Linguistica Silesiana 39, 2018
ISSN 0208-4228

JAROSŁAW WILIŃSKI
Siedlce University of Natural Sciences and Humanities
wilinskij@uph.edu.pl

DISTINCTIVE ADJECTIVES IN TWO VARIANTS OF THE *IT*-EXTRAPOSITION CONSTRUCTION: A QUANTITATIVE CORPUS-BASED INVESTIGATION

Adopting the standpoint of Construction Grammar (Goldberg 2006) and the corpus-based method known as *Distinctive-Collexeme Analysis* (Gries and Stefanowitsch 2004; Hilpert 2014), the investigation seeks to identify lexemes that indicate a strong preference either for the *it is ADJ to V*-construction or the *it is ADJ that*-construction. On the basis of data extracted from the academic sub-corpus of COCA, the study reveals that there are adjectives exhibiting a strong preference for one construction instead of the other, and that the constructions have a tendency to occur with adjectives evoking different semantic frames.

Keywords: *it-extraposition*, *COCA*, *distinctive collexeme analysis*, *Fisher exact test, construction*

1. Introduction

The past three decades have witnessed a remarkable proliferation of publications about *it*-extraposition in English (e.g. Seppänen 1999; Kaltenböck 2000, 2004). Some researchers have explored structural properties of different types of extraposition (e.g. Huddleston 1984; Quirk et al. 1985; Mair 1990; Collins 1994; Kaatari 2010), whereas others have drawn an elaborate comparison of extraposed constructions with right dislocation and *it*-clefts (Pérez-Guerra 1998; Calude 2008; McCawley 1988; Collins 1994). Some research studies have also concentrated on the use of adjectives in *it*-extraposition, their clausal complementation (Kaltenböck 2006; Van Linden and Davidse 2009; Kaatari 2010; Mindt 2008, 2011; Van Linden 2012) and different valency patterns (see Rivière 1983; Herbst et al. 2004). Finally, many studies have explored discourse-functional properties of various extraposed structures (Mair 1990; Collins 1994;

Gómez-González 1997; Herriman 2000a,b; Miller 2001; Rowley-Jolivet and Carter-Thomas 2005; Kaltenböck 2005).

To date, however, no research has sought to compare and contrast the use of adjectives in the *it is ADJ to V*-construction as opposed to the *it is ADJ that*-construction. In addition, little or no attention has been paid to the quantitative evaluation of adjectives in these two different variants of *it*-extraposition, the statistical corroboration of their occurrence in such closely-related constructions, or the empirical verification of previous assumptions and speculations about their use in academic discourse.

Previous work has been limited to the identification of several semantic categories for adjectival matrix predicates governing different syntactic types of extraposed subject clauses (Herriman 2000a; Kaltenböck 2005; Kataari 2010), to the comparison of the *it is ADJ to V*-construction with the predicative *X is ADJ*-construction (Hilpert 2014), and to the determination of the most strongly attracted adjectives of the *it is ADJ to V*-construction (Wiliński 2017) in academic discourse. Hilpert's (2014) case study, for example, contrasted the extraposed construction and the predicative construction in their respective collocational preferences. On the basis of the data extracted from the BNC corpus, Hilpert found that, in contrast to the predicative construction, the most distinctive collexemes of the *it is ADJ to V*-construction are adjectives referring to different semantic scales, such as possibility (*impossible*, *possible*), ease (*difficult*, *easy*, *hard*), advisability (*better*, *best*), and importance (*important*, *necessary*).

Wiliński's (2017) study, in turn, revealed that some adjectives are more strongly attracted to the *it is ADJ to V*-construction than others, and that the occurrence of certain adjectives in this construction is more significant than their use in different types of extraposed structures. Using the data extracted from the academic sub-corpus of COCA, Wiliński established that adjectives evoking different semantic frames, such as IMPORTANCE (*important*, *critical*, *crucial*, *imperative*, *vital*), DIFFICULTY (*difficult*, *easy*, *hard*), LIKELIHOOD (*possible*, *impossible*, *likely*, *unlikely*), NECESSITY (*necessary*, *essential*), MENTAL PROPERTY ATTRIBUTION (*reasonable*, *wise*), and USEFULNESS EVALUATION (*useful*, *helpful*, *instructive*) are the most significant lexemes of the *it is ADJ to V*-construction.

Given that no study has attempted to compare and contrast the *it is ADJ to V*-construction and the *it is ADJ that*-construction in their respective collocational preferences and that Wiliński's study was not specifically designed to uncover subtle distributional variations between these constructions, there is still a need for exploring the distribution of adjectives in these semantically related patterns and for comparing their frequency of occurrence, in view of the possible existence of slight variations in their use in academic discourse. This paper, therefore, seeks to ascertain consequential, if slight, distributional differences between the two semantically or functionally near-equivalent constructions and to demonstrate that, synonymous though these two constructions might appear to be at first glance, they do in fact display subtle differences.

The remainder of this paper is organized in the following order. Section 2 focuses both on the theoretical assumptions underpinning the semantic explanation of adjectives occurring in both constructions, and on the methodological framework underlying the quantitative analysis of adjectives. Section 3 describes the corpus, the data, and the tools. Section 4 sketches the statistical procedure adopted in this study. Section 5 describes the constructions under study with respect to their functional and structural properties. Section 6 combines the findings of the quantitative analysis with a semantic description of adjectives and elucidates subtle distributional differences between these two semantically near-equivalent constructions. Section 7 evaluates the results and puts forward some proposals for future research.

2. Theoretical and methodological assumptions

The study is founded upon the concepts and theory that pertain to usage-based Construction Grammar (Goldberg 2006, 2013) and Frame Semantics (Fillmore 1982). The usage-based model of grammar (Diessel 2015: 296) assumes that "grammar is a dynamic system of emergent categories and flexible constraints that are always changing under the influence of domain-general cognitive processes involved in language use." This approach (e.g. Bybee 2010; Goldberg 2006) rests on two specific assumptions about the nature of linguistic elements and the overall organization of the grammatical system: first, grammatical structure should be investigated with respect to simple and complex signs, i.e. constructions, pairings of a specific structural pattern with a particular function or meaning, and second, all linguistic signs (i.e. lexical units and grammatical patterns) combined with each other by different kinds of links so that grammar (or language in general) can be construed as a dynamic network of interconnected constructions.

Frame semantics assumes that meanings can be relativized to scenes, and that meanings have internal structure which is determined relative to background frames or scenes (cf. Fillmore, Lee-Goldman and Rhodes 2012). The term *frame* refers to a background of cultural beliefs and personal experiences that are fundamental to the understanding of the lexical meaning of a word (e.g. Fillmore 1982). Consequently, the meaning of a lexical unit cannot be interpreted and understood without access to all the encyclopaedic knowledge that is associated with that unit. For example, words such *theft*, *thief*, *steal* or *stealing* are assumed to evoke the frame THEFT, a coherent knowledge structure about a situation in which a perpetrator takes goods from a victim or a source: without knowledge of all of these related concepts, we are not capable of apprehending a complete knowledge of one of them.

The methodology used in this study is quantitative corpus linguistics. The method referred to as *Distinctive-Collexeme Analysis* (Gries and Stefanowitsch 2004; Stefanowitsch 2013; Hilpert 2014) is employed to determine those

adjectives that are distinctively associated with the *it is ADJ to V*-construction as compared to the *it is ADJ that*-construction. The application of this corpus-based approach was required by the semantic and functional focus of the current investigation, which made it necessary to explore the relevant context in which specific instances of both types of *it*-extraposition were found in different sections of the academic sub-corpus. With respect to the methodology, thus, the study combines both quantitative and qualitative analyses. The quantitative-statistical perspective concentrates on the frequency of the two types of *it*-extraposition and the distribution of adjectives in these constructions, while the qualitative dimension of the analysis seeks to capture all specific nuances of their meaning by determining the semantic frames with which the most strongly attracted adjectives are associated.

3. Corpora, data, and tools

The data used in this investigation were collected from the downloadable version of the academic sub-corpus of the Corpus of Contemporary American English (COCA), i.e. the full-text data corpus obtainable from Mark Davies. This academic section is composed of approximately 81 million words derived from nearly 100 different peer-reviewed journals. These comprise a wide spectrum of fields of study, including a certain percentage from philosophy, psychology, religion, world history, education, and technology.

One advantage of the corpus-based method over traditional approaches such as introspection and elicitation is that the use of corpus data enables the researcher to adopt a broader perspective of a particular syntactic construction by accounting for its fuzzy boundaries and gradient relations to other categories, hence placing the construction under scrutiny in a broader linguistic context.

The application of the downloadable version of COCA offered two further advantages. First, the inclusion of a wide range of written texts from academic discourse allowed the empirical study of distributional tendencies of adjectives in both constructions across a variety of text types, from linguistics, to history, to psychology, etc. Second, as the downloadable version, the academic sub-corpus of COCA allowed for easy retrieval of the constructions in question, viz. by means of MonoConc Pro, a concordance tool that automatically retrieved all the occurrences of adjectives in the constructions under study from the corpus. This computer program constructed concordances, i.e. concordance lines, which were manually examined to determine the frequencies of all adjectives occurring in the relevant pattern. Then, all these frequencies necessary to compute the mutual association between adjectives and the constructions were entered in a 2-by-2 table and submitted to the Fisher exact test. The p-value resulted from this test was used to gauge the strength of association, i.e. the degree of attraction to the *it is ADJ to V*-construction or the *it is ADJ that*-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 attraction between an adjective and one of the synonymous constructions in question (cf. Schmid and Küchenhoff 2013). This computation of statistical significance was performed by means of an online Fisher's exact test calculator for two-by-two contingency tables. The remaining values and expected frequencies were worked out in Microsoft Excel spreadsheets. The frequency lists resulted from the calculation of observed frequencies and then provided the input to the distinctive collexeme analysis.

At this stage of the discussion, a few problematic issues concerning the use of the Fisher exact p-value as a measure of significance need to be mentioned. First, according to Schmid and Küchenhoff, it is doubtful whether or not the Fisher exact p-value incorporates an effect size (2013: 539). In response to this point of criticism, 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." Second, p-values cause problems with the interpretability of scores. In other words, the Fisher exact test frequently provides an infinitely strong value of collostruction strength (Hilpert 2014: 397). This issue frequently arises in constructions with low type frequencies. However, this is not a problem for collostructional analysis per se, but for researchers who falsely interpret such results. Third, p-values decrease automatically with growing corpus size, and the size of the value in cell no. 4 of a contingency table has a strong impact on the p-values computed by the Fisher exact test. There are several other problematic issues concerning the use of the Fisher exact test as a measure of collostruction strength that have been raised and that would merit discussion. Interested readers are referred particularly to the critiques in Bybee (2010), Schmid and Küchenhoff (2013), Küchenhoff and Schmid (2015) and to the responses in Gries (2012, 2015).

Despite the controversial issues concerning the Fisher exact test, it is recommended to employ this measure when at least one expected value in the table is smaller than 5 (Levshina 2015: 29), when data is very unevenly distributed and/or infrequent (cf. e.g. Pedersen 1998; Gries and Stefanowitsch 2004: 101; see also Gries 2012 and Gries 2015 for further arguments), and when the total number of observations in all cells is smaller than 20 (Sheskin 2011: 646). In addition, as stated by Evert (2009: 1235), "Mathematicians generally agree that the most appropriate significance test for contingency tables is Fisher's exact test."

Considering the fact that the Fisher exact test is more appropriate when observed frequencies are low, and that other measures such as t-scores, z-scores, or variants of the chi-squared test have not been used in the context of the distinctive-collexeme analysis to date (such statistical tests have only been employed in the case of collexeme analysis), the author of this paper decided to use the Fisher exact test to gauge the strength of attraction between adjectives and the constructions under scrutiny.

4. Statistical procedure

The procedure adopted in this investigation is composed of four stages. These stages can be clearly exemplified by reference to the adjective *clear* in the adjective slot of the *it is ADJ to V*-construction and the *it is ADJ that*-construction. Table 1 shows the distribution of this adjective in both patterns and other frequencies required for a distinctive-collexeme analysis.

Table 1. Contingency table cross-tabulating frequency scores of the adjective *clear* and the constructions under study

	Adjectives (*clear*)	**All other adjectives**	**Total**
***It is ADJ to V*-construction**	Frequency of adjective (*clear*) in '*it is ADJ to V*-construction' a = 1 (595.5)	Frequency of all other adjectives in '*it is ADJ to V*-construction' b = 9833	Total frequency of '*it is ADJ to V*-construction' x = 9834
***It is ADJ that*-construction**	Frequency of adjective (*clear*) in '*It is ADJ that*-construction' c = 921 (326.5)	Frequency of all other adjectives in '*It is ADJ that*-construction' d = 4470	Total frequency of '*It is ADJ that*-construction' y = 5391
Total	Total frequency of adjective (*clear*) e = 922	Total frequency of all other adjectives f = 14303	Total frequency of both constructions z = 15225

The initial stage of this procedure entailed calculating the observed frequencies. These were worked out in the following order. First, all occurrences of the *it is ADJ to V*-construction were extracted from the corpus: x = 9834. Second, all occurrences of the *it is ADJ that*-construction were determined: y = 5391. These two frequencies were derived from the corpus by identifying all patterns containing the adjective *clear*. Finally, the frequency of the lemma *clear* in each construction was counted: a = 1 and c = 921, respectively. The figures (a, x, c, y) were obtained from the corpus directly while the remaining ones result from addition and subtraction.

At the second stage, these observed frequencies were employed to work out the expected frequencies of the adjective (*clear*) in the *it is ADJ to V*-construction versus the *it is ADJ that*-construction. This calculation was performed in Microsoft Excel in the following order: for the lemma *clear* in each pattern, its column total was multiplied by its row total, and this final score was divided by the overall table total. For illustrative purposes, the expected frequencies for the lemma *clear* in each construction are provided in parentheses (see Table 1). If the observed frequency of the adjective (*clear*) in

the *it is ADJ to V*-construction is significantly higher or lower than expected, the mutual association between the adjective *clear* and this construction is one of attraction or repulsion respectively (the adjective *clear* is then considered to be a significantly attracted or repelled *collexeme* of the *it is ADJ to V*-construction). Likewise, if the observed frequency of the adjective (*clear*) in the *it is ADJ that*-construction is significantly higher or lower than expected, then the adjective is a significantly attracted or repelled *collexeme* of the *it is ADJ that*-construction.

At the third stage, the direction of association (attracted or repelled), or the strength of association (the distinctiveness of *clear*), between the adjective (*clear*) and the constructions under investigation was estimated by means of the Fisher exact test. To this end, the following four frequencies were employed: (1) the frequency of the adjective (*clear*) in the *it is ADJ to V*-construction, (2) the frequency of all other adjectives in the *it is ADJ to V*-construction, (3) the frequency of the adjective (*clear*) in the *it is ADJ that*-construction, and (4) the frequency of all other adjectives in the *it is ADJ that*-construction. These were entered in a two-by-two table and submitted to the Fisher exact test. The p-value obtained from the computation of the Fisher exact test for this distribution is exceptionally small: 0, a maximum degree of collostruction strength which could only be rendered more precise by employing a more powerful calculator. This shows that the adjective *clear* is highly significant (distinctive) for one of the two extraposed constructions. However, it does not suggest for which one. In order to determine this, the observed frequencies of the adjective *clear* need to be compared with the expected ones. As this comparison reveals, the adjective *clear* occurs more frequently than expected in the *it is ADJ that*-construction and less frequently than expected in in the *it is ADJ to V*-construction. In other words, *clear* is a highly significant, very strongly distinctive collexeme of the *it is ADJ that*-construction as opposed to the *it is ADJ to V*-construction.

At the last stage, the results of the quantitative investigation were sorted according to the strength of attraction and then evaluated qualitatively and subjectively. More clearly, the quantitative findings were integrated with a frame-semantic description of adjectives, and subtle semantic differences between these two near-equivalent constructions were explained.

5. *It*-extraposed constructions with that-clauses and to-infinitives

It-extraposition is a syntactic mechanism that alters word order in a sentence in such a manner that a relatively "heavy" constituent, i.e. a clause in subject position, is shifted to the right of its canonical position while replacing it by the dummy *it* (cf. e.g. Quirk et al. 1985: 1391-1393). Two types of extraposition are examined in this study: *it*-extraposition with adjectives complemented by *that*-clauses and *to*-infinitives. Typical examples of such extraposed constructions are shown in (1), with their non-extraposed counterparts being provided in (2):

(1) *It*-extraposed structures with *that*- and *to*-clauses
   a. It is imperative *to annually assess the effectiveness of the program.*
   b. It is surprising *that he isn't saying anything more at this point.* (COCA).

(2) Non-extraposed structures
   a. *To annually assess the effectiveness of the program* is imperative.
   b. *That he isn't saying anything more at this point* is surprising.

These two types of *it*-extraposition and non-extraposition have been widely acknowledged in theoretical syntax as sharing apparent similarities in their structure and their propositional meaning. In particular, in the generative literature (cf. e.g. Rosenbaum 1967; Huddleston 1984; Emonds 1976) *it*-extraposition was treated as being a derivative of non-extraposition (its syntactic derivation or transformation) because the latter was frequently assumed to be the most basic form preserving the canonical SVC order (cf. e.g. Huddleston 1984: 451). In recent years, however, some corpus evidence (e.g. Biber et al. 1999; Kaltenböck 2000) has come to light that suggests that *it*-extraposed structures occur more frequently in English than their non-extraposed counterparts. For example, Biber et al.'s study (1999: 676, 724) revealed that the frequency of *it*-extraposition with *that*-clauses and *to*-clauses in spoken language substantially exceeds the frequency of occurrence of non-extraposition in this kind of discourse, whereas Kaltenböck's research (2000: 158) indicated that *it*-extraposition vastly outnumbers non-extraposition, with a ratio of 1:7.8 in the British section of the International Corpus of English. Similar observations were made by Collins (1994: 14), Herriman (2000a: 584), Mair (1990: 30), and Quirk et al. (1985: 1392), who noticed that, from a statistical standpoint, *it*-extraposition is more frequent than non-extraposition: that is, it appears to be the norm rather than the exception. Thus, it seems more rational and logical to posit the occurrence of two separate extraposed constructions (pairings of form and meaning/function) in their own right, and to investigate them accordingly, rather than treat them as two different variants of non-extraposed patterns that occur in corpora extremely infrequently.

This study, therefore, considers examples such as the ones in (1) as types of the English *it*-extraposition construction, i.e. partially lexically-filled constructions being composed of three fixed lexical items (it is […] to […] or it is […] that […]) and two flexible slots that can be filled by adjectives complemented by *to*-infinitives or by adjectives followed by *that*-clauses. These constructions can be represented structurally and schematically as [*it is ADJ to*-infinitive clause] and [*it is ADJ that*-clause], where an anticipatory subject *it* is followed by the third person singular form of the verb *be*, a predicative adjective, and either a *to*-infinitive clause or a *that*-clause. The following sentences extracted from the corpus can be used to illustrate the occurrence of both patterns in academic discourse:

(3) *However, it is important to bear in mind that a test should not only be technically adequate.*
(4) *It is difficult to determine with precision how widespread these abusive practices were.*
(5) *It is obvious that the surface of the fiber coating was well-distributed.*
(6) *It is unlikely that many programs address multiple constructs in a systematic manner.*

In respect of semantic and discourse-functional properties of examples such as those in (3), (4), (5), and (6), many studies (e.g. Huddleston 1984; Collins 1994; Gómez-González 1997; Herriman 2000a; Rowley-Jolivet and Carter-Thomas 2005) have pointed to two primary functions of extraposed constructions with *to*-infinitives and *that*-clauses. First, in accordance with the principles of end-weight and end-focus (Quirk et al. 1985: 863; Collins 1994: 15-16), the use of *it*-extraposition is motivated in part by a desire to avoid long and heavy subject clauses by placing them at the end of the sentence, hence allowing for easier processing of new pieces of information in both speech and writing (Huddleston 1984: 354, 453; see also Erdmann 1990: 137-8 for this view). This view is also confirmed by the research conducted by Mair (1990: 39), Miller (2001), and Kaltenböck (2005), who observed that extraposed clauses convey not only new but also given information. Second, *it*-extraposed constructions allow for an expression of beliefs, opinions, or personal preferences about some facts or states of affairs by presenting them as if they were generally accepted views rather than personal judgements, thereby placing 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). This evaluative comment is somewhat 'depersonalized' by the use of impersonal *it* and as such serves the function of interpersonal Theme (Whittacker 1995: 111).

While the syntax and discourse function of different kinds of extraposed patterns has received considerable attention in the literature, the role, use, and distribution of adjectives in extraposed constructions with *to*-infinitive clauses and *that*-clauses have largely been disregarded. Those studies that have discussed the occurrence of adjectives in *it*-extraposition are limited either in terms of their scope of interest or the data analysed (see e.g. Herriman 2000a; Kaltenböck 2005; Kataari 2010). Kaltenböck's (2005) study, for example, revealed that a thematic component of different syntactic types of the extraposed clause, which expresses some speaker's evaluation or stance, can convey a range of meanings. Kaltenböck (2005: 138) identified several semantic categories for adjectival matrix predicates governing extraposed subject clauses: likelihood/possibility (e.g. *likely*, *possible*), value judgement (e.g. *good*, *nice*), and necessity/desirability (e.g. *necessary*, *desirable*) for adjectives complemented by infinitive clauses; truth/transparency of a state of affairs (e.g. *clear*, *apparent*), existence of a state of affairs (e.g. *difficult*, *easy*), likelihood/possibility (e.g. *likely*, *possible*) and value judgement (e.g. *good*, *nice*) for adjectives followed by finite clauses.

These semantic categories correspond largely to the ones identified by Collins (1994: 19) and Gómez-González (1997: 102-3, 2001: 272-73). Kataari's (2010) study, in turn, concerned the use of epistemic, deontic, dynamic and evaluative adjectives in extraposed constructions and post-predicative ones complemented by *to*-infinitives and *that*-clauses. Kataari (2010: 22, 26) noticed, for example, that adjectives expressing epistemic modality (e.g. *clear*, *likely*, *true* or *unlikely*) in extraposed patterns are almost exclusively complemented by *that*-clauses (see Herriman 2000a: 592 and Mair 1990: 25 for a similar observation), whereas those expressing deontic modality (e.g. *essential*, *important*, *necessary*, *desirable*, or *useful*) are far more commonly complemented by *to*-clauses.

However, since these studies were limited in terms of their area of interest and the data examined, there is still a need to quantify adjectives in these types of *it*-extraposition and to determine subtle distributional variations in their use in academic discourse. The primary reason for carrying out such a quantitative investigation is that the meanings of adjectives contribute substantially to the meanings of the construction. For example, the adjectives in (3), (4), (5), and (6) directly affect the understanding of the illustrative sentences by attributing the meanings of importance, difficulty, obviousness, or likelihood to the construction under study.

Hence, the quantitative investigation of adjectives and their semantic description with respect to the semantic frames they evoke may allow for the identification of subtle distributional differences in their use and for the comprehension of their role in the constructions under study, as well as for the advancement of our knowledge and understanding of the meanings and functions of the patterns investigated.

Given that many adjectives occur in both of these constructions, we should expect that the two constructions are to some extent synonymous. However, there are significant differences between the *it is ADJ to V*-construction and the *it is ADJ that*-construction with respect to the semantic constraints they impose on the adjectives that can occur in them. The meaning of the adjectives is the primary factor precipitating the choice between these two patterns. Thus, the frame-semantic information on lexical units that occur in them may play a fundamental role in predicting the discrepancies between these constructions with respect to their preferred adjectives. For example, as noted above, adjectives that occur in the *it is ADJ to V*-construction may evoke the IMPORTANCE frame or the DIFFICULTY frame, whereas adjectives that occur in the *it is ADJ that*-construction may reflect not only the OBVIOUSNESS frame, but also the LIKELIHOOD frame. Consequently, on the basis of this frame-semantic information, we would predict that the *it is ADJ to V*-construction will prefer some adjectives denoting importance and difficulty, while the *it is ADJ that*-construction will prefer adjectives meaning likelihood or obviousness. These predictions will be tested below.

6. Results and discussion

The concordancer retrieved 9384 occurrences of the *it is ADJ to V*-construction and 5321 occurrences of the *it is ADJ that*-construction. In other words, the occurrence of the former turned out to be approximately twice as frequent as the latter in the academic section of COCA. The observed frequencies derived from the calculation of the tokens of adjectives in both constructions indicate that the pattern with *to*-infinitives collocates with 336 types of adjectives, out of which 139 types occurred only once in the construction in question. By contrast, the pattern with *that*-clauses combines with 167 types of adjectives, out of which 66 types were used only once with this construction. This in turn means that many adjectives are rather loosely associated with both patterns, and that the remaining ones are more strongly attracted to one of these near-synonymous constructions.

6.1. Findings for the *it is ADJ to V*-construction

The findings of this study empirically validate the hypothesis suggesting the existence of the collexemes distinguishing between the *it is ADJ to V*-construction and the *it is ADJ that*-construction. Furthermore, the specific hypotheses about the semantic discrepancies between the patterns and about the semantic constraints they impose on the adjectives co-occurring with them are also supported. Table 2 below displays the twenty most distinctive collexemes of the pattern with *to*-infinitives, the observed frequencies used to calculate the direction of association (attracted or repelled) and the strength of association (the distinctiveness of adjectives), the expected frequencies for each adjective: (a) and (c), as well as the findings of the distinctive-collexeme analysis ($P_{\text{Fisher exact}}$).

For the *it is ADJ to V*-construction, it was uncovered that the ten most distinctive adjectives are *difficult*, *important*, *necessary*, *easy/easier*, *hard/harder*, *impossible*, *useful*, *reasonable*, *good/better/best*, and *safe*. The p-values taken to be indicators of their distinctiveness are very small: 1.80E-214, 6.42E-119, 2.83E-96, 2.95E-90, 8.09E-88, 1.63E-83, 1.37E-39, 1.47E-31, 5.90E-28, and 2.13E-22, respectively. A comparison of the observed and the expected frequencies of each of these adjectives and each of the two extraposed constructions shows us that the adjectives occur more frequently than expected in the *it is ADJ to V*-construction and less frequently than expected in the *it is ADJ that*-construction. In other words, they are highly significant, very strongly distinctive collexemes of the former as compared to the latter. Note also that *difficult* is the strongest collexeme for the *it is ADJ to V*-construction, since its p-value resulting from the calculation of the Fisher exact test is exceptionally small (p = 1.80E-214), and the expected frequency is lower than the observed frequency in the pattern in question: that is, this lexeme occurs more frequently than by chance in the construction with *to*-infinitives and less frequently than predicted in the construction complemented by *that*-clauses.

Table 2. The twenty most distinctive collexemes of the *it is ADJ to V*-construction

rank	adjective	a	c	e	f	x	y	z	b	d	(a)	(c)	$P_{Fisher\ exact}$
1.	difficult	1076	0	1076	14149	9834	5391	15225	8758	5391	695.0	381.0	1.80E-214
2.	important	1946	350	2296	12929	9834	5391	15225	7888	5041	1483.0	813.0	6.42E-119
3.	necessary	823	57	880	14345	9834	5391	15225	9011	5334	568.4	311.6	2.83E-96
4.	easier/easy	463	0	463	14762	9834	5391	15225	9371	5391	299.1	163.9	2.95E-90
5.	hard/harder	451	0	451	14774	9834	5391	15225	9383	5391	291.3	159.7	8.09E-88
6.	impossible	483	6	489	14736	9834	5391	15225	9351	5385	315.9	173.1	1.63E-83
7.	useful	204	0	204	15021	9834	5391	15225	9630	5391	131.8	72.2	1.37E-39
8.	reasonable	228	10	238	14987	9834	5391	15225	9606	5381	153.7	84.3	1.47E-31
9.	good/better/best	217	12	229	14996	9834	5391	15225	9617	5379	147.9	81.1	5.90E-28
10.	safe	114	0	114	15111	9834	5391	15225	9720	5391	73.6	40.4	2.13E-22
11.	helpful	114	1	115	15110	9834	5391	15225	9720	5390	74.3	40.7	1.25E-20
12.	tempting	90	0	90	15135	9834	5391	15225	9744	5391	58.1	31.9	1.17E-17
13.	appropriate	99	7	106	15119	9834	5391	15225	9735	5384	68.5	37.5	4.89E-12
14.	fair	72	2	74	15151	9834	5391	15225	9762	5389	47.8	26.2	8.71E-12
15.	common	52	0	52	15173	9834	5391	15225	9782	5391	33.6	18.4	1.53E-10
16.	worthwhile	52	0	52	15173	9834	5391	15225	9782	5391	33.6	18.4	1.53E-10

17.	unrealistic	43	0	43	15182	9834	5391	15225	9791	5391	27.8	15.2	1.24E-08
18.	wise	33	0	33	15192	9834	5391	15225	9801	5391	21.3	11.7	9.61E-07
19.	wrong	36	1	37	15188	9834	5391	15225	9798	5390	23.9	13.1	2.64573E-06
20.	interesting	164	46	210	15015	9834	5391	15225	9670	5345	135.6	74.4	2.27408E-05

Note!

a = Observed frequency of adjectives (e.g. *difficult*) in the *it is ADJ to V*-construction; **b** = Frequency of all other adjectives in the *it is ADJ to V*-construction; **c** = Observed frequency of adjective (e.g. *difficult*) in the *it is ADJ that*-construction; **d** = Frequency of all other adjectives in the *it is ADJ that*-construction; **e** = Total frequency of adjective (e.g. *difficult*); **f** = Total frequency of all other adjectives; **x** = Total frequency of the *it is ADJ to V*-construction; **y** = Total frequency of the *it is ADJ that*-construction; **z** = Total frequency of both constructions; **(a)** = Expected frequency of adjective (e.g. *difficult*) in the *it is ADJ to V*-construction; **(c)** = Expected frequency of adjective (e.g. *difficult*) in the *it is ADJ that*-construction; $\mathbf{P_{Fisher\ exact}}$ = index of distinctive collostructional strength.

In accordance with the prediction, the group with the highest type frequency occupying the top positions in the collexeme ranking is constituted by adjectives denoting ease and difficulty. *Difficult, easy/easier*, and *hard/harder* in ranks 1, 4, and 5 fall into this significant category of distinctive collexemes of the *it is ADJ to V*-construction. As can be seen in Table 2, no occurrences of these adjectives were observed in the *it is ADJ that*-construction. Hence, the adjectives are most distinctive for the pattern with *to*-infinitives in direct comparison with the pattern with *that*-clauses. These lexemes evoke the DIFFICULTY frame, a situation in which a particular activity is evaluated by a speaker as being difficult or easy in certain circumstances, as in *It is difficult to* [*quantify the impact of arts education experiences in one year*] ACTIVITY.

Easy/easier in rank 4 is preceded by *important* and *necessary*, ranked second and third, thus confirming the prediction that the pattern with *ing*-clauses may exhibit a definite preference for some adjectives expressing importance. As the observed frequencies indicate, both adjectives occur much more frequently in the *it is ADJ to V*-construction than in the *it is ADJ that*-construction. The first adjective reflects the IMPORTANCE frame, whereas the latter instantiates the NECESSITY frame. The IMPORTANCE frame is concerned with a factor affecting a certain undertaking, a goal-oriented activity or the maintenance of a desirable state. This factor is judged by a speaker to be important or significant, as in *It is important to* [*examine the influence that peers have over college students' sartorial purchasing choices*] FACTOR. The NECESSITY frame refers to a requirement, i.e. a state of affairs, being viewed by a speaker as an important prerequisite for obtaining or gaining something, as in *It is necessary to* [*use valid and reliable assessment tools*] REQUIREMENT.

Hard/harder, ranked number 5, is followed by *impossible*, and *useful* in ranks 6 and 7. *Impossible* invokes the LIKELIHOOD frame concerning the probability of a hypothetical event occurring, i.e. a state of affairs or occurrence whose likelihood is assessed, as in *Unless the expectations are understood, it is impossible to* [*assess fitness relative to these demands*] HYPOTHETICAL EVENT. *Useful*, along with *helpful* in rank 11, can be described relative to the USEFULNESS/ASSISTANCE EVALUATION frame, a situation in which a particular goal (a desirable state of affairs) that a benefited party attempts to achieve is considered helpful or useful by a speaker, as in *It is helpful to* [*recall the central importance of O2 in metabolism*] GOAL.

Another set of the more distinctive collexemes of the pattern with *ing*-clauses consists of adjectives evoking the MENTAL PROPERTY ATTRIBUTION frame. *Reasonable*, its leading collexeme in rank 8, is accompanied by *wise* in rank 18. This semantic frame elaborates on a judgement (usually implicit) that attributes certain mental properties (e.g. rationality) to a person on the basis of that person's behaviour (any action, utterance, belief, or artifact), as in *It is reasonable to* [*suggest that it might take years for a fundamental reconstruction*] BEHAVIOUR.

The ninth position in the ranking list is occupied by the adjective *good*, along with its comparative and superlative forms, whose meaning can be interpreted

with reference to at least four semantic frames: MORALITY EVALUATION, SOCIAL INTERACTION EVALUATION, DESIRABILITY, and USEFULNESS EVALUATION. The MORALITY EVALUATION frame, which is also evoked by *wrong* in rank 19, refers to an evaluee being judged by a judge (usually implicit) with respect to the morality or rightness of his or her behavior, as in *This verse indicates that it is good to* [*eat all healthy and tasty foods, which God permits us to eat*] BEHAVIOUR. In the SOCIAL INTERACTION EVALUATION frame an evaluee is evaluated by a judger to be of a particular character based on her or his behavior, as in *It is good to* [*have them here together to provide us with a comparative perspective*] BEHAVIOUR. In the DESIRABILITY frame a certain state of affairs is implicitly judged good or bad relative to a set of circumstances, as in *I figured out that it is better to* [*calm down and talk to the student when they do something they are not supposed to do*] STATE OF AFFAIRS. In the USEFULNESS EVALUATION, as discussed above, a particular goal (a desirable state of affairs) is evaluated by a speaker as being useful: for example, as in *It is good to* [*use different types of feed readers because each reader has its advantages and disadvantages*] GOAL.

The next two highly distinctive collexemes of the pattern in question are adjectives denoting risk and suitability. *Safe*, ranked number 10, holds the highest position among the most distinctive adjectives belonging to these semantic categories. This adjective can be described with reference to a structured background of knowledge related to the RISK EVALUATION frame. This knowledge concerns an action evaluated with respect to how risky it is to the affected party in certain circumstances, as in *It is safe to* [*administer live vaccines to the children of pregnant women*] ACTION. The second adjective is *appropriate*, ranked number 13, whose meaning can be relativized to the SUITABILITY EVALUATION frame. In this semantic frame a certain action, or a state of affairs, is judged to be more or less suitable for some purpose or user, as in *It is appropriate to* [*employ accelerandos and ritardandos when the original performance has such tempo changes*] ACTION.

The ranking list of distinctive collexemes for the pattern under discussion also contains a range of adjectives evoking different semantic frames associated with various types of evaluation: CORRECTNESS (*tempting* in rank 12), FAIRNESS (*fair* in rank 14), TYPICALITY (*common* in rank 15), REALISM (*unrealistic* in rank 17), and LEGALITY (*wrong* in rank 19). The first frame concerns a piece of information that is judged correct or sensible but is probably wrong, as in *It is tempting to* [*call anti-Americanism a stereotype or a prejudice*] INFORMATION, *but it is much more than that.* The second frame refers to a coherent structure of concepts that are related to an action being evaluated with respect to how fairly, justly or equitably this action affects the affected party in a particular situation, as in *Accordingly, it is fair to* [*treat an offender who causes harm as if he had "assumed the risk" of a variable penalty*] ACTION *when he chose to behave culpably.* The third frame has something to do with a state of affairs being evaluated with regard to whether it exhibits the essential characteristics of a set of individuals, as in *It is common to* [*see many older visually handicapped individuals working in the factory*] STATE OF AFFAIRS.

In the REALISM frame, activated by *unrealistic*, a state of affairs is judged based either on hopes and wishes or on facts as they really are, as in *It is unrealistic to* [*envision a system where every degree completer obtains only the credits required*] STATE OF AFFAIRS. In the LEGALITY frame, invoked by *wrong*, the status of an action is described with respect to a set of rules laws or regulations determining whether this action is allowed or prohibited, as in *It is wrong to* [*discriminate on the basis of these particular ascriptive features*] ACTION.

Finally, apart from the adjectives forming impersonal judgements, the bottom of the ranking list in Table 2 includes the adjectives *worthwhile* and *interesting* in ranks 16 and 20. Both adjectives can be understood with respect to the MENTAL STIMULUS frame. This frame is concerned with a stimulus, i.e. an event or a state of affairs, that brings about a particular emotion or experience in an experiencer, as in *It is interesting to* [*reflect on what aspects of guided reading tend to be easiest or hardest for teachers to take on*] STIMULUS, or as in *It is worthwhile to* [*couch these results about social cues in terms of past research*] ACTION.

6.2. Findings for the *it is ADJ that*-construction

Concerning the *it is ADJ that*-construction, the results clearly confirm the hypothesis predicting adjectives denoting obviousness in the majority of the top ranks of the collexeme list. The results of the distinctive collexeme analysis for the twenty most strongly attracted collexemes of the *it is ADJ that*-construction in direct comparison with the *it is ADJ to V*-construction are rendered in Table 3. As can be observed, *clear* is the most distinctive collexeme of the pattern with *that*-clauses, as the p-value resulting from the calculation of the Fisher exact test for this adjective is small: $p = 0$ (i.e. infinite collostruction strength), and a comparison of the observed values with the expected ones shows that *clear* occurs more frequently than expected by chance in the *it is ADJ that*-construction as compared to the pattern with *ing*-clauses. This and other adjectives strongly associated with this construction, such as *evident*, *apparent*, and *obvious* in ranks 6, 7, and 8, evoke the OBVIOUSNESS frame. In this frame a certain phenomenon, i.e. an entity or facts, is understood in terms of the degree of likelihood that it will be perceived or known, given a particular piece of evidence, as in *However, it is clear* [*that the treaty right burdens not only governmental actions but also private property*] PHENOMENON.

In accordance with the prediction, the next group in the ranking is constituted by a range of adjectives denoting likelihood. *Likely* in rank 4, the most significant collexeme of this group, is accompanied by *unlikely*, *probable*, *conceivable*, and *possible* in ranks 5, 10, 12, and 19, respectively. These adjectives activate the LIKELIHOOD frame in which the probability of a hypothetical event, i.e. a state of affairs or occurrence, is indirectly evaluated by a judge, i.e. a speaker, as in *But it is likely* [*that more than half of the firms in the United States have experienced breaches*] HYPOTHETICAL EVENT.

In addition to the collexemes denoting obviousness and likelihood, the top of the table contains adjectives such as *certain*, *true* and *doubtful*, ranked second, third and ninth. These adjectives are highly distinctive for the pattern in question in direct comparison with the pattern with *ing*-clauses, in which they are used extremely infrequently. *Certain* and *doubtful* invoke the CERTAINTY frame, whereas *true* the ADMITTING FACT frame. The first frame concerns a person's certainty about the correctness of beliefs or expectations, as in *It is certain* [*that Lincoln saw the holding of slaves as being opposed to the total drift of morality in an extended free republic*] CONTENT. The content denotes the mental content that the person is certain or uncertain about. The latter frame assumes that a particular fact or statement is acknowledged as correct or true by a speaker, as in *It is true* [*that the metaphor had a somewhat wide usage in the culture*] FACT.

As shown in Table 3, the adjective *imperative* is also among the strongly attracted collexemes of the *it is ADJ that*-construction, occupying rank 11, hence being more distinctive for this construction as compared to the construction followed by *to*-infinitives. The findings for the *it is ADJ to V*-construction, however, have revealed that the collexeme list for this construction also includes one lexeme belonging to the same semantic category, i.e. *necessary* in rank 3 (see Table 2). *Imperative*, like *necessary*, can be understood relative to background knowledge concerning the NECESSITY frame, a situation in which a dependent state of affairs has a requirement as a prerequisite for obtaining or occurring, as in *It is imperative* [*that elders and their families understand their personal rights and responsibilities*] STATE OF AFFAIRS.

Moreover, in comparison to the pattern with *to*-infinitives, a further set of highly distinctive collexemes consists of *understandable* and *significant* in ranks 14 and 15, which follow *conceivable* and precede the adjectives *inevitable* and *undeniable* in the two subsequent ranks. The adjective *understandable* in turn can be interpreted with reference to the background information on UNDERSTANDABILITY. In this frame a state of affairs is evaluated as being normal, natural, expected or accepted under certain circumstances with regard to a particular feature or a set of individuals, as in *With more funds to spend and lower pricing, it is understandable* [*that a significant percentage of firm-ordered e-books fell into the HSS subject area*] STATE OF AFFAIRS. The adjective *significant* can be relativized to the IMPORTANCE frame in which the significance of a certain factor is assessed by a speaker, as in *It is significant* [*that the description of the graduation departs from a literary quotation*] FACTOR.

Inevitable, ranked number 16, invokes the UNAVOIDABILITY frame, while *undeniable*, ranked number 17, instantiates the UNQUESTIONABLE TRUTH frame. The first frame has something to do with a situation (usually undesirable) that cannot be avoided or prevented by an agent under certain circumstances, as in *It is inevitable* [*that another version of pseudo-conservatism will appear on the American political scene*] UNDESIRABLE SITUATION. The latter in turn is related to a fact, or a state of affairs, that cannot be denied, questioned, or disputed by an agent under certain circumstances, as in *It is undeniable* [*that the tourism industry has benefited many in Bali*] FACT.

Table 3. The twenty most distinctive collexemes of the *it is ADJ that*-construction

rank	adjective	a	c	e	f	x	y	z	b	d	(a)	(c)	$P_{Fisher\ exact}$
1.	clear/clearer	1	921	922	14303	9834	5391	15225	9833	4470	59553.02	326.4698	0
2.	certain	3	52	55	15170	9834	5391	15225	9831	5339	35.52512	194.7488	0
3.	true	8	300	308	14917	9834	5391	15225	9826	5091	198.9407	109.0593	1.44E-124
4.	likely	159	537	696	14529	9834	5391	15225	9675	4854	449.5543	246.4457	1.68E-117
5.	unlikely	44	334	378	14847	9834	5391	15225	9790	5057	244.1545	133.8455	2.33E-104
6.	evident	0	168	168	15057	9834	5391	15225	9834	5223	108.5131	59.4869	3.24E-77
7.	apparent	1	153	154	15071	9834	5391	15225	9833	5238	99.47034	54.52966	2.54E-68
8.	obvious	0	122	122	15103	9834	5391	15225	9834	5269	78.80118	43.19882	4.01E-56
9.	doubtful	0	93	93	15132	9834	5391	15225	9834	5298	60.06975	32.93025	6.97E-43
10.	probable	0	75	75	15150	9834	5391	15225	9834	5316	48.44335	26.55665	1.09E-34
11.	imperative	72	186	258	14967	9834	5391	15225	9762	5205	166.6451	91.35488	1.63E-33
12.	conceivable	0	69	69	15156	9834	5391	15225	9834	5322	44.56788	24.43212	5.83E-32
13.	noteworthy	1	59	60	15165	9834	5391	15225	9833	5332	38.75468	21.24532	7.98E-26
14.	understandable	0	43	43	15182	9834	5391	15225	9834	5348	27.77419	15.22581	3.67E-20
15.	significant	11	65	76	15149	9834	5391	15225	9823	5326	49.08926	26.91074	2.06E-19
16.	inevitable	0	38	38	15187	9834	5391	15225	9834	5353	24.54463	13.45537	6.75E-18

17.	undeniable	0	34	34	15191	9834	5391	15225	9834	5357	21.96099	12.03901	4.37E-16
18.	surprising	9	46	55	15170	9834	5391	15225	9825	5345	35.52512	19.47488	2.28E-13
19.	possible	1015	715	1730	13495	9834	5391	15225	8819	4676	1117.4	612.6	6.68E-08
20.	arguable	0	27	27	15198	9834	5391	15225	9834	5364	17.43961	9.560394	6.42E-13

Another group in the ranking is constituted by adjectives evoking the STIMULUS FOCUS frame. *Noteworthy*, its central collexeme in rank 13, is followed by s*urprising*, occupying rank 18. In this frame a certain stimulus brings about a particular emotion or experience in an experiencer in some circumstances, as in *Given these sources of variation, it is surprising* [*that little heterogeneity was found in many analyses*] STIMULUS, or as in *It is noteworthy* [*that coursework was not related to participants' perceptions in these areas*] STATE OF AFFAIRS.

Finally, the collexeme list includes the adjective *arguable* in rank 20, which is highly distinctive for the pattern under consideration as opposed to the pattern with *to*-infinitives. The adjective can be described in relation to the BEING POSSIBLY TRUE frame, a semantic frame in which a certain content, i.e. a course of action, a state of affairs, or a proposition that an addressee is to believe, is judged by an arguer as being possibly true because there are sound reasons for believing that this content is true, as in *From a philosophical standpoint it is arguable* [*that no such positive ethical right exists*] CONTENT.

7. Conclusion

In conclusion, the findings of the distinctive-collexeme analysis of the *it is ADJ to V*-construction and the *it is ADJ that*-construction point to a far-reaching effect of the frame-semantic knowledge associated with the adjectives under study upon the choice between the two nearly synonymous patterns. Moreover, the results confirm the hypothesis predicting the occurrence of highly distinctive collexemes for each of the two constructions in question, and they suggest that both constructions communicate subtle nuances of meaning, serve different functions in academic discourse, and place particular semantic restrictions upon the adjectives with which they co-occur.

It was found that some semantic categories of adjectives allow both *to*-infinitives and *that*-clauses in academic prose (e.g. *important*, *necessary*, *interesting*, *likely*, *unlikely*, *imperative*, *possible*), while others require only infinitival complements (e.g. *difficult*, *easy*, *hard*, *useful*, *safe*, *tempting*, *common*, *worthwhile*, *unrealistic*, *wise*), or solely finite ones (e.g. *evident*, *obvious*, *doubtful*, *probable*, *conceivable*, *understandable*, *inevitable*, *undeniable*). A possible explanation for this discrepancy may be that finite *that*-clauses generally express facts and as such may be evaluated for their truth, obviousness, likelihood, certainty, unavoidability, etc. By contrast, non-finite *to*-infinitive clauses frequently express actions, which, unlike facts, may be judged difficult, easy, important, necessary, useful, reasonable, safe, wrong, fair, etc.

Regarding the *it is ADJ to V*-construction, it was uncovered that the most distinctive collexemes are adjectives invoking the DIFFICULTY frame (*difficult*, *easy/easier*, *hard/harder*), the USEFULNESS/ ASSISTANCE EVALUATION frame (*useful, helpful*), and the MENTAL PROPERTY ATTRIBUTION frame (*reasonable, wise*). The other distinctive collexemes for this construction are adjectives

denoting necessity (*necessary*), importance (*important*), likelihood (*impossible*), morality (*good*), risk (*safe*), and suitability (*appropriate*), as well as adjectives activating different semantic frames pertaining to various forms of judgement: CORRECTNESS (*tempting*), FAIRNESS (*fair*), TYPICALITY (*common*), REALISM (*unrealistic*), and LEGALITY (*wrong*).

In contrast to the pattern with *that*-clauses, the *it is ADJ to V*-construction exhibits a strong preference for adjectives meaning difficulty, usefulness, reasonableness, or risk. However, despite this tendency, this construction appears not to impose too many semantic restrictions upon the adjectives with which it collocates, thereby occurring with a more extensive range of adjectives than the former. As noted in section 7, the pattern with *to*-infinitives collocates with 336 types of adjectives, whereas the pattern with *that*-clauses combines with 167 types. The simplest explanation for this may lie in the economy of language, i.e. a general tendency in grammar towards clarity and simplicity. Sentences with *to*-infinitives are generally shorter and less complex than those followed by *that*-clauses, as has been observed in the academic subcorpus. Thus, this construction allows speakers, or writers, to express a message more quickly and clearly as well as with greater precision. In addition, it is likely that speakers and writers choose the *it is ADJ to V*-construction as an alternative to the *it is ADJ that*-construction, and that they use a larger variety of adjectives with the former, in order to avoid confusion, lengthy explanations, and more complex sentences.

With regard to the *it is ADJ that*-construction, we find that the ten most distinctive collexemes are *clear/clearer*, *certain*, *true*, *likely*, *unlikely*, *evident*, *apparent*, *obvious*, *doubtful*, and *probable*. These and other adjectives significantly attracted to this pattern (such as *noteworthy*, *conceivable* or *true*) are used with *that*-clauses to serve a specific function: in other words, to convey new facts that have no direct link with the preceding context or states of affairs that have been explicitly evoked in the preceding verbal discourse. Similarly, the other distinctive collexemes for this construction (e.g. *imperative*, *noteworthy*, *undeniable*, *arguable*) seem clearly to confirm the function of the pattern mentioned above, and thus the claim that it is a functional construction that primarily prefers adjectives denoting obviousness, likelihood, certainty, necessity, unavoidability, or truth over adjectives pertaining to difficulty, usefulness, risk, or suitability.

In comparison with the pattern of *to*-infinitives, the *it is ADJ that*-construction appears to impose more semantic restrictions upon the adjectives with which it occurs, thus collocating with a more limited number of adjectives. For example, the pattern complemented by *that*-clauses exhibits a noticeable preference for adjectives evoking the OBVIOUSNESS frame (*clear*, *evident*, *apparent*, *obvious*), the LIKELIHOOD frame (*likely*, *unlikely*, *probable*, *conceivable*, *possible*), or the CERTAINTY frame (*certain* and *doubtful*) over adjectives reflecting the DIFFICULTY frame (*difficult*, *easy/easier*, *hard/harder*), the USEFULNESS frame (*useful* and *helpful*), or the SUITABILITY frame (*appropriate*). The primary reason

for this tendency may be that, as noted above, the *it is ADJ that*-construction expresses facts rather than actions. Thus, this function seems to be a crucial factor determining the construction's preference for adjectives expressing some speaker's or writer's evaluation of facts, i.e. their likelihood, transparency, truth or certainty. This claim can be also substantiated by indicating the tendency of the *it is ADJ that*-construction to co-occur very frequently with a group of adjectives expressing impersonal judgements, such as *inevitable*, *undeniable*, *surprising*, and *arguable*, coming from different domains: the UNAVOIDABILITY frame, the UNQUESTIONABLE TRUTH frame, the STIMULUS FOCUS frame, the BEING POSSIBLY TRUE frame, respectively. These evaluative adjectives allow speakers or writers to express beliefs, opinions, or preferences about some facts or states-of-affair by placing such evaluative comments at the beginnings of sentences.

The distinctive-collexeme analysis employed in this investigation has proved to be an effective technique for the identification of the most distinctive lexemes of the *it is ADJ to V*-construction and the *it is ADJ that*-construction, and hence could be adopted elsewhere for the determination of the most significant lexemes occurring in different types of *it*-extraposed constructions. Future research might, for example, focus on determining the adjectives distinctive for patterns complemented by *to*-infinitives, *for/to*-infinitives, *wh*-clauses, *-ing* clauses, or NP + relative clauses. Although the current investigation was limited to academic prose, it would also be interesting to explore the distribution of adjectives in the constructions under scrutiny across different types of both written and spoken registers, in view of the possible existence of additional slight variations in their occurrence. Future studies may even determine the most distinctive adjectives of these patterns 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. 2010. *Language, usage, and cognition*. Cambridge: Cambridge University Press.

Collins, P. 1994. Extraposition in English. *Functions of Language* 1: 7-24.

Diessel, H. 2015. Usage-based construction grammar. In E. Dąbrowska and D. Divjak (eds.), *Handbook of cognitive linguistics*, 295-321. Berlin: Mouton de Gruyter.

Emonds, J. 1976. *A transformational approach to English syntax*. New York: Academic Press.

Erdmann, P. 1990. *Discourse and grammar: Focusing and defocusing in English*. Tübingen: Niemeyer.

Evert, S. 2009. Corpora and collocations. In A. Ludeling and M. Kytö (eds.), *Corpus linguistics: An international handbook, Vol. 2*, 1212-1248. Berlin, Germany: Mouton de Gruyter.

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, C.J., R. Lee-Goldman and R. Rhodes 2012. The framenet construction. In H.C. Boas and I. Sag (eds.), *Sign-based construction grammar*, 309-372. Stanford: CSLI Publications.

Goldberg, A. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, A. 2013. Constructionist approaches to language. In T. Hoffmann and G. Trousdale (eds.), *Handbook of construction grammar*, 15-48. Oxford: Oxford University Press.

Gómez-González, M.Á. 1997. On subject *it*-extrapositions: Evidence from present-day English. *Revista Alicante de Estudios Ingleses* 10: 95-107.

Gómez-González, M.A. 2001. *The theme-topic interface. Evidence from English*. Amsterdam and Philadelphia: John Benjamins.

Gries, S.Th. 2012. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language* 36(3): 477-510.

Gries, S.Th. 2015. More (old and new) misunderstandings of collostructional analysis: on Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26(3): 505-536.

Gries, S.Th., and A. Stefanowitsch 2004. Extending collostructional analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics* 9(1): 97-129.

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.

Hilpert, M. 2014. Collostructional analysis: Measuring associations between constructions and lexical elements. In D. Glynn and J. Robinson (eds.), *Polysemy and synonymy. Corpus methods and applications in cognitive linguistics*, 391-404. Amsterdam: John Benjamins.

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*. A master's thesis. Uppsala University, Department of English.

Kaltenböck, G. 2000. *It*-extraposition and non-extraposition in English discourse. In C. Mair and M. Hundt (eds.), *Corpus linguistics and linguistic theory*, 157-175. Amsterdam and Atlanta: Rodopi.

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.

Kaltenböck, G. 2006. ...That is the question: complementizer omission in extraposed that-clauses. *English Language and Linguistics* 10(2): 371-396.

Küchenhoff, H., and H.-J. Schmid 2015. Reply to 'more (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff' by Stefan Th. Gries'. *Cognitive Linguistics* 26(3): 537-547.

Levshina, N. 2015. *How to do linguistics with r: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

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. 2008. Adjective complementation by that-clauses: The relation between the se-mantics of adjectives and the verb phrase in the that-clause. *International Journal of English Studies* 19(1): 141-155.

Mindt, I. 2011. *Adjective complementation: An empirical analysis of adjectives followed by that-clauses*. Philadelphia: John Benjamins.

Pedersen, T. 1998. Dependent bigram identification. *Proceedings of the fifteenth national conference on artificial intelligence* (AAAI-98), Madison, Wl, July 28-30.

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.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik 1985. *A comprehensive grammar of the English language*. New York and London: Longman.

Rivière, C. 1983. Modal adjectives: Transformations, synonymy, and complementation, *Lingua* 59: 1-45.

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. 1999. Extraposition in English revisited. *Neuphilologische Mitteilungen* 100: 51-66.

Sheskin, D.J. 2011. *Handbook of parametric and non-parametric statistical procedures*. 5th Edition. London: Chapman & Hall/CRC.

Stefanowitsch, A. 2013. Collostructional analysis. In T. Hoffmann and G. Trousdale (eds.), *Handbook of construction grammar*, 290-306. Oxford: Oxford University Press.

Van Linden, A. 2012. *Modal adjectives: English deontic and evaluative constructions in synchrony and diachrony*. Berlin and Boston: Walter de Gruyter.

Van Linden, A., and K. Davidse 2009. The clausal complementation of deontic-evaluative adjectives in extraposition constructions: A synchronic-diachronic approach. *Folia Linguistica* 43: 171-211.

Whittacker, R. 1995. Theme, processes and the realization of meanings in academic articles. In M. Ghadessy (ed.), *Thematic development in English texts*, 105-128. London: Pinter.

Wiliński, J. 2017. Normal and anomalous occurrences of adjectives in extraposed constructions with *to*-infinitive clauses: A quantitative corpus-based study. In J. Wiliński and J. Stolarek (eds.), *Norm and anomaly in language, literature and culture*, 89-106. Frankfurt am Main: Peter Lang.