

Santos, Diana (2004). *Translation-based corpus studies. Contrasting English and Portuguese tense and aspect systems*. Amsterdam/New York: Rodopi, pp.xii, 173. ISBN 904201751-1. 44 € / US \$ 55.

In this contribution to the growing volume of publications on translation and corpora, Diana Santos sets out to tackle the topic from two points of view, that of translation and that of language engineering or natural language processing, which she sees as synonymous. The book is aimed at a wide range of potential readers: corpus linguists, linguists researching tense and aspect, semanticists and translation scholars. In fact, there is also something in the book for those with some background in language engineering, although all could find something to broaden their horizons. Santos' overall goal is to emphasise the importance of linguistics for translation, specifically semantics, and conversely, the importance of translation for studying natural language. It is the latter aspect which is really followed through in her study (a revised PhD thesis), as it lies at the basis of her methodology. For reasons that should become clear, she demonstrates that a translation-based study, by comparison with a contrastive study of two monolingual representations of the problem in hand, can add value by raising additional problems of performance in mapping meanings across languages, or more precisely, texts. How linguistics can support the theory or practice of translation is less clear, although some suggestions are touched on towards the end of the book. The computational perspective shines through clearly when translation is defined 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' (p.68).

Hence, a word of warning for the reader with a translation background: the book is published in the series Language and Computers: Studies in Practical Linguistics (No. 50; editors: C. Mair, C.F. Meyer, N. Oostdijk), so there is a lot of technical detail from the perspective of a non-specialist and some general assumptions about relevant background knowledge. This is fair enough for a publication in this series, but the computational perspective only emerges gradually in the first part of the book, being assumed rather than explicitly stated. About half the book—its core—is concerned with the development on the one hand of a linguistic model for the description of tense and aspect in Portuguese and in English, and on the other hand of a computational model for mapping the Portuguese and English language models, the 'translation network', which Santos regards as her main contribution.

Chapter 1 sets out to discuss a number of key concepts, including the importance of the performance aspect of her study using literary texts in a 'small' corpus (of about 51k words in English and Portuguese as source languages, together with their translations in Portuguese and English: see chapter 5), rightly arguing that grammatical studies need smaller corpora than lexical studies to be convincing. In the case of tense and aspect,

relevant data are ubiquitous. Other topics touched upon include 'vagueness' (also related in some way to ambiguity and indeterminacy) and an introduction to aspect 'for beginners'. Chapter 2 attempts a broad sweep—occasionally in a somewhat polemical tone—through parallel corpora, contrastive studies, translation studies and translation theory, and translation data as linguistic data. The core claim is that no universal or *a priori* categories are assumed in the analysis undertaken i.e. Santos rejects the view that common meanings are shared across languages but simply expressed in different forms. Nevertheless, the categories of 'tense' and 'aspect' clearly do have some universality in natural languages: it is their realisation and relationship in particular languages that differs.

The third and fourth chapters discuss in some detail—using many examples from the corpus—the tense and aspect systems of both English and Portuguese, pointing out the difficulties of providing one-to-one mappings at both system and use levels. A language model is defined as: 'a set of categories which are linked, working as the input and output of a set of operators' (p.70). 'Operators', we learn later, are the actual forms such as tensed verbs appearing in text and bearing, or contributing to, the expression of aspect. Each language model is shown as a network of nodes (the types of aspect such as activity, state, achievement in English and *obra, estado, série* in Portuguese) linked by arcs; the models are then related in both translation directions. From a translation point of view, what is particularly interesting is the cases where the translator has been 'coerced' to specify in the translation something that was underspecified or vague in the original, has introduced vagueness in the target text, or has split one clause into two, making corpus comparisons problematic and drawing attention to the fact that such decisions, particularly in the case of vagueness, may have implications for the plot. All the Portuguese examples are closely translated, although a gloss would have been more helpful.

Chapter 5 has a lot to offer to those specifically interested in English-Portuguese contrasts and confirms the experience of many corpus linguists, highlighting the motivation for this method of studying language: the results may take you by surprise. The corpus was first aligned (semi-automatically), manually annotated based on tensed verbs in the source texts, and then occurrences of particular forms counted in preparation for the analysis, which was conducted manually. What is not clear is what came first: the language models developed in Chapter 3, or the corpus analysis. In chapter 5, Santos states that the translation network model provided a necessary framework without which she would have been 'lost in too much data' (p.150), but this begs the question of the basis of the network model.

The final chapter considers how the translation network model could be evaluated. The most realistic proposal, and one that promises rewards as an application, is the enhancement of what Santos calls a 'translation

browser', by which I understand a tool that allows a user to search for specific data in a targeted way in a parallel corpus.

This book will be of particular interest to anyone engaged in Portuguese-English-Portuguese translation and to translation scholars who want to delve into computational linguistics in the context of text corpora. The tone is a little defensive in places, particularly in the first two chapters, and some judicious editing would have improved the clarity of the argument in places. But writing for a multidisciplinary audience in any language is a challenge, and the work reported really does make a contribution to the contrastive study of tense and aspect in Portuguese and English. Whether it also makes a contribution to corpus studies, is, however, another question.

Margaret Rogers
Centre for Translation Studies
University of Surrey
M.Rogers@surrey.ac.uk