Introduction

Ever since corpus linguistics entered the mainstream, it has become increasingly difficult to keep track of its most recent developments due to the sheer volume of corpus-based, corpus-driven or corpus-informed studies. For twelve years, the conferences known as PALC (Practical Applications in Language and Computers) and organized at the University of Łódź in Poland have served the international community of corpus and computational linguists by providing a useful forum for the exchange of views and ideas on how corpora and computational tools can be effectively employed to explore and advance our understanding of language. The conferences and the ensuing volumes have attempted to reflect the widening scope and perspectives on language and computers. The present volume is no different in that it documents new developments and explorations in these areas encompassing an array of topics and themes, ranging from national corpora and corpus tools through cognitive processes, discourse and ideology, academic discourse, translation, and lexicography to language teaching and learning. In keeping with the PALC tradition, it is our policy to publish contributions from both seasoned researchers as well as from colleagues who are recently initiated members of the corpus linguistics community.

The contributions are drawn from papers presented at the 7th *Practical Applications in Language and Computers PALC* conference held at the University of Łódź in 2009. The plenary speakers were Khurshid Ahmad (Trinity College, Dublin), Mark Davies (Brigham Young University), Ken Hyland (then at University of London), Terttu Nevalainen (University of Helsinki) and Margaret Rogers (University of Surrey).

This volume is divided into nine Parts, each Part being further subdivided into chapters.

Part One NATIONAL CORPORA provides overviews of three national corpora. First, Mark Davies (Brigham Young University) in his plenary paper *Semantically-Based, Learner-Oriented Queries with the 400+ Million Word Corpus of Contemporary American English* demonstrates the unique ways in which language learners can use the new Corpus of Contemporary American English to carry out and use semantically-based queries. František Čermák (Charles University, Prague) in *The Case of the Czech National Corpus: Its Design and History* introduces readers to the methodological issues of data acquisition and corpus design involved in creating the Czech National Corpus. The next two contributions are related to the National Corpus of Polish. Rafał L. Górski (Institute of Polish Language, Polish Academy of Sciences) in *The Design of the National Corpus of Polish* describes the proposed design of the National Corpus of Polish arguing that the corpus should reflect the perception of the language by the Polish linguistic community. Finally, Adam

Przepiórkowski (Polish Academy of Sciences) and Piotr Bański (University of Warsaw) in *XML Text Interchange Format in the National Corpus of Polish* describe and provide rationale for employing the XML encoding of texts within the National Corpus of Polish.

Part Two CORPUS TOOLS, INFORMATION AND TERMINOLOGY EXTRACTION comprises eight contributions. The first paper of this part is a contribution from Natalia Kotsyba (Warsaw University), Andrij Mykulyak (A. Sołtan Institute for Nuclear Studies, Warsaw), Igor V. Shevchenko (ULIF NANU, Kyiv) *UGTag: Morphological Analyzer and Tagger for Ukrainian Language* in which the authors describe the UGTag, a programme for morphological analysis and tagging of Ukrainian texts developed within the Polish-Ukrainian Parallel Corpus (PolUKR) project to support morphosyntactic annotation for the Ukrainian part of the corpus. The authors of the next seven papers which follow are Michal Křen (Charles University, Prague) and Martina Waclawičová (Charles University, Prague), who in their contribution *Database Framework for a Distributed Spoken Data Collection Project*, look at the main features of database system that is used in the Czech National Corpus for collecting recordings and transcriptions of authentic spoken Czech used in informal situations, Adam Przepiórkowski (Institute of Computer Science, Polish Academy of Sciences) and Grzegorz Murzynowski: *Manual Annotation of the National Corpus of Polish with Anotatornia* present the procedure of the manual annotation of a 1-million-word subcorpus of the National Corpus of Polish using a purpose-built tool, Anotatornia, Danuta Karwańska (University of Warsaw) and Adam Przepiórkowski in their paper *On the Evaluation of Two Polish Taggers* discuss the results of the comparison of two Polish taggers and the implications they carry for future taggers of Polish, especially the tagger, developed within the National Corpus of Polish, Tomas By (Centro de Linguística da Universidade Nova de Lisboa) in *Additional Comments on the Prolog Version of the Tiger Dependency Bank* updates information and provides more details on some of the methods used to verify that the word order disambiguation produces accurate results, Marcin Miłkowski (Polish Academy of Sciences) in his contribution *Automating Rule Generation for Grammar Checkers* describes several approaches to automatic or semi-automatic development of symbolic rules for grammar checkers from the information contained in the corpora, Piotr Pęzik (University of Łódź) in *Providing Corpus Feedback for Translators with the PELCRA Search Engine for NKJP* introduces the PELCRA search engine for the National Corpus of Polish (PSEN) focusing on the usefulness of the tool in verifying the phraseology of translated texts, and finally Ewelina Kwiatek, Pius ten Hacken (Swansea University) in *Evaluating the Efficiency of MultiTerm Extract for Extraction of English and Polish Terms* investigate the efficiency of MultiTerm Extract, a component of SDL Trados 2007 to extract terms from English and Polish specialized corpora.

Part Three CORPUS-BASED LANGUAGE STUDIES includes the papers by Janusz Badio (University of Łódź), who in *What's in a "Duck"? A Corpus-based Study of Salience and Attention within Animal-Related, Denominal Verb,* attempts to find evidence for the claim that three processes of indexing, deriving affordances and meshing are used simultaneously in language understanding, Łukasz Grabowski (Opole University) in *Application of Parallel Corpora in Typological Investigations: the Case of Using English-Russian Parallel Subcorpus of the National Russian Corpus in Typological Study of Motion Verbs*, uses the English-Russian Parallel Subcorpus of the National Russian Corpus to explore the applications of this type of corpora in typological investigations, Milena Herbal-Jezierska (University of Warsaw) in *Corpus-based Morphology in the Czech Language* looks at various ways of using corpora in morpohological (especially inflexion) research in Czech; finally María Pérez Blanco (University of León) in her contribution *The Language of Evaluation in English and Spanish Editorials: A Corpus-based Study* discusses some of the linguistic features of newspaper editorials in English and Spanish.

Part Four COGNITIVE LINGUISTICS (Barbara Lewandowska-Tomaszczyk, University of Łódź and Paul Wilson, University of Łódź), *Culture Based Conceptions of Emotion in Polish and English* compares emotional dimensional structure of Polish and English native speakers and discusses the cultural bases to the findings.

Part Five groups papers which adopt A CORPUS-ASSISTED PERSPECTIVE ON DISCOURSE ANALYSIS AND IDEOLOGY. Cinzia Bevitori (University of Bologna at Forlì) in her paper *The Meanings of Responsibility in the British and American Press on Climate Change: A Corpus-Assisted Discourse Analysis Perspective* examines the ways in which the lemma 'responsibility' is used in a corpus of British and American press coverage of climate change in 2007. Mikołaj Deckert (University of Łódź): *Towards an Axiological Picture of the EU – Evidence from a Polish TV News Corpus* uses Polish broadcast news data to explore the axiological constructions of the personified European Union and its metonymic discursive forms. Katarzyna Fronczak (University of Łódź) in her contribution *Keywords and the Discourse of the Northern Ireland Peace Process, 1997–2007. A Case Study in the Election Manifestoes of the Democratic Unionist Party and Sinn Féin* adopts a keyword approach to carry out a textual analysis of language used in the election manifestoes focusing on the development and changes in the usage of words significant for the peace process. Joanna Kaim-Kerth (Jagiellonian University) in *Correspondence Analysis in Discourse Studies* employs multidimensional statistical analysis to demonstrate how such methods, particularly correspondence analysis, can be used in researching discourses focused on similar subject, i.e. values and authorities as well as visions of the family as perceived by different parliamentary fractions (understood as different

discourses). Finally, Inga Massalina (Kaliningrad State Technical University) in her study *Selected Cognitive and Discoursal Aspects of the LSP of the Navy* touches upon the cognitive and discourse aspects of the LSP of Navy.

ACADEMIC DISCOURSE is the topic of Part Six, which comprises three contributions. The first paper is a plenary contribution by Ken Hyland (University of Hongkong) *Corpora and EAP: Specificity in Disciplinary Discourses* in which the author draws on his own work conducted over several years into student and research genres to show how some familiar conventions of academic writing are employed by different fields. Silvia Cacchiani (University of Modena and Reggio Emilia) in her paper *Keywords and Key Lexical Bundles as Cues to Knowledge Construction in RAS in Economics* attempts to characterize disciplinary discourses by examining discourse signalling devices, and focussing on lexical bundles and keywords. In the last contribution of this Part *Conceptualising Spatial Relationships in Academic Discourse: A Corpus-Cognitive Account of Locative-Spatial and Abstract-Spatial Prepositions*, Christoph Haase (Chemnitz University of Technology) and Josef Schmied (Chemnitz University of Technology): investigate distributional properties of prepositions as heads of prepositional phrases that negotiate a mapping function between direct/literal and extended/metaphorical meaning when dealing with abstract concepts.

Part Seven deals with the theme of TRANSLATION and it includes eight contributions. It opens with a plenary contribution by Margaret Rogers (University of Surrey) *Translation Memory and Textuality: Some Implications*, who explores the use of computer-based tools in the translation of LSP (language for special purposes) texts and considers whether this has any implications for the nature of textuality. In the next paper, *Corpus Encoding and Integration for English-Spanish MT*, Alejandro Curado Fuentes (University of Extremadura) and Martin Garay Serrano (University of Extremadura) describe the compilation and encoding of the Spanish corpus and its integration on the database with the bilingual dictionary with a view to enhancing the Context-Based Machine Translation of English into Spanish. Dimitra Anastasiou (University of Limerick) in *Localisation, Centre for Next Generation Localisation and Standards* focuses on the description of a project called "Centre for Next Generation for Localisation" which currently runs in Ireland and she describes standards in terms of localization, especially the XLIFF standard. Hanne Eckhoff (University of Oslo), Dag Haug (University of Oslo) and Marek Majer (University of Oslo) in their contribution *Making the Most of the Data: Old Church Slavic and the PROIEL Corpus of Old Indo-European Bible Translations* report on ongoing work on the PROIEL corpus of old Indo-European New Testament texts, consisting of the New Testament in its Greek original and its earliest translations. Cécile Frérot (Université Stendhal Grenoble 3) in *Parallel Corpora for Translation Teaching and Translator Training*

Purposes explores the use of parallel corpora in designing a corpus-based translation course that is relevant from a translational standpoint and that is best suited to future professional translators, especially in terms of corpus-based translation tools. In the next paper, *The Combination of Comparable and Translation Corpora and How Translators May Benefit from it*, Marlén Izquierdo (University of Cantabria) continues the theme of parallel corpora by dealing with the combination of comparable and parallel corpora in the descriptive, functional analysis of languages in contrast and its positive impact for extending applications to translation. Marta Kajzer (Adam Mickiewicz University, Poznań) in *Translation of Eurojargon as a Source of Neologisms in Polish. A Corpus-based Study* presents a corpus-based analysis of Eurojargon translation as a potential source of neologisms in Polish. The last contribution in this section by Maria Tymczyńska (Adam Mickiewicz University, Poznań) *Community Interpreting in a Blended Environment – Student and Teacher Assessment* presents and discusses the application of Moodle, a free open-source online course management system, in the creation and implementation of practical interpreting courses in the Post-Graduate Programme in Community Interpreting offered at the Adam Mickiewicz University in Poznań, Poland.

Part Eight explores the theme of LEXICOGRAPHY. In the first contribution, Piotr Bański (University of Warsaw) and Beata Wójtowicz (University of Warsaw) in *New XML-encoded Swahili-Polish Dictionary: Micro- and Macrostructure* describe the structure of a new Swahili-Polish dictionary and some of the insights resulting from testing its electronic format. This is followed by Piotr Burmann's critical appraisal of a range of technical disctionaries in terms of their usefulness in the translator's work in *Insights into Selected Scientific and Technical Dictionaries Currently Available on the Polish Publishing Market.* Mirosława Podhajecka (University of Opole) in her contribution *Research in Historical Lexicography: Can Google Books Collection Complement Traditional Corpora?* demonstrates that the Google Books collection, a non-specialized textual resource can be applied successfully for research in historical lexicology and lexicography. The last paper in this section authored by Igor V. Shevczenko (ULIF NANU, Kyiv), Natalia Kotsyba (University of Warsaw), Kiryl Kurshuk (Hrodna University) *Towards the Creation of a Belarusian Grammatical Dictionary*, describes the process of creating the Belarusian grammatical word-inflexion dictionary, the first tool of the kind for this language, on the basis of linguistic similarities with the existing Ukrainian grammatical dictionary.

The last section in the volume, Part Nine LANGUAGE TEACHING AND LEARNING contains four contributions. First, Alex Boulton (CRAPEL–ATILF/CNRS, Nancy-Université) in his paper *Data-Driven Learning: the Perpetual Enigma* traces the evolution of DDL through the work of Tim Johns from 1984 up until his death in 2009, as well as in DDL studies by other

researchers. Then, Carmen Dayrell (University of São Paulo (USP) in *Anticipatory 'IT' in English Abstracts: A Corpus-based Study of Non-native Student and Published Writing,* explores the use of anticipatory *it* patterns (such as *it is found that* and *it is necessary to*) in English abstracts written by Brazilian graduate students from the disciplines of physics, pharmaceutical sciences and computing as opposed to abstracts of published papers from the same disciplines. Stefano Federici (University of Cagliari) in his contribution *Automatic Question Generation for E-learning* describes the "E-generation in E-learning" project that aims to evaluate and improve automatic question generation methodologies by means of analogy-based techniques to implement an automatic question generation system that best suites the needs of today e-learning platforms such as Moodle. Finally, Przemysław Krakowian (University of Łódź) in his paper *WEBCEF – The Project, Deliverables and Status Quo* looks at some of the issues and research questions which arose during the completion of WebCEF, a Socrates Minerva project in which web-based environments can be used for assisting the teacher in the process of evaluating spoken performance of language learners.

Acknowledgements

Stanisław Goźdź-Roszkowski