BOOK REVIEW
Open Access

Anne O'Keeffe, Michael McCarthy the Routledge handbook of corpus linguistics (London: Routledge, 2010), xxviii.682 pp., ISBN 13:978–0–415-46,489-5 (hbk), $ 66

Amare Tesfie Birhan

Correspondence: amaretesfie@
gmail.com
Department of English Language
and Literature, Bahir Dar Institute of
Technology, Bahir Dar University,
Bahir Dar, Ethiopia

Introduction

The idea of corpus linguistics is related with the nature of its occurrence and application. Historically, it existed since 1950's during the period of Chomskians (grammarians); it is neither a theory of linguistics nor a branch of linguistics rather it is a methodology. It is defined as an authentic spoken and written data or a mix of both that assembled according to explicit design criteria and exploited through user friendly software (Babanoglu, 2013; Nesselhauf, 2005; O'Keeffe, McCarthy, & Carter, 2007). Similarly, Granger (2002, p, 4) defines corpus as "a methodology which is found on the use of electronic collections of naturally occurring texts." It is also one of the outcomes of the advancement of technology in language study, and language teaching. Through computer program analysis of large bodies of text, it supports language teaching and language analysis. It is large language database which is designed for specific linguistic or socio-pragmatic purposes (Flowerdew, 2012).

Currently, corpora are designed with a distinct set of criteria to use for classroom pedagogy and language analysis (O'keeffe, et al., 2007). It has been widely recognized emerging language instruction tool to address language learners' communicative language needs and to enhance their engagement (Yoon & Hirvela, 2004). This has been contributing a lot to teaching language skills, research, language analysis, classroom pedagogy, language testing, and material design.

Review

The Routledge Handbook of Corpus Linguistics, was edited by Anne O'Keeffe and Michael McCarthy and published in 2010 by Routledge. It encompasses 45 individual chapters and organized into eight major sections. The hand book was contributed by 52 researchers. It raises one of current and rapidly growing issue in language teaching and language analysis and approached with clear theoretical and applied methodology. This hand book addresses the historical and theoretical over view of corpus linguistics, the key considerations in building and designing a corpus, basics to analysis the corpus, using a corpus for language research, using corpus for language pedagogy and methodology, designing corpus-based materials for the language classroom, using corpora to study literature and translation, and applying corpus to other areas of research.

The first section elucidates the historical perspectives of corpora and the theoretical overview of the evaluation of corpus linguistics. McCarthy and O'Keeffe and Bonelli were contributors of the section. Particularly, these researchers explicate the historical origins of corpus, reasons that drove the creation of modern corpora, the influence of technology in creation and corpus development, quantitative and a qualitative revolution of corpora, theoretical shift from text-linguistics to corpus linguistics, and corpora typology. The intent of the second section is to introduce the key features that need to be considered to build and to design corpora. This chapter starts with the work of Reppen, the key considerations to build corpus. According to the author corpus size, techniques of collecting texts, markup and annotation are the important considerations to build and to design corpus. In relation to size of corpus, Reppen asserts that representativeness and practicality are very important considerations.

Likewise, Adoplhs and Knight discuss basic considerations to build spoken corpus. Accordingly, they stated (p. 38) that spoken corpora are smaller in size and unable to offer the same level of recurrence of individual items with that of written corpus. They discussed the importance of designing spoken data. The authors mentioned (cited in Sinclair's study 2005) 10 guidelines which are very important to design corpus linguistics in general. They also cover important considerations to design spoken corpora. Nelson introduces the basics ideas to build written corpus. Hence planning, sampling, balancing, representative, gathering, computerizing, and organizing written texts are important considerations. Koester explains building specialized corpora. According to the author, small corpora allows closer link between the corpus and the context in which the texts in the corpus were produced; it gives insights in to patterns of language use in particular setting; it has a high degree of familiarity with the context to set up for specific research or pedagogical purpose (P.67). Furthermore, the author discusses the extent of its size and considerations to build small corpora.

In addition, Clancy elaborates building a corpus to represent a variety of a language. Clancy defines variety as it is variant of a language that differs from another variant of the same language systematically and coherently (cited in McEnery, et al., 2006, p.90 idea). Accordingly, corpus designer has to consider size, diversity of texts, text length and number, representativeness, and balance of corpus for its varieties. The last chapter of this section is about building a specialized audio-visual corpus by Thompson. According to the author, building audio-visual is a challenging task in the area of corpus linguists. In addition, collecting raw data, preparing transcriptions, and annotating are identified as the crucial issues to design specialized corpora.

The third section set out the basics of corpus analysis. It starts with Lee's work available corpora. Hence, the author first discusses text collections, archives, and corpus distribution sites. Second, he raises issues such as accessing and categorizing corpora, major types of English corpora, developmental, learner and lingua franca corpora, non-English corpora, and multilingual corpora. Likewise, Evison elucidates the basics of analyzing corpus. Thus, the important issues to analyze corpus, exploring word frequency lists, exploring concordances lines, and discourse are stressfully discussed. Similarly, Scott explains the importance of corpus software such as to sorting data and to showing concordances, word lists, and key word lists.

The other issue that is stated in this section is how a corpus is used to explore patterns by Hunston. Hunston defines the idea of patterns and the difficulty of patterns to

spot and how to read concordances are mentioned. Tribble also explains the idea of concordances. In line with this, concordances before the computer age, approaches, tools, and resources for computer generated concordance, and working with corpus data are discussed. The other important issue that is discussed by Lu is corpus software and language development. According to Lu (p.184) "language development is the process in which the language faculty develops in human being. Thus, Lu addresses measuring language development, using corpus to find more about first and second language development.

The intent of section four is to introduce the use of corpus for a language research. Moon explains the way to learn and to search lexis, phraseologies, and different kinds of meanings (context meaning, polysemy, metaphor, connotation, and ideology), sets and synonyms (lexical sets, synonyms, annotations, and opposite) through corpus. In addition, lexis in spoken language (phraseology, meaning, and usage) is also highlighted.

The idea of multi-word units and the reasons of studying multi-word units meaning and n-gram to phraseology variations are also discussed by Greaves and Warren. They explain the idea of multi-word unit and the reasons of studying multi-word units. It is also mentioned that corpus research tells us about phraseology that we do not before. Conrad also discusses the contribution of corpus to grammar. Hence understanding grammar through patterns and contexts, types of grammatical patterns, investigating multiple features, the grammar of speech and new challenges for judging acceptability are mentioned. Similarly, Biber explains corpus to registers and genres. According to Biber, corpus has register perspectives and genre perspectives. In addition, it is also stated corpus studies and multi-dimensional studies show linguistic variations.

Handford explains corpus to specialist genre. According to the author, general corpus linguistics is criticized because of its decontextualized, its approach, and its size. Besides, the rational of genre approach, methodological advantage of specialized corpora in analyzing genre, corpus to academic genres, professional genres, non- institutional genres are briefly covered by the author. Likewise, Thornbury explicates the limitations of using a corpus in the study of discourse, the limitations of corpus in the study discourse and the kind of data we need to study discourse. Ruhleman addresses the issue of pragmatics and related issues such as semantic prosody and pragmatic phenomena. The last issue of this section is corpus to creativity by Anh Vo and Carter. Hence, it is explained that corpus linguistics facilitates creative writing and creative studies.

The main intent of section five is to explain using corpus for language pedagogy and methodology. This section starts with the work of Cheng, what can a corpus can tell us about language teaching. Hence the author demonstrates corpora and language teaching, corpus-driven form and function, corpus evidence as teaching materials, tasks for language learning, bridging corpus linguistics and language teaching. Walsh explains features of spoken and written corpora in creating language teaching materials and syllabus. Thus, integrating corpus-based approaches in syllabus, corpus based materials to teach speaking and listening skills, using corpus-based materials to teach reading and writing, and exploiting learner corpora are explained. Chambers devotes on data-driven learning. The author provides detail examples and activities that can be applied in language teaching classrooms through data driven learning approach. Similarly, Gilquin and Granger elucidate the pedagogical function of data driven learning (DDL).

Hence, assessing the effectiveness of DDL and the problems and limitation of DDL are mentioned. Furthermore, Sripicharn discusses preparing learners for using language corpora. Therefore, finding out what students know and providing general information, identifying task, objectives and types of corpora, preparing corpus data, corpus analysis tools and interpreting corpus results are explained.

The sixth section, designing corpus-based material for the language classroom, starts by the work of Jones and Durrant, what can a corpus tell us about vocabulary teaching material. The authors explain that corpus offers list of key vocabulary that are used to course design and material writing. Thus, vocabulary importance is recognized as reliable data for language learners. In the same way, corpus also plays a key role to develop material to teach grammar. Hence, Hughes discusses corpus to grammar teaching materials, and the author considers benefits of using corpus to teach grammar, and future development perspectives.

The other issue of this section is about corpus informed course book design by McCarten. It is stated that corpus is crucial to design a course book and syllabus by considering various issues. Walter also investigates using corpora to write dictionaries. Lexicographers are highly interested to use corpus linguistics to prepare dictionary. Flowerdew also explains corpus-based and corpus-driven approaches to design writing instruction. Coxhead deals with using corpus for academic purpose. He mentioned that corpus influences English for academic purpose (EAP) pedagogy, EAP language learners, and EAP materials. Vaughan also explores teacher's usage of corpus linguistics for their own research. It is explained that teachers can build and use corpus linguistics for research purpose and language teaching profession.

The seventh section aims at describing using corpora to study literature and translation. It is started by Kenning's work parallel and comparable corpora and how we use them? The main aim of this section is to explore how to design, use and analyze parallel and comparable corpora. Similarly, Kubler and Aston discuss the application of corpora to translation. Corpus also has implication to poetry and drama. Hence, Mclntyre and Walker state the application corpus to language of poetry and drama. The last issue of this section, how can corpora be used to explore literary speech representation is introduced by Amador-Moreno. The author explores similarities and difference of real and fictional speech and the role of corpora to compare real represented speech.

The eighth section introduces applying corpus linguistics to other areas of research. This section starts with the work of Andersen, how to use corpus linguistics in sociolinguistics. Accordingly, it is mentioned that corpus and sociolinguistics are the two related and historical distinct research traditions. The author claims that corpora and corpus methods can be helpful to researcher who wants to pursue a sociolinguistics research questions. In the same way, O'Halloran focuses on critical discourse analysis and corpus based approach to critical discourse analysis. The other topic which is explained in this section is the use of corpus linguistics in forensic linguistics by Cotterill. The author explains that corpus can be a potential tool to analysis forensic linguistics. Besides, the author discusses methodological challenges and limitations in forensic linguistic analysis. Adel also covers the use of corpus linguistics in the study of political discourse. The interrelation between corpora and political genres, and corpora techniques for exploring political discourses are also explained. Health communication

studies start to recognize corpus methodology. Thus, Atkins and Harvey elucidate how to use corpus linguistics in the study of health communication. The authors focus on building a corpus of adolescent health language and techniques that are used to explore health language patterns. Farr also explains the use of corpora in teacher education. The author mentions corpora play a great role to teacher education. Hence, the author covers corpora application in teacher education, pedagogical application of pedagogical corpora, learner corpora, and corpora of classroom language. The last chapter of this section and the hand book is the work of Baker, using corpora in language testing. Hence, the author discusses the development of corpus use in language testing, the use of learner corpus, and native speaker's corpus to inform language testing.

Advantage of the handbook

This handbook is user friendly and effective handbook. It raises one of potentially pivotal issues in English language pedagogy, second and foreign language research, and language analysis. The main contribution of this handbook is that it addresses issues step-by step on how to use corpus linguistics in language classroom, material and syllabus design, testing and language analysis accompanied by cases studies, implications and clear examples. In addition, the authors demonstrate how corpus linguistics application is widening to computational linguistics, discourse analysis, forensic linguistics, creative writing and translation. The handbook could be considered as marvelous starting point for further in depth investigation and discussion on corpus linguistics.

Moreover, detail reference books are provided in each chapter for further reading. These help readers to consider various issues of corpus linguistics application. Besides, the handbook provides a step by step guide how to assemble and to design different kinds of corpora. These could give the opportunity to language teachers and researchers easily duplicate or apply corpus linguistics in their research and classrooms. Hence, this handbook fills the gap that exist using corpus linguistics application in a wider contexts. The handbook could be invaluable resources for undergraduate students, postgraduate students, language teachers and researchers who want to engage in corpus related issues. It would also be a primary interest of researchers and language teachers who have less exposure on corpus linguistics and its application.

Limitation

This handbook is an introductory handbook and I would like to suggest some points. Firstly, the handbook lacks inconsistency on its organization. I observed that similar issues that have similar concepts have addressed indifferent sections. For instance, instead of putting using corpus to language testing and teacher education in section VIII, they should be placed on section V, using corpus for language pedagogy and methodology since these issues are very interrelated. Though the organization is approached thematically, identical concepts are placed in different sections. In addition, the two sections in the introduction part have similar concepts. The authors address similar concepts which could be mentioned in one section. Despite this short coming, this comprehensive volume has achieved its purpose and can make very practical contribution to corpus linguistics.

References
Babanoglu, M. P. (2013). A corpus-based study on the use of pragmatic markers as speech-like features in Turkish EFL learners' argumentative essays. *Procedia- Social and Behavioral Sciences, 136,* 186–183. https://doi.org/10.1016/j.sbspro.2014.05.312.
Flowerdew, L. (2012). *Corpora and language and education.* New York: Palgrave Macmillan.
Granger, S. (2002). A bird's-view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *computer corpora, second language acquisition and foreign language teaching* (pp. 3–33). Johns Benjamins Publishing company.
Nesselhauf, N. (2005). *Studies in corpus linguistics: Collocations in a learner corpus.* John Benjamins Publishing Company.
O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge: Cambering University Press.
Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257–283. https://doi.org/10.1016/j.jslw.2004.06.002.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.