

# Co-varying verbs and adjectives of *it*-extraposed constructions with *to*-infinitive clauses in academic discourse: a quantitative corpus-driven study

Jarosław Wiliński (Siedlce)

## ABSTRACT

This paper employs the background assumptions of usage-based Construction Grammar (Goldberg 1995, 2006, 2013), Frame Semantics (Fillmore 1982), and a quantitative corpus-driven method for investigating the reciprocal interaction between lexical items occurring in two different slots of a grammatical construction. The method, referred to as *co-varying collexeme analysis* (Stefanowitsch and Gries 2005; Stefanowitsch 2013; Hilpert 2014), is applied to the determination of strongly attracted and repelled pairs of adjectives and verbs occurring in the extraposition construction with *to*-infinitive clauses in American English. Using the data extracted from the academic sub-corpus of COCA, the author seeks to indicate that some pairs of adjectives and verbs co-occur significantly more frequently than expected in the *it is ADJ to V*-construction. Furthermore, the results of the analysis of the co-variation of collexemes in two different slots of the same construction seem to suggest that such strong correlations between these slots can be determined by frame-semantic knowledge and/or discourse-functional properties of the construction under study.

## KEYWORDS

Construction Grammar, Frame Semantics, extraposition, co-varying collexeme analysis, COCA

## DOI

<https://doi.org/10.14712/18059635.2019.1.1>

## 1. INTRODUCTION

Extraposition in English has received much attention over the last three decades (Quirk et al. 1985; Seppänen, Engström and Seppänen 1990; Seppänen 1999; Kaltenböck 2000, 2003, 2004; Kaatari 2010). Some research studies have compared extraposed constructions with clefts (Pérez-Guerra 1998; Calude 2008) as well as extraposition to right dislocation (McCawley 1988; Collins 1994), while others have focused on discourse functions of different extraposed structures (Mair 1990; Herriman 2000a,b; Hoey 2000; Hewings and Hewings 2002). Some researchers have also explored the occurrence of epistemic, deontic, or evaluative adjectives in a variety of extraposed constructions (Biber et al. 1999; Van Linden 2012), their clausal complementation (Mindt 2011), and various valency properties (Herbst et al. 2004).

Thus far, however, scant attention has been paid to the quantitative evaluation of adjectives and verbs in the extraposed construction complemented by *to*-clauses, the statistical validation of their occurrence in academic discourse, or the empirical confirmation of previous assumptions and speculations about their use. Hilpert's (2014)

case study of different pairs of adjectives and verbs occurring preferentially in the *it is ADJ to V*-construction and Wiliński's (2017) quantitative investigation of adjectives complemented by *to*-clauses in academic discourse are notable exceptions. Using the data retrieved from the BNC corpus, Hilpert established that there are certain pairs of verbs and adjectives closely related semantically that display a strong preference for the construction in question. On the basis of the data extracted from the academic sub-corpus of COCA, Wiliński in turn found that some adjectives are more strongly attracted to this construction than others, and that the occurrence of certain adjectives in this construction is more significant than their use in different types of extraposed constructions.

Given that Hilpert's study solely concerned the use of the *it is ADJ to V*-construction in British English and Wiliński's study was not specifically designed to capture interdependencies between adjectives and verbs, there is still a need for the quantitative determination of strongly attracted and repelled combinations of verbs and adjectives in the *it is ADJ to V*-construction in American English and for the qualitative analysis of their usage in this kind of extraposition, in view of the widespread occurrence of the construction in question in academic discourse. Thus, using data extracted from the academic section of the Corpus of Contemporary American English, the author seeks to determine pairs of verbs and adjectives strongly and loosely associated with the pattern under investigation.

The remainder of this paper is structured in the following fashion. Section 2 explains both the theory fundamental to the semantic explanation of pairs of verbs and adjectives occurring in the *it is ADJ to V*-construction, and the methodology underlying the quantitative analysis of these combinations. Section 3 describes the corpus, the data, and the tools. Section 4 outlines the statistical procedure employed in this study. Section 5 defines the construction under scrutiny and discusses its function and usage. Section 6 combines the findings of the quantitative analysis with a semantic description of adjectives and verbs and elucidates the contribution of various semantic frames to the constructional meaning. Section 7 assesses the findings and formulates some proposals for future research.

## 2. THEORY AND METHOD

This study rests on the theoretical foundation provided by the usage-based Construction Grammar (Goldberg 1995, 2006, 2013) and Frame Semantics (Fillmore 1982; Fillmore and Atkins 1992; Fillmore and Baker 2010). A constructional approach to grammar assumes that there is no strict division between grammar and lexicon. Grammar consists of constructions or symbolic units, pairings of a form and a meaning/function, i.e. conventionalized associations of a phonological structure and a semantic/conceptual structure. Crucially for the current study, the notion of construction is not restricted to morphemes or words, but encompasses more complex and schematic constructions such as partially filled structures, lexically unspecified patterns, argument structure constructions, extraposed constructions, idioms, etc. This theory places special emphasis on actual frequencies of usage or occurrence and hence it is

explicitly usage-based (see Bybee 2010, 2013) in the sense that exposure to, or use of, constructions is deemed to influence the linguistic system of speakers and hearers, while sufficient frequency is a necessary condition for the entrenchment (Langacker 1987) and the achievement of construction status of a linguistic expression.

Frame Semantics is a theory of linguistic meaning formulated by the American linguist Charles Fillmore that aims to explain the meaning of lexical units and grammatical constructions in terms of frames or prototypical scenes relating to structured — but to some extent individually and culturally varying — background knowledge (Fontenelle 2003). For example, the word *sell* cannot be understood without access to all the essential knowledge associated with the COMMERCIAL TRANSACTION frame (cf. Petruck 1996), i.e. the situation of commercial transfer involving, among other things, a seller, a buyer, goods, money, and the relations between them. Thus, words and constructions activate, or evoke, frames of encyclopedic knowledge providing the background and motivation for their meanings.

Frame Semantics has been put into practice in the Berkeley FrameNet project (Fillmore, Johnson and Petruck 2003; Fillmore and Baker 2010). The primary aim of FrameNet is to systematically describe syntactic and semantic valency patterns of lexical units based on extensive corpus annotation. In this respect, the syntactic properties of lexical units in corpora are systematically aligned with the semantic frames evoked by the words. The description of each frame in the FrameNet database includes the following components: the name of the frame, a definition of the situation the frame is supposed to represent, the set of core and non-core frame elements (semantic roles) associated with the frame, and the corresponding word senses (lexical units) that activate the frame.

Practically all semantic frames and their modified definitions discussed in this study are taken from the Berkeley FrameNet project: IMPORTANCE, DIFFICULTY, MENTAL STIMULATION, STATEMENT, BECOMING AWARE, REMEMBERING INFORMATION, GRASP, COGITATION, AWARENESS, COMING TO BELIEVE, SEPARATING, MENTAL PROPERTY ATTRIBUTION, EXPECTATION, FAIRNESS EVALUATION, USEFULNESS EVALUATION, CORRECTNESS EVALUATION, FREQUENCY EVALUATION, LIKELIHOOD, and PREDICTION. The remaining semantic frames along with their descriptions, implemented in the description of semantic properties of verbs and adjectives, are created by the author himself: in other words, RISK EVALUATION and REALISM EVALUATION.

The methodology of quantitative corpus linguistics is applied in this study. The method called *co-varying collexeme analysis* (Stefanowitsch and Gries 2005; Stefanowitsch 2013; Hilpert 2014) is aimed at determining combinations of elements that occur more often than would be expected by chance in the *it is ADJ to V*-construction, i.e. identifying pairs of adjectives and verbs that are significantly attracted to, or biased towards, the investigated pattern in the academic section of COCA through a statistical evaluation of the observed frequencies of the lexemes in question in relation to the overall frequency of the construction in the corpus. The output is a ranking list of the so-called *co-varying collexemes*, i.e. of those pairs of lexemes that exhibit a stronger preference for the investigated construction than others. Although this technique is quantitative, the results of this analysis are evaluated qualitatively and subjectively. For example, the meanings of the lexemes that are strongly associated

with the construction can be interpreted with respect to the semantic frames to which they are relativised.

### 3. CORPUS, DATA, AND TOOLS

The data were gathered from the downloadable version of the academic part of the Corpus of Contemporary American English (COCA), i.e. the full-text data corpus purchased from Mark Davies. This academic part contains approximately 81 million words coming from nearly 100 different peer-reviewed journals. These encompass a wide range of academic disciplines: e.g. a certain percentage from philosophy, psychology, religion, world history, education, and technology.

The observed frequencies were retrieved from the corpus by means of MonoConc Pro, a concordance program. This tool was used to search through the corpus for all the occurrences of adjectives and verbs in the construction under study. Each concordance line was manually inspected to determine the frequencies of all pairs of adjectives and verbs co-occurring in the relevant pattern. Then, all these frequencies required for the computation of the mutual association between combinations of elements in the *it is ADJ to V*-construction were entered in a 2-by-2 table and submitted to the Fisher exact test. The p-value provided by this test was used to gauge the strength of association, i.e. the degree of attraction to or repulsion from the *it is ADJ to V*-construction: the smaller the p-value, the higher the probability that the observed distribution is not due to chance and the higher the strength of the association between two slots of the same construction (cf. Schmid and Küchenhoff 2013). This calculation of statistical significance was performed by means of an online Fisher's exact test calculator for two-by-two contingency tables. The rest of the values and expected frequencies were calculated by means of Microsoft Excel spreadsheets. The resulting frequency lists then provided the input to the co-varying collexeme analysis.

Generally, p-values are so low that their significance lies only in the number of decimal places. These values are precisely expressed in numbers of the type "1.31E-10" (see for example the result provided for the combination *difficult to ascertain* in Table 2 below), which reads "1.31 times 10 to the power of minus 10", i.e. 0.00000000131. To simplify things, a log transformation of these p-values is frequently given in some studies (e.g. Hilpert 2014; Perek 2014) which apply *Coll.analysis 3*, an R script written by Stefan Gries. This script uses a log transformation to the p-values yielded by the Fisher exact test, and turns the sign into a plus if the association is one of attraction (i.e. the actual frequency of occurrence of a verb and an adjective exceeds the expected frequency) and into a minus in the case of repulsion (i.e. the actual frequency of the combination is lower than the expected frequency). This provides a more readable value than p-values, expressed in powers of ten (cf. Hilpert 2014: 402; Perek 2014: 69). The values of the strength of association between adjectives and verbs larger than 1.301 mean that particular combinations are significantly attracted to the construction, whereas the values lower than -1.301 mean that combinations are significantly repelled by the construction (cf. Hilpert 2014: 402).

It is worth noting that the use of the p-value as a significance measure has come under heavy criticism from Schmid and Küchenhoff in their last publication (2013: 539). This criticism centres on the issue of whether or not the Fisher exact p-value incorporates an effect size. Gries (2015: 520) argues that although “p-values are not effect sizes, p-values by their very nature reflect a combination of different things including the size of the sample(s), the variability of the sample(s), and the effect size.” The rationale for the use of the Fisher exact as a significance test is that, in comparison to other statistical tests, this measure can be used to assess the interaction among variables when data is very unevenly distributed and/or infrequent (cf. e.g. Stefanowitsch and Gries 2003: 9; Gries and Stefanowitsch 2004a: 101; see also Gries 2012 and Gries 2015: 508 for further arguments).

#### 4. STATISTICAL PROCEDURE

The procedure followed in this study consisted of four stages. This procedure can be illustrated with the aid of the adjective *important* in the adjective slot along with the verb *note* in the verb slot of the *it is ADJ to V*-construction.

	<b>Important in adjective slot of the <i>it is ADJ to V</i>-construction</b>	<b>All other adjectives in adjective slot of the <i>it is ADJ to V</i>-construction</b>	<b>Total</b>
<b>Note in verb slot of the <i>it is ADJ to V</i>-construction</b>	a: Frequency of adjective ( <i>important</i> ) and verb ( <i>note</i> ) in the <i>it is ADJ to V</i> -construction 318 (97.91)	b: Frequency of all other adjectives and verb ( <i>note</i> ) in the <i>it is ADJ to V</i> -construction 168	x: Total frequency of verb ( <i>note</i> ) in the <i>it is ADJ to V</i> -construction 486
<b>Other verbs in verb slot of the <i>it is ADJ to V</i>-construction</b>	c: Frequency of adjective ( <i>important</i> ) and other verbs in the <i>it is ADJ to V</i> -construction 1628	d: Frequency of all other adjectives and other verbs in the <i>it is ADJ to V</i> -construction 7545	y: Total frequency of all other verbs in the <i>it is ADJ to V</i> -construction 9173
<b>Total</b>	e: Total frequency of adjective ( <i>important</i> ) in the <i>it is ADJ to V</i> -construction 1946	f: Total frequency of all other adjectives the <i>it is ADJ to V</i> -construction 7713	z: Total frequency of the <i>it is ADJ to V</i> -construction 9659

TABLE 1. Co-occurrence table for a co-varying-collexeme analysis

At the initial stage of this procedure, the observed frequencies were calculated on the basis of the data extracted from the corpus. First, all occurrences of the construction under study were identified from the corpus: 9659. Second, the frequency of the adjective *important* in the adjective slot was determined: 1946. Third, the frequency of the verb *note* in the verb slot was calculated: 486. Finally, the frequency of the adject-

tive *important* and the verb *note* appearing together was counted: 318. These four values were derived from the corpus directly, while the remaining ones resulted from subtraction in Table 1. For example, in order to calculate the frequency of all other adjectives and the verb *note* in the *it is ADJ to V*-construction, the frequency of the adjective *important* and the verb *note* in the same construction (318) was subtracted from the total frequency of the verb *note* in the *it is ADJ to V*-construction (486), giving the result (168). Table 1 above displays the actual frequencies necessary to carry out a co-varying collexeme analysis of the adjective *important* and the verb *note* in the construction under scrutiny (for expository purposes, it also gives the expected frequencies for the adjective *important* and the verb *note* in parentheses).

At the second stage, these observed values were used to calculate the expected frequency of the adjective (*important*) and the verb (*note*) in the investigated construction. This calculation was performed in Microsoft Excel in the following fashion. For this combination of elements, its column total was multiplied by its row total, and the result was divided by the overall table total. For example, for the top cell containing the figure 318, its column total (1946) was multiplied by the row total (486), giving the figure (945756). Then this figure was divided by the table total (9659), yielding the result (97.91). If the observed frequency of the adjective (*important*) and the verb (*note*) together in the construction is significantly higher or lower than expected, the relation between this pair of lexemes is one of attraction or repulsion respectively (the adjective *important* and the verb *note* are then assumed to be significantly attracted or repelled *collexemes* of this construction).

At the third stage, the degree of attraction, or the association strength, between the adjective (*important*) and the verb *note* was estimated by means of the Fisher exact test. To this end, the following four frequencies were employed: the frequency of the adjective (*important*) and the verb (*note*) in the *it is ADJ to V*-construction, the frequency of all other adjectives and the verb (*note*) in this construction, the frequency of the adjective (*important*) and other verbs in the pattern in question, and the frequency of all other adjectives and other verbs in the investigated construction. The p-value resulting from the computation of the Fisher exact test for this combination is exceptionally small: 5.21E-111. This means that the adjective (*important*) and the verb *note* share strong mutual attraction in the investigated construction, but this can only be determined by comparing the observed frequencies of the adjective (*important*) and the verb *note* with the expected ones. As this comparison indicates, the adjective (*important*) occurs more frequently than expected with the verb *note* in the construction. In other words, *important* and *note* are highly significant, very strongly attracted collexemes of the construction.

This procedure was employed for all pairs of adjectives and verbs in the investigated construction. At the next stage, the results were arranged, first, according to the direction of association (attracted or repelled), and second, according to their association strength. Finally, the data were interpreted qualitatively and subjectively. More specifically, the results of the quantitative analysis were integrated with a semantic description of the most strongly attracted pairs of adjectives and verbs, and the contribution of frame-semantic knowledge to the meaning of the pattern under study was explained.

## 5. *IT*-EXTRAPOSITION WITH *TO*-INFINITIVE CLAUSES

*It*-extraposition with *to*-infinitive clauses refers to a syntactic process by which a *to*-infinitive clause is shifted (extraposed) from its initial position (i.e. its subject position) to the end of a sentence. This usually involves the use of the dummy pronoun *it* as a subject. Classic examples of this type of extraposition are given in (1), with the non-extraposed counterparts being provided in (2):

- (1) *it*-extraposition with *to*-infinitive clauses
- a. It is impossible *to buy a flat here*
  - b. It is difficult *to find a good wife*
  - c. It is important *to be able to speak English*
- (2) non-extraposition
- a. *To buy a flat here* is impossible
  - b. *To find a good wife* is difficult
  - c. *To be able to speak English* is important

Given that *it*-extraposition is similar to non-extraposition in respect to its structure and logico-semantic properties, many researchers from the formal school of generative linguistics (e.g. Rosenbaum 1967; Huddleston 1971; Emonds 1972) treated the sentences in (1) as syntactic derivations or transformations of the sentences in (2). Recently, however, a body of empirical evidence (e.g. Francis 1993; Biber et al. 1999; Kaltenböck 2000) obtained from naturally occurring data in corpora has suggested that examples of non-extraposition are extremely rare in corpora. For example, Kaltenböck's research (2000: 158) revealed that instances of *it*-extraposition considerably outnumber those of its non-extraposed counterpart with a ratio of 1:7.8 in the British section of the International Corpus of English. Biber et al. (1999: 676, 724), in turn, noticed that occurrences of *that*-clauses and *to*-clauses in non-extraposition are extremely infrequent in spoken language, and that *it*-extraposition is much more preferred. In addition, as noted by Quirk et al. (1985: 964–965), some extraposed examples do not allow for reversion to non-extraposed constructions, in either writing or speech. Hence, it is debatable whether sentences such as those in (1) are indeed transformations of the sentences in (2), and it seems to be more acceptable to treat *It is impossible to buy a flat here* as a construction (a pairing of form and meaning/function) in its own right, and to examine it accordingly, rather than consider it as a version of something that is used extremely infrequently.

In this study, therefore, examples such as the ones in (1) are assumed to be a type of the English *it*-extraposition construction, a partially lexically-filled pattern consisting of three fixed lexical items (*it is* [...] *to* [...]) and two flexible slots that can be filled by adjectives and verbs. This pattern can be represented structurally and schematically as [*it is* ADJ *to*-infinitive clause], where a dummy subject *it* is followed by the third person singular form of the verb *be*, a predicative adjective, and a *to*-infinitive clause. The use of this construction can be exemplified by the following sentences retrieved from the corpus:

- (3) *It is important to note that this assessment is criterion based and leveled by grade*
- (4) *It is hard to imagine that nearly 300,000 men died or were wounded here almost a century ago*
- (5) *It is reasonable to expect a significant reacceleration of inflation in the near future*

Regarding discourse-functional properties of extraposed constructions with *to*-infinitive clauses, much research (Huddleston 1984; Collins 1994; Gómez-González 1997; Herriman 2000a; Hoey 2000; Hewings and Hewings 2002; Rowley-Jolivet and Carter-Thomas 2005) has shown that *it*-extraposition in examples such as those in (3), (4) and (5) can serve two crucial functions. First, it is commonly used in both speech and writing to avoid long and heavy subject clauses because they sound awkward, thereby placing them at the end of the sentence, in accordance with the principles of end-weight and end-focus. This function allows speakers to convey new pieces of information in a way that is easier to process (cf. Huddleston 1984: 453; Quirk et al. 1985: 863; Erdmann 1990: 137–8; Collins 1994: 15–16). A slightly different view is expressed by Mair (1990: 39), Miller (2001), and Kaltenböck (2005), who found that extraposed clauses convey not only new but also given information. Second, *it*-extraposition allows a speaker/writer to express subjective opinions about some state-of-affairs by presenting them as if they were generally accepted views rather than his/her personal judgement, hence introducing evaluative comments at the beginning of a sentence (cf. Herriman 2000b: 211; Gómez-González 2001: 272; Rowley-Jolivet and Carter-Thomas 2005: 51; Kaltenböck 2005: 137).

Although discourse-functional and structural properties of various types of extraposed constructions have received systematic treatment in the literature, the role of the adjectives and verbs in extraposed patterns with *to*-infinitive clauses has largely been ignored and neglected. Hence, research into interdependences between adjectives and verbs in this kind of *it*-extraposition deserves more attention. The rationale for undertaking such an empirical study is that the meanings of adjectives and verbs enormously influence the constructional meaning. For example, the combination of elements (*important to note*, *hard to imagine*, *reasonable to expect*) in (3), (4), and (5) contribute substantially to the understanding of the illustrative sentences by assigning different meanings to the constructions under scrutiny. In these cases, adjectives denoting importance, difficulty, or a specific mental property co-occur with verbs that can be used to introduce a statement, to denote awareness or knowledge about a fact, or to express the belief that some phenomenon will take place in the future. Thus, the quantitative investigation of such pairs and their semantic description with respect to the semantic frames they activate may enable us to find subtle distributional differences in their use and understand their role in the investigated pattern, as well as to broaden our knowledge and understanding of the meaning and function of the construction.

Given the semantic and discourse-functional properties of the examples mentioned above and the results of the study conducted by Hilpert (2014), it is possible to predict roughly what adjectives and verbs are likely to occur in both slots of this construction in academic discourse. The adjectival slot should prefer adjectives expressing the speaker's or writer's evaluative judgement, whereas the verbal slot should

prefer verbs denoting cognitive processes, introducing a statement, and/or conveying new facts and states of affairs. These predictions will be tested below.

It is important, however, to note that even the detailed description of the construction's semantics does not allow us to predict whether adjectival and verbal slots in this pattern are related semantically and in what way. It follows from the principle of semantic compatibility (Stefanowitsch and Gries 2005: 11) that co-occurrences of adjectives and verbs are expected to be semantically coherent, but it does not specify what kind of semantic coherence we could expect for the construction under investigation. In this study, it is assumed that the semantic coherence between two different slots of this construction can be determined by frame-semantic knowledge, i.e. a relationship between semantic frames evoked by adjectives and verbs.

## 6. FINDINGS AND DISCUSSION

The concordancer extracted 9659 occurrences of the *it is ADJ to V*-construction containing 3956 different combinations of adjectives and verbs, out of which 2713 occurred only once in the investigated construction. Because of the limitation of space, however, this section will only interpret the findings for the 30 most strongly attracted and repelled co-varying collexemes of this pattern. Table 2 provides the results of a co-varying collexeme analysis ( $P_{\text{Fisher exact}}$ ) for the 30 most strongly attracted combinations of adjectives and verbs. It also displays the observed and the expected frequencies for each pair of lexical items occurring in two different slots of the investigated construction. The figures (a, e, x, z) were derived from the corpus directly, while the remaining figures (b, c, d, f, y) result from addition and subtraction.

The results support the prediction that the semantic coherence between two different slots of the construction under study is based on a relationship between two frames, i.e. on frame-semantic knowledge. Furthermore, the specific suggestions concerning the meaning of this construction are also confirmed. For this construction, we find that combinations of lexemes evoking the **IMPORTANCE** frame and **DIFFICULTY** frame constitute the bulk of the most strongly associated pairs of co-varying collexemes in the ranking list. The former frame is evoked by combinations such as *important to note*, *important to recognize*, *important to remember*, *important to understand*, *important to keep in mind*, *important to acknowledge*, and *important to consider* in ranks 1, 9, 10, 15, 19, 20 and 23. The p-values resulting from the calculation of the Fisher exact test for these collexeme pairs are exceptionally small: 5.21E-111, 2.11E-36, 2.43E-34, 8.04E-16, 4.77E-13, 7.16E-13, 7.16E-13, respectively. A comparison of the observed and the expected frequencies of these pairs of lexical items occurring in two different slots indicates that these pairs occur more frequently than expected in the construction. In other words, they are highly significant and very strongly attracted to each other in this construction.

Note also that *important to note* is the most strongly associated co-varying-collexeme pair in the construction, since its p-value is exceptionally small (5.21E-111) and the observed frequency is much higher than the expected one. This combination is

rank	pairs of collexemes	a	x	e	z	b	c	y	f	d	(a)	$P_{\text{Fisher exact}}$
1.	important to note	318	486	1946	9659	168	1628	9173	7713	7545	97.91	5.21E-111
2.	interesting to note	99	486	164	9659	387	65	9173	9495	9108	8.25	3.15E-88
3.	hard to imagine	90	202	424	9659	112	334	9457	9235	9123	8.87	7.82E-70
4.	fair to say	47	222	72	9659	175	25	9437	9587	9412	1.655	5.54E-61
5.	reasonable to assume	53	148	228	9659	95	175	9511	9431	9336	3.49	9.83E-50
6.	reasonable to expect	47	105	228	9659	58	181	9554	9431	9373	2.479	1.39E-49
7.	safe to say	45	222	101	9659	177	56	9437	9558	9381	2.321	6.75E-48
8.	unrealistic to expect	27	105	43	9659	78	16	9554	9616	9538	0.467	5.91E-44
9.	important to recognize	98	143	1946	9659	45	1848	9516	7713	7668	28.81	2.11E-36
10.	important to remember	89	127	1946	9659	38	1857	9532	7713	7675	25.59	2.43E-34
11.	easy to see	59	277	330	9659	218	271	9382	9329	9111	9.46	3.45E-31
12.	hard to see	57	277	424	9659	220	367	9382	9235	9015	12.16	1.60E-23
13.	safe to assume	21	148	101	9659	127	80	9511	9558	9431	1.548	1.68E-18
14.	unreasonable to expect	11	105	20	9659	94	9	9554	9639	9545	0.217	2.27E-17
15.	important to understand	106	252	1946	9659	146	1840	9407	7713	7567	50.77	8.04E-16
16.	reasonable to conclude	18	71	228	9659	53	210	9588	9431	9378	1.676	2.71E-14
17.	useful to recall	13	32	204	9659	19	191	9627	9455	9436	0.676	2.77E-14
18.	difficult to determine	54	170	1076	9659	116	1022	9489	8583	8467	18.94	2.50E-13

followed by *interesting to note* ( $P_{\text{Fisher exact}} = 3.15\text{E-}88$ ), instantiating a relationship between the MENTAL STIMULATION frame and the STATEMENT frame. In the sentence *It is interesting [to note that very often bullies exhibit rudimentary organizational skills]* <sub>STIMULUS</sub> for example, the speaker (an experiencer) not only introduces the statement by using the verb *note*, but also expresses his/her evaluative judgement about the stimulus, i.e. an event or state of affairs bringing about a particular emotion or experience in the experiencer under certain circumstances.

rank	pairs of collexemes	a	x	e	z	b	c	y	f	d	(a)	$P_{\text{Fisher exact}}$
19.	important to keep in mind	48	87	1946	9659	39	1898	9572	7713	7674	17.53	4.77E-13
20.	important to acknowledge	31	44	1946	9659	13	1915	9615	7713	7700	8.865	7.16E-13
21.	true to say	7	222	7	9659	215	0	9437	9652	9437	0.161	3.09E-12
22.	hard to believe	19	62	424	9659	43	405	9597	9235	9192	2.722	8.08E-12
23.	important to consider	88	224	1946	9659	136	1858	9435	7713	7577	45.13	7.16E-13
24.	easy to forget	9	13	330	9659	4	321	9646	9329	9325	0.444	3.61E-11
25.	rare to find	7	141	11	9659	134	4	9518	9648	9514	0.161	3.82E-11
26.	difficult to ascertain	15	21	1076	9659	6	1061	9638	8583	8577	2.339	1.31E-10
27.	difficult to imagine	55	202	1076	9659	147	1021	9457	8583	8436	22.50	1.35E-10
28.	difficult to separate	19	34	1076	9659	15	1057	9625	8583	8568	3.788	2.41E-10
29.	difficult to know	49	177	1076	9659	128	1027	9482	8583	8455	19.72	7.49E-10
30.	impossible to predict	15	44	483	9659	29	468	9615	9176	9147	2.2	1.48E-09

**TABLE 2.** The results of co-varying collexeme analysis for the 30 most strongly attracted pairs of adjectives and verbs

**Note!**

**a** = Observed frequency of adjective (e.g. *important*) and verb (e.g. *note*) in the construction; **b** = Frequency of all other adjectives and verb (e.g. *note*) in this construction; **c** = Observed frequency of adjective (e.g. *important*) and other verbs in the construction; **d** = Frequency of all other adjectives and other verbs in the construction; **e** = Total frequency of adjective (e.g. *important*) in the construction; **f** = Total frequency of all other adjectives in the construction; **x** = Total frequency of verb (e.g. *note*) in the construction; **y** = Total frequency of other verbs in the construction; **z** = Total frequency of the construction; **(a)** = Expected frequency of adjective (e.g. *important*) and verb (e.g. *note*) in the construction;  $P_{\text{Fisher exact}}$  = index of co-varying collostructional strength

The interdependence between the adjective *important* and verbs such as *note*, *recognize*, *remember*, *understand*, *keep in mind*, *acknowledge*, and *consider* in academic discourse is evidence of a significant correlation between the IMPORTANCE frame and several other frames evoked by these verbs, e.g. STATEMENT (*note*, *acknowledge*), BECOMING AWARE (*recognize*), REMEMBERING INFORMATION (*remember*, *keep in mind*), GRASP (*understand*), and COGITATION (*consider*). The IMPORTANCE frame refers to the speaker's evaluation of the importance of a particular undertaking, which can be an

activity, a state, or a process, as in the sentence: *It is important [to remember how politics and economics color and distort honest religious engagement]* <sub>UNDERTAKING</sub>. The STATEMENT frame contains the verbs *note* and *acknowledge* communicating the act of a speaker to address a message to some addressee, as in *It is important to note [that this assessment is criterion based and leveled by grade]* <sub>MESSAGE</sub>. The BECOMING AWARE frame describes a process by which a cognizer becomes aware of some phenomenon, an entity or a situation in the world, as in *It is important to recognize [that not all essential words will be abstract or complex in meaning]* <sub>PHENOMENON</sub>.

In the REMEMBERING INFORMATION frame, a cognizer retains factual information in memory and is able to extract it for a certain purpose, as in *It is important to remember, however, [that eating disorders are psychological in nature, not physical]* <sub>MENTAL CONTENT</sub>. This mental content is derived from experience or activity, or from being informed of it. In the GRASP frame, a cognizer possesses deep insight about a certain phenomenon, i.e. about the importance or meaning of an idea or object, and is able to make predictions about the behaviour or occurrence of this phenomenon, as in *It is important to understand [these assumptions and that Tomlinson's model originated through nonmusic curricular areas]* <sub>PHENOMENON</sub>. In the COGITATION frame, a cognizer thinks about a certain topic over a period of time, as in *It is important to consider [the potential impact of treatment fidelity on study outcomes]* <sub>TOPIC</sub>.

In addition to the combinations activating the IMPORTANCE frame, the table is dominated by pairs of adjectives and verbs instantiating a close relationship between the DIFFICULTY frame and frames such as AWARENESS, GRASP, COMING TO BELIEVE, REMEMBERING INFORMATION, and SEPARATING. In the case of *hard to imagine*, *hard to believe*, *difficult to imagine*, and *difficult to know* in ranks 3, 22, 27, and 29, we can observe a strong correlation between the DIFFICULTY frame and the AWARENESS frame. The former frame is related to the experiencer's assessment of the ease or difficulty associated with an activity, as in *It is difficult [to determine the sources of his ideas]* <sub>ACTIVITY</sub>. The latter in turn is concerned with a piece of mental content presupposed by a cognizer in their model of the world. This content is not perceived directly or immediately, but usually, rather, deduced from something perceivable, as in *It is hard to imagine [that nearly 300,000 men died or were wounded here almost a century ago]* <sub>CONTENT</sub>.

A significant correlation also exists between the DIFFICULTY frame and the GRASP frame as well as between the DIFFICULTY frame and the COMING TO BELIEVE frame. The first relationship is instantiated by the combinations *easy to see* and *hard to see*, occupying ranks 11 and 12, while the second one by *difficult to determine* and *difficult to ascertain* in ranks 18 and 26. *Easy to see* and *hard to see* are used not only to assess the ease or difficulty of an activity, but also to make predictions about the behaviour or occurrence of a phenomenon on the basis of knowledge possessed by a cognizer, as in *It is easy to see [why Elephanta is considered one of the greatest Hindu monuments]* <sub>PHENOMENON</sub>. *Difficult to determine* and *difficult to ascertain* evaluate the difficulty of finding out the true or correct information about something. The verbs *determine* and *ascertain* seem to have something to do with a situation in which a cognizer changes his/her belief about a piece of content, usually after a process of reasoning. This change in belief is usually initiated by a person or a piece of evidence: e.g., in *It is difficult to ascertain [whether response rates have any effect on prevalence rates for two reasons]* <sub>CONTENT</sub>.

Another close correlation can be found between adjectives and verbs instantiating a relationship between the DIFFICULTY frame and the REMEMBERING INFORMATION frame, as well as between the DIFFICULTY frame and the SEPARATING frame. The first relationship is directly reflected in the co-occurrence of *easy* and *forget*, whereas the second one is seen in the combination of lexemes such as *difficult* and *separate*. *Easy to forget* is used to evaluate the ease or difficulty of retaining pieces of factual information (mental content) in memory and retrieving them, as in *It is easy to forget [how hard life was for our ancestors]*<sub>MENTAL CONTENT</sub>. *Difficult to separate* is employed for the assessment of the difficulty of separating a whole into parts, or of separating one part from another, as in *It is difficult to separate [the formation of language]*<sub>PART1</sub> *from [the formation of basic cognitive concepts]*<sub>PART2</sub>.

The top of the ranking list also comprises combinations of elements in which the first slot is filled by adjectives pertaining to the MENTAL PROPERTY ATTRIBUTION frame or the RISK EVALUATION frame, while the second one is occupied by verbs whose meanings can be interpreted in relation to the following semantic frames: STATEMENT, EXPECTATION, and COMING TO BELIEVE. *Reasonable to assume*, ranked fifth, instantiates a relation between the MENTAL PROPERTY ATTRIBUTION frame and the STATEMENT frame, as in *It is reasonable [to assume a parent may find it difficult to read a 30 page evaluation report during the course of a results meeting]*<sub>BEHAVIOUR</sub>. In this sentence, a judge (the speaker) indirectly addresses the message to an addressee using the extraposed construction, and attributes certain mental properties (in this case, rationality) to a person on the basis of that person's behaviour (any action, utterance, belief, or artifact).

*Reasonable to expect* and *unreasonable to expect* in ranks 6 and 14 are evidence of the semantic coherence determined by the MENTAL PROPERTY ATTRIBUTION frame and the EXPECTATION frame. For example, in the sentence *It is reasonable to expect [a significant reacceleration of inflation in the near future]*<sub>PHENOMENON</sub>, a speaker (a cognizer) not only evaluates the action in terms of rationality, but also believes that the phenomenon will take place in the future. In the case of *reasonable to conclude* in rank 16, we can observe a strong connection between the MENTAL PROPERTY ATTRIBUTION frame and the COMING TO BELIEVE frame, as is clearly evident in *It is reasonable to conclude [that use of proxy information from multiple sources may also increase data accuracy]*<sub>CONTENT</sub>. In this sentence, a speaker (a cognizer) comes to believe, after a process of reasoning, that the use of this kind of information may be beneficial for its users.

The RISK EVALUATION frame, mentioned above, is evoked by *safe to say* and *safe to assume* in ranks 7 and 13. The frequent co-occurrence of the adjective *safe* with the verbs *say* and *assume* is evidence of a direct association between this frame and the STATEMENT frame, as the sentence *This time, it is safe [to assume the regulators will pay more continuing attention]*<sub>CONTENT/ACTION</sub> seems to suggest. In this sentence, a speaker evaluates the action or the piece of mental content with respect to how risky it is to a person in certain circumstances.

Among the most strongly associated co-varying-collexeme pairs of the investigated construction, we can also find co-occurrences of verbs with adjectives evoking semantic frames related to different forms of evaluation: FAIRNESS EVALUATION (*fair to say*), REALISM EVALUATION (*unrealistic to expect*), USEFULNESS EVALUATION (*useful*

rank	pairs of col-lexemes	a	x	e	z	b	c	y	f	d	(a)	P <sub>Fisher exact</sub>
1.	difficult to be	8	516	1076	9659	508	1068	9143	8583	8075	57.48	1
2.	important to say	3	222	1946	9659	219	1943	9437	7713	7494	44.73	1
3.	difficult to consider	2	224	1076	9659	222	1074	9435	8583	8361	24.95	1
4.	important to imagine	1	202	1946	9659	201	1945	9457	7713	7512	40.7	1
5.	possible to note	1	486	1015	9659	485	1014	9173	8644	8159	51.07	1
6.	necessary to note	4	486	823	9659	482	819	9173	8836	8354	41.41	1
7.	possible to be	10	516	1015	9659	506	1005	9143	8644	8138	54.22	1
8.	important to be	45	516	1946	9659	471	1901	9143	7713	7242	104	1
9.	important to find	3	141	1946	9659	138	1943	9518	7713	7575	28.41	1
10.	difficult to have	3	237	1076	9659	234	1073	9422	8583	8349	26.4	1
11.	important to do	4	123	1946	9659	119	1942	9536	7713	7594	24.78	0.999999996
12.	necessary to see	3	277	823	9659	274	820	9382	8836	8562	23.6	0.999999995
13.	possible to understand	5	252	1015	9659	247	1010	9407	8644	8397	26.48	0.999999981
14.	easy to note	1	486	330	9659	485	329	9173	9329	8844	16.6	0.999999971
15.	impossible to be	5	516	483	9659	511	478	9143	9176	8665	25.8	0.999999951
16.	possible to assume	1	148	1015	9659	147	1014	9511	8644	8497	15.55	0.999999936

to recall), CORRECTNESS EVALUATION (*true to say*), FREQUENCY EVALUATION (*rare to find*), and LIKELIHOOD (*impossible to predict*). The adjective *fair* co-occurs with the verb *say*, thus instantiating a relationship between the FAIRNESS EVALUATION frame and the STATEMENT frame, as in *It is fair [to say that few people expected this]*<sub>ACTION</sub>. In this sentence, a speaker makes a statement about people's expectation and evaluates the action with respect to how fair, just, equitable, or appropriate it is to the affected party in a particular situation. The adjective *unrealistic* tends to occur with the verb

rank	pairs of collexemes	a	x	e	z	b	c	y	f	d	(a)	P <sub>Fisher exact</sub>
17.	important to conclude	1	71	1946	9659	70	1945	9588	7713	7643	14.3	0.999999892
18.	hard to be	4	516	424	9659	512	420	9143	9235	8723	22.65	0.999999867
19.	difficult to remember	1	127	1076	9659	126	1075	9532	8583	8457	14.15	0.999999724
20.	impossible to see	1	277	483	9659	276	482	9382	9176	8900	13.85	0.999999455
21.	possible to remember	1	127	1015	9659	126	1014	9532	8644	8518	13.35	0.999999318
22.	necessary to assume	1	148	823	9659	147	822	9511	8836	8689	12.61	0.999998302
23.	reasonable to be	1	516	228	9659	515	227	9143	9431	8916	12.18	0.99999686
24.	necessary to say	4	222	823	9659	218	819	9437	8836	8618	18.92	0.999996314
25.	difficult to do	2	123	1076	9659	121	1074	9536	8583	8462	13.7	0.999992596
26.	important to go	1	52	1946	9659	51	1945	9607	7713	7662	10.48	0.999991983
27.	possible to say	7	222	1015	9659	215	1008	9437	8644	8429	23.33	0.999991248
28.	important to believe	2	62	1946	9659	60	1944	9597	7713	7653	12.49	0.999986038
29.	possible to know	5	177	1015	9659	172	1010	9482	8644	8472	18.6	0.999976107
30.	difficult to take	2	106	1076	9659	104	1074	9553	8583	8479	11.81	0.999950877

**TABLE 3.** The results of co-varying collexeme analysis for the 30 most strongly repelled pairs of adjectives and verbs

*expect*, thereby reflecting a connection between the REALISM EVALUATION frame and the EXPECTATION frame, as in *It is unrealistic [to expect the problem to be solved today at this point]*<sub>SITUATION</sub>. In this sentence, the situation is judged based either on hopes and wishes or on facts as they really are.

*Useful* displays a tendency to collocate with *recall* in the pattern under study, and hence this combination represents a relationship between the USEFULNESS EVALUATION frame and the REMEMBERING INFORMATION frame. For example, in the sentence

*It is useful [to recall that, at its start, Poland's powerful Solidarity movement lacked clear and cohesive leadership]*<sup>ACTION</sup>, some action or desirable state of affairs (in this case, the recollection of facts) is considered by a speaker as useful for a benefiting party.

*True to say*, ranked number 21, can be described with reference to the background knowledge associated with the CORRECTNESS EVALUATION frame and the STATEMENT frame. For example, the speaker of the sentence *It is true [to say, however, that even his earliest horses and riders had an unsettled or unsatisfactory partnership]*<sup>INFORMATION</sup> introduces the statement by means of the verb *say* and judges a piece of information to be correct or true. The pair *rare to find* in rank 25 is a concrete instance representing a relationship between the FREQUENCY EVALUATION frame and the BECOMING AWARE frame. The first frame refers to an action, event or salient entity evaluated by a cognizer as being frequent or rare, while the second one concerns a cognizer becoming aware of some phenomenon, an entity, or a situation in the world. Both frames are evoked by the combination of *rare* and *find* in the sentence *It is rare [to find a stein entirely made of ivory]*<sup>EVENT</sup>

Finally, the bottom of the ranking list contains *impossible to predict*, a pair of lexemes invoking the LIKELIHOOD frame and the PREDICTION frame. The first frame is concerned with the likelihood of a hypothetical event (the state of affairs or occurrence) being evaluated by a judge. The second one, in turn, has something to do with an event or state that is predicted by a speaker to occur or hold true at a future time. The co-occurrence of *impossible* and *predict* in the sentence *It is impossible [to predict which students will be future bullies, victims, or bystanders]*<sup>HYPOTHETICAL EVENT</sup> is evidence of a strong correlation between these two frames, i.e. a mutual connection affecting the semantic coherence between the two slots of the construction under investigation.

At the last stage of the interpretation, it is also worth pointing out pairs of adjectives and verbs that are not significantly attracted to the construction in academic discourse: that is, co-varying collexemes that occur less frequently than expected in the investigated pattern. The results of a collexeme analysis for the 30 most strongly repelled pairs of the *it is ADJ to V*-construction are shown in Table 3. The top of the ranking list in this table is dominated by pairs of adjectives and verbs such as *difficult to be*, *important to say*, *difficult to consider*, *important to imagine*, *possible to note*, *necessary to note*, *possible to be*, *important to be*, *important to find*, *difficult to have* that are not strongly attracted lexemes, since their p-values resulting from the calculation of the Fisher exact test are very high: 1, in all of these cases. In addition, a comparison of the observed and the expected frequencies for each of this pair shows us that these collexemes occur less frequently than expected in this construction and hence they are loosely associated with the pattern under scrutiny in academic discourse.

A cursory look at Table 3 already reveals that adjectives such as *difficult*, *possible*, *important*, *impossible*, *hard*, and *reasonable* demonstrate a loose association with the verb *be*, that the adjectives *important*, *necessary* and *possible* occur less frequently than expected with the verb *say*, that the adjectives *possible*, *necessary* and *easy* are loosely associated with the verb *note*, and that *difficult* and *possible* co-occur extremely rarely with the verb *remember* in the investigated pattern. In addition, some of these adjectives have a weak correlation with verbs such as *have*, *do*, *go*, and *take*, since these verbs occur extremely infrequently in the construction in academic discourse. A pos-

sible explanation for their loose association in the investigated pattern may lie in the function and usage of the *it is ADJ to V*-construction in academic discourse. The results of the analysis for the 30 most strongly attracted combinations have revealed that the verbal slot of this construction exhibits a strong preference for verbs conveying new facts and information, i.e. verbs introducing a statement, denoting awareness and expectation, and evoking semantic frames such as GRASP, REMEMBERING INFORMATION, COMING TO BELIEVE, OR BECOMING AWARE. These verbs, in turn, have a stronger tendency to occur with particular types of adjectives than with others. For example, the verb *say* co-occurs more frequently with the adjectives *fair*, *safe* and *true* than with *important*, *necessary* and *possible*, the verb *note* tends to collocate more often with *important* and *interesting* than with *possible*, *necessary* and *easy*, and the verb *remember* prefers the adjective *important* to *difficult* and *possible*. The interdependence between these adjectives and verbs is strongly determined by specific semantic frames, the construction's function, the speaker's or writer's communicative intention, and the context in which such constructions are used.

## 7. CONCLUDING REMARKS

In conclusion, the findings of this investigation have indicated that the semantic coherence between the most strongly associated co-varying-collexeme pairs of the *it is ADJ to V*-construction is based on frame-semantic knowledge, i.e. a relationship between semantic frames evoked by adjectives and verbs co-occurring in the construction under study in a specific situational, discourse, and conceptual-cognitive context. The co-varying collexeme analysis has revealed not only the high degree of semantic coherence that exists between different adjectival and verbal slots of the pattern in question, but also systematic relationships between semantic frames that determine this semantic coherence. These relationships are clearly not the exception, but the rule for this construction.

It has been found, for example, that the interaction between the adjective *important* and the verbs *note*, *recognize*, *remember*, *understand*, *keep in mind*, *acknowledge*, and *consider* in academic discourse is determined by a reciprocal relationship between the IMPORTANCE frame and several other frames activated by these verbs, e.g. STATEMENT (*note*, *acknowledge*), BECOMING AWARE (*recognize*), REMEMBERING INFORMATION (*remember*, *keep in mind*), GRASP (*understand*), and COGITATION (*consider*). The interdependence between the pairs *hard to imagine*, *hard to believe*, *difficult to imagine*, and *difficult to know*, in turn, is based on a close relationship between the DIFFICULTY frame and the AWARENESS frame.

The results also confirm previous predictions about types of adjectives and verbs preferred by both slots of this construction in academic discourse. The adjectival slot seems to show a marked preference for adjectives expressing the speaker's or writer's evaluative judgement, whereas the verbal slot prefers verbs denoting cognitive processes, introducing a statement, and/or conveying new facts and states of affairs. For example, adjectives denoting importance (e.g. *important*), difficulty (*hard*), or a specific mental property (e.g. *reasonable*) co-occur with verbs that can be used to

introduce a statement (*important to note*), to denote awareness of or knowledge about a fact (*hard to imagine*), or to express the belief that some phenomenon will take place in the future (*reasonable to expect*).

A logical explanation as to why such combinations are preferred by writers and speakers may lie in the nature and specificity of academic discourse. In this kind of register, researchers aim to present, interpret and comment on the findings of their studies. To this end, they seek to convey new factual information about the current state of their research by expressing their evaluative opinions on the importance of their results, the difficulties encountered in the process of their interpretation, the correctness of their predictions, the meaning of an idea, the occurrence of a phenomenon, the likelihood of a hypothetical event, etc.

All these findings support the specific suggestions concerning the semantic and discourse-functional properties of the *it is ADJ to V*-construction. The illustrative examples, discussed in section 6, show two main and partially related functions of this type of extraposition in academic discourse. First, a speaker or a writer attempts to express his/her evaluative opinion in an indirect way by introducing the evaluative comments in the form of the dummy *it*, the verb form *is*, and adjectives such as *important*, *difficult*, *hard*, *reasonable*, *fair*, or *true* at the beginning of a sentence, as in *It is important [to note that data suggest mixed results for the success of anti-bullying programs]* UNDERTAKING. Second, a speaker or a writer aims to introduce a completely new idea into the discourse, a new topic that is linked to the previous context or has no direct link with the preceding context. This new idea is introduced at the end of a sentence, as in *It is interesting [to note that they correspond to different stabilizing control laws]* STIMULUS.

These findings are in agreement with earlier studies into the discourse function of extraposition (e.g. Collins 1994; Herriman 2000a; Hoey 2000; Hewings and Hewings 2002; Kaltentböck 2005; Rowley-Jolivet and Carter-Thomas 2005), and with the results of Hilpert's (2014) co-varying-collexeme analysis of adjectives and verbs in the *it is ADJ to V*-construction. Using the data extracted from the BNC corpus, Hilpert found, for example, that adjectives denoting ease and difficulty (*difficult*, *easy*, *hard*) co-occur with verbs pertaining to cognitive processes (*see*, *imagine*, *believe*), while the adjective *important* co-occurs with verbs introducing a statement (*note*, *remember*). He also states that combinations such as *it is interesting to note* or *important to remember* "do not carry focal information in themselves and are usually less prominently stressed than the material that follows", thereby "setting the stage for a new piece of information in discourse" (Hilpert 2014: 402).

Hilpert's analysis, however, was restricted to the indication of the 20 most strongly attracted combinations, as its primary aim was to demonstrate the application of the quantitative method for a semantic analysis of the *it is ADJ to V*-construction. Surprisingly, apart from pairs such as *reasonable to suppose*, *important to realise*, *important to stress*, *interesting to compare*, and *good to be*, the ranking list of the most strongly associated co-varying-collexeme pairs in the current study contains the same fifteen combinations interpreted by Hilpert (cf. 2014: 402) as the most significant pairs of the investigated construction. These combinations, however, hold various positions in both lists. In Hilpert's (see 2014: 402) ranking the top nine positions are occupied by *interesting to note*, *fair to say*, *important to remember*, *true to say*, *reasonable to assume*,

*hard to believe, hard to imagine, important to note, and unrealistic to expect*, while in the present study the nine most strongly attracted pairs are *important to note, interesting to note, hard to imagine, fair to say, reasonable to assume, reasonable to expect, safe to say, unrealistic to expect, and important to recognize*.

This suggests, for example, that pairs reflecting the relationship between the IMPORTANCE frame and the STATEMENT frame, between the MENTAL STIMULATION frame and the STATEMENT frame, between the DIFFICULTY frame and the AWARENESS frame, between the FAIRNESS EVALUATION frame and the STATEMENT frame, and between the MENTAL PROPERTY ATTRIBUTION frame and the STATEMENT frame are the five most strongly attracted pairs of this construction in the academic section of COCA, while the combinations instantiating the relationship between the MENTAL STIMULATION frame and the STATEMENT frame, between the FAIRNESS EVALUATION frame and the STATEMENT frame, between the IMPORTANCE frame and the REMEMBERING INFORMATION frame, between the CORRECTNESS EVALUATION frame and the STATEMENT frame, and between the MENTAL PROPERTY ATTRIBUTION frame and the STATEMENT frame co-occur more frequently with this pattern in the corpus of general British English. In other words, out of the five relationships between frames listed above, three occur in both studies at the top of the ranking list.

Five notable exceptions, listed by Hilpert but not included in the ranking list of the current research, are *reasonable to suppose, important to realise, important to stress, interesting to compare, and good to be. Reasonable to suppose* (ranked number 10 in Hilpert's list) and *important to stress* (ranked 13 in Hilpert's table) are also among the most attracted pairs of this construction in this study but occupy lower positions: *reasonable to suppose*, with 7 occurrences, is in rank 46, while *important to stress*, with 15 occurrences, is in rank 33. The remaining three combinations occur very rarely in the academic register, thus being among the least strongly associated pairs of the construction in academic discourse: *important to realise* (1 occurrence, in rank 2503), *interesting to compare* (4 occurrences, in rank 531), and *good to be* (2 occurrences, in rank 928).

A possible explanation for their loose association in the construction under scrutiny may lie in the influence of academic discourse on the preferred combinations of semantic frames. For example, this kind of register allows speakers or writers to present their evaluative opinions about the importance of the current state of their studies by introducing the adjective *importance*, evoking the IMPORTANCE frame, and verbs activating several other frames, e.g. STATEMENT (*note, acknowledge*), BECOMING AWARE (*recognize*), REMEMBERING INFORMATION (*remember, keep in mind*), GRASP (*understand*), and COGITATION (*consider*), rather than by introducing the adjective *important* and the verb *realise*, activating the IMPORTANCE frame and the COMING TO BELIEVE frame, or the combination *important to be*, reflecting the relationship between the IMPORTANCE frame and the EXISTENCE frame. The interdependence between the adjective *important* and these verbs is strongly determined by the speaker's or writer's communicative intention and the academic context in which such combinations are used.

The co-varying collexeme analysis applied in this study has proved to be an effective technique for the determination of the most strongly associated co-varying-collexeme pairs of the *it is ADJ to V*-construction, and hence may be employed for the identification of the most significant pairs of lexemes co-occurring in other types of *it-ex-*

traped constructions. Future research, for example, might focus on determining interdependencies between adverbs and adjectives found in two different slots of *it*-extraped constructions complemented by *to*-infinitive clauses or *that*-clauses. Such a quantitative analysis could reveal those combinations that occur more often than would be expected by chance, considering the respective frequencies of their participating elements. This in turn may be accompanied and supported by an analysis of semantic frames associated with these participating elements. Given that the current research was confined to the academic register, it would also be interesting to explore the distribution of adjectives and verbs in the investigated pattern across different types of both written and spoken registers, in view of the possible existence of slight variations in their occurrence. Future research, therefore, may determine the most strongly attracted co-varying collexeme pairs of the construction in other sections of COCA.

## REFERENCES

- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999) *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Bybee, J. J. (2010) *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.
- Bybee, J. J. (2013) Usage-based theory and exemplar representations of constructions. In: Hoffmann, T. and G. Trousdale (eds) *Handbook of Construction Grammar*, 49–92. Oxford: Oxford University Press.
- Calude, A. S. (2008) Clifing and extraposition in English. *ICAME Journal* 32, 7–34.
- Collins, P. (1994) Extraposition in English. *Functions of Language* 1, 7–24.
- Emonds, J. (1972) A reformulation of certain syntactic transformations. In: Peters, S. (ed.) *Goals of Linguistic Theory*, 21–62. Englewood Cliffs, NJ: Prentice-Hall.
- Erdmann, P. (1990) *Discourse and Grammar: Focusing and Defocusing in English*. Tübingen: Niemeyer.
- Fillmore, Ch. J. (1982) Frame semantics. In: The Linguistic Society of Korea (eds) *Linguistics in the Morning Calm*, 111–137. Seoul: Hanshin Publishing Company.
- Fillmore, Ch. J. and B.T. Atkins (1992) Toward a frame-based lexicon: The semantics of RISK and its neighbors. In: Lehrer, A. and E. F. Kittay (eds) *Frames, Fields and Contrasts*, 75–102. Hillsdale, New Jersey: Lawrence Erlbaum Assoc.
- Fillmore, Ch. J. and C. Baker (2010) A frames approach to semantic analysis. In: Heine, B. and H. Narrog (eds) *The Oxford Handbook of Linguistic Analysis*, 313–340. Oxford: Oxford University Press.
- Fillmore, C. J., C. R. Johnson and M. R. L. Petruck (2003) Background to FrameNet. *International Journal of Lexicography* 16(3), 235–250.
- Fontenelle, T. (ed.) (2003) Special issue on FrameNet and frame semantics. *International Journal of Lexicography* 16(3), 231–385.
- Francis, G. (1993) A corpus-driven approach to grammar: principles, methods and examples. In: Baker, M., G. Francis and E. Toginini-Bognelli (eds) *Text and Technology: In Honour of John Sinclair*, 137–156. Amsterdam: John Benjamins.
- Goldberg, A. (1995) *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.
- Goldberg, A. (2006) *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, A. (2013) Constructionist approaches to language. In: Hoffmann, T. and G. Trousdale (eds) *Handbook of Construction Grammar*, 15–48. Oxford: Oxford University Press.
- Gómez-González, M. A. (2001) *The Theme-Topic Interface. Evidence from English*. Amsterdam and Philadelphia: John Benjamins.

- Gómez-González, M. A. (1997) On subject *it*-extrapositions: Evidence from present-day English. *Revista Alicante de Estudios Ingleses* 10, 95–107.
- Gries, S. Th. and A. Stefanowitsch (2004a) Extending collocation analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics* 9 (1), 97–129.
- Gries, S. Th. (2012) Frequencies, probabilities, association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language* 36, 477–510.
- Gries, S. Th. (2015) More (old and new) misunderstandings of collocation analysis: on Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26 (3), 505–536.
- Herbst, T., D. Heath., I. F. Roe and D. Götz (2004) *A Valency Dictionary of English: A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives*. Berlin: Mouton de Gruyter.
- Herriman, J. (2000a) Extraposition in English: A Study of the Interaction between the Matrix Predicate and the Type of Extraposed Clause. *English Studies* 6, 582–599.
- Herriman, J. (2000b) The Functions of Extraposition in English Texts. *Functions of Language* 7(2), 203–230.
- Hewings, M. and A. Hewings (2002) It is interesting to note that ...: A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes* 21, 367–383.
- Hilpert, M. (2014) Collocation analysis: Measuring associations between constructions and lexical elements. In: Glynn, D. and J. Robinson (eds) *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 7–38. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Hoey, M. (2000) Persuasive rhetoric in linguistics: A stylistic study of some features of the language of Noam Chomsky. In: Hunston, S. and G. Thompson (eds) *Evaluation in Text. Authorial Stance and the Construction of Discourse*, 28–37. Oxford: Oxford University Press.
- Huddleston, R. (1971) *The Sentence in Written English*. Cambridge: Cambridge University Press.
- Huddleston, R. (1984) *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.
- Kaatari, H. (2010) *Complementation of Adjectives: A Corpus-Based Study of Adjectival Complementation by that- and to-Clauses*. Unpublished MA thesis. Uppsala University, Department of English.
- Kaltenböck, G. (2000) *It*-extraposition and non-extraposition in English discourse. In: Mair, C. and M. Hundt (eds) *Corpus linguistics and linguistic theory*, 157–175. Amsterdam and Atlanta: Rodopi.
- Kaltenböck, G. (2003) On the syntactic and semantic status of anticipatory *it*. *English Language and Linguistics* 7(2), 235–255.
- Kaltenböck, G. (2004) *It-Extraposition and Non-extraposition in English: A Study of Syntax in Spoken and Written Texts*. Wien: Braumüller.
- Kaltenböck, G. (2005) *It*-extraposition in English: A functional view. *International Journal of Corpus Linguistics* 10 (2), 119–159.
- Langacker, R. (1987) *Foundations of Cognitive Grammar. Theoretical Prerequisites*. Volume I. Stanford, CA: Stanford University Press.
- Mair, C. (1990) *Infinitival Complement Clauses in English*. Cambridge: Cambridge University Press.
- McCawley, J. D. (1988) *The Syntactic Phenomena of English*. Volume II. Chicago: The University of Chicago Press.
- Miller, P. H. (2001) Discourse constraints on (non)extraposition from subject in English. *Linguistics* 39 (4), 683–701.
- Mindt, I. (2011) *Adjective Complementation: An Empirical Analysis of Adjectives Followed by that-Clauses*. Philadelphia: John Benjamins.
- Pérez-Guerra, J. (1998) Integrating right-dislocated constituents: A study on cleaving and extraposition in the recent history of the English language. *Folia Linguistica Historica* XIX, 7–25.
- Perek, F. (2014) Rethinking constructional polysemy: The case of the English conative construction. In: Glynn, D. and J. Robinson

- (eds) *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 61–86. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Petruck, M. R. L. (1996) Frame semantics. In: Verschueren, J., J.-O. Östman, J. Blommaert and C. Bulcaen (eds) *Handbook of Pragmatics*, 1–13. Philadelphia: John Benjamins.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1985) *A Comprehensive Grammar of the English Language*. New York and London: Longman.
- Rosenbaum, P. S. (1967) *The Grammar of English Predicate Complement Constructions*. Cambridge, MA: M.I.T. Press.
- Rowley-Jolivet, E. and S. Carter-Thomas (2005) Genre awareness and rhetorical appropriacy: Manipulation of information structure by NS and NNS scientists in the international conference setting. *English for Specific Purposes* 24, 41–64.
- Schmid, H.-J. and H. Küchenhoff (2013) Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3), 531–577.
- Seppänen, A., C. G. Engström and R. Seppänen (1990) On the so-called anticipatory It. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43, 748–776.
- Seppänen, A. (1999) Extraposition in English revisited. *Neophilologische Mitteilungen* 100: 51–66.
- Stefanowitsch, A. and S. Th. Gries (2003) Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8, 209–243.
- Stefanowitsch, A. and S. Th. Gries (2005) Co-varying collexemes. *Corpus Linguistics and Linguistic Theory* 1 (1), 1–43.
- Stefanowitsch, A. (2013) Collostructional analysis. In: Hoffmann, T. and G. Trousdale (eds) *Handbook of Construction Grammar*, 290–306. Oxford University Press.
- Van linden, A. (2012) *Modal Adjectives: English Deontic and Evaluative Constructions in Synchrony and Diachrony*. Berlin and Boston: Walter de Gruyter.
- Wiliński, J. (2017) Normal and anomalous occurrences of adjectives in extraposed constructions with *to*-infinitive clauses: A quantitative corpus-based study. In: Wiliński, J. and J. Stolarek (eds) *Norm and Anomaly in Language, Literature and Culture*, 89–106. Frankfurt am Main: Peter Lang.

## SOURCES AND TOOLS

The Corpus of Contemporary American English (COCA). The full-text data (1990–2012).

Available from <https://www.corpusdata.org/purchase.asp>

The FrameNet project. Available from <https://framenet.icsi.berkeley.edu/fndrupal/>

MonoConc Pro (MP 2.2). Available from <http://www.athel.com/mono.html>

Fisher's Exact Test. Available from <http://www.langsrud.com/fisher.htm>

### Jarosław Wiliński

Siedlce University of Natural Sciences and Humanities

Wydział Humanistyczny Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach

ul. Żytnia 39, 08-110 Siedlce

ORCID ID: 0000-0002-3136-6529

e-mail: [wilinskij@uph.edu.pl](mailto:wilinskij@uph.edu.pl)