*RESLA* 24 (2011), 73-94

# A CORPUS ANALYSIS OF PROTOTYPICAL CAUSATION IN WRITTEN SCIENTIFIC AND TECHNICAL ENGLISH

PEDRO A. FUERTES-OLIVERA*
*Universidad de Valladolid*

BEATRIZ PÉREZ CABELLO DE ALBA*
*UNED, Madrid*

ABSTRACT. *In an attempt to come to a better understanding of the concept of prototypicality, this article compares two different tools for explaining prototypes, viz. the theoretical constructs found in the literature on cognitive linguistics and claiming to describe cognitively salient models, and frequency of linguistic usage as evidenced in the UVaSTECorpus. In particular, this article focuses on the periphrastic causative constructions of the verbs* allow, cause, *and* enable, *and investigates two research questions: (i) are prototypes the most frequently corpus-attested items? (ii) Are there significant differences among these verbs, considering that* cause *is an implicative causative, whereas* allow *and* enable *are non-implicative causatives? Results are commented on making our case for indicating that the notion of prototypicality is fuzzy and must be interpreted in terms of language varieties. This paper follows Gilquin's work (2006) but explains the results obtained in terms of medium (written) and genre (scientific and technical English).*

KEY WORDS. *Prototypical causation, corpus linguistics, cognitive linguistics, written scientific and technical English, inclusive 'we', metaphor.*

RESUMEN. *En este artículo se analiza el concepto de prototipicalidad comparando la frecuencia de construcciones perifrásticas causativas en un corpus de ingles escrito científico y técnico con la propuesta teórica sobre prototipicalidad descrita por la Lingüística Cognitiva. Aunque se sigue el método y la caracterización llevada a cabo por Gilquin (2006), hemos adaptado nuestra propuesta a las características de los verbos analizados* –allow, cause, *y* enable– *con el fin de responder a dos preguntas concretas: (1) ¿podemos decir que los prototipos son los más frecuentes en el corpus estudiado?; (ii) ¿existe alguna diferencia significativa entre los tres verbos teniendo en cuenta que* cause *es un verbo causativo implicativo mientras que* allow *y* enable *no lo son? El estudio de los dos nos permite reformular el concepto de prototipicalidad y adaptarlo a las características del medio (inglés escrito) y el género (inglés científico y técnico).*

PALABRAS CLAVE. Construcción causativa perifrástica, lingüística del corpus, lingüística cognitiva, inglés escrito científico y técnico, 'we' inclusivo, metáfora.

## 1. Introduction

Most cognitive and functional linguists share some assumptions about the nature of language, and the way different language issues can be investigated. Gries (2006: 2), for example, mentions that within semantics and syntax much work was concerned with offering and elaborating new approaches to the characterization of both word meanings in general as well as word senses of polysemous words and constructions, and with identifying meaningful syntactic structures whilst characterizing and explaining their semantic and distributional properties. He goes on to indicate that in order to carry out the above-mentioned analyzes, cognitive linguists have mostly resorted to qualitative approaches, whereas functional linguists have preferred approaches based on performance data from natural settings.

Considering the existence of common assumptions, common methods of inquiry must also be expected. Hence, during the general upsurge of corpus-based and/or corpus-driven work, the text-based orientation of a large amount of work within Functional Linguistics "has also carried over to Cognitive Linguistics where, especially with reference to the notion of usage-based approaches, many studies are now also based on the analysis of naturally-occurring language in the form of electronic corpora". (Gries 2006: 4)

There are different research methods to locate linguistic forms in a corpus. Ädel and Reppen (2008: 2-4) enumerate some of them: the *one-to-one searching*; the *sampling*; the *sifting;* and the *frequency-based listing*. One-to-one searching "involves investigating a linguistic form through a search term that only yields relevant hits." (Ädel and Reppen 2008: 2). Gilquin (2006) has used this research method in her exploration of the relationship between prototypicality and corpus frequencies. Gries (2006: 11) classifies her approach as:

> (i) of intermediate granularity since she is mainly interested in the verbs as instantiations of the respective lemmas, but also takes into consideration variants such as active and passive, and (ii) as moderately quantitative, since most of her results are based on frequency data and their implications.

Gilquin (2006) analyzed 3,574 constructions with the main periphrastic causative verbs (*cause, get, have,* and *make*) in 10m words from the British National Corpus (BNC) (5m words of spoken English and 5m words of written English). Each of the constructions was examined in terms of the parameters of the models of prototypical causation discussed by Lakoff and Johnson (1980), Lakoff (1987) and Langacker (1991), who have described prototypical causative constructions according to the ordering and nature of the participants.

Gilquin bases her definition of the prototypical ordering of the participants in a causative construction on the principle of iconic sequencing which indicates a relationship between the ordering of the linguistic elements and the sequence of events as we mentally conceive them. This principle, for example, accounts for the oddity of a sentence such as *He poured himself a glass of wine and opened the bottle*, quoted by Gilquin (2006: 162) and taken from Ungerer and Schmid (1996: 251). In a word, according to the principle of iconic sequencing the prototypical ordering of the linguistic elements in a causative construction is CAUSER, CAUSEE, and PATIENT, if any.

Concerning the nature of the elements making up the causative construction, Gilquin analyzes the constructions in terms of the parameters associated with two prototypical causation models: the "billiard-ball model" and the "direct manipulation model". The former has been defined by Gilquin as:

> A single, specific, physical CAUSER transmits energy to a single, specific, physical CAUSEE, which can absorb the energy or transmit it further to a single, specific, physical PATIENT.
> E.g. *The rolling circle causes the central circle to rotate.*
> *The tree falling on it made the lorry lose its loading*
> (Gilquin 2006: 164).

The direct manipulation model contains a cluster of 12 interactional properties which Gilquin adapts and defines in the following way:

> A single, definite, human CAUSER manipulates a single, definite, human CAUSEE, distinct from the CAUSER, into producing a volitional and material EFFECT, which can affect, or not, a single, definite and distinct PATIENT
> E.g. *I'll **make** her go up there.*
> ***I got** John to repaint the wall.*
> (Gilquin 2006: 166).

Table 1 summarizes the parameters of the prototypical causation models used by Gilquin (2006: 171) in her exploration of the relationship between prototypicality and corpus frequencies.

| | Iconic Sequencing | Billiard-Ball Model | Direct Manipulation |
|---|---|---|---|
| CAUSER | 1st participant | – single<br>– specific<br>– physical | – single<br>– definite<br>– human |
| CAUSEE | 2nd participant | – single<br>– specific<br>– physical | – single<br>– define<br>– human<br>– distinct from CAUSER |
| (PATIENT) | (3rd participant) | (– single<br>– specific<br>– physical) | (– single<br>– definite<br>– human<br>– distinct from CAUSER and CAUSEE) |
| EFFECT | | – material | – material<br>– volitional |
| CLAUSE | | – affirmative | – affirmative |

Table 1. *Parameters of the prototypical causation model.*

Gilquin found that the theoretical models are not reflected in corpus data. On relaxing the parameters of prototypicality, however, Gilquin obtained results conforming more closely to the theoretically motivated expectations. She (2006: 176) introduced four modifications. The first modification concerns the structure of the construction considered. For example, an "infinitive structure such as *He got me to open the door* is more likely to reflect one of the models of prototypicality than a past participle structure like *He had the door opened*." This is true for the "model of iconic sequencing, but also, to a certain extent, for the other two models, since in the second sentence, the CAUSEE is left unmentioned and so cannot participate in the elaboration of the model."

The second modification advocates that instead of a strict definition of prototypicality requiring the presence of *all* (her emphasis) the prototypical features already commented on, a looser one should be accepted where the prototype possesses the greatest number of features (but not necessarily all of them).

The third modification refers to the introduction of a "weight" factor, which would have to be assigned to each parameter defining the prototype. Gries (2003), for example, uses corpus data and the notion of "cue validity" to identify the attributes that most strongly support the choice of the prepositional construction.

Finally, Gilquin claims that *medium* also seems to play a role in the establishment of prototypical causation, at least as far as data for the model of direct manipulation with *make* are concerned. For example, human CAUSERS and human CAUSEES are significantly more frequent in speech than in writing, and the "single CAUSEE" and "definite CAUSEE" parameters display a marginally significant difference. No parameters rate significantly higher in writing than in speech (Gilquin 2006: 177).

Following Gilquin's work, this article addresses the joint influence of *medium* and *genre* by analyzing the relationship between the parameters of prototypical causation summarized in Table 1 and corpus frequencies extracted from a corpus of written scientific and technical English (see Section 3 below). In particular, the paper addresses two research questions:

1. Are prototypes the most frequently corpus-attested items? Some linguists have elevated the link between corpus data and prototypicality to the status of principle. Schmid (2000: 39), for example, proposed the 'From-Corpus-to-Cognition Principle', according to which "frequency in text instantiates entrenchment in the cognitive system".
2. Are there significant differences among the different verbs studied? If so, the analysis will support the idea that prototypicality is a very fuzzy concept. In addition, it will demand a re-elaboration of the models, perhaps by explaining the results in terms of the joint influence of medium and genre.

Although we have copied Gilquin's parameters (Table 1) and have followed her method of inquiry, the selection of the verbs analyzed has been modified in order to adapt the analysis of prototypical periphrastic causation to the possible influence of *medium* (i.e., written English) and *genre* (i.e., *scientific and technical English*). (In section 3, we explain the rationale for the verbs chosen in this analysis).

## 2. Medium and Genre in prototypical periphrastic causatives

Causation refers to the set of all particular cause-and-effect relations. In very broad terms, causation is defined as a relationship that holds between events, properties, variables, or state of affairs. English grammars (Biber *et al*. 1999; Huddleston and Pullum 2002; Quirk *et al*. 1985), for example, indicate that causal relationships can be expressed by three causative expressions: a set of causative verbs (for example, *cause*, *make*); a set of causative names (e.g., *actor*, *author*); and a set of effective names (e.g., *consequence*, *creation*). Causative verbs are very adequate for highlighting the distinction between causation as a scientific notion or as a semantic one (Talmy 2000: 475). The former explains causation as an autonomous event (1), whereas the latter mentions the existence of the causer (2):

(1) Water poured from the tank. (Talmy 2000: 475)

(2) Gravitation caused water to pour from the tank. (Hollmann 2003: 10)

As already indicated, prototypical periphrastic causation – i.e., constructions such as *he makes me laugh* or *I have my watch repaired,* quoted in Gilquin (2006: 160) – have been described in the cognitive literature in terms of the ordering and nature of the participants. Consequently, prototypical models of periphrastic causation (see Table 1, above) have been taken for granted, although the sources of prototypical causation remain rather obscure and seem to be based on intuition and on an attested cognitive principle, viz. the primacy of the concrete over the abstract in neural representations (MacLennan 1998). However, empirical analyses of causative constructions have concluded that there is no correlation (or only a very slight one) between frequency and prototypicality, i.e., the occurrences of causative verbs in real texts do not support the existence of prototypical causative constructions naturally produced in accordance with the models of prototypical causation, which hinge on the ordering and nature of the participants (Gilquin 2006; Nordquist 2004; Roland and Jurafsky 2002).

The above analyses have focused on general language, without investigating the possible influence of *medium* and *genre* on the results obtained. Only Gilquin (2006: 177) introduced the idea that medium "seems to play a role in the establishment of prototypical causation, at least as far as the nature of the participants is concerned (iconic sequencing shows no such variation)." She mentioned that for the model of direct manipulation with *make* there is a tendency for speech to come closer to prototypicality than writing.

Since the mid 1990s, the concept of prototype has also been central to the framework for genre analysis proposed by scholars such as Swales (1990), McCarthy and Carter (1994), and Paltridge (1995). Paltridge (1995: 394), for instance, claims that "[i]f the categorization of individual language items and concepts is based on a system of relations between instances and their models with properties of the model being inherited by their instances, the same may be said for genres", and adds that prototypicality allows for the inclusion of umbrella cases which are correctly identified and assigned to particular categories, not because they "involve an exact match in terms of characteristics or properties. Rather, it draws upon the notion of 'sufficient similarity'" (Paltridge 1995: 396). It allows for inclusion on the basis of linguistic, pragmatic and perceptual aspects of communicative events. For example, 'research articles', review articles, 'letters to the Editor' (in "Nature", for example), etc., published in English, are all examples of research genres which should contain similar linguistic, pragmatic, and perceptual aspects of communicative events.

To sum up, the above review of the state of play of prototypical periphrastic causation has, to the best of our knowledge, revealed that the possible influence of *medium* and *genre* has not been explored in connection with prototypical periphrastic causation described in the cognitive literature. Such an analysis is carried out in this paper, which uses an *in-house specialized corpus* compiled to further our understanding of professional and academic English (Fuertes-Olivera and Rodrigues-Rodrigues 2010).

## 3. Material and method

### 3.1. *The University of Valladolid Corpus of Written Scientific and Technical English (UVaSTECorpus)*

The University of Valladolid Corpus of Written Scientific and Technical English (UVaSTECorpus) stands at our original target of 3,000,000 words of scientific and technical English. It was designed by Fuertes-Olivera and Rodrigues-Rodrigues, collected at the University of Valladolid, and enriched with information on text type, genre, author(s), geographical variety, subject matter, year of production, purpose of the text, and origin of the documents (Table 2).

| Sub-corpus | Genre | Number of words | % of sub-corpus | % of corpus |
|---|---|---|---|---|
| Sub-corpus: | Learned Informational: | 2,137,171 | 100 | 70.5 |
| 'Research' | Research Articles | 1,529,595 | 71.5 | 50.4 |
| | Review Articles | 607,576 | 28.5 | 20.1 |
| | Sub-total | 2,137,171 | 100 | 70.5 |
| Sub-corpus: | Informational (Reportage): | 727,876 | 81.3 | 24 |
| 'Official' | Report | 727,876 | 81.3 | 24 |
| | Instructional: | | | |
| | Sub-total | 167,078 | 18.7 | 5.5 |
| | | 894,954 | 100 | 29.5 |
| | TOTAL CORPUS | 3,032,125 | | 100 |

Table 2. *Documents in the UVaSTECorpus.*

From the standpoint that a corpus should be compiled with recurrence to non-probability sampling techniques involving judgment and convenience (Meyer 2002: 44), the Corpus was randomly selected from English collections of written documents comprising scientific and technical informational and instructional genres, suitable for investigating particular grammatical, lexical, lexico-grammatical, discourse or rhetorical features of written scientific and technical English.

The documents are the product of three types of users: native researchers; non-native researchers using English as a lingua franca; and official bodies in the UK and the USA producing different types of reports and/or instructional documents on 'science and technology'. Hence, the corpus is subdivided into two main sub-corpora of documents: informational learned 'research' texts (70.5% of the corpus) and informational reportage and instructional administrative documents (29.5% of the corpus).

The former comprise texts produced by researchers in different fields published in peer-reviewed international journals. This sub-corpus groups publications either as 'research' or as 'review', depending on the purpose of the paper. Where the paper presents new findings or methodologies, these were classified as "research". Where the paper summarizes the state of the art in a particular topic, it was grouped as "review". Considering the number of pages usually devoted in journals to the two categories, a 3:1 selection criterion was used for collecting them; this resulted in 71.5% of the sub-corpus for 'research articles' and 28.5% for 'review articles' (including 'PhD summaries', 'discussion notes', 'letters', etc.). All the texts were taken from leading journals indexed in the ISI Science Citation Index, covering the most important domains in science and technology. For example, in the genre 'research article' included in the sub-corpus 'research', the following domains and number of words per domain were compiled: (i) applied mathematics and statistics (237,911 words), (ii) physics, X-Rays and radiology (138,615 words), (iii) chemistry and food processing (174,921 words), (iv) engineering (154,784 words), (v) materials, composites, and minerals (180,574 words), (vi) environment (206,437 words), (vii) computing (147,399 words), (viii) biotechnology (137,221 words), and (ix) information systems (151,382 words) (Figure 1).

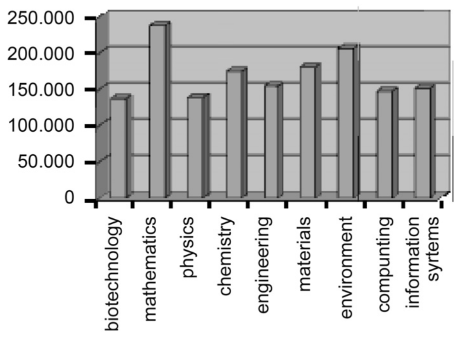


Figure 1. *Number of words per domain of the research articles included in the sub-corpus 'Research'.*

The latter consists of official documents produced by government bodies and professional societies. These were grouped into two main categories: informational reports and instructional documents. Informational reports are the work of national and/or international committees. For example, the National Science and Technology Committee is an American public body that produces reports on national and/or international policies on science and technology, or informs on research priorities and related issues.

Instructional documents are also produced by official bodies devoted to possible future courses of action regarding scientific and technical policies: for instance, guidelines and bulletins to universities, professional societies and research bodies commenting on future allocation of funds for Science and Technology programs. Considering that reports are more widely disseminated (for example, they are discussed in parliaments), we selected these according to a 5:1 criterion: reports comprise around 81% of the sub-corpus and instructional documents amount to a figure just below 19% of the sub-corpus (see Table 2 above).

Briefly, the corpus can be considered prototypical of the medium and genre here investigated (Paltridge 1995), and consequently adequate for exploring the relationship between prototypicality and the corpus frequencies of three verbs that are considered prototypical exemplars in written scientific and technical English: *allow*, *cause* and *enable*.

### 3.2. *Periphrastic causative constructions with* allow*,* cause*, and* enable

As regards the formal dimension, periphrastic causative constructions consist of a noun phrase, a causative verb, another noun phrase and an infinitive. The first noun phrase is the subject of the causative verb (e.g., *cause*), the second of the infinitive (see example (3) below). Functionally, these constructions describe a causative situation. Talmy (2000: 478-9) indicates that a causative situation can be analyzed in two sub-events: a causing and a caused event. He adds that the caused event must follow causally from the causing event. In example (3), the causing event consists of the *rolling circle* interacting in some way with the *RO circle*:

(3) They found that the rolling circle caused the RO circle to rotate.
(UVaSTECorpus: *Environmental Science and Technology: 40: 1726-1736)*

Some causative verbs entail the occurrence of the lower cause events, whereas some other verbs do not. The former are called "implicative" causatives whereas the latter are "non-implicative" causatives (Talmy 2000: 478-9). Implicative causatives indicate core causation by showing that the caused event is entailed by the causing event. The verb *cause* is an example of an implicative causative and is very appropriate for this study due to its association with the natural sciences (Lakoff 1987: 55) and as it is typically used to describe scientific experiments. Gilquin (2006: 173) reports a previous study (Gilquin 2004) confirming that *cause* is predominantly used in scientific and technical genres (over 50% of its occurrences).

*Allow* and *enable* are predicates of "enabling causation" (Talmy 2000: 478-9), which is force-dynamically different from core causation, indicating that the caused event is not entailed by the causing event. They are also suitable for this study as they are frequent in academic prose: Biber et al. (1999: 710-12) claim that these two verbs occur around 100 times per million words in academic prose.

Periphrastic causatives, in English and elsewhere, have been extensively studied (see Hollmann 2003 for a review). Large though the body of scholarly work on the causative is, various holes and shortcomings still exist. For example, do they behave in written scientific and technical English as hypothesized in the literature on cognitive linguistics, and do implicative and non-implicative causatives behave differently in written scientific and technical English? These questions will be addressed in this paper in connection with the two research questions mentioned in section 1.

### 3.3. *Method and procedure of the present study*

The concordancing software package *WordSmith Tools 4.0* (Smith 2004) was used to generate concordances of the lemmas *allow*, *cause* and *enable* present in the UVaSTECorpus (Table 3):

| Lemma | N | Relative Frequency |
|---|---|---|
| ALLOW | 676 | 222 |
| CAUSE | 453 | 149 |
| ENABLE | 624 | 205 |

Table 3. *Absolute frequency (N) and relative frequency (per 1,000,000 words) of the lemmas* cause, allow *and* enable *in the UVaSTECorpus*.

Subsequently, all the concordances of the three verbs were manually analyzed, the first selection being all the occurrences of the lemmas as verbs (Table 4), which were then studied in order to be assigned to the semantic domain 'causation' (Table 5). So that results were meaningful, we performed a careful corpus-based semantic analysis by differentiating between causative instantiations of the three verbs (example 4) and other meanings (example 5). Each concordance is considered an example of causative verbs if they indicate that some person or inanimate entity causes a new state of affairs.

| Verb | N | Relative Frequency |
|---|---|---|
| ALLOW | 656 | 216 |
| CAUSE | 293 | 96 |
| ENABLE | 456 | 150 |

Table 4. *Absolute frequency (N) and relative frequency (per 1,000,000 words) of the lemmas* allow, cause *and* enable *as verbs in the UVaSTECorpus*.

| Causative verb | N | Relative Frequency |
|---|---|---|
| ALLOW | 591 | 195 |
| CAUSE | 243 | 80 |
| ENABLE | 516 | 170 |

Table 5. *Absolute frequency (N) and relative frequency (per 1,000,000 words) of the lemmas* allow, cause *and* enable *as causative verbs in the UVaSTECorpus.*

(4) Improving our ability to assess potential vulnerability and resilience to future variations and changes in climate and environmental conditions could enable governments, businesses, and communities to reduce negative impacts and seize opportunities to benefit from changing conditions by adopting infrastructure, activities, and plans.
<UVaSTECorpus: *Official: Report by the Climate Change Science Program*: 173).

(5) The following goals and criteria developed by the NITRD Program are intended to enable agencies considering participation to assess whether their research and development activities fit the NITRD framework.
<UVaSTECorpus: *Official: Supplement to the President's Budget for Fiscal year* 2007: 7).

Finally, all the causative verbs of Table 5 were studied in terms of their syntactic forms and functions. Basically, we differentiated between non-periphrastic causative constructions and periphrastic causative constructions (Table 6), described by Gilquin (2006: 162-163) as an action chain where a "head", the CAUSER, transmits its energy to a second entity, the CAUSEE, which can consume the energy or transmit it further to a third entity, the PATIENT, which absorbs the energy and thus represents the "tail" of the action chain. Example:

(6) The excess variables euv and evw cause the objective function to increase.
(UVaSTECorpus: *Research: 40R*: 4).

| Periphrastic causatives | N | Relative Frequency |
|---|---|---|
| ALLOW | 356 | 117 |
| CAUSE | 56 | 18 |
| ENABLE | 242 | 80 |

Table 6. *Absolute frequency (N) and relative frequency (per 1,000,000 words) of the lemmas* allow, cause *and* enable *as periphrastic causatives in the UVaSTECorpus.*

Below we will comment on the analysis of the periphrastic causative constructions of Table 6. This implies both indicating the main results obtained and showing how they were obtained. Our analysis will be divided into two sub-sections, each referring to the two research questions formulated in section 1.

## 4. Results

### 4.1. *From-Corpus-to-Cognition-Principle*

The 'From-Corpus-to-Cognition-Principle' proposed by Schmid (2000) assumes a direct correlation between prototypicality and frequency. In this study this means that a significant number of the periphrastic causative constructions found in the UVaSTECorpus (Table 6) should agree with the models of prototypical causation described in the literature on cognitive semantics, and adapted by Gilquin (2006; see Table 1, above). To prove this assumption, we analyzed all the occurrences of the periphrastic causative constructions reported in Table 6 and recorded them appropriately if they fulfilled all the parameters of the three models summarized in Table 1. In other words, each occurrence was deemed 'prototypical' if it agreed with the obligatory parameters accorded to each model: 2 parameters for 'iconic sequencing' (i.e., 1st participant as CAUSER; 2nd participant as CAUSEE); 8 parameters for the 'billiard-ball model' (i.e., 'single', 'specific' and 'physical CAUSER; 'single', 'specific', 'physical CAUSEE, material EFFECT and affirmative CLAUSE); and 10 parameters for the 'direct-manipulation model' (i.e., 'single', 'specific', 'human' CAUSER; 'single', 'definite', 'human' CAUSEE distinct from CAUSER; 'material' and 'volitional EFFECT; affirmative CLAUSE). The results are given in Table 7: both the absolute frequency ('N') and the percentage (%) indicate the number of examples of periphrastic causative constructions fulfilling all the parameters assigned to each model. Before going on to a discussion of these results, let us underline a point that does not appear from the table, viz. the complete degree of overlap between iconic sequencing on the one hand and the models of billiard-ball causation and direct manipulation on the other. This overlap has also been commented on by Gilquin (2006: 170), who indicates that a causative construction displaying either the billiard-ball model or the direct-manipulation model always displays iconic sequencing if the sentence is in the active.

| | Iconic sequencing | | Billiard-Ball Model | | Direct Manipulation | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| ALLOW | 322 | 90.44 | 3 | 0.84 | 1 | 0.28 |
| CAUSE | 42 | 75 | 2 | 3.57 | 0 | 0.00 |
| ENABLE | 227 | 93.80 | 0 | 0.00 | 1 | 0.41 |
| TOTAL | 591 | 90.36 | 5 | 0.76 | 2 | 0.30 |

Table 7. *Models of prototypical causation in corpus data.*

Table 7 shows three interesting results. One of them will be discussed in section 4.2 in connection with differences and similarities between the three verbs. The other two refer to the quite clear difference between the ordering of the participants and the nature of the participants. They indicate a direct correlation between prototypicality and frequency in the case of the iconic sequencing model but no correlation at all regarding the other two prototypical causation models: billiard-ball and direct manipulation. Example:

(7) The effective droplet radius retrieved by TRMM VIRS does not exceed 14 µm in polluted clouds within Area 2, lending credence to the hypothesis that pollution causes effective cloud water droplet size to drop below this precipitation threshold.
<UVaSTECorpus: Official: *A Report by the Climate Change Science Program and the Subcommittee on Global Change Research*>

This difference, which is very significant, indicates that written scientific and technical English assumes the parameters of the iconic sequencing model proposed in the literature on cognitive linguistics.

Turning to the billiard-ball model of causation, it should be emphasized that the model is a very marginal organizing principle in naturally-occurring language (only 0.76%). While the verbs studied share a number of characteristics with this model, the main problems have to do with the semantic category of the CAUSER and CAUSEE and the material EFFECT: most of the CAUSERS and CAUSEES do not refer to physical entities, force or events, (examples 8 and 9) and many of the effects are not material (example 10). In (8), 'HEC capabilities' is not a physical CAUSER, nor 'service logic for multiple application' a physical CAUSEE in (9), nor does the verb 'choose' indicate material EFFECT in (10). Table 8 summarizes the findings indicating the number and percentage of occurrences ('N' and %, respectively) that do not satisfy the conditions summarized in Table 1. The percentage is calculated in relation to the total number of occurrences in Table 6.

| | Non-physical CAUSER | | Non-physical CAUSEE | | Non-material EFFECT | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| ALLOW | 249 | 69.94 | 315 | 88.44 | 154 | 43.25 |
| CAUSE | 41 | 73.21 | 39 | 69.64 | 17 | 30.35 |
| ENABLE | 231 | 95.45 | 236 | 97.52 | 127 | 52.47 |

Table 8. *Main problems with the parameters of the billiard-ball model.*

(8) HEC (High End Computing) capabilities enable researchers in academia, Federal laboratories, and industry to model and simulate complex processes in biology, chemistry, climate, and weather, environmental science, nanoscale science and technology, physics and other areas to address Federal agency mission needs.
<UVaSTECorpus: Offical: *Report by Subcommittee on Networking and Information technology*>.

(9) The IMS CC enables service logic for multiple applications to be executed in a single node. <UVaSTECorpus: Research: *IP Multimedia Subsystem Online Session: Charging Call Control*>.

(10) These options allow instructors to choose between alternative paths to the program. <UVaSTECorpus: Research: *The Geotechnical Virtual Laboratory*>.

The situation is almost identical for the direct manipulation model: only 2 occurrences (0.30% of incidences in the UVaSTECorpus) fulfill all the parameters of the model; the rest shows the following: inanimate CAUSER or CAUSEE or both (example 11), non-material (example 12), non-volitional (example 13) and non-material and non-volitional EFFECTS (example 14). In (11) the CAUSER and CAUSEE are non-human ('This' and 'UK', respectively). In (12) the verb 'determine' does not show a material EFFECT; in (13) the verb 'permeate' does not lead to a volitional EFFECT; in (14) the verb 'create' does not show a material and volitional EFFECT. Table 9 shows the number ('N') and percentage (%) of the occurrences which do not fulfill the required parameters summarized in Table 1. The percentage is calculated in relation to the total number of occurrences given in Table 6.

| | Non-human CAUSER | | Non-human CAUSEE | | Non-material EFFECT | | Non-volitional EFFECT | | Non-material/ non-volitional EFFECT | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % | N | % |
| ALLOW | 341 | 95.78 | 152 | 42.69 | 93 | 26.12 | 8 | 2.24 | 59 | 16.57 |
| CAUSE | 46 | 82.14 | 36 | 64.28 | 3 | 5.35 | 23 | 41.07 | 14 | 25.00 |
| ENABLE | 234 | 96.69 | 117 | 48.34 | 40 | 16.42 | 4 | 1.65 | 85 | 35.12 |

Table 9. *Main problems with the parameters of the direct manipulation model.*

(11) Finally, subject to satisfactory completion of Gateway reviews, NERC will receive a capital allocation of up to £23.1 million from the large facilities fund towards the cost of a new oceanographic research vessel. This will allow the UK to maintain its lead in internationally competitive programmes aimed at

understanding the role of oceans and shallow seas in the functioning of the earth system. <UVaSTECorpus: Official: 2003-04 to 2005-2006 Science Budget>.

(12) Controlled laboratory experiments were, therefore, conducted so that the influence of changes in soil water chemistry on DOC could be seen over changes in DOC driven by variations in temperature and water table. This allowed us to determine whether drought-induced changes in soil-water chemistry suppressed DOC concentrations in a more conclusive way than was possible from our earlier analysis. <UVaSTECorpus: Research: Suppression of Dissolved Organic carbon by Sulfate Induced Acidification during Simulation Droughts>.

(13) The BCR consisted of a 304 stainless steel cube casing enclosing 30 permeable, polyethylene, removable modules that allow water to permeate under the force of gravity through the walls into their cavities while keeping the biomass outside. <UVaSTECorpus: Research: Bioremediation of Groundwater Contaminated with Gasoline Hydrocarbons and Oxygenates Using a Membrane-Based Reactor>.

(14) For example, Masterman and Rogers (2002) developed a number of online activities that allowed children to create their own cognitive traces when learning about chronology using an interactive multimedia application. <UVaSTECorpus: Research: New Theoretical Approaches for Human-Computer Interaction>.

What this analysis indicates, then, is that the billiard-ball model and the direct manipulation model described in the cognitive literature are not frequently used in the corpus analyzed in this paper. As a whole, both models account for a very low proportion of the corpus data, and merit an explanation.

Our assumption is that the social environment in which science and technology is developing in the 21st Century may offer some clues regarding the reasons for not fulfilling certain parameters. Today's scientific and technical experiments can use virtual simulations, relying on the Internet and computers (they create virtual worlds). In addition, scientists receive press coverage (an ideal situation to promote the 'I' and 'we'), etc. In short, there are many more occasions where the physical environment has been replaced by a virtual one and where an impersonal situation is changed into a personal one, as a means of highlighting who is making the headlines regarding a particular scientific and/or technical experiment.

This new social environment has produced many transformations that might affect the parameters summarized in Table 1. Thus, it might be necessary to reformulate the parameters of the billiard ball model and the direct manipulation model to cope with this new social environment. In particular, we think that this reformulation has to make room for two modifications which can be explained syntactically, discursively and semantically. In Section 4.2, we elaborate on a possible reformulation of the parameters in connection with the semantic differences of the three verbs analyzed.

4.2. *Implicative versus non-implicative causatives*

In the functional-typological literature the term *causative* is sometimes used to refer to implicative causatives and non-implicative ones (Hollmann 2003: 11; Talmy 2000: 478-9). Functionally, an implicative causative indicates two things, whereas a non-implicative causative signifies only the first of them:

(i) the caused event would not occur if the causing event (CAUSER) did not occur;

(ii) the caused event does indeed occur;

The verbs analyzed in this paper are predicates of "implicative causation" (*cause*) and of "enabling causation" or "non-implicative causation" (*allow*, *enable*). Considering that it has been assumed that their semantics is different (Talmy 2000: 478-9), the results commented on in section 4.1 merit a reinterpretation which allows for the difference between them. This can be done by subjecting the results to two mathematical tests: a chi-square test and a binomial proportion test.

The chi-square test indicates the significance distribution for each model so that we can determine which model of causation is correlated with each verb. By performing a chi square test on the results of Table 7, two results are highlighted. First, the billiard-ball model and the direct manipulation model exhibit very low values and show no significance ($\chi^2 = 5.29$ and no value, respectively): the three verbs do not show any significant difference regarding the billiard-ball model and the direct manipulation model.

Second, the distribution of iconic sequencing is significant ($\chi^2$ = 15.45)[i], and shows that the verb *cause* contributes the most to rejecting the null hypothesis (Table 10, below). In other words, there is a direct relationship between each of the verbs and the fulfillment of the iconic sequencing parameters, thus indicating that variation is mostly representative in the verb *cause* (which contributes 10.05 to the chi-square test), whereas it is much lower in the verbs *enable* (contributing 4.07 to the chi-square test), and *allow* (with a contribution of 1.32 to the chi-square test). Therefore, regarding the model of iconic sequencing, the verb *cause* is different from the other two verbs. This does not only confirm Talmy's distinction between 'implicative causatives ' and 'non-implicative causatives', but also suggests distinctions between both groups of causative verbs, which may lead to a re-elaboration of the theoretical models to cope with the semantic distinctions between implicative and non-implicative causatives. Table 10 shows the results of the chi-square test for the model of iconic sequencing.

| | Allow | Cause | Enable | Total |
|---|---|---|---|---|
| Row 1 | 90* <br> 86.00** <br> 0.19*** | 75* <br> 86.00** <br> 1.41*** | 93* <br> 86.00** <br> 0.57*** | 258 <br> 86.00% |
| Row 2 | 10* <br> 14.00** <br> 1.14*** | 25* <br> 14.00** <br> 8.64*** | 7* <br> 14.00** <br> 3.50*** | 42 <br> 14.00% |
| Column Total | 100 <br> 33.33% | 100 <br> 33.33% | 100 <br> 33.33% | 300 <br> 100% |

Legend: * observed frequency. ** expected frequency. *** contribution to chi-square test.

Table 10. *Table of contingence for the model of iconic sequencing*.

By subjecting the results of the parameters 'non-physical CAUSEE' (Table 8) and 'non-human CAUSEE' (Table 9) to a binomial proportion test, we can discover whether these parameters contribute in the same proportion to confirming or rejecting the null hypothesis, which maintains that the three verbs should behave similarly, i.e., the parameter analyzed behaves similarly in each verb studied. Interestingly, this test shows that the proportion rate is totally different for the verbs *allow*, and *enable* ($p < 0.0001$), but exactly the same for *cause* ($p = 0.5465$). This means that the null hypothesis is completely rejected for the verbs *allow* and *enable*, (i.e., the proportions for 'non-physical' CAUSEE and 'non-human' CAUSEE in Tables 8 and 9 are significantly different for these two verbs) but confirmed for the verb *cause* (i.e., the proportion for 'non-physical' CAUSEE and 'non-human' CAUSEE in Tables 8 and 9 is significantly similar). This result offers two interesting conclusions. First, the behavior of the 'non-implicative causatives' is similar between themselves but very different from the implicative causative. Second, as a consequence, we may confirm not only that implicative and non-implicative causatives behave differently, but also that the idea of prototypical causation is so fuzzy that it cannot be assumed as it stands, and needs therefore some re-elaboration in order for it to adapt to the different language varieties. This fuzziness stems from the fact that verbs having the same formal properties and performing the same syntactic and semantic functions behave very differently.

Our hypothesis is that *inclusive we* and *metaphorical mappings* equating physical and/or human worlds to virtual worlds and simulations must be considered parameters of prototypicality in written scientific and technical English, thus affecting the conditions summarized in Table 1 for the billiard ball model and the direct manipulation model of prototypical causation.

*Inclusive we* as CAUSEE, for example, signals who the author(s) is, perhaps emphasizing the idea that different marketing strategies (for example, commenting in the media on the benefits of a particular scientific and/or technological breakthrough) are necessary to receive funds for science and technology. This trend for including *inclusive we* (examples 15 and 16) as CAUSEE partly explains the difference observed in Figures 2 and 3: whereas it is found that the condition 'human CAUSEE' is fulfilled in around 52% of the occurrences of the verb *enable* (Figure 3), the condition 'physical CAUSEE' is only fulfilled in less than 3% of the incidences of the same verb (Figure 2). If this assumption holds true, it could be hypothesized that passive and ergative constructions might be less frequent in written scientific and technical English than previously reported. Future research will confirm or reject this hypothesis.

(15) These include carbon sequestration, which might enable us to continue to burn fuel fossils, collecting the carbon dioxide and sequestering them safely: energy efficiency gains; hydrogen usage; nuclear, solar photovoltaics; and wave and tidal.
<UVaSTECorpus: Official: *The Ninth Zuckerman Lecture*>

(16) This allows us to set the lower bound to 1, i.e., zmin = 1 or lf = 1, in the above MS-FAP model. <UVaSTECorpus: Research : *40R 1: Models and solutions techniques for frequency assignments problems*>

In addition, the concept of *metaphor* as a discursive phenomenon grounded in metaphorical mappings can be used for explaining some of the problems reported in Tables 8 and 9 regarding the semantics of the participants. Considering that in today's scientific and technical experiments the role of CAUSER and/or CAUSEE may be played by governing bodies, virtual worlds and virtual simulations instead of physical and/or human ones, we may assume that very often all of the latter are metaphorically used instead of physical and/or human CAUSERS and CAUSEES. Under this assumption, the billiard-ball and the direct manipulation model will account for a much larger proportion of the corpus data, thus paving the way for our inferring that some of the metaphorical CAUSERS and CAUSEES participating in the periphrastic causative construction substitute for physical and human CAUSERS and CAUSEES, respectively. In other words, we assume that on some occasions the use of virtual CAUSERS and CAUSEES refer metaphorically to human or physical participants (Table 11). For example, the 'IM tab' in (17) signals the presence of a human participant.

(17) The tabs near the top of the screen represent actions that the user can take; for example, clicking the IM tab would cause the IM screen to pop up.
<UVaSTECorpus: Research: Friends Night Out: A working prototype of a blended lifestyle service enabled through IMS>.

| | Iconic sequencing | | Billiard-Ball Model | | Direct Manipulation | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| ALLOW | 322 | 90.44 | 17 | 4.77 | 4 | 1.12 |
| CAUSE | 42 | 75 | 5 | 8.92 | 0 | 0.00 |
| ENABLE | 227 | 93.80 | 7 | 2.89 | 3 | 1.23 |
| TOTAL | 591 | 90.36 | 27 | 4.5 | 7 | 1.2 |

Table 11. *Models of prototypical causation in corpus data. Calculations made after assuming the presence of 'metaphors' and 'inclusive we'.*

Although the results from Table 11 do not change the overall picture referred to in Table 7, the former brings to the fore results more in line with the hypothesis formulated in the cognitive literature, especially those that maintain that the billiard-ball model is characteristic of natural sciences (Lakoff 1987) and that the verb *cause* is the most salient verb occurring in this syntactic pattern (Gilquin 2004). For example, fulfillment of the parameters of the billiard-ball model with the verb *cause* increases 60%. In addition, Table 11 also shows that more parameters respect the billiard-ball model and the direct manipulation model, and particularly the former, whose fulfillment results increase almost 6 times.

The study also shows that the models of prototypical causation rely on such a strict definition of the prototypicality of the billiard-ball and direct manipulation models of causation (i.e. 19 obligatory parameters and 6 optional parameters), that it is almost impossible for these verbs to respect all of them, although they do in fact comply with many of them (Gilquin 2006 reached a similar conclusion). As an illustration, Figures (2) and (3) show how the verb *enable* fares with respect to the different obligatory parameters defining the billiard-ball model (Figure 2), and the model of direct manipulation (Figure 3), assuming that virtual worlds, simulations, and corporate persons refer metaphorically to physical and/or human participants, respectively.

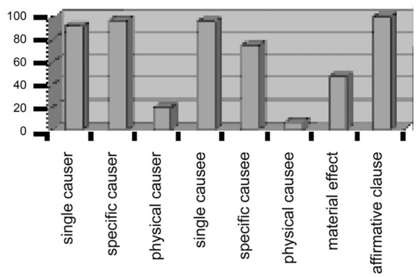


Figure 2. Enable *and the billiard-ball model of causation. Fulfillment of the parameters under the assumptions formulated in this section.*

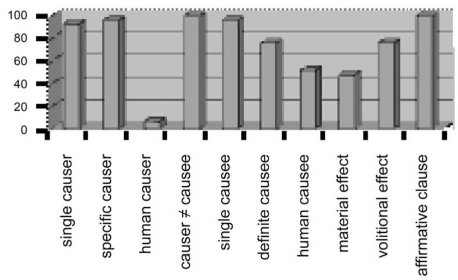


Figure 3. Enable *and the direct manipulation model of causation. Fulfillment of the parameters under the assumptions formulated in this section.*

Figures (2) and (3) show that only three parameters (physical CAUSER; physical CAUSEE in the billiard-ball model; and human CAUSER in the direct manipulation model) do not behave according to what has been hypothesized. These findings, then, together with those reported for the iconic sequencing model (Table 7, above), favor the interpretation that the notion of prototypical causation is fuzzy, that it differs in terms of the semantic nature of the causative, and should be explained by paying attention to language varieties. For example, the parameters that define the CAUSEE increase by 17.5% if we assume that governing bodies and virtual worlds are metaphorical representations of physical or human participants. In (18), 'agencies' is a collective noun that refers to a group of humans working for an agency; in (19), 'hospitals' is a plural noun that means something like 'professionals working in these facilities'.

(18) Frameworks and environments: Advances that enable agencies to more efficiently develop and certify high-quality software that is critical to Federal agency missions.
<UVaSTECorpus: Official: Supplement to the President's Budget for fiscal year 2007>.

(19) A cheaper and more effective Positron Emission Tomography (PET) camera will enable more hospitals to detect cancers and manage their treatment.
<UVaSTECorpus: Official: Together in research delivers science>.

## 5. Conclusion

In an attempt to come to a better understanding of the concept of prototypicality, this article compares two different tools for explaining prototypes, viz. the theoretical constructs found in the literature on cognitive linguistics and claiming to describe cognitively salient models, and frequency of linguistic usage as evidenced in the UVaSTE Corpus. In particular, this article focuses on the periphrastic causative constructions of the verbs *allow*, *cause*, and *enable*, which are common in written scientific and technical English. It analyzes the occurrences of these three verbs in terms of the parameters associated with iconic sequencing, billiard-ball causation and direct manipulation causation used by Gilquin in her analysis of prototypical causation in a balanced (written and spoken) corpus of general English.

The findings reveal that the frequency of periphrastic causative constructions in the corpus is lower (11.7%) than in the BNC (see Gilquin 2006). They also show that there is a complete degree of overlap between iconic sequencing and the billiard-ball and direct manipulation models of causation, and that the model of iconic sequencing accounts for about 90% of the causative constructions, a 100 per cent increase on Gilquin's BNC-based study that stood at 45% (2006: 172).

The findings also show that both the billiard-ball model and the direct manipulation model are very marginal organizing principles in naturally-occurring languages. They account for 0.76% and 0.30% of the occurrences. This finding, however, does not mean that causation is not widespread in scientific and technical English, but that the parameters of the three models are so strict that it is almost impossible to fulfill all of them. Hence, prototypicality is to be understood as a fuzzy phenomenon whose contours are subject to the presence of language varieties; the adaptation of the linguistic features to the pragmatic characteristics of the genre offers results more in line with the hypotheses. For instance, fulfillment of the conditions regarding the CAUSEE in the direct manipulation model for the verb *enable* increases 17.5%.

The study also shows that implicative causatives and non-implicative causatives behave differently in written scientific and technical English. For example, non-implicative causatives more frequently fulfill the conditions of the iconic sequencing model (between 15% and 18% more). In addition, the parameters 'physical' and 'human' of the CAUSEE behave similarly in the verb *cause* but differently in the other two verbs studied.

Finally, a caveat should be mentioned. This study has been carried out in a balanced written corpus of scientific and technical English. It follows that the findings are explained in terms of medium and genre, thus confirming our first assumption: the notion of prototypicality is so fuzzy that we should make room for adapting it to social conditions. Under this assumption this paper has re-elaborated some of the conditions of the billiard-ball model and the direct manipulation model. In particular, we have assumed that *inclusive we* and *metaphor* should be incorporated into a description of the parameters that explain prototypicality in terms of medium and genre. To conclude, today's scientific and technical developments have forced researchers to present themselves as 'actants', and virtual worlds are replacing real worlds when scientific experiments are performed.

## Acknowledgments

Thanks are due to Dr. Manuel de Prada and Dr. Ángel Sanz for their help in performing the two mathematical tests.

## NOTES

* Correspondence to: Pedro A. Fuertes Olivera. Escuela Universitaria de Estudios Empresariales. Paseo del Prado de la Magdalena s/n. 47005. Valladolid. E-mail: *pedro@tita.emp.uva.es*.

i *Chi-square test for iconic sequencing*

| *Chi-square* | *Degree of Freedom (DoF)* | *P-value* |
|---|---|---|
| *15.45* | *2* | *P<0.0004* |

## 6. REFERENCES

Ädel, A. and R. Reppen, eds. 2008. *Corpora and Discourse. The Challenges of Different Settings*. Amsterdam/Philadelphia: John Benjamins.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Fuertes-Olivera, Pedro A. and J. M. Rodrigues-Rodrigues. 2010. "Working with English specialized corpora: Lexical bundles in written scientific English". *A Mosaic of Corpus Linguistics. Selected Approaches*. Eds. A. Sánchez and M. Almela. Bern: Peter Lang 119-134.

Gilquin, G. 2004. *Corpus-based cognitive study of the main English causative verbs. A syntactic, semantic, lexical and stylistic approach*. PhD Dissertation. Centre for English Corpus Linguistics, Université Catholoque de Louvain.

Gilquin, G. 2006. "The place of prototypicality in corpus linguistics: Causation in the hot sea". *Corpora in Cognitive Linguistics. Corpora-based Approaches to Syntax and Lexis*. Eds. S. T. Gries and A. Stefanowitsch. Berlin/New York: Mouton de Gruyter. 159-191.

Gries, S. T. 2003. "Towards a corpus-based identification of prototypical instances of constructions". *Annual Review of Cognitive Linguistics* 1: 1-28.

Gries, S. T. 2006. "Introduction". *Corpora in Cognitive Linguistics. Corpora-based Approaches to Syntax and Lexis*. Eds. Stefan Th. Gries and A. Stefanowitsch. Berlin/New York: Mouton de Gruyter. 1-17.

Hollmann, W. B. 2003. *Synchrony and diachrony of English periphrastic causatives: A cognitive perspective*. PhD Dissertation. University of Manchester. [Internet document available at http://www.lancaster.ac.uk/staff/hollmann/WBH_PhD_causatives.pdf]

Huddleston, R. and G. K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago/London: The Chicago University Press.

Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. Chicago/London: Chicago University Press.

Langacker, R. W. 1991. *Foundations of Cognitive Grammar. Descriptive Applications*. Stanford, CA: Stanford University Press.

MacLennan, B. J. 1998. "Finding order in the world: The primacy of the concrete in neural representations and the role of invariance in substance reidentification". *Behavioral and Brain Sciences* 21 (1): 78-79.

McCarthy, M. and R. Carter. 1994. *Language as Discourse: Perspectives for Language Teaching*. London: Longman.

Meyer, C. F. 2002. *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge University Press.

Nordquist, D. 2004. "Comparing elicited data and corpora". *Language, Culture, and Mind*. Eds. M. Achard and S. Kemmer. Stanford, CA: CSLI Publications. 221-224.

Paltridge, B. 1995. "Working with genre: A pragmatic perspective". *Journal of Pragmatics,* 24: 393-406.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Roland, D. and D. Jurafsky. 2002. "Verb sense and verb subcategorization probabilities". *The Lexical basis of Sentence Processing: Formal, Computational, and Experimental Issues*. Eds. P. Merloand, S. Stevenson. Amsterdam/Philadelphia: John Benjamins. 325-346.

Schmid, H-J. 2000. *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin/New York: Mouton de Gruyter.

Scott, M. 2004. *Wordsmith Tools. Version 4*. Oxford: Oxford University Press.

Swales, J. 1990. *Genre Analysis*. Cambridge: Cambridge University Press.

Talmy, L. 2000. *Towards a Cognitive Semantics. Concept Structuring Systems*. Cambridge, MA/London: MIT Press.

Ungerer, F. and H. J. Schmid. 1996. *An Introduction to Corpus Linguistics*. London/New York: Longman.