DE GRUYTER OPEN

ICAME Journal, Volume 40, 2016, DOI: 10.1515/icame-2016-0001

An interview with Joybrato Mukherjee, the Chair of the ICAME Board

Merja Kytö (Uppsala University), Anna-Brita Stenström (University of Bergen) and Ilka Mindt (University of Paderborn)

To mark the 40th issue of the *ICAME News / ICAME Journal*, the Editors interviewed Joybrato (Danny) Mukherjee, the current Chair of the ICAME Board, about issues of interest to corpus linguistics and developments in this field. Danny has served in the ICAME Board since 2006 and been the Chair since 2011.

- *Digital Humanities and Corpus Linguistics – how is the current connection/ relation between them and how can these two fruitfully profit from each other?*

— I feel that corpus linguistics has been one of the major fields in which what has come to be known as digital humanities has been advanced and developed. In fact, many of the data-oriented approaches and the empirical methodologies that are characteristic of digital humanities – in the linguistic sciences as well as in other disciplines – go back to and have profited from the contribution of corpus linguistics ever since the 1970s. For example, an increasing number of literary and cultural studies today make unprecedented use of corpus data and corpus technology. Similar trends can be observed in economics and history, to mention just two more areas that have benefited from the empirical footing that corpus-based techniques provide. In essence, therefore, it strikes me that it is important for corpus linguistics not to be absorbed by the digitalization of humanities. Quite on the contrary: the corpus-linguistic community should provide more of a plat-

form for an interdisciplinary exchange about how to best apply state-of-the-art corpus-based methods to answer research questions that are related to language and communication in all fields of humanities.

• *Hosting and Archiving of corpora: should ICAME be more present in the CLARIN-initiative and ask other corpus compilers to seriously consider offering their corpus to CLARIN?*

— ICAME has long been the only institution to provide a safe harbor for archives and corpora compiled by a multitude of research teams around the world. Thus, the major aims of CLARIN had been at the heart of ICAME long before this large-scale European initiative was launched. The Executive Board of ICAME welcomes this enterprise, and we are very happy about the fact that all ICAME corpora have found a new home at the Norwegian component of CLARIN in the meantime. We will continue to strongly encourage current and future corpus compilers to offer their data to CLARIN as it provides an open and stable platform for long-term storage, distribution and analysis.

• *Is there a "pragmatic turn" in corpus linguistics?*

— I am not so sure whether I would call it a "pragmatic turn" and restrict a general trend in corpus linguistics to the realm of pragmatics only. The general trend I see in much recent work in corpus linguistics is an increasing awareness that corpus data and their description and analysis are not an end but a means – to find out more about how and why language users speak and write in specific ways under certain communicative, contextual, sociocultural and historical conditions. The utilization of corpus data in pragmatic approaches to language description is, thus, part of an overarching trend, namely an increasing functionalization of corpus data and corpus analyses in a broader setting – and the "pragmaticization" of corpus linguistics certainly plays an important role in this development of the field at large: for example, more and more attempts are being made to annotate corpora with pragmatic tags, and various elaborated statistical approaches are being developed to model the pragmatic complexity of the underlying factors guiding language users in their choice of linguistic forms and structures. Against this background, it was no coincidence, by the way, that the ICAME Conference in Ascona 2008 was explicitly dedicated to the emerging field of "corpus pragmatics".

- *How can ICAME encourage young researchers to conduct research with corpora?*

— The best thing ICAME can do is, on the one hand, to provide an attractive and welcoming habitat for excellent young linguists to present and discuss their findings at ICAME conferences and in the ICAME Journal, and, on the other hand, to foster excellence in research in corpus linguistics in general. It is important for ICAME to continue to place special emphasis on the descriptive relevance, the analytical plausibility and the methodological soundness of corpus-linguistic research – early-career corpus linguists should be encouraged to use corpus technology to address relevant and unanswered research questions about actual language use. The unprecedented scale of available corpora does not mean that the linguist's expertise and intuition is no longer needed – the opposite is true: Today, it seems even more important than in the early days of our discipline for corpus linguists to define their research interests without simply restricting themselves to what can be easily and speedily done with available data and methods.

- *Peer review and quality management systems are current issues. How does ICAME care for its standards and ensure high quality in research?*

— By example. ICAME conferences have developed from small-scale meetings of relatively few researchers to openly announced conferences with up to 200 participants, with a rigid double-blind peer review regime in place for the selection of abstracts (which is inevitable in light of the huge number of submissions), with conference proceedings including a rigidly selected group of papers. We have always been lucky with our conference organizers as they always found the right balance between an academic program focusing on excellent research and allowing for intense discussions on the one hand, and a social program on the other, bringing together corpus linguists from all over the world, integrating early career researchers in particular and, thus, fostering a spirit of collaboration and an open exchange of ideas. The ICAME Journal, too, has a rigid peer-review system in place.

- *What are the prospects of compiling new corpora of naturally occurring spontaneous conversation in the near future (from an economical and technical point of view)?*

— Corpus-based research into spontaneous spoken speech has always suffered from the enormous resources that corpus compilers have to invest in the transcription and annotation of the raw audio data – not many research

teams were able to invest the necessary manpower. Of course, the Internet offers a nearly unlimited amount of spoken data from a range of registers, but these data, too, would have to be transcribed (or the available transcriptions would have to be checked carefully) and annotated according to corpus-linguistic standards, at least for a range of speech-related research topics, e.g. intonation and fluency. In this regard, I believe and fear that – in spite of all the advances in software development and in spite of the potential automatization of individual transcription and annotation steps – not much will change in the foreseeable future if the spoken corpora are to live up to established corpus-linguistic standards. On the other hand, we witness a dangerous development because many people outside our field mistakenly assume that spoken language data were a low-cost input for corpus analyses in the year 2015. In Germany, for example, funding agencies are increasingly reluctant to financially support corpus-compilation projects including the collection, transcription and annotation of spoken corpora. This situation poses a serious threat to the initiation of new corpus projects which we definitely need, especially with regard to spontaneous spoken language.

- *ICAME has developed from a small group to a large community of researchers. What were the main aims of ICAME 25 years ago? What are ICAME's aims today?*

— In my view, the central aims of ICAME have not changed over the past 25 years: to bring together linguists interested in the collection and analysis of corpora and to provide a platform for the exchange and discussion of corpus-linguistic research work, in particular through ICAME conferences and the ICAME Journal. However, the field is now much broader, with various other, partly competing, corpus-linguistic organizations in existence, with many more outlets for excellent corpus-linguistic research than 25 years ago, with the Internet as a convenient source for language data of an unprecedented size, with a much stronger focus on quantitative rigor and the plausibility of statistical models, with certain areas – such as World Englishes and Learner Englishes – figuring much more prominently nowadays. All this makes it all the more necessary for ICAME to reflect and foster state-of-the-art research and to represent the entire range of English corpus linguistics.

The Editors thank Danny Mukherjee for granting them this interview.