

**Review of Gaëtanelle Gilquin. *Corpus, cognition and causative constructions*.** Amsterdam / Philadelphia: John Benjamins, 2010, xvii + 326 pp., ISBN: 9789027223135.

Reviewed by Dylan Glynn, Lund University, Sweden. E-mail: dylan.glynn@englund.lu.se

The title of this book raises what is arguably one of the most important questions in contemporary linguistics—what is the relationship between frequency-based corpus research and salience-based ‘cognitive’ research? Wisely, the author does not attempt to answer this question directly. Such a difficult issue will not be resolved by a single work, or perhaps even by a single generation of linguists. However, the author does exactly what needs to be done—develop empirical methodology so that, one day, we will be able to answer this question. Using a combination of elicitation and quantitative corpus-driven techniques, the author maps the use of the periphrastic causative constructions in English. The work is detailed, empirical, and accurate; yet it never loses sight of the crucial theoretical issues its results inform. Seen in this light, the book is an excellent contribution to both empirical descriptive linguistics and theoretical research. Three lines of enquiry are developed—theoretical, methodological, and descriptive. The review considers each in turn.

The descriptive aim of this book is to account for the onomasiological field of Causative Constructions in British English. Before this study, no such description had attempted to offer a comprehensive and empirical account of English causation. If only for this reason, the work is invaluable and deserves to become a point of reference for all scholars working on causation in English or periphrastic causatives more generally. In its efforts at comprehensiveness, the study examines all four causative lexemes (*cause*, *get*, *make*, *have*) in each of their syntactic patterns, making a total of ten periphrastic Causative Constructions ([X MAKE Y V<sub>to-inf</sub>], [X BE made V<sub>to-inf</sub>], [X MAKE Y V<sub>pp</sub>], [X HAVE

Y V<sub>inf</sub>], [X *HAVE* Y V<sub>pp</sub>], [X *HAVE* Y V<sub>pp</sub>], [X *GET* Y V<sub>pp</sub>], [X *GET* Y V<sub>pp</sub>], [X *GET* V<sub>to-inf</sub>], [X *CAUSE* Y V<sub>to-inf</sub>]). A sample of 3,500 examples, extracted from the British National Corpus (BNC), forms the basis of the study. The sample is submitted to a usage-feature analysis typical of corpus-driven Cognitive Linguistics research (Gries and Stefanowitsch 2006; Glynn and Fischer 2010 *inter alia*). Following the norm in usage-feature analysis, the manual annotation of the dataset covers a wide range of usage factors from purely formal to semantico-pragmatic and sociolinguistic. These corpus data and their analysis are then supplemented with elicitation data collected in the form of 50 questionnaires, completed by native speakers of British English.

The bulk of the descriptive value is found in Chapters 5 to 10, where the empirical results of the onomasiological analysis are presented. These descriptive chapters begin with an examination of each of the usage-factors, considered relative to the construction's argument structure. The individual features are explained briefly and examples are given. A linear regression analysis determines which of the 47 factors are significant and influential in distinguishing the use of constructions.

Chapter 6 tests the intuition-based hypotheses about the prototype structures of causation in Cognitive Linguistics. The chapter is important because it raises the methodological problem of discrepancy between corpus results, elicitation results, and results obtained through individual intuition. However, it leaves the question open with only a short discussion as to any possible explanation for the discrepancies. Chapter 7 presents a collocation analysis of the constructions. It follows traditional corpus methods and its results reveal clear lexical patterning. A supplement study using elicited data confirms many of the findings. Some of these findings suggest sociolinguistic factors are at play. However, since the traditional corpus methods do not readily permit multivariate analysis, no information on how these results interact with other semantic and syntactic factors is offered. Chapter 8 employs a more complex form of collocation analysis identifying the collexemes of the second verb. This gives us the distinctive lexical-constructional association for each of the ten constructions. Again, relatively clear usage patterns are identified, but this time elicitation data do not confirm the findings. Reasons for this are proposed, but no conclusion is reached.

Chapter 9 zooms in on the role of register in the semantic structure of the constructions. Again, collocation analysis is used and the results demonstrate unequivocally that text type and register directly impact upon the use of the constructions, especially the second verb slot. Although the effects of sociolinguistic dimensions are investigated, the results are not submitted to multivariate analysis and so the interactions of some of these dimensions are left for future research. A final descriptive chapter on the place of periphrastic causatives in language learning moves us towards applied linguistics. It is

shown that these constructions pose problems even for advanced language learners and the study offers pedagogical solutions to this problem.

Descriptively, there exist two possible shortcomings. Neither of these shortcomings detracts from the value of the work, but they do represent important limitations to the completeness of the results. As a first limitation, we have the question of representativity. The author restricts the analysis to British English. Although one cannot deny that a careful in-depth examination of one dialect is better than a less careful analysis of several, the descriptive completeness would be much greater if more than one of the major dialects were included. Similarly, all data are taken from the BNC, which the author acknowledges is unbalanced (p. 30). The well-known bias towards written and rather ‘conservative’ language is compensated for by taking equal samples of both written and spoken data. However, today, the options for supplementing the data in the BNC are improving and, ideally, the author could have compensated for some of the BNC’s biases with other sources. This said, for some of the statistical methods employed, one needs accurate tagging and on that front, the BNC is excellent and other sources may have introduced noise into the results.

This brings us to the second limitation—the role of genre and register could have been more properly integrated into the main of the analysis. The role of these language dimensions, categorised under the rubric of “various” (p. 100), seems, at times, a little *ad hoc*. Although Chapter 9 is devoted entirely to such issues, and it covers them well, the previous analyses could have benefited from including the social factors more systematically. For example, in Chapter 5, although the medium (spoken versus written) is found to be a statistically significant factor in the analysis, it is not investigated. In Chapters 7 and 8, the role of such issues is entirely absent, even though register effects appear to be relevant (p. 189). The reason for treating such issues separately, in Chapter 9, might be methodological—such dimensions of language are notoriously difficult to capture using elicitation methods and the quantitative method used in Chapters 7 and 8 offers no possibilities for including multifactorial analysis.

Despite these more than reasonable limitations on the descriptive completeness, the tremendous care and detail devoted to such a wide range of data is nothing short of excellent. The descriptive merits of the study assure the book’s value for a wide range of scholars.

Turning to the methodological goals of the book, we again see the sheer scope of this work. Not one of a recent string of influential monographs in the field of corpus-driven Cognitive Linguistics, to which this book belongs, even attempts at comparing other empirical methods with corpus results. Although the avant-garde of the field, represented by Szmrecsanyi (2006), Hilpert (2008), Wulff (2009), and Divjak (2010) *inter alia*, have all advanced statistical

analysis, none of these studies have considered any evidence obtained from elicitation or experimentation. Moreover, although the quantitative research is based largely on collocation analysis, the author employs the latest and most precise methods in the form of collocation analysis (Stefanowitsch and Gries 2003).

The study takes the first vital steps towards bringing elicited and found data techniques together, just as the author stresses that converging methods are an essential path for future research. However, the study remains biased towards corpus data, elicited data being considered *post hoc* and experimental data not being included. This should not be seen as a criticism, not only because the book goes further than any others, but because undertaking a truly fully-fledged experimental study, as well as a corpus-driven study, presents tremendous technical hurdles and vast amounts of labour that, arguably, only an established multidisciplinary research team could achieve. That this single work takes these first steps in this direction is more than commendable.

A second methodological concern lies in the quantitative techniques employed. The author seems to draw heavily upon the British corpus tradition, sometimes at the expense of relevant techniques developed by comparable work in the Cognitive Linguistic corpus tradition. Gilquin's study joins the ranks of a vibrant and growing empirical Cognitive Linguistics research community, yet the study seems to largely ignore much of this comparable work. This is, perhaps, regrettable in many instances since the methods used in the study could have been supplemented with other techniques prevalent in the field.

The choice and reporting of statistical techniques is a minefield that goes well beyond the interest or knowledge of most readers. However, the hierarchical selection in Chapter 5 was performed using a linear regression (p. 137) without reporting the model or the statistics of the actual regression. Moreover, why was a McNemar's Test employed over other comparable tests? In later chapters, why was not an analysis of variance or a configural frequency analysis performed? Although the use of collocation analysis offers many important insights, the results suggest that one cannot fully capture the lexical-constructional interaction without considering the impact of other usage factors. Current corpus-driven cognitive research offers many examples of techniques for capturing this interaction.

At the theoretical level, the book is highly relevant. The theory of Cognitive Linguistics and how it relates to the use of corpus data is explained—a well-trodden path, but one that is kept pertinent and clear. A model is proposed, the 'corpus-cognition integrated model'. Although this model does not add anything to already established understandings of the use of corpus data in Cognitive Linguistics, it clearly explains fundamental notions. Perhaps the only improvement to the theoretical discussion would be that the author could go

further than she does in fleshing out the problems of relating salience-based results to frequency-based results (even if answering such a question is beyond the purview of the book). Indeed, on this very issue, the author has other publications that tackle the debates surrounding the questions of frequency, salience, convention, entrenchment and how elicitation or corpus data permit different operationalisations of Langacker's and Lakoff's models of grammar and categorisation (Gilquin and Gries 2009; Arppe et al. 2010). However, I wish to stress that although the book does not advance theoretical discussion on such problems, it still advances our understanding. The empirical revolution that is taking place in linguistics is arguably a result of a desire to look for ways of testing our hypotheses and adequately operationalising our analytical concepts. I believe that many would agree with the idea that it is through careful data-driven analysis that we will come to understand such issues, rather than through further theoretical research.

This work is an excellent linguistic study, advancing the science through its methodological rigour and its attempts to understand how elicited data inform corpus linguistics. The volume joins a growing community of empirical and quantitative research at the heart of Cognitive Linguistics. The theoretical implications of combining elicitation and corpus methodology remain to be understood, but this is a task that we may expect to keep language science busy for some years to come and it would be unrealistic to ask a single work to answer such questions. Finally, its descriptive depth and thoroughness make this work an incontestable piece of linguistic research deserving a place on any linguist's shelf.

## References

- Arppe, A., G. Gilquin, D. Glynn, M. Hilpert, & A. Zeschel. 2010. Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora* 5: 1–27.
- Divjak, D. 2010. *Structuring the lexicon: A clustered model for near-synonymy*. Berlin: Mouton de Gruyter.
- Gilquin, G. & St. Th. Gries (eds.). 2009. Corpora and experimental methods (Special edition. *Corpus Linguistics and Linguistic Theory* 5). Berlin: Mouton de Gruyter.
- Glynn, D. & K. Fischer (eds.). 2010. *Quantitative cognitive semantics. Corpus-driven approaches*. Berlin: Mouton de Gruyter.
- Gries, St. Th. & A. Stefanowitsch. (eds.). 2006. *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*. Berlin: Mouton de Gruyter.
- Hilpert, M. 2008. *Germanic future constructions. A usage-based approach to language change*. Amsterdam: John Benjamins.
- Stefanowitsch, A. & St. Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209–243.
- Szmrecsanyi, B. 2006. *Morphosyntactic persistence in spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin: Mouton de Gruyter.
- Wulff, S. 2009. *Rethinking idiomaticity. A usage-based approach*. London: Continuum.