

Introduction

Hanna Christ, Daniel Klenovšak, Lukas Sönning and Valentin Werner
University of Bamberg

This volume presents selected contributions to MaLT 2015, the first *Methods and Linguistic Theories Symposium*, which was organized by the *Bamberg Graduate School of Linguistics* and hosted at the University of Bamberg from November 27th–28th 2015. The focus of the event was on bringing together research methods and theory, and – we were glad to realize – it struck a timely theme. On the one hand, this is due to the fact that over the past few decades, linguistics as a field has undergone a major transformation, evolving from a predominantly descriptive to an increasingly empirical discipline. Even a cursory glance at current linguistic journal volumes clearly reveals this shift. At the beginning of the 21st century, it is fair to acknowledge that research on language relies heavily on empirical and quantitative evidence.

Many would agree that linguistic theorizing has benefited from this transformation. Once largely dominated by introspective methodology, we now have at our disposal additional tools for inductive reasoning, and, perhaps more importantly, for assessing the adequacy of established models and theories. In recent decades, our field has gained new insights by *directly* turning to the object of knowledge – that is, language and how it is used and processed by humans. As such, data may suggest new routes toward understanding and question familiar paths. Upon confrontation with quantitative evidence, our formalized state of knowledge may have to be refined or rethought.

In the course of this transformation, methodological know-how has become one of the key qualifications for researchers, especially young academics at the beginning of their careers. New skills are required to find a way through the quantitative maze in the literature and to choose a sensible approach for the particular phenomenon on one's desk. Learning from data is both an art and a science. In a field where widespread use of empirical methodologies is a relatively recent development, it may at times be difficult to acquire the relevant (statistical and other) literacy.

However, it is a fact that linguistics curricula at the tertiary level commonly lack training in empirical methods, so that young researchers more often than not have to resort either to external offers (for instance workshops organized by professional organizations) or become self-taught (provided their institutional libraries contain adequate resources). Further, the empirical turn in linguistics has gone hand in hand with a considerable diversification of research methods. While this diversity has come to be seen as a strength of linguistics as a field, the plethora of procedures may puzzle even the seasoned researcher. Still, ignoring methodological developments is not an option if meaningful linguistic research is to be conducted. In the light of the current vibrancy of the interplay between research methods and theory building, the aim of MaLT was twofold:

- (i) to provide a forum for researchers to meet peers from other branches of linguistics;
- (ii) to provide a venue to look beyond specific disciplinary boundaries and draw inspiration from neighboring fields.

The emphasis on cross-disciplinary exchange offered researchers the opportunity to expand their repertoire of theoretical approaches and methods within and beyond those typically adopted in their subfields.

The conference was thus conceptualized as an ensemble of talks and practical workshops, which offered hands-on advice in two broad fields currently taking center stage in the empirical study of linguistic structures: corpus linguistics and experimental linguistics. In the former area, Samantha Laporte (University of Louvain) introduced the hows, whats and whys of corpus linguistics in her workshop *What corpora can do for you: An introduction to corpus methods and corpus tools*. Quantitative methods for handling corpus data were discussed in a practical *Introduction to statistics for corpus linguistics* by Stefan Evert (University of Erlangen-Nürnberg). Two workshops focused on experimental linguistics. Franziska Günther (Ludwig-Maximilians-University of München) discussed the fundamentals of experimental work in *How to collect (and combine) linguistic and behavioral data: A practical workshop on experiments in linguistics*. Participants also had the opportunity to delve deeper

into the state-of-the-art toolbox of psycholinguists, with Franziska Hartung's (MPI Nijmegen) workshop *Experimental methods in discourse processing*. The general program was rounded off by two topical plenary talks by Alexander Ziem (University of Düsseldorf) titled *From discourses to corpora: Cognitive approaches to (lexical) meaning-making* and by Martin Hilpert (University of Neuchâtel) on *How to blend MaLT: Bringing methods and linguistic theory together*. We owe heavily to the latter for inspiration for the title of this volume and would like to express our gratitude to all workshop conveners and plenary speakers for the time and effort invested.

With close to 100 participants from more than 10 countries around the globe, MaLT can be considered a great success. With the program being aimed at early-career researchers, one main concern of the organizers was to grant participation in the conference and the workshops free of charge. It thus goes without saying that MaLT 2015 would not have been possible without generous financial support. In particular, we would like to thank the German Academic Exchange Service (DAAD) for supporting MaLT through its IPID4all scheme. We also received considerable funding from the University of Bamberg (FNK) and the alumni represented by the Universitätsbund Bamberg e.V.

As an event such as MaLT is very much a collective effort, we would like to extend our gratitude to all those who supported the symposium in various ways, first and foremost to Marion Hacke and Simone Treiber from the *Trimberg Research Academy* (TRAc) of the University of Bamberg. Further we would like to thank Geoffrey Haig and Hans-Ingo Radatz as speakers of the *Bamberg Graduate School of Linguistics* for their input, all those involved chairing the individual sessions (Hanna Budig, Romina Buttafoco, and Ole Schützler), and the student helpers (Carolin Cholotta and Katharina Scheiner), who ensured a smooth running of the whole event.

Further, as regards the preparation of this volume, we would like to say thank you to all authors for their efforts in preparing and revising their manuscripts, and for their feedback as internal reviewers for other papers. In addition, Romina Buttafoco, Jiří Milička, Jochen Podelo, Ole Schützler, and Fabian Vetter acted as external referees and provided helpful suggestions to improve the overall quality of the individual papers.

That said, in this book we are proud to present a selection of the contributions. They can be seen as the essence of what MaLT was about, and nicely illustrate the range of topics covered as well as the various concerns and approaches that featured during the event.

The first part is predominantly oriented toward crucial aspects relating to linguistic methodology in terms of data types, collection, presentation and analysis.

Alexander Ziem opens the volume with a paper titled *From discourses to corpora: (lexical) meaning-making as a challenge for cognitive semantics*, which discusses the use of corpus-linguistic tools in cognitive-linguistic discourse analysis. Navigating within a cognitive-functional framework, the analyses are grounded in the assumption that linguistic meaning emerges from language use. The primary object of study is what Ziem refers to as “U-relevant knowledge”: language users’ cumulative and collective knowledge about linguistic signs. The paper exemplifies empirical procedures for the investigation of how U-relevant knowledge is shaped in discourse. Methods at different levels of analysis are illustrated with data from discourses on sociopolitical and economic crises in Germany over the past 40 years. First, Ziem demonstrates the use of exploratory lexicometric techniques, which not only serve to provide a birds-eye perspective on lexical patterns across discourses, but also yield insights for hypothesis generation. To add substance to the abstract numbers provided by multifactorial analyses, word frequency distributions are compared to identify lexis that is specific to a particular discourse, or shared across two or more discourses. Ziem then shows how analytical categories borrowed from functional-cognitive grammar allow the researcher to “zoom in” further to uncover different conceptualizations of shared lexical items. A frame-based analysis of lexical meaning may thus detect more fine-grained differences in the way concepts are used and essentially shaped in discourse. Throughout the paper, Ziem’s central concern is to illustrate how discourse analysis can benefit from the use of corpus-linguistic methods.

The two following chapters deal with methodological aspects of sociolinguistic fieldwork. Adina Staicov sets the scene with her paper *Methodologies in sociolinguistic fieldwork*, in which she provides practical

advice on a wide range of issues involved in collecting sociolinguistic data in the field. The author discusses essential steps in planning and carrying out field research and gives valuable insights based on her own experience, which she gathered in research projects on different varieties of English, including the Fiji islands in the South Pacific, British Asians in London and the San Francisco Chinatown community. Her advice carefully balances technical, cultural, and personal reflection. Throughout her contribution, she stresses the importance of knowledge and awareness of the target community, which may have critical implications for the researcher's conduct. Staicov's contribution is valuable for anybody planning to enter the field, as her advice and experience reports sensitize the reader to potential challenges along the way.

In another contribution relating to linguistic fieldwork, this time for the study of variation in a non-native variety of English, Sofia Rüdiger introduces an innovative approach to the elicitation of conversational material. In *Cuppa coffee? Challenges and opportunities of compiling a conversational English corpus in an Expanding Circle setting* she first contrasts written and spoken linguistic data and discusses both why spoken data – despite their often-cited primacy – are understudied, and why data allegedly “spoken” often underlie certain constraints (for instance in terms of a formal setting during sociolinguistic interviews) that preclude an analysis as “conversational” and “naturalistic”. Rüdiger continues to argue that many of these constraints can be avoided if a truly informal interview setting is established, and she proposes what she labels the “cuppa coffee method”, where interviewer and interviewee engage in mutual exchange over a cup of coffee (used as “social lubricant”) in a public space. Like in traditional interview approaches, parts of the conversation are recorded and thus can be subject to linguistic analysis. Rüdiger also points out potential drawbacks of the “cuppa coffee method” such as increased transcribing time or potential recording quality issues. That the method developed by her is not merely an intellectual game is shown in the last part of her chapter where she details how her approach resulted in the compilation of the *Spoken Korean English Corpus* (SPOKE) used to analyze naturalistic speech.

Collecting linguistic data via online experiments is a mixed blessing, as is shown by Jana Häussler and Tom Juzek. In their contribution *Detecting and discouraging non-cooperative behavior in online experiments using an acceptability judgment task*, the authors point out that recruiting participants via crowdsourcing platforms like *Amazon's Mechanical Turk* is a cheap and easy way of collecting quantitative data. However, focusing on the observations in their acceptability judgment tasks, the authors also cast doubt on the reliability of these data, and show that participants often exhibit non-cooperative behavior in the sense of merely “clicking their way through”. Thus, participants are negligent of actually performing the task, which potentially influences the quality of the overall results in a negative way. Through careful examination of their methodology, Häussler and Juzek provide the reader with ways of detecting such behavior based on response times. They further discuss some ideas on how to circumvent and discourage unaccommodating ratings like the implementation of booby trap items and the tracking of response times in order to keep the data as “clean” as possible.

Moving on to visualization methods in quantitative research, the last contribution of part one presents a relatively unfamiliar type of display – *The dot plot: A graphical tool for data analysis and presentation*. In his paper, Lukas Sönning introduces and illustrates the dot plot and argues for its routine usage in quantitative research. Based on principles of graph construction and empirical evidence from research into visual perception, advantages of dot plots over other commonly used chart types such as the bar chart are demonstrated. The paper outlines design options and extensions and illustrates the application of this chart type in linguistic data analysis, including examples from corpus linguistics and meta-analysis. Sönning also reflects on its limitations and provides Microsoft Excel spreadsheet templates for the production of dot plots.

The papers in the second part aim to show how varying methodological approaches or changing methodological parameters can affect the interpretation of results, which may yield different implications for linguistic theory building.

In the field of word formation, Chiara Naccarato illustrates how the notoriously vague concept of productivity can be assessed using quanti-

tative diachronic data. Her paper *A corpus-based quantitative approach to the study of morphological productivity in diachrony: The case of samo-compounds in Russian* investigates the changing productivity of the Russian prefixoid *samo-* from 1700 to the present day. In her concise analysis, the author applies Baayen's measure of "potential productivity" and discusses its major disadvantage: unreliability due to artifacts when it is applied to corpora of different sizes, yielding a result of supposedly decreasing productivity. This problem is overcome through the use of a Large Number of Rare Events model estimating the expected number of types and *hapax legomena* with *samo-*. Thus, Naccarato is able to demonstrate the increase in productivity of *samo-* over time. She goes on to analyze the productivity of different compound patterns with *samo-* in detail, confirming the frequently assumed interrelation of lexicalization and productivity: highly lexicalized words of high frequency form a small group and are based on less productive patterns (*samolet* 'aircraft', *samogon* 'moonshine') whereas productive patterns produce a large number of low-frequency items.

The volume is completed by an empirical assessment of the German pluralization system by Eugen Zaretsky and Benjamin P. Lange, in which they argue that *No matter how hard we try: Still no default plural marker in nonce nouns in Modern High German*. In their paper, the authors analyze how various intralinguistic factors, such as grammatical gender, word-final phonemes, plural markers of the rhyming real words, unusual orthography, final-obstruent devoicing, etc., condition the choice of plural allomorphs in nonce words, such as *Pind* → *Pinder*, in a sample of German native speakers. Comparing their findings to an earlier study with the same test items, their main methodological aim is to show that (i) the sample size, (ii) the type of regression, and (iii) particularly the study design, mainly in terms of task types (plausibility rating vs. production) used, may markedly influence the overall results. Based on their quantitative evidence, they identify a number of weaknesses of earlier approaches and eventually suggest that, instead of dual-route models, which have been advocated in previous studies, single-route models best account for the distribution of plural markers.