Book review

On Invisible Language in Modern English: A Corpus-based Approach to Ellipsis. By Evelyn Gandón-Chapela. London: Bloomsbury Academic. 2020.

Zheyuan Dai[1*] ⬤

[1] Linguistics Department, Zhejiang University
[*] Corresponding author's email: zydai@zju.edu.cn

Instead of directly addressing "ellipsis" in the main title, Evelyn Gandón-Chapela describes this linguistic phenomenon as "invisible language" colloquially and vividly. However, readers may be misled by the title of the book "On Invisible Language in Modern English – A Corpus-based Approach to Ellipsis", which in fact only concentrates on the post-auxiliary ellipsis (PAE) rather than the overall ellipsis in Modern English. PAE covers ellipsis cases in which a verb phrase, prepositional phrase, noun phrase, adjective phrase, or adverbial phrase is omitted after one of the following licensors: modal auxiliaries, auxiliaries *be*, *have*, and *do*, and infinitival marker *to*. Concretely, the book discusses two main sub-types of PAE, namely verb-phrase ellipsis (VP-Ellipsis) and pseudo-gapping (PG). See examples (1) and (2) as follows for further illustration, respectively:

(1) I have [eaten an apple] this morning, but Mary *hasn't* ~~eaten an apple~~.

(2) Peter [kissed] Daisy, and Paul *did* ~~kiss~~ Nancy.

Example (1) of VP-Ellipsis shows the omission of the VP triggered by the licensor *have*. The elided VP antecedent (eaten an apple) is highlighted by square brackets. Though example (2) is close to the structure of example (1), in PG, there would be a complement left after the auxiliary, as the directly object Nancy after the licensor *did*.

In all, the book stands out among other ellipsis-related volumes for its qualitative discussions accompanied by quantitative analysis. With this work, Dr. Gandón-Chapela conducted the first sustained diachronic corpus investigation towards PAE. Aiming to provide an empirical account for PAE, this book not only reports the descriptive overview of PAE in the Penn Parsed Corpus of Modern English (PPCME) (1700-1914) but also compares its results with former corpus studies. What is noteworthy is that the book presents a new series of algorithms for automatically detecting and retrieving PAE from the Penn Parsed Corpus of Modern English. Retrieving a considerable amount of PAE accurately has always been a premise for quantitative research on the ellipsis phenomenon. Nonetheless, it is not easy

due to their highly-liberalized structures. The methodology Dr. Gandón-Chapela updated was stated in a detailed and specific style, making it quite applicable to a wide range of parsed corpora, contributing to the more efficient realization of this premise.

The book is very much written in the traditional academic style. With four main chapters, this book presents literature review in the first Chapter, methodology in the second, and then they are followed by descriptive results and the analyses. In the end, it goes with concise and clear conclusions. One of the most striking advantages is that the illustrative instances are widely scattered around the book for clarifying some concepts and the author's opinions. For this reason, this book is well suited for those who have relatively weak theoretical backgrounds as well as linguists with an intense interest in English ellipsis, corpus linguistics, and language evolution.

The beginning of the introduction sets the scene for the whole book by showing the necessity for the research with vivid explanations of illustrative sentences. The methodology used in the book is briefly mentioned, along with the employed corpus and the variables to be analyzed. Ellipsis is a unique linguistics phenomenon at the semantics-syntax-pragmatics interfaces (Merchant 2010). The mismatches between the invisible information (the elided structure with intended meaning) and the visible elements (what is actually pronounced) can cause an ambiguity. As the consideration of ellipsis is a complex linguistic phenomenon, in Section 1.2, along with the three most influential theoretical accounts, that is, Comprehensive Grammar of English, Systematic Functional Grammar, and Transformational Generative Grammar (TGG), the book helps the readers to have a broad understanding of ellipsis. Take the TGG part as an example. TGG mainly focuses on the formal characteristics of ellipsis with the endeavor of answering three relevant questions about the structure, the identity, and the licensing. The book follows Bîlbîie's (2011: 129) clear-cut criteria for the identification of elliptical structure. However, the criteria proposed by Bîlbîie concentrate mainly on the syntactic and semantic sides. Discourse-related influences such as context and thematic structure matter as well. To supplement the theoretical analyses, the empirical findings of ellipsis from the psycholinguistic perspective were also provided.

As stated in Liu (2018), language can be considered as a human-driven complex adaptive system. The syntactic structure of PAE is highly-liberalized owing to both grammatical and interactional reasons. To detect and retrieve such a linguistic phenomenon as accurately and completely as possible in a corpus, the query should be designed with careful consideration for all possible PAE scenarios. In light of that, Chapter 2 further primes the reader with the literature review on studying English PAE with corpus-based approaches. Section 2.1 introduces the research target, namely, the data source, PPCME (1700-1914). For two main branches of English PAE, VP-Ellipsis and PG, a number of reviews with empirical methods are revisited. For academic disputes over some structural details, the author posits herself clearly. For instance, the book follows Sag's (1976: 53) correction on the misnomer of so-called "VP-Ellipsis", and reclassifies it into a subtype of PAE. VP-Ellipsis together with PG, these two main subtypes of PAE, is dissected in terms of syntactic characteristics. Because of that, all types of PAE cases

are listed in Section 2.2.3.2. The difficulty with retrieving PAE structures is that, as the author states, the omission words are *invisible* in a sentence. Fortunately, the problem can be solved through retrieving algorithms designed in concise and basic formal logical running with the Corpus Search 2, a java-based software assorted with PPCMBE. In the annotation scheme of the PPCMBE, the asterisk * thus could correspondingly represent those invisible elements. This symbol is a wild card for any combinations of ellipsis resources and targets. With the assistance of the symbol, the invisible part of the sentence can be visualized. Take the query ((MD* iPrecedes HV*) AND (HV* iPrecedes [.,])) as an instance, the query returns to examples where any modal auxiliary immediately precedes the auxiliary *have*, and in turn, immediately precedes any punctuation mark. For the technical operability, corresponding algorithms for detecting PAE in the corpus are outlined one by one. From my perspective, one of the most outstanding merits of the monograph is that with as many PAE structural scenarios considered, the recall rate has been raised from 0.89 to 0.97. The increase in recall rate surely would optimize the automatic corpus retrieving work, guaranteeing the quantity as well as the quality of the database for further analysis. Appendix 1 shows the basic query language of Corpus Search 2, making the retrieving process accessible for linguists as well as beginners.

In Chapter 3, based on the retrieved corpus, 32 variables are divided into three main groups for different research purposes, "Core defining variables", "Usage variables" and "Processing variables". Each group is accompanied by illustrative sentences for definition interpretation, statistical analysis, comparison with results of Present-Day English, and a succinct summary. What is noteworthy is that some variables, such as type of anaphora, sloppy identity, remnants, etc., have been studied empirically in Present-Day English (e.g., Bos and Spenader 2011; Miller 2014), which paves the way for the analysis of the diachronic evolution of PAE.

Defining variables focuses on grammatical and discursive aspects. The present monograph focuses on the licensors of PAE in PPCME. The licensor, a grammatical element that triggers the appearance of ellipsis, is fairly useful in helping language learners or researchers quickly detect the ellipsis. It discovers that modal auxiliaries, such as can/could, will/would, may/might, etc., are the most frequent licensors of PAE in Late Modern English. This conclusion remains valid after comparing with that of Present-Day English (Bos and Spenader 2011). Moreover, through the diachronic exploration, the book discovered that some licensors employed in Late Modern English such as *shouldest*, *shalt*, *durst*, *dost*, and *ought*, are no longer used in the Present-Day English. As for the discursive perspective, for example, the frequencies of four clausal types under the framework of discourse conditions were calculated for comparing the types of clause of the source of ellipsis versus the target ellipsis. Usage variables concentrate on the dynamic description of PAE, especially for their diachronic evolutions and genre distributions. To explore the possible diachronic variations of PAE, examples were sorted arbitrarily into 5 groups with every 50 years as nodes for classification. As for the genre part, 18 genres were equally divided into speech-related and writing-related genres (Fiction is treated as a mixed type). Processing

variables, as the name implies, process the connection between the resource and the target of ellipsis, primarily concerning the co-textual aspects of PAE. Two types of processing distances were estimated. The lexical distance measures the words between the resource and the target of ellipsis, while the syntactic distance is measured in the number of IPs. Take the syntactic distance for illustration, in the vast majority (around 76.88 %) of the VP-Ellipsis examples, there are no clauses intervening between the source and target of ellipsis. The concept of lexical distance is close to that of dependency distance between two words in a sentence (Heringer et al. 1980; Hudson 1995; Liu 2008) in Dependency Grammar. Dependency distance is a measurement of syntactic complexity as well as human cognitive load (Hudson 1995; Liu 2008). Coincidentally, according to the Dependency Locality Theory proposed by Gibson (1998 & 2000), the longer the syntactic dependency is, the higher the possibility for a sentence to bear larger syntactic complexity. Therefore, the calculation of the lexical distance variable may be helpful empirical assistance to further analysis for measuring the syntactic complexity of PAE. Moreover, Popescu et al. (2014) proposed that the length distribution of all types of language units would conform to the Zipf-Alekseev distribution. Following their steps, Jiang & Liu (2015) discovered that the DD distribution of natural human languages also conforms to the right-truncated Zipf-Alekseev distribution. With the empirical contribution of the present book, the regularity of the PAE phenomenon can also be captured based on the processing variables.

Chapter 4, the conclusion, summarizes the main research results with a brief presentation of the research targets, methods and steps once again. With all the quantitative descriptions of PAE, the book verifies many hypotheses and theoretical claims raised by former research. For example, VP-Ellipsis can be licensed by more than one auxiliary, whereas, as a general rule, PG cannot. Also, the research has established the syntactic linking between the antecedent and the ellipsis site in PAE. All the discoveries may assist both the researchers and language learners to get acquainted with PAE more.

There are, however, a few points that could be possibly improved, though the book is generally well-organized and informative. First, this book quantitatively describes the PAE of Late Modern English but less consideration has been given to the linguistic interpretation behind these statistics. Second, for the diachronic analysis of the PAE, the monograph mainly takes Bos and Spenader (2011)'s work on PAE of Present-Day English for reference, which is based on the data collected from the Wall Street Journal sections of the Penn Treebank. As mentioned by Dr. Gandón-Chapela herself, Bos and Spenader's data may be biased due to their genre or theme monotonicity in the database. In all, this book is an excellent contribution to the corpus-based study on English PAE and also a valuable empirical addition to theoretical research. It provides a complete set of quantitative research methodology that can be applied to traditional research on a language phenomenon for reference, namely, the construction of theoretical background, the selection of appropriate corpus, the retrieval of the target structures, and the analysis of quantitative results. Besides, readers may also find that it can be referred to as an encyclopedia for PAE in Late Modern English. As a pioneering study of English PAE with corpus-

based approaches, this book has laid a solid foundation for the follow-up diachronic and synchronic research. Moreover, with the process of detecting and retrieving highly-liberalized structures in a corpus, the book has demonstrated good operability of employing corpus for its readers, from amateur to professional. The valuable experience has great reference significance for the quantitative research of specific linguistic structures.

Acknowledgements

References

Bîlbîie, G. (2011). *Grammaire des Constructions Elliptiques. Une étude Comparativedes Phrases sans Verbe en Roumain et en Français*. Université Paris Diderot-Paris 7, Paris. (PhD thesis)

Bos, J., Spenader, J. (2011). An Annotated Corpus for the Analysis of VP Ellipsis. *Language Resources and Evaluation*, 45(4), pp. 463-94.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, pp. 1-76.

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In: Marantz, A., Miyashita, Y., O'Neil, W. (Eds.). *Image, language, brain*. Cambridge MA: MIT Press, pp. 95-126.

Heringer, H. J., Strecker, B., Wimmer, R. (1980). *Syntax: Fragen-Lösungen- Alternativen*. München: Wilhelm Fink Verlag.

Hudson, R. A. (1995). *Measuring Syntactic Difficulty*. Manuscript. London: University College London.

Jiang, J., Liu, H. (2015). The effects of sentence length on dependency distance, dependency direction and the implications-based on a parallel English-Chinese dependency treebank. *Language Sciences* 50, pp. 93-104.

Liu, H. (2008). *Dependency Grammar: From Theory to Practice*. Beijing: Science Press.

Merchant, J. (2001). *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*. Oxford: Oxford University Press.

Miller, P. (2014). A Corpus Study of Pseudogapping and its Theoretical Consequences. In: Piñón, C. (ed.). *Empirical Issues in Syntax and Semantics 10*, pp. 73–90. Available online: http://www.cssp.cnrs.fr/eiss10/ (accessed 29 November 2018).

Popescu, I.-I., Best, K. H., Altmann, G. (2014). *Unified Modeling of Length in Language*. Lüdenscheid: RAM-Verlag.

Sag, I. A. (1976). *Deletion and Logical Form*. MIT: Cambridge, MA. (PhD thesis)