

Preface

AMTA 2002: From Research to Real Users

Ever since the showdown between Empiricists and Rationalists a decade ago at TMI 92, MT researchers have hotly pursued promising paradigms for MT, including data-driven approaches (e.g., statistical, example-based) and hybrids that integrate these with more traditional rule-based components.

During the same period, commercial MT systems with standard transfer architectures have evolved along a parallel and almost unrelated track, increasing their coverage (primarily through manual update of their lexicons, we assume) and achieving much broader acceptance and usage, principally through the medium of the Internet. Webpage translators have become commonplace; a number of online translation services have appeared, including in their offerings both raw and postedited MT; and large corporations have been turning increasingly to MT to address the exigencies of global communication. Still, the output of the transfer-based systems employed in this expansion represents but a small drop in the ever-growing translation marketplace bucket.

Now, 10 years later, we wonder if this mounting variety of MT users is any better off, and if the promise of the research technologies is being realized to any measurable degree. In this regard, the papers in this volume target responses to the following questions:

- Why aren't any current commercially available MT systems primarily data-driven?
- Do any commercially available systems integrate (or plan to integrate) data-driven components?
- Do data-driven systems have significant performance or quality issues?
- Can such systems really provide better quality to users, or is their main advantage one of fast, facilitated customization?
- If any new MT technology could provide such benefits (somewhat higher quality, or facilitated customization), would that be the key to more widespread use of MT, or are there yet other more relevant unresolved issues, such as system integration?
- If better quality, customization, or system integration aren't the answer, then what is it that users really need from MT in order for it to be more useful to them?

The contributors to this volume have sought to shed light on these and related issues from a variety of viewpoints, including those of MT researchers, developers, end-users, professional translators, managers, and marketing experts. The jury appears still to be out, however, on whether data-driven MT, which seems to have meandered along a decade-long path of evolution (instead of revolution, as many thought it might be), will lead us to the holy grail of high-quality MT. And yet, there is a sense of progress and optimism among the practitioners of our field.

I extend my sincere thanks to the members of the AMTA 2002 program committee, who sacrificed time and effort to provide detailed analyses of the papers submit-

ted to the conference. Many of the authors expressed gratitude for the insightful and helpful comments they received from the reviewers as they prepared their papers for publication.

Many thanks also go to the organizers of AMTA 2002 who spent untold hours to ensure the success of the conference:

Elliott Macklovitch, General Chair
Violetta Cavalli-Sforza, Local Arrangements Chair
Robert Frederking, Workshops and Tutorials
Laurie Gerber, Exhibits Coordinator
Jin Yang, Webmaster
Debbie Becker, Registrar

In particular, I am grateful to Elliott and Laurie, who provided me with a constant and sustaining flow of guidance and wisdom throughout the conception and assemblage of the program.

Final and special thanks go to Deborah Coughlin, who assisted me in managing the submissions to the conference and in overseeing all substantial aspects of the production of this volume, and to my other colleagues at Microsoft Research, who supported us during this process.

August 2002

Stephen D. Richardson