

Introduction

 <https://doi.org/10.1075/btl.79.01int>

Pages vii–xii of

Topics in Language Resources for Translation and Localisation

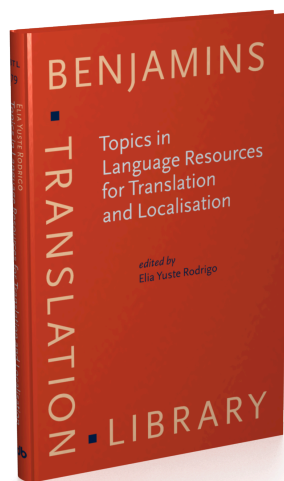
Edited by Elia Yuste Rodrigo

[Benjamins Translation Library, 79] 2008. xii, 220 pp.

© John Benjamins Publishing Company

This electronic file may not be altered in any way. For any reuse of this material written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

For further information, please contact rights@benjamins.nl or consult our website at benjamins.com/rights



Introduction

ELRA, the European Language Resources Association, defines Language Resources (LRs) as sets of language data and descriptions in machine readable form, which are specifically used to create, optimise or evaluate natural language processing and speech algorithms and systems, and generally, as core resources in the language services industries and localisation, for language studies, subject-area research, etc. Examples of LRs include but are not limited to written and spoken language corpora, terminology databases, computational lexica and dictionaries, software tools, etc. developed for different types of Human Language Technologies (HLT) applications, with their varied end-users in mind.

When Translation is understood as *process* rather than as *product* only, LRs play an indispensable role. Language resources such as the ones mentioned above may be of outstanding usefulness in the process of creating, standardising, leveraging, adapting... content for more than one language and culture. However, it has not been until recent years¹ that Language Resources for Translation (LR4Trans, for short) have been given the necessary attention. Since this has been the case mainly in academic and research circles, some efforts ought yet to be made to raise further awareness about LRs in general, and LRs for translation and localisation, in particular, to a wider audience in all corners of the world. Hence, the motivation number one behind this book.² The volume focuses on language resources from

1. Elia Yuste Rodrigo brought scholars and industry players from the areas of translation and corpus and computational linguistics together in a workshop held on 28th May 2002 in conjunction with LREC 2002 (Third International Language Resources and Evaluation Conference, Las Palmas de Gran Canaria, Spain). The workshop goal was to explore new avenues relating to language resources and technology-enhanced multilingual work and research. The term 'language resources for translation' was first formally used here. Given the attendants' interest, she would organise and chair two other workshops (second edition celebrated on 28th August 2004 as a satellite event of the 20th International Conference on Computational Linguistics – COLING at the University of Geneva, Switzerland; third edition, LR4Trans-III, held on 28th August 2006 under the auspices of LREC 2006 in Genoa, Italy).

2. Even if *Topics in Language Resources for Translation and Localisation* is the logical inheritance of the workshop series initiated by the editor, this is not a conference proceedings book and its novelty has to be emphasised. The selected contributions capture the current state-of-art in terms of research, work practices and industry standards. Much attention has been given to

several angles and in relation to current trends of multilingual content processes, being appealing to the heterogeneous readership of the Benjamins Translation Library (BTL)³ series worldwide.

Students, educators, researchers and professionals related to the translation and localisation arena will remember that in a not so distant past there seemed to be two extremes, represented at one end by those exclusively preoccupied with the then new market tools (essentially, commercial translation memories products), and at the other end by the ones that felt somehow threatened by an increasing degree of translation automation and kept defending human translation as the only possible alternative. What LR4Trans does is to underline the interaction of all the electronic language resources, applications and technologies that may be used in (learning about) the process of translating or localising – without forgetting about all the human agents that may also be involved therein, from technical writers, domain specialists, and corporate linguists of various kinds to computational linguists and future translators, to name a few. In other words, the approach behind LR4Trans is integrative and includes data, tools and human agents, allowing for targeted yet varied discussion points. This is one of the main features of the book you are now holding in your hands, its array of innovative topics for the language professional.

A truly practical and applied linguistic book in nature that is highly connected with multilingual technologies as used in translation and localisation processes must be written in as current a fashion as possible. Nevertheless, the principles behind this volume will not outdate so rapidly; only those aspects intrinsically dependant on a technological update or a new industry standard would require a content revision in the future. It is the potential of the flexible concept of language resources for translation and localisation what the readership has to adapt to their own working scenario and set of needs. The authors and the editor remain at the reader's disposal to clarify or expand on any of the issues presented here.

What follows are the rationale behind and a summary of the twelve selected contributions that make up this book. After going through the first half of the volume, the reader might get the impression that it focuses primarily on corpora, in one way or another. Yet there are two important things to note here: First, this is a reflection of a current yet maturing research trend and, secondly, this should be a good incentive to get to know more about different types of and aspects surrounding corpora (e.g. parallel and comparable corpora, treebanks, exploitation tools, interface and other design points, potential for teaching future translators in less

expand and update the information presented in any of the workshops. Some contributions have in fact been written from scratch to better serve the needs of the here targeted wider readership.

3. http://www.benjamins.com/cgi-bin/t_seriesview.cgi?series=BTL

resourced languages, etc.). As the reading progresses, the audience will find other exciting paths along the way to satisfy their curiosity about, for example, the perception of languages resources for translation in international organisations, how language resources could coadyuvate in addressing the needs of volunteer translators and the role of language resources in localisation (both to tackle specific challenges and from the latest industry and research joint efforts).

Chapter 1, *A Comparative Evaluation of Bilingual Concordancers and Translation Memory Systems*, is a thorough analytical study by **Bowker & Barlow** of two complementary (not competing) technologies that inform the language professional about the potential advantages that each offers so that they can choose the better tool for them. Opening the book with this chapter is intentional. The less acquainted with corpora readership will benefit from a first-hand introduction to what concordancing tools and parallel corpora can do for them.

In Chapter 2, however, the reader will learn that Translation problems derived from the specificity of a language or a register require a detailed linguistic analysis, which cannot always be accomplished with the use of parallel corpora. In *Interactive Reference Grammars: Exploiting Parallel and Comparable Treebanks for Translation*, **Hansen-Schirra** points out that the grammatical slant has not yet been addressed in corpus-based translation work. She argues that monolingual and multilingual treebanks may assume the role of grammatical reference resources for professional translators and translators-to-be. A translation corpus should be annotated with more abstract kinds of linguistic information, such as semantic and discourse information.

Chapter 3 by **Bernardini** and **Castagnoli**, *Corpora for translator education and translation practice*, aims to promote an educational rather than a training perspective of corpora for student translators. These should be educated to explore the role of the corpus. E-learning materials should foster this raising-awareness factor, ideally being contrastive in focus, i.e., corpora against other resources, such as dictionaries or translation memories (TMs). Other important aspects relate to corpus construction and corpus searching, which should be made faster and more user-friendly, and ideally integrated with CAT tools. Concerning the exploitation technologies discussed, we see a complementary or integrative standpoint once more, rather than an exclusive or imposing one.

Also willing to represent an educational rather than a training voice for translators, **Maia's** contribution and the book's Chapter 4, *Corpógrafo, V.4 – tools for educating translators*, describes the history, motivation and latest developments of a flexible tool suite, the Corpógrafo, which integrates and responds to principles of corpus linguistics, extraction and management of terminology and knowledge engineering. An online, freely available suite, this research environment for autonomous study also offers various possibilities for education in translation. It

should also be an incentive for self-discovery and improvement in other translation operational settings.

In Chapter 5, *Corpus exploitation for translation teaching and research: state of the art*, **Colominas & Badia** examine the weaknesses and strengths of current corpora interfaces and exemplify search types that can be relevant in translation training contexts or for translation research purposes, with a view to identify the basic requirements a corpora interface should satisfy. The lack of sufficiently large corpora representative of modern languages is currently being solved by means of web corpora, their analysis providing evidence of work done regarding this matter but also of the need for further work.

Sometimes not only how to better access corpora is at stake but also how to create translation corpora for less resourced languages. In Chapter 6, *The Use of Corpora in Translator Training in the African Language Classroom: A Perspective from South Africa*, **Gauton** draws on her expertise with electronic text corpora, corpus query tools and translation memory tools to enable the African language translator to mine target language texts for possible translation equivalents, coin terms in the absence of standardised practices, and re-use existing translation to attain terminology standardisation. Future action lines include further standardisation work and the transformation of the multilingual student output site into a large and comprehensive language resource available to external parties.

In a radically different work setting but equally aware of local constraints and requirements, **de Saint Robert** pinpoints in Chapter 7, *CAT tools in international organisations: lessons learnt from the experience of the Languages Service of the United Nations Office at Geneva*, that the usefulness of such tools does not have to be taken for granted. Here translation is seen as a highly interrelated activity which has to go hand in hand with and become closer to other internal business processes. Much attention has to be given to less sophisticated tools, but suitable for the organisation's *modus operandi*, as well as to the way internal language resources are built and integrated in the workflow.

Shedding some light on *Global Content Management – Challenges and Opportunities for Creating and Using Digital Translation Resources* (Chapter 8), **Budin** discusses the convergence of content management and cross-cultural communication. After exploring the concept of content, he goes on to explain that specialised translation is currently taking place within the wider, integrative paradigm of global content management. Translation resources (e.g., translation memories and other aligned corpora, multilingual terminological resources, reference resources, etc.) are typical examples of content that needs to be managed in global action spaces.

Bey, Boutet and Kageura then present *BEYTrans: A Wiki-based Environment for Helping Online Volunteer Translators* (Chapter 9). Following major Web 2.0 advances, this research project reflects new collaborative work patterns among

volunteer translators that could open new avenues for all communities involved in translation. Leveraging the Wiki technology, BEYTrans aims at empowering online translators through system components for producing a quick, yet high-quality translation in several languages. A range of system functionalities allow them to manage the language resources themselves online, in the *wiky* fashion.

Chapter 10 is also the result of innovative research. Cruz Lara et al. are concerned with *Standardising the management and the representation of multilingual data: the Multilingual Information Framework*. The MLIF framework, based on a methodology of standardisation resulting from the ISO (International Standards Organisation), is being designed with a high-level common conceptual model of multilingual content in mind and as a platform allowing interoperability among several translation and localisation standards and their related tools. This interoperability is the main benefit of MLIF, which also facilitates the evaluation and comparison of multilingual resources and tools.

Kato & Arisawa's Chapter 11, *Tagging and Tracing Program Integrated Information*, introduces the reader to a software internationalisation challenge by focusing on the translation of Program Integrated Information (PII). PII is normally separated from the computer programs themselves and brought into text resource files that are translated outside the program development lab. How can this decontextualisation be compensated during the translation verification test (TVT)?

In Chapter 12, *Linguistic Resources and Localisation*, Schäler provides the reader with the essential definitions surrounding localisation (L10N). The discussion is followed by a real-life case study showing the use of language resources in localisation that laid the foundations for the "translation factory". The commonalities found in the L10N process, in terms of frequent updates, repetitive material, etc. may facilitate standard approaches to L10N problems. Yet the mark-up and formatting of source material and the complexity of L10N processes hinder localisation automation efforts. These have to be tackled when developing innovative L10N frameworks. We encourage every reader to keep on reading till the very end of the book and find out what IGNITE is all about.

Acknowledgements

I am exceptionally lucky to have drawn on the outstanding expertise of all the contributors who have enthusiastically believed in this exciting book. My special thanks must also go to Prof. Sue Ellen Wright (Kent State University), who did see a need for such a volume and encouraged me to submit the book proposal.

I would also like to thank all the professors and fellow lecturers and researchers from the University of Valencia, UMIST (now University of Manchester), Staffordshire University and the University of Zurich, who have forged my career as a

professional linguist and researcher. I also feel indebted to the fellow translators/localisers, terminologists, project managers and language service providers, with whom I work on a regular basis, for having made me capable of keeping abreast of what the real world does need from university scholars (and the way round) and helping me keep a *sane* balance between academia and industry in the last fourteen years.

Last but not least, I am especially thankful to the publishers' team at John Benjamins Publishing Company, in particular to Prof. Yves Gambier (University of Turku), for his valuable time and attentive responses and to Isja Conen, for her help from the very beginning.

Elia Yuste Rodrigo
(Zurich / Dublin, Winter-Spring 2008)

N.B. All the online pointers and links mentioned in this book were last consulted in January 2008.