Chapter 1

Editorial

Georg Rehm

Felix Sasaki

Daniel Stein

Andreas Witt

The roots of this special issue of *Translation: Computation, Corpora, Cognition* go all the way back to 2011. At the end of September of that year, the guest editors organised a workshop at the Conference of the German Society for Computational Linguistics and Language Technology (GSCL), which took place in Hamburg. The topic of the GSCL 2011 conference – "Multilingual Resources and Multilingual Applications" – had already set the stage for our pre-conference workshop on September 27, 2011, which put special emphasis on "Language Technology for a Multilingual Europe".

Our intention behind this workshop was to bring together various groups concerned with the umbrella topics of multilingualism and language technology, especially multilingual technologies. This encompassed, on the one hand, representatives from research and development in the field of language technologies, and on the other hand users from diverse areas such as, among others, industry, administration and funding agencies. Two examples of language technologies that we mentioned in the call for contributions were Machine Translation and processing of texts from the humanities with methods drawn from language technology, such as automatic topic indexing, and text mining, as well as integrating numerous texts and additional information across languages.

What these kinds of application areas and research and development in language technology have in common is that they either rely – critically – on language resources (lexicons, corpora, grammars, language models etc.) or produce

Georg Rehm, Felix Sasaki, Daniel Stein & Andreas Witt. Editorial. In Georg Rehm, Felix Sasaki, Daniel Stein & Andreas Witt (eds.), *Language technologies for a multilingual Europe: TC3 III*, 1–4. Berlin: Language Science Press. DOI:10.5281/zenodo.1291922

these resources. A multilingual Europe supported by language technology is only possible if an adequate and interoperable infrastructure of resources (including the related tooling) is available for all European and other important languages. It is necessary that the aforementioned groups and other communities of developers and users of language technology stand as a single homogenous community. Only if all members of our (quite heterogeneous and hitherto mostly fragmented) community stand together and speak with one voice, it will be possible to assure the long-term political acceptance of the "Language Technology" topic in Europe.

The Workshop "Language Technology for a Multilingual Europe" was co-organised by two GSCL working groups (*Text Technology* and *Machine Translation*) and META-NET (http://www.meta-net.eu). META-NET, an EU-funded Network of Excellence, is dedicated to building the technological foundations of a multilingual European information society. To this end, META-NET is forging META, the Multilingual Europe Technology Alliance.

This special issue of *Translation: Corpora, Computation, Cognition* includes the majority of the papers presented at the GSCL 2011 Workshop "Language Technology for a Multilingual Europe", held at the University of Hamburg on September 27, 2011, along with several additional contributions.

The first article, "Machine Translation – Past, Present and Future", provides an overview of what must be considered the essential core of multilingual technologies. Setting the stage, Daniel Stein looks at the history of MT and discusses current approaches and future perspectives. The backgrounds of the next two articles are two interlinked EU-funded initiatives. Georg Rehm describes the Network of Excellence META-NET, which consists of 60 research centres in 34 European countries, and its goal to build the technological foundations of a multilingual Europe. He provides a summary of one of the key outcomes of the initiative, "The META-NET Strategic Research Agenda for Language Technology in Europe". In his article "Metadata for the Multilingual Web", Felix Sasaki provides an overview of the Internationalization Tag Set (ITS) which will become a W3C recommendation later in 2013. ITS 2.0 is one of the key results from the European Union-funded Multilingual Web project.

The second part of this special issue contains six full research papers. First is Uwe Reinke with a paper on the "State of the Art in Translation Memory Technology", that focused upon technologies applied by human translators. He takes a detailed look at major concepts and recent trends in research and also in commercial Translation Memory (TM) systems, with an emphasis on integrating MT into TM, data exchange formats, and approaches of improving the information retrieval performance of TM systems. As a complement to the technologies

used by translators, Melanie Siegel examines "Authoring Support for Controlled Language and Machine Translation", i. e., language technologies that help and assist authors to produce high quality documents. She concludes that it is necessary to combine methods from authoring support and MT and to make them integrated tools in the production and translation process. The paper "Integration of Machine Translation in On-line Multilingual Applications" by Mirela-Stefania Duma and Cristina Vertan takes a look at a difficult and challenging problem that MT, especially statistical MT, is confronted with domain adaptation. The method employed by the authors for this task is language model interpolation, which produces good results even when only sparse domain-specific training data is available. This, in turn, is an advantage for less-resourced languages. The next article concentrates on a specialised application that provides help for users of monolingual or crosslingual search. In "Disambiguate Yourself – Supporting Users in Searching Documents with Query Disambiguation Suggestions", Ernesto William De Luca and Christian Scheel describe a semantic approach and a corresponding architecture and prototype for making more sense of queries as they are typed in by the user. The penultimate article, "Multilingual Knowledge in Aligned Wiktionary and OmegaWiki for Translation Applications", goes back to the topic of MT. Michael Matuschek, Christian M. Meyer, and Iryna Gurevych take a look at multilingual lexical-semantic resources and their role in translingual technologies. They focus on two crowd-sourced resources and present methods for aligning these resources in order to combine them on the level of word senses, this way providing increased coverage and improved interoperability. In the final article, Igor Leturia and colleagues present "The BerbaTek project for Basque: Promoting a less-resourced language via language technology for translation, content management and learning". In this joint project between companies and research centres, the partners developed several technologies for the Basque language which is, as the META-NET study "Europe's Languages in the Digital Age" pointed out, among the 21 European languages in danger of digital extinction.

Since we held the workshop, there have been quite a few very positive developments in the area of multilingual language technologies *from* Europe *for* Europe. Among those developments are a new series of projects funded by the European Commission such as, for example, QTLaunchPad, or additional projects around the open source machine translation system Moses. In addition, META-NET organised its third META-FORUM conference in June 2012, which was attended by more than 250 participants from the domains of research, industry, administration, and politics. Important milestones for the work of META-NET were the publica-

tion of the *META-NET Language White Papers* (September 2012) and the *META-NET Strategic Research Agenda for Multilingual Europe 2020* (January 2013). While the first funded phase of the initiative came to an end on January 31, 2013, there will be a fourth META-FORUM conference later this year (http://www.meta-forum.eu). Among the topics of META-FORUM 2013 are upcoming opportunities for multilingual technologies in the frameworks of Connecting Europe Facility (CEF) and Horizon 2020. On February 29, 2012, the Common Language Resources and Technology Infrastructure (CLARIN) received EU-ERIC legal status, as the second European Research Infrastructure overall. These initiatives clearly demonstrate the emphasis the European Commission lays on the further development of language resources and language technology.

This special issue would not have been possible without the help of several colleagues. First of all, the guest editors would like to thank all authors who contributed articles to this special issue and those who presented papers at the workshop back in September 2011. We would like to thank the reviewers who provided valuable and helpful feedback to all authors. Many thanks are also due to our colleague Sarah Weichert (DFKI) who supported us in a critical phase during the preparation of this special issue. Finally, we would like to express our gratitude towards the editors of *Translation: Computation, Corpora, Cognition*, especially Oliver Čulo, who not only made it possible that we could publish the results of our workshop in this journal but also constantly supported us whenever necessary.

Georg Rehm, Felix Sasaki, Daniel Stein, Andreas Witt June 21, 2013