

Hoffman, Sebastian. *Grammaticalization and English complex prepositions: A corpus based study*. (Routledge advances in corpus linguistics.) New York: Routledge. (ix, 214).

The topic of English complex prepositions, preposition-noun-preposition (PNP) structures that seem to function as single prepositional units (discussed in detail 26, f) is not new to the literature, nor are opinions about their function, even their reality, consistent. Not only does the account of the development and use of the 30 most frequently used PNPs in this book, in itself, make an excellent contribution in empirical linguistics, the present study also provides insights into the fascinating possibilities offered in computational, corpus-based research arising from the search capabilities of numerous electronic databases and the capability of searching for items which occur not only as lexemes listed in dictionaries but are used in example sentences, such as in the Oxford English Dictionary Quotations Database. In this volume the texts of the Gutenberg Corpus are listed (Appendix 1), and the David Lee genre classification scheme for the British National Corpus is reproduced (Appendix 2). Abundant tables enhance readability of the conclusions formed, and a useful index makes reference convenient.

In Ch. 2 H discusses several of the major corpora available for English: the Oxford English Dictionary quotation database, The Gutenberg Corpus, The British National Corpus (BNC, a 100 million word corpus including both written [90%] and spoken [10%] data), and the Bank of English (more than 450 million words), among other more specialized corpora. These are described in detail, and along the way are delineated criteria a corpus must satisfy — selection criteria for data included, representativeness and balance, reliability of data, and quantification of results (10). The author offers a thorough critical discussion of the challenges involved in utilizing any given corpus for a certain purpose. These include the problem of searchability due to spelling irregularities or changes (one recalls Dr. Samuel Johnson's famous

plaint about the spelling conventions of writers of his time) (8); unbalanced representation of writers (9) (the Gutenberg Corpus includes, by my count, eight works by D. Defoe, eight by J. Austen, the six volumes of E. Gibbon's *Decline and Fall of the Roman Empire*, the seventeen volumes of T. Carlyle's *History of Friedrich II of Prussia*, along with four other of his works, and 45 works of C. Dickens); unbalanced representation, in sheer volume of material available, of time periods — the BNC is designed to serve as a 'representative sample' of English as heard on the British Isles near the end of the twentieth century (13), while corpora covering earlier periods, such as the Helsinki Corpus, are very much smaller (7, f); and reliability, such as the observation (from Berg, D.L. 1991. *A User's Guide to the Oxford English Dictionary*. Oxford: Oxford University Press.) that in the first edition of the OED, where no examples of current usage were found, compilers 'made up' illustrations, indicated in the text as 'Mod.' (11). H also describes the search methods employed for retrieving data (20, f), which demonstrates how the thirty most frequent items were located and tabulated (see Table 2.4, 23).

Contrasting views of the role and reality of PNPs in the grammar are discussed in Ch. 3, and the argument is developed that such PNP units as discussed in the present text "are generally conceptualized as constituting single syntactic units" (25). Data from the BNC, for example, are employed to demonstrate that among tokens of the 30 most frequently occurring complex prepositions, 96% are compatible with analysis as a single unit, while 4% appear to be "stylistically restricted variant[s] of coordination" (35-39). We note in passing that such conclusions are made possible not by examining an isolated datum uttered by a single speaker but by a comprehensive search of a large corpus. Likewise made available is the observation in the BNC oral section that filled pauses (*er, uh*) "overwhelmingly occur just before or just after" PNPs, i.e., in 300 instances (45), while only five instances each, in a 10 million word database,

occur internally, before or after the nominal element (45). These data contribute to the conclusion that “PNP-constructions are stored in the memory as single chunks” (ibid.) and not assembled from scratch in the process of speech production. It can also be pointed out that this level of analysis contravenes the presupposition in theoretical linguistics that syntax can be isolated from the rest of language. The central pillar of cognitive grammar (R. Langacker. 1987. *Foundations of cognitive grammar: Theoretical prerequisites*. Vol. I. Stanford, CA: Stanford University Press) denies the “separability of syntax and semantics” (1). Accordingly, we cannot characterize these structures and analyze their functions without attending to their meaning or the context of their utterance, be it stylistic or social.

The development of *in view of* as a complex preposition, ranking number 16 in frequency in the BNC (Table 2.4, 23), is traced from examples in which *view* is clearly analyzed as a nominal element in a composition, used with its “literal” meaning, “In this manner, we came *in view of* the entrance of a wood ...” (from Defoe, *Robinson Crusoe*, quoted 53) to its use as a grammatical element, “*In view of* your comments ...” (BNC, 53), where the semantics have become generalized (54), with a primarily grammatical meaning. This development is consistent with Meillet’s well known pronouncement, “*le passage d’un mot au rôle d’élément grammatical,*” thus a prototypical example of grammaticalization.

Another benefit of a computational search of a corpus is the ability to discover examples of “layering” of older and more recent forms. An anecdotal approach may seize one of the nine instances in the BNC in which the *view* element occurs with its “literal” meaning, just as used by Defoe above (54), but when considered against the 1,507 total instances (Table 2.4, 23), it seems clear that these represent not counterexamples to a theory of grammaticalization of the item, but

instances where a fossilized or compositional usage occurs alongside the newer innovation, much like the co-occurrence of *shined* and *shone* in many varieties.

In Chapter 5, the thirty most frequent PNP units are discussed in greater or lesser detail, organized according to chronology of development as complex preposition units. Some of these put in question both the notions typical of theories of grammaticalization that semantics shift from the more concrete to the more abstract and that grammaticalization necessarily involves a long, gradual process. The unit *by way of* is shown to contradict the ‘typical’ path of concrete to abstract meaning. The first OEC entry for this item, “Þan may Ðe saules in purgatory, *By way of* grace specialy, be delivered of pyn” (67) dates from 1340. Not until the eighteenth century are found such uses as, “I ... left the city *by way of* the Bowery” (OED, 1787, 69), in which the meaning is concrete, as a ‘path’ or ‘route’. This item demonstrates how the corpus approach in this study make possible accessing data which show the seemingly backward development of the unit with the concrete meaning of *way*.

Likewise, the unit *by virtue of* appears in clearly compositional usage in the fourteenth century in the OED, “He was i-bore of Ðe mayde Marie *by vertu of* God wiÐ oute mannys mone (71), but appears as a unit rather suddenly in the eighteenth, “The piston continues to descend *by virtue of* the expansion of the steam” (OED, 72), when the concrete meaning of the Fr. stem appears to have dissolved. The rapid advent of this item can be accounted for quite likely by the fact that in French, *par la vertu de* is attested from the thirteenth century; the output in English is most likely a calque (72).

We can make a similar observation about *by way of* in its seeming shift from abstract to concrete meaning. After the eleventh century a very high degree of bilingualism existed among the literate classes, who constituted, nonetheless, a very small proportion of the English speaking

population. Subsequently, the Renaissance occasioned a wide familiarity among the increasingly numerous literate classes, many of whom also achieved bilingualism in an era when foreign-born tutors and touring Europe were commonplace. By the middle of the eighteenth century, literacy among the middle and trade classes had risen greatly, and every schoolboy was familiar with Latin and Greek classics. Thus the Latin preposition *via*, glossed ‘by way of’, is quite likely the source of the concrete semantic representation of the PNP found late in the eighteenth century and afterward. This influence of bilingualism will figure later in another aspect of the development of some PNPs as well.

Another property of studies involving corpus searches is the ability to generate algorithms that retrieve collocates for units that are the objects of search. Applying this to *in need of* permits the observation that in the written portion of the BNC, the unit collocates (in a window of one to four words to its left) with *be* more than 3.5 million times, and with *may* and *might* another quarter million times, which “exclusively” co-occur with *be* (N. 18, 193). The next most frequent collocate represented is *see*, with 150,000 tokens and *look*, at 88,000 (79). We should bear in mind, too, that the latter does frequent service as a linking verb, analogous to *be*. Thus a strong case can be made that *in need of* seems to constitute part of a lexicalized sentence stem rather than a complex preposition (79). Furthermore, the computational search reveals that *in need of* exhibits frequent variation, i.e., with qualifiers before the nominal element, in contrast to items like *with regard to* and *in relation to*, which exhibit virtually none, suggesting a much lower degree of grammaticalization (85).

The range of collocation demonstrated by a unit offers another clue to grammaticalization. In the Early Modern English period *in conjunction with* typically was used with nouns concerning astrological bodies or religious entities, as in this 1603 example from the OED,

“Neither could Venus *in conjunction with* Mars cause any to mistresse another mans wife” (82). Later usages are seen to generalize in collocation (83), an example of the principle described by Nikolaus P. Himmelmann, that in grammaticization an item participates with “a *set*” of items (*italics his*) in its newly reanalyzed status, usually involving the expansion of the “syntagmatic context” (“Lexicalization and grammaticization: Opposite or orthogonal?” (*What makes grammaticalization? A look from its fringes*. Bisang, Walter, Himmelmann, Nikolaus P., & Wiemer, Bjorn, (eds.), 2004. Berlin, New York: Mouton de Gruyter.21 – 42). Enumerating collocates and contexts using computational methods applied to the BNC permits us to sketch the semantic range of collocates of the generalized PNP unit.

The item *in favor of* demonstrates a further function as a grammaticalized unit, that of text structuring. As an analyzable unit, *favor* was used in the sixteenth and seventeenth centuries as in the following OED example from 1600, (The Tuscans ... got hart and were very iolie, saying that the Gods were *in favor of* them (84). This example from the BNC, “Perrier’s French television advertisements eschew weighty health talk *in favor of* Gallic schmaltz” (*ibid.*), demonstrates the text structuring function (*ibid.*) of forming a comparison conveying a subjective evaluation in the speaker.

Further development of the text structuring function of grammaticalized PNPs is seen in the discussion of *in terms of*, from its origins as a Latinic loan to its modern use as a topicalizing and a hedging device. Data developed from the oral section of the BNC demonstrate that co-occurrence of *in terms of* with filled pauses, repetitions, discourse markers (*I mean, you know, sort of*), “odd syntax,” and pauses, constitute 30% of all uses (Table 7.2, 132). The availability of this data suggests new possibilities for studying the stylistic and sociolinguistic distribution of such items. The evidence from Table 7.3 (133) demonstrates, for example, a very high

frequency of *in terms of* in contexts such as meetings, interviews, and the like. It strikes me that the speaker in such examples as “*in terms of* when you’re taking so much into consideration, the actual loading or weighting of that factor would actually turn into something very insignificant *in terms of* weighting I think” (BNC, 127) uses *in terms of* as a focalizer, a hedge for her idiosyncratic inability or contextual unwillingness to come to the point, and as a sociolinguistic lever, a kind of *inkhorn* expression, to elevate the status of her otherwise unintelligible committee-speak. Plentiful evidence is provided to demonstrate discourse marking functions, but it is hard to see whether “time pressures” (138) or other idiosyncratic factors motivate these discourse-focused uses of *in terms of* in the oral domain. Given the well known principle that among the rising middle classes also arose linguistic insecurity and a tendency for hypercorrection, especially in learned or professional contexts, the data offered here suggest further research focusing on the sociolinguistic marking that they seem to offer.

In that regard we do see evidence that certain PNPs occur at least 20 times more frequently in written text (Table 6.2, 105). The author suggests “emphasis on clarity and precision from the seventeenth and eighteenth centuries” as a factor in the rise of PNPs as complex prepositions during that era (101-2). But here one must take into account also the high degree of familiarity with the Latin classics among the literate classes of the period and the very high degree of prestige invested in Latin and her daughter tongues during that time, and recognize that many of the stems of the nominal elements of these PNPs are of Latin origin. Also contributing impetus was the influence exerted by such writers as Hugh Blair, who insisted that among the criteria for good writing is the requirement that the language be “local,” praising “Dean Swift, one of our most correct writers, [who] valued himself much on using no words but such as were of native growth” (1783, *Rhetoric and the Belle Lettres*). The obvious outcome of

these twin influences is the calquing of these units from Latin and French and employing them in written, i.e., learned, contexts.

This conception leads to a new suggestion for a grammaticalization *cline*, in connection with which the case of *in contrast to*, *in response to*, and *in accordance with* offers an intriguing analysis. Unlike *in front of*, which demonstrates a gradual increase in frequency from the first half of the eighteenth century to the time of the BNC, the use of *in accordance with* increases rapidly from 1850 forward, with very little attested previous use of the noun *accordance* on its own (contrary to typically observed grammaticalization processes), while *in contrast to* and *in response to* occur as late as 1870 (91-2). The author accounts for this development by the process of *analogy* — as PNP constructions were grammaticalized, the PNP pattern suggested further like innovations (92-3).

Here we can appeal again to the comparatively extensive bilingualism that occurred among the lettered classes as a source of comparisons that contributed both to an environment where grammaticalized PNPs were frequent and as a source for loans and calques of such items as *par la vertu de / in virtue of* and *via / by way of*, among numerous others, noting the frequency of Latinic stems that occur in these PNP items. Thus the observation that these items seem to have developed rather suddenly by analogy offers the intriguing prospect of reapplying one of the central principles of the neogrammarians to the process of grammaticalization *via* calquing.

It is unlikely that a comprehensive discussion of an item undergoing grammaticalization can occur in the absence of some mention of its semantics, and here we see the references to *semantic bleaching* and *generalization* typical of the topic. The case of *in terms of*, *by virtue of*, and *in view of*, though, suggest the process of *semantic entropy*. Certainly, by the time such

items have reached the status of discourse markers, their erstwhile semantic conceptualization has not merely faded, but undergone randomization.

Some comments that touch on semantic development merit further attention. It is hard to see how *depend upon* is more explicit than *depend on* (102), although I do agree that it is bulkier and may result from an effort to elevate the style. Likewise it is difficult to aver that *depend upon* is “cognitively more complex” (ibid.). These notions recall the comment about “clarity and precision” above. If we accept on the one hand the conclusion that such items are “are stored in the memory as single chunks”, we cannot claim, on the other, that owing to their (compositional?) structure they somehow require greater processing complexity. A unit is a unit, regardless of the length of the string used to realize it. And as M. Pienemann points out, morphological units that do not cross phrasal boundaries fall lower on a hierarchy of processing complexity than those which cross phrasal boundaries (1998, *Language processing and second language development: Processability theory*. Amsterdam, Philadelphia: John Benjamins Publishing Company). If complex prepositions are conceptualized as “single syntactic units,” and the evidence marshaled here indicates that they are, and they participate as such in phrases, they offer no more “clarity and precision,” nor do they demand greater cognitive complexity, than any other substitutable “chunk.”

The discussion of 132 “low-frequency” PNPs (Table 8.1, 142) raises very apt questions about their state of grammaticalization, but also about the salience of *frequency* in grammaticalization (144, f). H concludes that the “relative frequency” of an item, among the various possibilities available, “has an influence on the cognitive representation this ... item receives” (164). The discussion of “textual frequency” and “conceptual frequency” (171, f) leads as well to the suggestion that the salience of the item chosen is a function of its stylistic and

sociolinguistic selection, among suitable possibilities in the grammar. The author makes the observation that in empirical linguistics the emphasis is on “the assessment of probability” of the likelihood of an item occurring (ibid.). Quantitative analysis may not offer the last word, but even in the case of low frequency items, such an approach certainly can make possible insights into authentic language use which are simply not available otherwise.

In a few spots glosses could be improved. *Mete and drynk* in an OED quotation from 1440 is glossed ‘meat and drink’ (89), where ‘food and beverages’ would be more accurate. These points do not diminish, though, the thorough, multi-faceted treatment of the development of English complex prepositions and the insights into the techniques and possibilities of computational, corpus-based studies that this work offers.

Robert D. Angus
California State University, Fullerton