Corpus Linguistics Applications For The Study Of English Linguistic Insights

This textbook introduces the fundamental concepts and methods of corpus linguistics for students approaching this topic for the first time, putting specific emphasis on the enormous linguistic diversity represented by approximately 7,000 human languages and broadening the scope of current concerns in general corpus linguistics. Including a basic toolkit to help the reader investigate language in different usage contexts, this book: Shows the relevance of corpora to a range of linguistic areas from phonology to sociolinguistics and discourse Covers recent developments in the application of corpus linguistics to the study of understudied languages and linguistic typology Features exercises, short problems, and questions Includes examples from real studies in over 15 languages plus multilingual corpora Providing the necessary corpus linguistics skills to critically evaluate and replicate studies, this book is essential reading for anyone studying corpus linguistics.

This volume showcases original, agenda-setting studies in the field of learner corpus research of both spoken and written production. The studies have important applications for classroom pedagogy. The volume brings readers up-to-date with new written and spoken learner corpora, often looking at previously under-examined variables in learner corpus investigations. It also demonstrates innovative applications of learner corpus findings, addressing issues such as the effect of task, the effect of learner variables and the nature of learner language. The volume is of significant interest to researchers working in corpus linguistics, learner corpus research, second language acquisition and English for Academic and Specific Purposes, as well to practitioners interested in the application of the findings in language teaching and assessment.

This volume assesses the state of the art of parallel corpus research as a whole, reporting on advances in both recent developments of parallel corpora – with some particular references to comparable corpora as well– and in ways of exploiting them for a variety of purposes. The first part of the book is devoted to new roles that parallel corpora can and should assume in translation studies and in contrastive linguistics, to the usefulness and usability of parallel corpora, and to advances in parallel corpus alignment, annotation and retrieval. There follows an up-to-date presentation of a number of parallel corpus projects currently being carried out in Europe, some of them multimodal, with certain chapters illustrating case studies developed on the basis of the corpora at hand. In most of these chapters, attention is paid to specific technical issues of corpus building. The third part of the book reflects on specific applications and on the creation of bilingual resources from parallel corpora. This volume will be welcomed by scholars, postgraduate and PhD students in the fields of contrastive linguistics, translation studies, lexicography, language teaching and learning, machine translation, and natural language processing.

This book will be of particular interest to anyone interested in the application of corpus linguistic techniques to language study and instruction. This volume includes selected papers from the Fourth North American Symposium, held in Indianapolis and hosted by the Indiana Center for Intercultural Communication at Indiana University Purdue University in Indianapolis (IUPUI) in November, 2002. These papers – from authors representing eight countries including the U.S., Belgium, China, France, Germany, Ireland, the Netherlands, and Spain – provide a wide range of views of and approaches to corpus linguistic. Topics range from theory and analysis to classroom application, and include the study of oral discourse as well as the study of written discourse, including internet-based discourse. Consequently, this volume is divided into two sections. The first section focuses on the use of corpus linguistics in the analysis of spoken and written discourse; the second section focuses on the direct pedagogical application of corpus linguistics, reflecting the applied foundation of this branch of linguistics.

Seminar paper from the year 2010 in the subject English Language and Literature Studies - Linguistics, grade: 2, University of Innsbruck (Anglistik), course: Synchronic and/or Diachronic English Linguistics: English Syntax and Morphology, language: English, abstract: This paper is dealing with the topic of corpus linguistics. It tries to give an overview about this topic, as I think that, although it already gained popularity, not everyone is familiar with it. First, I want to explain what a corpus actually is, and what it is useful for. The different types of corpora will be described and also potential risks of depending to much on computer-processable corpora. Then the focus shifts to the fields of application of corpus linguistics, and also the use in syntax and morphology is discussed. I will also try to illustrate the opportunities a corpus provides by using an example for better understanding. The main aim of this paper is to give an overview about corpus linguistics and the fields of application, with attention to syntax and morphology.

The Routledge Handbook of Corpus Linguistics provides a timely overview of a dynamic and rapidly growing area with a widely applied methodology. Through the electronic analysis of large bodies of text, corpus linguistics demonstrates and supports linguistic statements and assumptions. In recent years it has seen an ever-widening application in a variety of fields: computational linguistics, discourse analysis, forensic linguistics, pragmatics and translation studies. Bringing together experts in the key areas of development and change, the handbook is structured around six themes which take the reader through building and designing a corpus to using a corpus to study literature and translation. A comprehensive introduction covers the historical development of the field and its growing influence and application in other areas. Structured around five headings for ease of reference, each contribution includes further reading sections with three to five key texts highlighted and annotated to facilitate further exploration of the topics. The Routledge Handbook of Corpus Linguistics is the ideal resource for advanced undergraduates and postgraduates.

Corpus Linguistics for Grammar provides an accessible and practical introduction to the use of corpus linguistics to analyse grammar, demonstrating the wider application of corpus data and providing readers with all the skills and information they need to carry out their own corpus-based research. This book: explores the kinds of corpora available and the tools which can be used to analyse them; looks at specific ways in which features of grammar can be explored using a corpus through analysis of areas such as frequency and collocation; contains exercises, worked examples and suggestions for further practice with each chapter; provides three illustrative examples of potential research projects in the areas of English Literature, TESOL and English Language. Corpus Linguistics for Grammar is essential reading for students undertaking corpus-based research into grammar, or studying within the areas of English Language, Literature, Applied Linguistics and TESOL.

This volume provides an overview of four currently booming areas in the discipline of corpus linguistics. The first section is concerned with studies of the history and development of morphological and syntactic phenomena in English, Spanish, and Mandarin Chinese. The second section contains case studies investigating the functions and contexts of use of different morphological and syntactic forms in English, Spanish, Russian, and Mandarin Chinese. The third section contains studies in the field of genre and register from settings as diverse as health, call center, academic, and legal discourse. The final section features papers refining existing, and exploring new, corpus-linguistic methods: dispersions, text mining, corpus similarity, as well as the development of extraction patterns and the evaluation of tagging methods.

Corpus Linguistics Applications for the Study of EnglishPeter Lang

This book pinpoints the impact of new technologies on language and communication, highlights the evolution and changes undergone by humanities in conjunction with technological innovation, and looks at how language has adapted to the challenges of today’s digitized world.
Corpus linguistics is one of the most exciting approaches to studies in applied linguistics today. From its quantitative beginnings it has grown to become an essential aspect of research methodology in a range of fields, often combining with text analysis, CDA, pragmatics and organizational studies to reveal important new insights about how language works. This volume captures some of the most stimulating and significant developments in the field, including chapters on language teaching, institutional and professional discourse, English as an International Language, translation, forensics and media studies. As a result it goes beyond traditional, limited presentations of corpus work and shows what is involved in designing and building a corpus; reviews cutting-edge studies of online communication using corpus linguistics, foregrounding different analytical components to facilitate studies in professional discourse, online learning, public understanding of health issues and dating apps; showcases both freely-available corpora and the innovative tools that students and researchers can access to carry out their own research. Corpus Linguistics for Online Communication supports researchers and students in generating high quality, applied research and is essential reading for those studying and researching in this area.

Corpus linguistics is a key area of applied linguistics and one of the most rapidly developing. Winne Cheng’s practical approach guides readers in acquiring the relevant knowledge and theories to enable the analysis, explanation and interpretation of language using corpus methods. Throughout the book practical classroom examples, concordance based analyses and tasks such as designing and conducting mini-projects are used to connect and explain the conceptual and practical aspects of corpus linguistics. Exploring Corpus Linguistics is an essential textbook for post-graduate/graduate students new to the field and for advanced undergraduates studying English Language and Applied Linguistics. The Cambridge Handbook of English Corpus Linguistics (CHECL) surveys the breadth of corpus-based linguistic research on English, including chapters on collocations, phraseology, grammatical variation, historical change, and the description of registers and dialects. The most innovative aspects of the CHECL are its emphasis on critical discussion, its explicit evaluation of the state of the art in each sub-discipline, and the inclusion of empirical case studies. While each chapter includes a broad survey of previous research, the primary focus is on a detailed description of the most important corpus-based studies in this area, with discussion of what those studies found, and why they are important. Each chapter also includes a critical discussion of the corpus-based methods employed for research in this area, as well as an explicit summary of new findings and discoveries.

Stefan Th. Gries provides an overview on how quantitative corpus methods can provide insights to cognitive/usage-based linguistics and selected psycholinguistic questions as well as introductory examples of how to use R for such research. Paradoxically, doing corpus linguistics is both easier and harder than it has ever been before. On the one hand, it is easier because we have access to more existing corpora, more corpus analysis software tools, and more statistical methods than ever before. On the other hand, reliance on these existing corpora and corpus linguistic methods can potentially create layers of distance between the researcher and the language in a corpus, making it a challenge to do linguistics with a corpus. The goal of this Element is to explore ways for us to improve how we approach linguistic research questions with quantitative corpus data. We introduce and illustrate the major steps in the research process, including how to: select and evaluate corpora, establish linguistically-motivated research questions, observational units and variables, select linguistically interpretable variables, understand and evaluate existing corpus software tools, adopt minimally sufficient statistical methods, and qualitatively interpret quantitative findings.

With an ever-growing body of corpus linguistic tools, resources and applications, it becomes increasingly important to reflect critically on the underlying assumptions that corpus linguistics is based on. Focusing on meaning and methods, this book tackles fundamental concepts and approaches that define the discourse of the field. Internationally renowned contributors address topics that range from the history of corpus linguistics to contrastive perspectives between languages, to interpreting patterns in corpora as evidence of both mainstream discourses and individual voices within them. This collection not only adds to our understanding of the fundamentals of corpus linguistics, it also brings innovative meanings to the corpus linguistics discourse. It has been edited in honour of Wolfgang Teubert, who for decades has been a significant voice in this discourse. Corpus Linguistics for Online Communication provides an instructive and practical guide to conducting research using methods in corpus linguistics in studies of various forms of online communication. Offering practical exercises and drawing on original data taken from online interactions, this book: introduces the basics of corpus linguistics, including what is involved in designing and building a corpus; reviews cutting-edge studies of online communication using corpus linguistics, foregrounding different analytical components to facilitate studies in professional discourse, online learning, public understanding of health issues and dating apps; showcases both freely-available corpora and the innovative tools that students and researchers can access to carry out their own research. Corpus Linguistics for Online Communication supports researchers and students in generating high quality, applied research and is essential reading for those studying and researching in this area.

The aim of this volume is to present a state-of-the-art view on corpus studies. This collection of papers, presented at the XII Susanne Hübner Seminar in November 2003 at the University of Zaragoza, comprises both quantitative and qualitative analyses and studies on both written and oral corpora. Structured in seven sections, the book covers a wide range of approaches and methodologies and reflects current linguistic research. The papers have been written by scholars from a large number of universities, mainly from Europe, but also from the USA and Asia. The volume offers contributions on diachronic studies, pragmatic analyses and cognitive linguistics, as well as on translation and
English for Specific Purposes. The book includes several papers on corpus design and reports on research on oral corpora. At a more specific level, the papers analyse aspects such as politeness issues, dialectology, comparable corpora, discourse markers, the expression of evidentiality and writer stance, metaphor and metonymy, conditional sentences, evaluative adjectives, delexicalised verbs and nominalization.

This book is about investigating the way people use language in speech and writing. It introduces the corpus-based approach to linguistics, based on analysis of large databases of real language examples stored on computer. Each chapter focuses on a different area of linguistics, including lexicography, grammar, discourse, register variation, language acquisition, and historical linguistics. Example analyses are presented in each chapter to provide concrete descriptions of the research methods and advantages of corpus-based techniques. Ten methodology boxes provide clear and concise explanations of the issues in doing corpus-based research and reading corpus-based studies and there is a useful appendix of resources for corpus-based investigation. This lucid and comprehensive introduction to the subject will be welcomed by a broad range of readers, from undergraduate students to professional researchers.

Corpus Linguistics for English Teachers: New Tools, Online Resources, and Classroom Activities describes Corpus Linguistics (CL) and its many relevant, creative, and engaging applications to language teaching and learning for teachers and practitioners in TESOL and ESL/EFL, and graduate students in applied linguistics. English language teachers, both novice and experienced, can benefit from the list of new tools, sample lessons, and resources as well as the introduction of topics and themes that connect CL constructs to established theories in language teaching and second language acquisition. Key topics discussed include: • CL and the teaching of English vocabulary, grammar, and spoken-written academic discourse; • new tools, online resources, and classroom activities; and • focus on the "English teacher as a corpus-based researcher." With ready-to-use teaching vignettes, tips and step-by-step guides, case studies with practitioner interviews, and discussion of corpora and corpus tools, Corpus Linguistics for English Teachers is a thoughtfully designed and skillfully executed resource, bridging theory with practice for anyone looking to understand and apply corpus-based tools dynamically in the language learning classroom.

The use of large, computerized bodies of text for linguistic analysis and description has emerged in recent years as one of the most significant and rapidly-developing fields of activity in the study of language. This book provides a comprehensive introduction and guide to Corpus Linguistics. All aspects of the field are explored, from the various types of electronic corpora that are available to instructions on how to design and compile a corpus. Graeme Kennedy surveys the development of corpora for use in linguistic research, looking back to the pre-electronic age as well as to the massive growth of computer corpora in the electronic age.

The aim of this volume is to present a state-of-the-art view on corpus studies. This collection of papers, presented at the XII Susanne Hubner Seminar in November 2003 at the University of Zaragoza, comprises both quantitative and qualitative analyses and studies on both written and oral corpora. Structured in seven sections, the book covers a wide range of approaches and methodologies and reflects current linguistic research. The papers have been written by scholars from a large number of universities, mainly from Europe, but also from the USA and Asia. The volume offers contributions on diachronic studies, pragmatic analyses and cognitive linguistics, as well as on translation and English for Specific Purposes. The book includes several papers on corpus design and reports on research on oral corpora. At a more specific level, the papers analyse aspects such as politeness issues, dialectology, comparable corpora, discourse markers, the expression of evidentiality and writer stance, metaphor and metonymy, conditional sentences, evaluative adjectives, delexicalised verbs and nominalization.

Corpus Linguistics in Literary Analysis provides a theoretical introduction to corpus stylistics and also demonstrates its application by presenting corpus stylistic analyses of literary texts and corpora. The first part of the book addresses theoretical issues such as the relationship between subjectivity and objectivity in corpus linguistic analyses, criteria for the evaluation of results from corpus linguistic analyses and also discusses units of meaning in language. The second part of the book takes this theory and applies it to Northanger Abbey by Jane Austen and to two corpora consisting of 1) Austen's six novels and 2) texts that are contemporary with Austen. The analyses demonstrate the impact of various features of text on literary meanings and how corpus tools can extract new critical angles. This book will be a key read for upper level undergraduates and postgraduates working in corpus linguistics and in stylistics on linguistics and language studies courses.

The second edition of this successful text provides an ideal introduction for university students of English at the intermediate level. Students planning papers, dissertations or theses will find the book a particularly valuable guide. After introducing corpora and the rationale and basic methodology of corpus linguistics, the authors present a number of recent case studies providing new insights into vocabulary, collocations, phraseology, metaphor and metonymy, syntactic structures, male and female language, and language change. A final chapter shows how the web and social media can be used as a source for linguistic investigations and contains information on how to compile your own corpus. Each chapter includes study questions, exercises and updated suggestions for further reading.

This volume will be of particular interest to readers interested in expanding the applications of corpus linguistics techniques through new tools and approaches. The text includes selected papers from the Fifth North American Symposium, hosted by the Linguistics Department at Montclair State University in Montclair New Jersey in May 2004. The symposium papers represented several areas of corpus studies including language development, syntactic analysis, pragmatics and discourse, language change, register variation, corpus creation and annotation, and practical applications of corpus work, primarily in language teaching, but also in medical training and machine translation. A common thread through most of the papers was
the use of corpora to study domains longer than the word. Not surprisingly, fully half of the papers deal with the computational tools and linguistic strategies needed to search for and analyze these longer spans of language while most of the remaining papers examine particular syntactic and rhetorical properties of one or more corpora.

The use of corpora has conventionally been envisioned as being either corpus-based or corpus-driven. While the formal definition of the latter term has been widely accepted since it was established by Tognini-Bonelli (2001), it is often applied to studies that do not, in fact, fulfill the fundamental requirement of a theory-neutral starting point. This volume proposes the term pattern-driven as a more precise alternative. The chapters illustrate a variety of methods that fall under this broad methodology, such as the extraction of lexical bundles, POS-grams and semantic frames, and demonstrate how these approaches can uncover new understandings of both synchronic and diachronic linguistic phenomena.

The book offers a combined discussion of the main theoretical, methodological and application issues related to corpus work. Thus, starting from the definition of what is a corpus and why reading a corpus calls for a different methodology from reading a text, the underlying assumptions behind corpus work are discussed. The two main approaches to corpus work are discussed as the “corpus-based” and the “corpus-driven” approach and the theoretical positions underlying them explored in detail. The book adopts and exemplifies the parameters of the corpus-driven approach and posits a new unit of linguistic description defined systematically in the light of corpus evidence. The applications where the corpus-driven approach is exemplified are language teaching and contrastive linguistics. Alternating between practical examples and theoretical evaluation, the reader is led step-by-step to a detailed understanding of the issues involved in corpus work and, at the same time, tempted to explore for himself some of the major applications where a corpus-driven methodology can reveal unprecedented insights into linguistic patterning.

Corpora are used widely in linguistics, but not always wisely. This book attempts to frame corpus linguistics systematically as a variant of the observational method. The first part introduces the reader to the general methodological discussions surrounding corpus data as well as the practice of doing corpus linguistics, including issues such as the scientific research cycle, research design, extraction of corpus data and statistical evaluation. The second part consists of a number of case studies from the main areas of corpus linguistics (lexical associations, morphology, grammar, text and metaphor), surveying the range of issues studied in corpus linguistics while at the same time showing how they fit into the methodology outlined in the first part. Corpus Linguistics for English for Academic Purposes (EAP) shows how corpus analyses can enhance students’, practitioners’ and researchers’ knowledge of academic language. The book provides a reader-friendly discussion of the key concepts, practices and research applications of Corpus Linguistics which are relevant to the EAP community. The volume: empowers readers to compile and analyse EAP-relevant corpora to support their practice; draws on open-access resources, allowing readers in all contexts to engage in corpus analyses; examines how corpus studies have advanced the description of spoken, written and computer-mediated academic discourses; contains numerous reflective and hands-on tasks. Corpus Linguistics for English for Academic Purposes is an essential book for EAP students, practitioners and researchers who wish to develop corpus analytical skills to support their learning, teaching and research practice. It is equally important to novice corpus linguists who wish to find out how they can contribute to the ever-expanding area of EAP.


Corpus Linguistics for Education provides a practical and comprehensive introduction to the use of research-methods in the field of education. Taking a hands-on approach to showcase the applications of corpora in the exploration of educationally relevant topics, this book: • covers 18 key skills including corpus building, the role of frequency, different corpus methods, transcription and annotation; • demonstrates the use of available corpora and desktop and online corpus analysis tools to conduct original analyses; • features case studies and step-by-step guides within each chapter; • emphasises the use of interview data in research projects. Corpus Linguistics for Education is an essential guide for students and researchers studying or conducting their own corpus-based research in education.

The explosion of information technology has led to substantial growth of web-accessible linguistic data in terms of quantity, diversity and complexity. These resources become even more useful when interlinked with each other to generate network effects. The general trend of providing data online is thus accompanied by newly developing methodologies to interconnect linguistic data and metadata. This includes linguistic data collections, general-purpose knowledge bases (e.g., the DBpedia, a machine-readable edition of the Wikipedia), and repositories with specific information about languages, linguistic categories and phenomena. The Linked Data paradigm provides a framework for interoperability and access management, and thereby allows to integrate information from such a diverse set of resources. The contributions assembled in this volume illustrate the band-width of applications of the Linked Data paradigm for representative types of language resources. They cover lexical-semantic resources, annotated corpora, typological databases as well as terminology and metadata repositories. The book includes representative applications from diverse fields, ranging from academic linguistics (e.g., typology and corpus linguistics) over applied linguistics (e.g., lexicography and translation studies) to technical applications (in computational linguistics, Natural Language Processing and information technology). This volume accompanies the Workshop on Linked Data in Linguistics 2012 (LDL-2012) in Frankfurt/M., Germany, organized by the Open Linguistics Working Group (OWLG) of the Open Knowledge Foundation (OKFN). It assembles contributions of the workshop participants and, beyond this, it summarizes initial steps in the formation of a Linked Open Data cloud of linguistic resources, the Linguistic Linked Open Data cloud (LLOD).

These specially-commissioned studies cover corpus-informed approaches to researching, teaching and learning English for Specific Purposes (ESP). The corpora used range from very large published corpora to small tailor-made collections of written and spoken text, as well as parallel and contrastive corpora, in both the hard and softer sciences. Designed to tackle the problems faced by a variety of first- and second-language ESP users (specialised translators, undergraduates, junior and experienced researchers, and language trainers), the breadth of approaches enables treatment of issues central to ESP and corpus research, from corpus compilation and analysis to new applications and data-driven learning. The first full-length book on applied corpus use in France, Corpus-Informed Research and Learning in ESP will be of interest not only to those working in the French context, but to a wide variety of language professionals – teachers, researchers or course designers – in many countries looking at ESP from...
different linguistic, cultural and educational perspectives.

Corpus linguistics has become one of the most widely used methodologies across the different linguistic subdisciplines; especially the study of world-wide varieties of English uses corpus-based investigations as one of the chief methodologies. This volume comprises descriptions of the many new corpus initiatives both within and outside Africa that aim to compile various corpora of African Englishes. Moreover, it contains cutting-edge corpus-based research on African Englishes and the use of corpora in pedagogic contexts within African institutions. This volume thus serves both as a practical introduction to corpus compilation (Part I of the book), corpus-based research (Part II) and the application of corpora in language teaching (Part III), and is intended both for those researchers not yet familiar with corpus linguistics and as a reference work for all international researchers investigating the linguistic properties of African Englishes.

Corpus Linguistics for ELT provides a practical guide to undertaking ELT-related corpus research. Aimed at researchers, advanced undergraduate and postgraduate students of ELT and TESOL, and English language teachers, this volume: covers corpus research in the main areas of language study relevant to ELT: grammar, lexis, ESP, spoken grammar and discourse; presents a review of relevant corpus research in these areas, and discusses the implications of this research for ELT; suggests potential ELT-focused corpus research projects, and equips the reader with all the required tools and techniques to carry them out; deals with the growing area of learner corpora and direct classroom application of corpus material. Corpus Linguistics for ELT empowers and inspires readers to carry out their own ELT corpus research, and will allow them in turn to make a significant contribution to corpus-informed ELT pedagogy.

Perspectives on Corpus Linguistics is a collection of interviews with fourteen well-known researchers in the field of linguistics. Each interview consists of a set of ten questions: the first seven are common to all contributors while the last three are connected to the research experience of each guest. In the general questions, the invited scholars explore (sometimes controversial) topics such as the concept of representativeness, the role of intuition and the status of Corpus Linguistics. In the specific questions, they provide a thorough discussion of materials and methods in corpus research as well as theoretical and applied perspectives on the use of corpora in language studies. Whether experts or novices, the volume should be of interest to all those who want to learn about corpus linguistics and carry out research in this fascinating and growing area.

Corpus Linguistics: An Introduction will appeal to a wide spectrum of scholars, researchers, and particularly to students of linguistics. It offers guidelines for the creation and usage of corpora in the form of empirical language databases with direct functional and theoretical interpretation of a natural language. Drawn from original research and written in an accessible language and style, this book will create avenues for further advancements in mainstream and applied linguistics and language technology.