1990. "Toward Memory-based Translation". In Hans Karlgren (ed.), *Proceedings of COLING'90* (Helsinki, August 1990), Vol. 3, pp. 247-252.
- Sato, Satoshi. 1992. "CTM: An Example Based Translation Aid System". In *Proceedings of COLING'92* (Nantes, 23-28/7/98), Vol. IV, pp. 1259-1263.
- Schmied, Josef. 2002. "A translation corpus as a resource for translators: the case of English and German prepositions". In Maia, Belinda, Johann Haller & Margherita Ulrych (eds.), *Training the language services provider for the New Millenium, Proceedings of the III Encontros de Tradução de Astra-FLUP*, Porto: FLUP, pp. 223-238.
- Schmied, Josef & Hildegard Schäffler. 1996. "Approaching translationese through parallel and translation corpora". In Carol E. Percy, Charles F. Meyer & Ian Lancashire (eds.), *Synchronic corpus linguistics: papers from the sixteenth International Conference on English Language Research on Computerized Corpora (ICAME 16)*. Amsterdam & Atlanta, GA: Rodopi, pp. 41-56.
- Schubert, K. 1988. "Implicitness as a guiding principle in machine translation". *Proceedings of COLING'88* (Budapest, 22-27 August 1988), pp. 599-601.
- Slobin, Dan I. 1987. "Thinking for Speaking". In Jon Aske, Natasha Beery, Laura Michaelis & Hana Filip (eds.), *Berkeley Linguistics Society: Proceedings of the Thirteenth Annual Meeting February 14-16, 1987: General Session and Parasession on Grammar and Cognition*. Berkeley: Berkeley Ling. Soc., 1987, pp. 435-45.
- Slobin, Dan I. 1996. "Two ways to travel: Verbs of motion in English and Spanish". In M. Shibatani & S. A. Thompson (eds.), *Grammatical constructions: their form and meaning*. Oxford: Oxford University Press, pp. 195-219.

- Slobin, Dan I. 1997a. "Mind, code and text". In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on Language Function and Language Type, Dedicated to T. Givón*. Amsterdam: John Benjamins, pp. 437-467.
- Slobin, Dan I. 1997b. "In a Manner of Going: Influences of lexicalization Patterns on Descriptions of Motion Events". Paper presented at *Second meeting of the Association for Linguistic Typology (ALT II)*, University of Oregon, Eugene, OR, September 11-14, 1997.
- Smith, Carlota S. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.
- Snell-Hornby, Mary. 1995 [1988]. *Translation Studies: An integrated approach*. Revised edition. Amsterdam & Philadelphia: John Benjamins.
- Steiner, George. 1992 [1975]. *After Babel: aspects of language and translation*. Oxford: Oxford University Press, 1992.
- Talmy, Leonard. 1985. "Lexicalization patterns: semantic structure in lexical forms". In Timothy Shopen (ed.), *Language Typology and Semantic Description, Vol.3: Grammatical Categories and the Lexicon*. Cambridge: Cambridge University Press, pp. 57-149.
- Taylor, John R. 1989. *Linguistic categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press.
- Teubert, Wolfgang. 2002. "The role of parallel corpora in translation and multilingual lexicography", in Altenberg, Bengt & Sylviane Granger (eds.), *Lexis in Contrast: Corpus-based approaches*. Amsterdam & Philadelphia: John Benjamins Publishing Company, pp. 189-214.
- Tobin, Yishai. 1993. *Aspect in the English verb: process and result in language*. London & New York: Longman.
- Tobin, Yishai. 1994. *Invariance, Markedness and Distinctive Feature Analysis: A Contrastive Study of Sign Systems in English and Hebrew*. Amsterdam & Philadelphia: John Benjamins.
- des Tombe, Louis & Susan Armstrong-Warwick. 1993. "Using function words to measure translation quality". In *Making Sense of Words: Proceedings of the Ninth Annual Conference of the UW Centre for the New OED and Text Research* (Oxford, September 27-28, 1993), pp. 1-18.
- Toury, Gideon. 1995. *Descriptive translation studies and beyond*. Amsterdam & Philadelphia: John Benjamins.
- Tsujii, Jun-ichi. 1986. "Future Directions of Machine Translation". In *Proceedings of COLING'86* (Bonn, 25-29 August 1986), pp. 655-668.
- Utsuro, Takehito, Yuji Matsumoto & Makoto Nagao. 1992. "Lexical Knowledge Acquisition from Bilingual Corpora". In *Proceedings of COLING'92* (Nantes, 23-28/7/98), Vol. II, pp. 581-587.
- Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca: Cornell University Press.
- Verkuyl, Henk. 1972. *On the Compositional Nature of the Aspects*. Dordrecht: D. Reidel.
- Verkuyl, H.J. 1989. "Aspectual classes and aspectual composition", *Linguistics and Philosophy* 12, 1989, pp. 39-94.

- Verkuyl, H.J. 1993. *A theory of aspectuality: The interaction between temporal and atemporal structure*. Cambridge: Cambridge University Press.
- Véronis, Jean (ed.). 2000. *Parallel Text Processing*. Dordrecht: Kluwer Academic Publishers.
- Vinay, J.-P. & J. Darbelnet. 1977 [1958] *Stylistique Comparée du Français et de l'Anglais: Méthode de traduction*. Paris: Didier, 1977, Nouvelle édition revue et corrigée.
- Wikberg, Kay. 1996. "Questions in English and Norwegian: Evidence from the English Norwegian Parallel Corpus". In Carol E. Percy, Charles F. Meyer & Ian Lancashire (eds.), *Synchronic Corpus Linguistics: Papers from the 16th International Conference on English Language Research on Computerized Corpora*. Amsterdam and Atlanta, GA: Rodopi, pp. 17-28.
- Zadeh, L.A. 1975. "Fuzzy logic and approximate reasoning". *Synthese* **30** (1975), pp. 407-428.
- Øverås, Linn. 1996. "In Search of the Third Code: An Investigation of Norms in Literary Translation". Cand. Phil. Degree thesis, University of Oslo.

