

Contents

List of Figures	ix
List of Tables	xi
Introduction <i>Steven Coats and Veronika Laippala</i>	1
Part I Methods for Data Collection, Analysis and Visualization	
1 Noisy Data: Using Automatic Speech Recognition Transcripts for Linguistic Research <i>Steven Coats</i>	17
2 Low-code Data Science Tools for Linguistics: Swiss Army Knives or Pretty Black Boxes? <i>Jukka Tyrkkö and Daniel Ihrmark</i>	40
3 The Visualization and Evaluation of Semantic and Conceptual Maps <i>Gerold Schneider</i>	67
Part II Corpus Construction, Registers and Genres	
4 Towards Automatic Register Classification in Unrestricted Databases of Historical English <i>Liina Repo, Brett Hashimoto, Aatu Liimatta, Lassi Saario, Tanja Säily, Iiro Tiihonen, Mikko Tolonen and Veronika Laippala</i>	97
5 The Topical Landscape of Web Registers: Exploring the Interplay of Registers and Topicality in a Web-scale Corpus <i>Valtteri Skantsi, Veronika Laippala and Aki-Juhani Kyröläinen</i>	127
6 Towards ‘Large and Tidy’: Establishing Internal Structure in Megacorpora <i>Axel Bohmann</i>	157
Part III Social Media, Discourse and Meanings	
7 Multi-modal Considerations for Social Media Discourse Analysis: A Specialized Corpus of Twitter Commentary on ‘Working from Home’ <i>Christopher Fitzgerald, Geraldine Mark, Anne O’Keeffe, Dawn Knight, Justin McNamara, Svenja Adolphs, Benjamin Cowan, Tania Fahey Palma, Fiona Farr and Sandrine Peraldi</i>	187

8	Exploring Self-identification and the Functions of the <i>Identify as</i> Construction in the LGBTQ+ Reddit Corpus Laura Hekanaho, Turo Hiltunen, Minna Palander-Collin and Helmiina Hotti	213
---	--	-----

	List of Contributors	243
--	----------------------	-----

	Index	249
--	-------	-----

