Chapter 1

Introduction

Oliver Czulo

Silvia Hansen-Schirra

University of Mainz, Germersheim

Contrastive Linguistics (CL), Translation Studies (TS) and Machine Translation (MT) have common grounds: They all work at the crossroad where two or more languages meet. Recently, all three have shown a strong affinity towards using multilingual (parallel and comparable) corpora. In MT, for instance, parallel data collections serve as training material for translation models, as well as for related issues from computational linguistics like multilingual grammar induction, automatic lexicography, etc. Translation scholars use corpora and strive for empirical models of the translation process (including translation strategies or specific properties of translated text). For professional translators, multilingual corpora serve as reference works that enable quick interactive access and information processing. Contrastive linguistics uses corpora both to ground its findings empirically and to uncover differences in linguistic features that have not been studied before. Furthermore, multilingual corpora have found their way into lexicography and grammar writing.

Despite their inherent relatedness, methodological exchange between the three disciplines is rare. For instance, when parallel corpora are used in CL or MT, factors like translation direction or translation properties and strategies are largely ignored. Also, MT in particular is agnostic about dimensions like text type or register. At the same time, the use of multilingual annotation and query techniques is often restricted to the most basic techniques in CL and TS - if applied at all.

This special issue touches upon areas where the three fields converge. It results directly from a workshop at the 2011 German Association for Language Technology and Computational Linguistics (GSCL) conference in Hamburg where

Oliver Czulo & Silvia Hansen-Schirra. Introduction. In Oliver Czulo & Silvia Hansen-Schirra (eds.), *Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation: TC3 II*, 1–3. Berlin: Language Science Press.
DOI:10.5281/zenodo.1019679

researchers from the three fields presented and discussed their interdisciplinary work.

The volume begins with a contribution by Steiner who takes a broad perspective on the topic of cross-fertilisation between CL and TS on the one hand and what he refers to as "relevant sub-fields of Computational Linguistics", extending to MT, on the other. He discusses three methodological approaches to inherently multilingual tasks and how they could serve as valuable blueprints for other disciplines, including operationalization of hypothesis testing on lexicogrammatical data from parallel corpora, study of textual cohesion in originals and translations, and integration of product- and process-based data. Steiner concludes with suggestions about how the above-mentioned disciplines could profit from each other.

Korzen and Gylling present a study on the structure of Italian and Danish texts both from a contrastive and a translational perspective. Using samples from Europarl, Korzen and Gylling reveal differences in information density and clause linkage between the two languages in a contrastive fashion. They then go on to formulate some simple translation rules based on their findings.

The progression from contrastive investigations to translation-oriented observations is also characteristic of the contribution by Zinsmeister, Dipper and Seiss. Taking examples from German-English bitexts from the Europarl corpus, the authors contrast the realisation of abstract anaphors in the two languages. They then examine translation-specific differences in the realisations and how the findings from the contrastive analysis may help improve translation procedures.

Thunes aims at linking perspectives from TS and MT explicitly. Based on the crucial to TS theory notion of text type, she develops a classification of translation complexity for two text types. Thunes then applies her classification to MT by proposing to use it as a diagnostic for the feasibility of MT for text types in general.

Kremer, Hartung, Padó and Riezler present a study on how human translation could benefit from MT-generated data. In their experiment, translators were asked to translate adjective-noun pairs. Based on a phrase table created by an SMT system, possible adjective translations were suggested to the translators. The results of the study show significant improvement in translation quality.

The concluding contribution of this volume by Carl and Dragsted reports on a process-oriented study, a field of research from which both TS and MT can benefit. The authors investigate the "monitor model", a hypothesis by which a monitor disrupts the default mode of literal translation in case a problem occurs. In their experiment, subjects either copied or translated a source text. In the contribution, the results from the reading and writing process of copyists and

translators are contrasted, and insights into the role of the decoding task are presented.

While the studies contained in this volume draw from a wide variety of objectives and methods, and various areas of overlaps between CL, TS and MT are addressed, the volume is by no means exhaustive with regard to this topic. Further cross-fertilisation is not only desirable, but almost mandatory in order to tackle future tasks and endeavours, and TC3 remains committed to bringing these three fields even closer together.

Germersheim, June 2012
Oliver Czulo, Silvia Hansen-Schirra