Linguistica Silesiana nr 44/1, 2023
ISSN 0208-4228
DOI: 10.24425/linsi.2023.144822

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

IT-EXTRAPOSITION WITH PAST PARTICIPLES AND *THAT*-CLAUSES: A QUANTITATIVE CORPUS-BASED ANALYSIS

Applying frame semantics, usage-based construction grammar, and quantitative corpus-based methodology, this article seeks to explore the nature of the extraposed construction with past participles complemented by *that*-clauses. To this end, the author extracts the occurrences of the *It BE* V_{en} *that*-construction from the Corpus of Contemporary American English (COCA), determines its structural, semantic, distributional, and discourse-functional features, and identifies verbs that are strongly associated with the construction in question. The study meaningfully contributes to a growing body of research on *it*-extraposition by conducting a qualitative and quantitative analysis of one of its variants, a grammatical pattern with past participles that has not been hitherto investigated in much detail from a quantitative corpus-based perspective.

Keywords: *it*-extraposition, construction grammar, frame semantics, COCA, corpus-based analysis

1. Introduction

Over the past few decades, much research has explored the nature of different extraposed constructions in English (Quirk et al. 1985; Seppänen 1999; Biber et al. 1999; Herriman 2000a,b; Kaltenböck 2004, 2005; Van Linden 2012). Some studies have identified their structural and discourse-functional properties (Collins 1994; Gómez-González 1997; Herriman 2000a,b; Kaltenböck 2005), while others have revealed the occurrences of many epistemic, deontic, and evaluative adjectives in *it*-extraposition (Biber et al. 1999; Herriman 2000a; Kaltenböck 2005; Van Linden 2012). Some research has also focused on

quantifying such adjectives in extraposed constructions complemented by *that*-clauses and/or *it*-clauses (Hilpert 2014; Wiliński 2018a, 2018b).

However, scant attention has hitherto been paid to the *It BE* V_{en} *that*-construction, a specific variant of the extraposed construction in English. To the best of the author's knowledge, the nature of this construction has not been exhaustively studied thus far: some researchers have solely mentioned this pattern in passing or have given one or two examples of its usage (e.g., Biber et al. 1999: 1020; Kaltenböck 2005: 146; Bierwiaczonek 2016: 179). Even *Collins Cobuild English Grammar* (2017: 370) based on the Collins Corpus has solely provided an alphabetical list of some reporting verbs (e.g., *accept, assume, believe, consider, discover, find, know, note,* and several others) that are used in the passive with *it* as their subject. However, this list does not include any frequency information about their occurrences in the construction. To date, no single research has been specifically designed to examine different features of the *It BE* V_{en} *that*-construction in greater detail, to quantify past participles in this construction, and to identify the most frequent patterns of its usage.

Thus, because of the limitations in the previous research, this study seeks to attain two goals: first, to identify structural, semantic, pragmatic, distributional, and discourse-functional properties of this construction, and second, to determine those past participles that are strongly attracted to the pattern under consideration. More specifically, on the basis of the data retrieved from the Corpus of Contemporary American English (COCA), the research aims to show that the construction at issue has a specific structure and meaning, serves various functions in discourse, is distributed differently across various registers in COCA, and occurs with specific categories of verbs evoking different semantic frames.

The rest of the article is organized as follows. The theory and corpus-based methodology are briefly discussed in Section 2. Data sources, their nature, and the tools and procedures employed for their retrieval and evaluation are elucidated in Section 3. Structural, semantic, distributional and discourse-functional properties of the *It BE* V_{en} *that*-construction are considered in Section 4. The findings of the quantitative investigation of participles occurring in this construction are presented and assessed in Section 5. Finally, concluding remarks are provided in Section 6.

2. Theory and methodology

In this study, usage-based construction grammar and the theory of frame semantics are used for identifying the structural and semantic properties of the construction in question. Usage-based models of grammar (e.g., as defined by Barlow and Kemmer 2000; Tummers, Heylen, and Geeraerts 2005; Bybee and

Beckner 2010; see also Szcześniak 2014; Bierwiaczonek 2016; Hoffman 2022) are built upon the assumption that grammar is composed of constructions, pairings of form and meaning/function, which can be classified along a continuum of complexity and schematicity, from morphemes and words through fixed phrases and idioms to the most schematic patterns (e.g., ditransitive constructions) and associated rules of their semantic, pragmatic and discourse-functional interpretation. Hence, for instance, the *it*-extraposition (e.g., *It was argued that…*) is a construction since it has a complex form (e.g., [*It BE* V_{en} *that …*]) and a specific reporting function/meaning (generally, the pattern is used to report that something was said, shown, believed or found, without mentioning who said, believed, or discovered it) that are conventionally associated with each other. All units in grammar can be stored and represented as constructions as long as they are used frequently enough to become deeply rooted in the linguistic system and firmly entrenched in the speaker's mind (e.g., Schmid 2016).

Frame semantics (Fillmore and Baker 2010) identifies lexical meanings with mental representations or schematic knowledge structures referred to as *semantic frames*. This theory assumes that the meaning of a word can be defined relative to the background knowledge associated with that word. For example, we would not be able to comprehend the verb *to steal* without familiarity with the situation of stealing goods, which involves four core elements, such as a perpetrator, source, victim, and goods, and several other peripheral elements, such as a place, purpose, manner, instrument, and means. In this study, this theory is used to define the semantics of the construction in question and the meanings of verbs colligating with this construction. Apart from the HYPOTHESIZING frame and the ACCEPTING_TRUTH frame invented by the author himself, the names of all semantic frames and their modified descriptions were taken from the FrameNet lexical database (Boas 2017; see The FrameNet project in data sources).

Furthermore, this study applies quantitative corpus-based methodology. The method known as *the attraction-reliance measure* (Schmid 2000) is used to gauge the mutual association between past participles and the *It BE* V_{en} *that*-construction. In other words, it is specifically tailored to identify past participles that are more strongly associated with this construction than others and those whose occurrences in this construction are more significant than their uses in other contexts in the corpus. *Attraction* is intended to quantify the degree to which the *It BE* V_{en} *that*-construction attracts a particular past participle, while *reliance* is designed to measure the degree to which a past participle appears in the construction under study versus other patterns or contexts in the corpus (cf. Wiliński 2018a). Although the method involves quantifying data and evaluating them statistically, the findings are interpreted qualitatively and subjectively. In this case, past participles are grouped into semantic sets, and their specific contextual uses in the construction are explained relative to the semantic frames they invoke.

3. Corpus, data extraction, and statistical evaluation

The primary source of data in this study is an earlier version of the well-balanced Corpus of Contemporary American English (COCA). This version covers the years between 1990 and 2017 and includes over 560 million words. It is composed of five registers: fiction (short stories and plays), academic papers (from roughly 100 peer-reviewed journals), spoken transcripts (from a wide variety of TV and radio programs), popular magazines (different magazines covering a wide array of domains such as finance, health, news, sports, religion, fashion, or travel), and newspapers (ten American newspapers: USA Today, San Francisco Chronicle, New York Times, etc.). In March 2020, COCA was updated. It currently contains more than one billion words of text from eight genres: fiction, spoken transcripts, academic texts, newspapers, popular magazines, TV and Movies subtitles, blogs, and other web pages. This updated version of COCA was the source of the illustrative examples analysed in subsection 4.1 and the source of quantitative data for a distributional analysis in subsection 4.3.[1] This distributional analysis included data collected between 1990 and 2019, and it was based on the five registers (spoken discourse, fiction, magazines, newspapers, and academic journals).

The procedure adopted for data retrieval and their statistical evaluation consisted of several steps. First, the observed frequencies of past participles were extracted from the corpus. More specifically, the software program installed in COCA was used for carrying out corpus searches and extracting all the occurrences of past participles in the *It BE* V_{en} *that*-construction. To this end, more than ten different patterns with wildcards for each form of the verb BE (i.e., be, is, 's, was, has been, or had been) were entered separately into the search engine. For example, the pattern (It is * _v?n) was used to extract the combination *It is often said* (…), while the pattern (It has been * _v?n) to find *It has been widely reported* (…). The application of such wildcards allowed for retrieving relatively long combinations, such as *It is well known at my Walmart store that they will fire you* (…), *It has been generally accepted by almost everyone that* (…), or *It is often said of our best actors that they disappear* (…), in which *that*-clauses are not used directly after participles.

Then, all the occurrences of past participles were manually inspected to identify authentic combinations, i.e., the occurrences that precisely reflect grammatical patterns of usage indicated by the wildcards specified in a search query. All false hits (i.e., the occurrences which did not correspond with the *It BE* V_{en} *that*-construction) were discarded from further quantitative analysis. Thus,

[1] The reason why this new version of COCA was used in subsection 4.3 was that the author decided to determine the distribution of the *It was* V_{en} *that*-construction across different registers based on a much larger data set than it had been originally planned.

for instance, the strings *It is highly recommended to use* (…) or *It is well known what* (…), in which past participles are followed by *to*-clauses or *wh*-clauses, were excluded from the study. The observed frequencies of occurrence (e.g., a: the frequency of the past participle *noted* in the *It BE* V_{en} *that*-construction and x: the total frequency of all past participles in the construction), as rendered in Table 1 below, were calculated manually by inspecting concordance lines, whereas the total frequencies of past participles in COCA (e.g., e: the total frequency of the past participle *noted*) were computed automatically by the software program.

Table 1. Cooccurrence table for a quantitative analysis

participle	**a**	**x**	**e**	**attraction**	**reliance**
noted	2136	25820	4018	8.27 %	53.16%

Note: a= The frequency of the past participle *noted* in the *It BE Ven that*-construction; x= The total frequency of all past participles in the construction; e= The total frequency of the past participle *noted* in COCA

After the observed frequencies had been estimated, the numbers (a, x, and e) in Table 1 above were next entered into an Excel worksheet and employed for computing Schmid's measures of attraction and reliance (cf. Wiliński 2018a). *Attraction* was estimated by dividing the raw frequency of a past participle in the *It BE* V_{en} *that*-construction by the total frequency of all past participles in this construction, whereas *reliance* was calculated by dividing the frequency of occurrence of a past participle in the construction in question by its frequency of occurrence in COCA (cf. Schmid 2000: 54). The scores resulting from this calculation were expressed as percentages by multiplying the raw frequency of each past participle in the construction by one hundred. The percentage was deemed to be an index of attraction or reliance: the higher the percentage, the stronger the attraction to, and reliance on, the *It BE* V_{en} *that*-construction (cf. Wiliński 2018a).

A cursory examination of Table 1 reveals that the scores of attraction and reliance for the past participle *noted* are 8.27 % and 53.16%, respectively. This means that the participle occurs in 8.27 % of the occurrences of past participles in the *It BE* V_{en} *that*-construction: in other words, *noted* is a highly significant lexeme, which is very strongly attracted to the construction under consideration. In addition, 53.16% of the occurrences of the same participle are found in this construction, which means that the verb *noted* relies on other constructions in a proportion of 46.84%. In other words, it occurs less frequently in other contexts or grammatical patterns. Finally, the quantitative results were ranked according to the measure of attraction and then assessed qualitatively and subjectively.

4. It-extraposition with past participles

4.1. Properties

Although extraposed constructions have received much treatment in grammar, *it*-extraposition with past participles complemented by *that*-clauses has been relatively neglected. Hence, this section seeks to analyse one specific variant of the extraposed constructions, referred to as the *It BE* V_{en} *that*-construction, the use of which can be illustrated by the following examples retrieved from COCA:

(1) [It is widely understood by the general populace] *that emergency measures and major changes are required*. (BLOG: http://www.theoildrum.com/node/6116)
(2) [It can be said of both] *that they were warnings that the American system of government is more fragile than we think.* (NEWS: Washington Post)
(3) [It was found in this study] *that teachers saw relationships with students as central to their jobs.* (ACAD: American Secondary Education)
(4) [For many years it had been thought] *he was born ten years or so before that.* (MAG: Smithsonian)

The examples given above can be divided into two main components: one is backgrounded and placed in the passive voice construction; the other is foregrounded (focused on) and functions as the complement of the past participle. The specially foregrounded element is provided in italics in the sentences above, while the clause with passive voice is placed in square brackets.

A brief inspection of Table 2 shows that this type of information-packaging construction (IP construction) consists of one fixed lexical item (*it*) and at least three flexible slots that can be filled by various forms of the verb *be,* past participle verb forms, and the complement of the past participle introduced by *that*-clauses:

Table 2. Properties of the *It BE* V_{en} *that*-construction

form	(adverbial) +IT + backgrounded element: BE + (not or AdvP) + past participle + (AdvP or PP)	foregrounded element: *that*-clauses
meaning	something is/was/has been, etc said, shown, believed or found, without signalling who performed this speech act	states/activities/achievements/ accomplishments
IP/function	presupposition/topic/new/background information/reporting function	focus/comment/new or given

The properties of this construction can be represented structurally and schematically in the form of Table 2, where *It* is followed by the backgrounded component [*be* + (not or AdvP) + past participle + (AdvP or PP)] and the foregrounded component (finite *that*-clauses). Occasionally, some optional elements can be inserted in the passive voice construction: e.g., *widely*, *of both*, *in this study*, or *for many years*, as in (1), (2), (3), and (4). As the corpus data reveals, it is possible to omit *that* in the foregrounded part, as in (4). The foregrounded component of the construction is placed in the final position, in accord with the topic-comment structure and the information principles of end-weight and end-focus (Quirk et al. 1985: 863; Biber et al. 1999: 896–898).

The purpose of this construction is to signal explicitly what is considered as foreground information and the main communicative point (i.e., the primary focus of attention):

(5) [It has been mentioned earlier] *that Olfert Dapper had more than a little of the romantic in his nature.* (FIC: Fantasy & Science Fiction)
(6) [It was said about Gladstone, Prime Minister Gladstone,] *that when you had dinner with him, you* […] (SPOK: PBS_Newshour).
(7) [It can be said now, perhaps,] *that Kellie Ann Mann's first crime was falling in love with a guy like Patrick.* (MAG: Time)

In both sentences, the main idea (the most important information) is given at the end, in accordance with the principles of end-weight and end-focus: i.e., short and simple clauses that provide background information are more frequently used in the initial position, while long and complex clauses are generally used in the final position to enable users to process information more easily and efficiently. The passive clauses in (5) and (6) begin with a topic (*it*) and background information by carrying a reference to the preceding context: *It has been mentioned* and *It was said about Gladstone, Prime Minister Gladstone*, respectively. They end with foreground information by adding comments on the topic. In these examples, the foregrounded component does not provide new information, but rather is used to highlight or sum up what has been said or written in the preceding context. By contrast, in (7), both the backgrounded component and the foregrounded one seem to introduce new pieces of information into the discourse. Thus, this corpus evidence confirms that the extraposed construction with past participles can convey given as well as new information. This new piece of information is in particular introduced by the variants with modal verbs (*can, must* or *should*).

Regarding the semantics of the construction in question, the subject position, which is filled by 'dummy' or ‘anticipatory’ *it*, has no lexical meaning here, but merely serves a grammatical function. It is used to report that something is, was, has been, etc said, demonstrated, thought, or found, without indicating who did this. The contextual implication is that this can be a group of persons, people in

general, or sometimes an unspecified person. The past participles that occur in the construction take on a range of meanings associated with saying, knowing, showing, or thinking. They are occasionally used with modals such as *can* or *must* to indicate that the user of a particular sentence is able or obliged to say something, rather than that someone else has mentioned something in the preceding context. The whole passive clause functions as a discourse opener or a starting point for an utterance, and it also introduces the speaker's or writer's evaluative judgment about a particular person, entity, or situation by providing the impersonal comments in the form of dummy *it*, followed by the appropriate form of the verb *be*, past participles, and occasionally prepositional phrases. Lastly, *that*-clauses following the past participles denote situations: states or occurrences (activities, accomplishments, or achievements). They serve an explanatory function by providing new or given information about a particular person, an entity or a state of affairs.

4.2. Variants of the it BE V_{en} that-construction

The corpus evidence reveals that the construction occurs in five main variant forms (*It is* V_{en} *that*-clause, *It was* V_{en} *that*-clause, *It has been* V_{en} *that*-clause, *It had been* V_{en} *that*-clause, and *It modal verb be* V_{en} *that*-clause), each of which tends to be used with specific types of reporting verbs conveying a specific message. Table 3 below lists the twenty most common verbs colligating with the four variants in five sections of COCA (1990-2017: spoken, fiction, magazine, newspaper, and academic), while Table 4 indicates the twenty most frequent combinations with modal verbs.

As shown in Table 3, the first variant exhibits a strong preference for verbs denoting a guess or calculation (*estimated*), expressing a statement (*said, reported, noted,* and *suggested*), designating awareness (*believed, known, thought,* and *understood*), stating an opinion (*expected, believed, assumed,* and *thought*), indicating a desire (*hoped*), denoting advice (*recommended*), referring to reasoning (*argued*), postulating a hypothesis (*hypothesized*), denoting expectation (*expected* or *anticipated*), meaning 'to become aware' (*noted, recognized,* and *seen*), denoting 'to come to believe' (*concluded*), designating 'understanding' (*seen* and *understood*), and pertaining to a prediction (*predicted*). By contrast, the variant in past simple displays a strong tendency to combine with verbs evoking the BECOMING_AWARE frame (*found, discovered, noted,* and *observed*), the ACHIEVING_FIRST frame (*found* and *discovered*), the COMING_TO_BELIEVE frame (*found, determined,* and *concluded*), the STATEMENT frame (*said, reported, announced, noted,* and *suggested*) and the semantic frames of HYPOTHESIZING (*hypothesized*), DECIDING (*decided*), REVEAL_SECRET (*reveal*), AWARENESS (*assumed, thought,* and *believed*), OPINION (*assumed, expected, thought, believed,* and *felt*), EXPECTATION (*expected*), ESTIMATING (*estimated*), and PREDICTING (*predicted*).

Table 3. The twenty most common verbs occurring in the five variants

Variants	Verbs and their frequencies of occurrence
It is V_{en} *that*-clause	estimated (684), said (415), believed (346), expected (340), assumed (317), hoped (304), recommended (282), known (203), suggested (198), thought (165), argued (137), hypothesized (89), reported (87), understood (86), anticipated (83), noted (70), recognized (66), concluded (66), seen (65), predicted (57)
It was V_{en} *that*-clause	found (641), hypothesized (414), reported (308), said (283), discovered (271), decided (268), determined (254), revealed (249), assumed (236), expected (233), announced (227), thought (176), concluded (160), noted (151), believed (141), suggested (130), observed (126), estimated (125), predicted (123), felt (120)
It has been V_{en} *that*-clause	suggested (381), estimated (185), argued (181), shown (173), said (164), reported (151), found (104), noted (81), proposed (64), observed (55), demonstrated (51), assumed (33), determined (32), established (26), hypothesized (24), speculated (23), proven (22), recommended (18), stated (17), known (17)
It had been V_{en} *that*-clause	decided (14), assumed (12), determined (9), reported (9), thought (8), suggested (8), expected (8), hoped (6), revealed (4), agreed (4), established (4), proposed (3), discovered (3), said (3), shown (3), rumored (2), told (2), hypothesized (2), understood (2), designed (2)

The pattern used in present perfect in turn tends to colligate with verbs invoking the following semantic frames: STATEMENT (*suggested, said, reported, noted, proposed*, and *stated*), ESTIMATING (*estimated*), REASONING (*argued, shown, demonstrated*, and *proven*), BECOMING_AWARE (*found, noted,* and *observed*), COMING_TO BELIEVE (*found, determined, established*, and *speculated*), AWARENESS (*assumed* and *known*), HYPOTHESIZING (hypothesized), and ATTEMPT-SUASION (*recommended*). Lastly, the pattern in past perfect is mainly restricted to verbs instantiating semantic frames, such as DECIDING (*decided*), AWARENESS (*assumed* and *thought*), COMING_TO BELIEVE (*determined* and *established*), STATEMENT (*reported, suggested, proposed, said,* and *told*), OPINION (*thought* and *expected*), EXPECTATION (*expected*), DESIRING (*hoped*), REVEAL_SECRET (*revealed*), BE_IN_AGREEMENT_ON_ASSESSMENT (*agreed*), BECOMING_AWARE (*discovered*), REASONING (*shown*), UNATTRIBUTED_INFORMATION (*rumored*), HYPOTHESIZING (*hypothesized*), and COMING_UP_WITH (*designed*).

In addition to these four variants, the study also identified the combinations with modal verbs. As can be seen in Table 4, the most frequent combinations are *should be noted, could be argued*, *can be argued*, *can be said*, *must be noted, can be seen*, *must be remembered, might be argued,* and several others. The modal verbs are used here to indicate that a particular state of affairs is desirable, required, possible or likely to happen, while participles report that something

must, should, can, could, might, or may be noted, argued, said, emphasized, etc, without indicating any person who must, should, can, or may do it. This variant imposes certain constraints upon the verbs with which it colligates. Thus, it is mainly confined to verbs such as *noted, argued, said, seen, remembered, emphasized, concluded, assumed, recognized, shown* and *stressed.*

Table 4. The 20 most frequent combinations with modal verbs

Variant	**The twenty most common combinations with modal verbs**
It modal verb be V_{en} *that*-clause	should be noted (1302), could be argued (308), can be argued (262), can be said (164), must be noted (141), can be seen (141), must be remembered (100), might be argued (94), must be said (93), should be emphasized (92), could be said (87), can be concluded (81), should be remembered (78), may be argued (76), can be assumed (72), must be recognized (56); might be said (55), can be shown (52), should be stressed (46), must be emphasized (43)

4.3. Distribution across different registers

As for distributional properties of the *It BE* V_{en} *that*-construction, Table 5 presents the frequency distribution of the most significant verbs occurring in one specific instantiation of this construction, i.e., *It was* V_{en} *that*-clause. The reason why this variant was selected for distributional analysis is that it is the most frequent pattern of occurrence of the construction in the five sections of COCA, covering the years between 1990 and 2019: i.e., it occurs 7,066 times in total.

Table 5. Distribution of *It was* V_{en} *that*-clause

rank	**patterns of use**	**ALL**	**SPOK**	**FIC**	**MAG**	**NEWS**	**ACAD**
1.	It was found that	737	16	14	51	16	640
2.	It was hypothesized that	426	-	-	1	-	425
3.	It was reported that	363	69	7	78	39	170
4.	It was said that	292	38	152	40	21	41
5.	It was discovered that	286	36	40	75	49	86
6.	It was decided that	283	29	78	43	24	109
7.	It was revealed that	283	73	11	74	74	51
8.	It was determined that	270	27	23	27	33	160
9.	It was announced that	250	78	21	55	58	38

Table 5. cont.

rank	patterns of use	ALL	SPOK	FIC	MAG	NEWS	ACAD
10.	It was assumed that	245	14	13	38	13	167
11.	It was expected that	244	13	8	18	13	192
12.	It was thought that	190	31	13	47	18	81
13.	It was concluded that	172	-	4	4	1	163
14.	It was observed that	162	-	1	5	2	154
15.	It was noted that	158	5	5	17	11	120
16.	It was believed that	154	17	18	31	7	81
17.	It was suggested that	148	12	5	18	24	89
18.	It was estimated that	129	7	5	18	13	86
19.	It was agreed that	127	10	28	22	16	51
20.	It was predicted that	125	1	-	3	1	120

A cursory examination of frequency data in Table 5 reveals that the variant under study predominates in academic registers and tends to occur frequently in written texts, such as magazines and newspapers. However, it is relatively less frequent in fiction and spoken discourse. These results are partially congruent with the findings of the studies into extraposed constructions with *that*-clauses conducted by Biber et al. (1999) and Wiliński (2018a). Wiliński observed that *it*-extraposition with adjectives is commonly used in academic texts, while Biber et al. found that *it*-extraposition with *that*-clauses is most frequent in news and academic prose but less common in fiction and conversation.

As can be observed in Table 5, several exceptions to this rule are patterns of occurrence with the participles *said, decided, revealed, announced,* and *agreed.* For example, *revealed* and *announced* are much more common in spoken discourse than in academic prose, while *said* is more frequent in fiction than in other registers. *Decided* and *agreed* in turn have high frequencies not only in academic corpora, but also in fiction. In addition, the observed frequencies in Table 5 show that the patterns of occurrence with *hypothesized, concluded, observed,* and *predicted* are mainly restricted to academic registers. The close association between the verbs and these registers may have something to do with their formality and textuality. In other words, scholars tend to use these verbs in scientific discourse to postulate hypotheses, report findings, judge or decide something after reasoning, or predict something on the basis of data and observation. By contrast, verbs such as *said, reveal* ('tell a secret'), and *announced* are more commonly used by speakers in everyday communication rather than in scientific discourse.

5. Quantitative findings and their discussion

Type frequencies observed in the corpus proved to include 1911 unique participle verb forms, out of which 70 occurred only once in the construction in question. However, this section will solely examine the findings for the 60 most strongly attracted participles of the construction, since it is impossible to evaluate the quantitative results for all the verb forms in the space here reserved. Table 6 provides the scores of attraction and reliance for the 30 most significant participles, including the raw frequency of participles in the construction, the total frequency of all participles in the construction, and the total frequency of these participles in the whole corpus. The scores for the next thirty significant participles are presented in Table 7 in Appendix 1.

Both tables group the findings according to the score of attraction. They indicate that there are past participles that are strongly attracted to the *It BE* V_{en} *that*-construction. On closer inspection, we can notice that the top of Table 6 lists relatively frequent verbs, such as *noted, said, argued*, *estimated*, and *assumed.* The primary reason why these verbs hold the top positions in the ranking list is that their high frequencies observed in the construction are likely to affect their strength of attraction to the construction. For example, *noted* (attraction score 8.27%) and *said* (attraction score 7.96%) obtained much higher results for attraction than *recommended* (attraction score 1.64%) and *understood* (attraction score 1.62%), because they occurred much more frequently in the *It BE* V_{en} *that*-construction than *recommended* and *understood*, as shown in Table 6. By contrast, the ranking list for reliance lists much higher results for less common participles colligating with the construction, such as *hypothesized* (reliance score 90.13%), *rumored* (reliance score 100%), *presumed* (reliance score 59.35%), and *inferred* (reliance score 74.71%), because the statistical formula employed for calculating the score of reliance considers the total frequency of each participle in

Table 6. The 30 most strongly attracted participles of the *It BE* V_{en} *that*-construction

rank	participles	a	x	e	attraction	reliance
1.	noted	2136	25820	4018	8.27%	53.16%
2.	said	2056	25820	11825	7.96%	17.39%
3.	argued	1378	25820	4677	5.34%	29.46%
4.	estimated	1318	25820	1799	5.10%	73.26%
5.	assumed	1105	25820	2277	4.28%	48.53%
6.	suggested	962	25820	4778	3.73%	20.13%
7.	found	904	25820	8336	3.50%	10.84%

Table 6. cont.

rank	participles	a	x	e	attraction	reliance
8.	known	871	25820	3688	3.37%	23.62%
9.	expected	865	25820	1365	3.35%	63.37%
10.	reported	877	25820	2672	3.40%	32.82%
11.	believed	814	25820	2623	3.15%	31.03%
12.	hypothesized	630	25820	699	2.44%	90.13%
13.	thought	552	25820	3475	2.14 %	15.88%
14.	shown	483	25820	6764	1.87%	7.14%
15.	hoped	469	25820	1166	1.82%	40.22%
16.	recommended	423	25820	876	1.64%	48.29%
17.	understood	417	25820	1157	1.62%	36.04%
18.	determined	415	25820	3062	1.61%	13.55%
19.	concluded	394	25820	2342	1.53%	16.82%
20.	recognized	376	25820	1189	1.46%	31.62%
21.	revealed	373	25820	1293	1.44%	28.85%
22.	decided	362	25820	2636	1.40%	13.73%
23.	discovered	357	25820	1681	1.38%	21.24%
24.	observed	336	25820	1008	1.30%	33.33%
25.	agreed	311	25820	1196	1.20%	26.00%
26.	seen	307	25820	3390	1.19%	9.06%
27.	announced	279	25820	1600	1.08%	17.44%
28.	accepted	268	25820	671	1.04%	39.94%
29.	predicted	243	25820	797	0.94%	30.49%
30.	remembered	235	25820	464	0.91%	50.65%

Note: **a**= The frequency of the past participle *noted* in the *It BE* V_{en} *that*-construction; **x**= The total frequency of all past participles in the construction; **e**= The total frequency of the past participle *noted* in COCA

COCA. In other words, although *noted* occurs much more frequently in the construction than *hypothesized*, the latter obtains a much higher result for reliance because its total frequency of occurrence in COCA is much lower (699 occurrences). Hence, the reliance of *hypothesized* on the construction is greater (90.13%), which in turn means that *hypothesized* as a past participle is used in other constructions in a proportion of 9.87%.

The participles ranked in Tables 6 and 7 can be grouped into several semantic categories. The first semantic group consists of verbs evoking the BECOMING_AWARE frame. In this frame, a certain cognizer (a speaker) becomes aware of some phenomenon, an entity or a situation in the world, as in [*From*

Figure 7] evidence, *it can be seen* [*that foreign-bound patenting by Canadians has grown very rapidly*] phenomenon. *Noted*, the most significant verb of this group, is followed by *found, recognized, discovered, observed*, and *seen* in ranks 7, 20, 23, 24, and 26. *Noted* accounts for 8.27% of the occurrences of the construction in COCA, and that 53.16% of occurrences of the same participle can be observed in that specific pattern. By comparison, the five other verbs are much less relevant slot fillers for the construction (attraction scores: 3.50%, 1.46%, 1.38%, 1.30% and 1.19%) and rest on this pattern to a lesser degree (reliance scores: 10.84%, 31.62%, 21.24%, 33.33% and 9.06%).

Some of these participles can also invoke other semantic frames. For example, *noted* and *observed* can be understood relative to the STATEMENT frame (see below), while *found* and *discovered* with reference to the ACHIEVING_FIRST frame. In the latter frame, a cognizer introduces a new idea into society, as in *It was found that* [*crocodiles have one of the slowest-evolving genomes*] idea. *Seen* in turn may evoke the GRASP frame, in which a cognizer acquires new knowledge about a certain phenomenon. This frame is also evoked by the participle *understood* in rank 17, as in *It can be understood from Fig. 12 that* [*forensic undetectability decreases*] phenomenon.

The second category is constituted by participles invoking the STATEMENT frame, a frame in which a certain speaker addresses a message to some addressee, as in *It is often said* [*that one of the qualities of a great leader is visibility*] message. *Said*, ranked second, is one of the most significant verbs belonging to this category. It is attracted to the pattern in a proportion of 7.96%, and relies on the pattern in a proportion of 17.39%. It is accompanied by a range of verbs conveying a message, such as *noted, suggested, reported, stated, observed, announced, acknowledged, proposed, claimed, mentioned, alleged, confirmed,* and *asserted.*

The third semantic set includes verbs instantiating the REASONING frame. In this frame, an arguer (a speaker) presents a certain content, along with its support, to an addressee. This content can be a proposition to be believed or a course of action to be taken, as in *It was argued* [*that special admissions program was based on quota*] content. *Argued*, ranked third, is the most strongly attracted participle evoking this frame. Others are *shown* in rank 14, *demonstrated* in rank 39, and *proven* in rank 43. *Argued* is followed by the verb *estimated* in rank 4 that reflects the ESTIMATING frame. In this frame, a cognizer considers some evidence or performs calculations to arrive at an approximate value for some feature of an item, as in *It is estimated* [*that 8,000 people die from cancer*] estimation.

The fourth category comprises participles invoking the AWARENESS frame, a frame in which a cognizer deduces a piece of content (the object of the cognizer's awareness) from some evidence. The content is presupposed to be true. The deduction of the content is implicitly based on logic (*think*), the source

of knowledge (*know*), or confidence in sources of information (*believe*), as in *It is known* [*from other studies of cadets*] $_{\text{evidence}}$ [*that they have difficulty choosing honesty*] $_{\text{content.}}$ The most significant member of this category is *assumed*, ranked fifth. It is followed by *known* in rank 8, *believed* in rank 11, *thought* in rank 13, *understood* in rank 17, *presumed* in rank 46, *suspected* in rank 57, and *supposed* in rank 58.

The next semantic set consists of participles whose use in the construction can be relativized to the OPINION frame, in which a cognizer holds a particular opinion about a particular topic. This opinion is not presupposed to be true; rather it is something that is perceived as a potential point of difference, as in *It was held* [*that Nigeria could achieve technological advancement attained already by the US*] $_{\text{opinion.}}$ This frame seems to be evoked by the following verbs in the ranking list: *assumed, expected, believed, thought, felt,* and *held.*

Another significant category includes *found, determined, concluded, established, speculated, learned, realized,* and *inferred.* These verbs activate the COMING_TO_BELIEVE frame. In this frame, a cognizer comes to believe something (the content), frequently after a process of reasoning. This change in belief is usually based on a piece of evidence, as in *It can be concluded* [*from the evidence above*] $_{\text{evidence}}$ [*that none of these experiments contributed to a serious process of mutual social learning*] $_{\text{content.}}$

Among the most significant verbs in the ranking list, there are also five pairs of participles, such as *expected/anticipated*, *revealed/disclosed*, *predicted/projected*, *emphasized/stressed*, and *accepted/admitted*. The first pair evokes the EXPECTATION frame, a scene in which a cognizer believes that some phenomenon will take place in the future, as in *It is expected* [*that the ballads will be published in book form*] $_{\text{phenomenon.}}$ The second pair pertains to the situation of revealing secret information to an addressee. Hence, both verbs reflect the REVEAL_SECRET frame, as in *It was revealed* [*that he had lied about his conversation with the Russian ambassador*] $_{\text{information.}}$

The third pair is associated with the PREDICTING frame, in which an event or state (a future eventuality) is predicted to occur or hold at a future time on the basis of some evidence: [*By 2035*] $_{\text{time}}$ *it is projected* [*that people 65 years and over will make up 23% of the total population*] $_{\text{eventuality.}}$ The next pair is identified with the EMPHASIZING frame. In this frame, an agent who participates in an undertaking ascribes a certain degree of significance to some consideration (i. e., an entity, event, or state of affairs) that determines the success of this undertaking: *It must be stressed* [*that some training is needed before implementation*] $_{\text{consideration.}}$ The last pair invokes the ACCEPTING_TRUTH frame, in which some content (a proposition) is believed to be true or right, as in *It is generally accepted* [*that Alfred Hart was hired by Edwin Crocker*] $_{\text{content.}}$

Finally, the last group is constituted by single verbs, such as *hypothesized, hoped, recommended, decided, agreed, remembered, rumored, documented,*

recalled, and *feared,* which invoke the following semantic frames: HYPOTHESIZING, DESIRING, ATTEMPT_SUASION, DECIDING, BE_IN_AGREEMENT_ON_ASSESMENT, REMEMBERING_INFORMATION, UNATTRIBUTED_INFORMATION, RECORDING, REMEMBERING_EXPERIENCE, and EXPERIENCER_ FOCUSED_EMOTION, respectively. The HYPOTHESIZING frame describes a situation in which a person proposes an idea or a possible explanation for something (a hypothesis) that is based on a few known facts but that has not yet been tested or proved to be true or correct: *It was hypothesized* [*that BTE would improve recovery from an acute bout of intense exercise*] hypothesis. The DESIRING frame concerns a scene in which an experiencer wishes for a certain event (a desired change) to occur, as in *It is hoped* [*that this will lead to improvements in the patient experience of end-of-life care*] event.

The third frame pertains to a situation in which a speaker expresses his wish for action on the addressee's part in order to bring about events or states mentioned in the content, as in *It is recommended* [*that tickets only be purchased from a trusted source*] content. The DECIDING frame describes a decision about an entity or course of action that is made by a cognizer, as in *It has been decided* [*that your services are no longer required*] decision. The next frame refers to a situation in which cognizers hold a similar or differing opinion about a particular topic: *It is agreed* [*that no decision can be reached*] opinion. The REMEMBERING_INFORMATION frame is concerned with cognizers retaining and retrieving factual information (the mental content) stored in memory: *It is remembered* [*that he was a man of considerable means and influence*] mental content.

In the UNATTRIBUTED_INFORMATION frame, a fact (a state, event, or existence) is reported by a speaker as deriving from stories or statements of third parties, as in *It was rumored* [*that she held a black belt in Brazilian jiu jitsu*] fact. In the RECORDING frame, an agent creates a permanent record of a certain phenomenon (an entity or situation), as in *It was documented* [*that she had been raped and physically abused*] phenomenon. In the next frame, a cognizer recalls memories of past experience or impressions, as in *It should be recalled* [*that Caravaggio died in his late 30s*] experience. In the last frame, an experiencer's emotions are described relative to some content (a current state of affairs), as in *It is feared* [*that any reduction in river flows would affect Texas agriculture*] content.

6. Concluding remarks

This study has aimed at identifying different properties of the *It BE V_{en} that*-construction and the most common past participles occurring in this pattern. The findings of the corpus-based investigation have indicated that the construction has several different variants demonstrating their own specific preferences for participles, is relatively more frequent in academic prose and written registers (magazines and newspapers) than in fiction and spoken discourse, and reveals

a strong tendency to occur with specific categories of verbs evoking, in addition to several others, the following semantic frames: BECOMING_AWARE, STATEMENT, REASONING, ESTIMATING, AWARENESS, OPINION, and COMING_TO_BELIEVE.

The common use of the past participles invoking these frames is directly related to information-packaging and discourse-functional properties of the *It BE Ven that*-construction, though the high frequencies of these verbs in this pattern may also reinforce and influence those properties. As noted in section 4, the construction in question is an information-packaging construction that commonly occurs with reporting verbs conveying a specific message, and that the pattern is regularly used in academic registers to comment on some topics mentioned previously or to convey new facts and states of affairs by introducing the evaluative comments in the form of the dummy *it,* different forms of the verb *be*, and various past participles followed by *that*-clauses or occasionally preceded by adverbs. In particular, the pattern is used for expressing beliefs and opinions about some facts or states of affairs by presenting them as if they were generally accepted views rather than personal judgments, thereby placing impersonal comments at the beginning of each sentence and long and complex *that*-clauses at the end (cf. Herriman 2000b: 2011; Kaltenböck 2005: 137; Wiliński 2018a: 98). Hence, this impersonal construction serves a reporting function and introduces new or old pieces of information at the end of a sentence.

After a summary of the results, several caveats should be borne in mind when interpreting these findings. First, the study was limited to data derived from an American variety of English. Second, it was impossible to evaluate the quantitative findings for all the verbs occurring in the construction, because of little space reserved for this article. Thirdly, the study was unable to compare the distribution of this construction with the distribution patterns of other types of extraposed constructions across different corpora. Future research might, therefore, focus on comparing and contrasting different types of *it*-extraposition. Such a quantitative study would uncover minor distributional differences in their use. In addition, in view of the limitation of this research to COCA and the potential existence of subtle variations in the use of this construction in other varieties of English, future research might explore the occurrence of the *It BE* V_{en} *that*-construction in other corpora (e.g., the British National Corpus) across various kinds of both written and spoken registers.

References:

Barlow, M., and S. Kemmer (eds.). 2000. *Usage-based models of language*. Chicago: The University of Chicago Press.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Bierwiaczonek, B. 2016. *An introductory English grammar in constructions.* Częstochowa: Wydawnictwo AJD.

Boas, H. 2017. Computational resources: FrameNet and constructicon. In B. Dancygier (ed.), *The Cambridge handbook of cognitive linguistics*, 549-573. Cambridge: Cambridge University Press.

Bybee, J., and C. Beckner 2010. Usage-based theory. In B. Heine and H. Narrog (eds.), *The Oxford handbook of linguistic analysis*, 827-855. Oxford: Oxford University Press.

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

Fillmore, C.J., and C. Baker 2010. A frames approach to semantic analysis. In B. Heine and H. Narrog (eds.), *The Oxford handbook of linguistic analysis*, 791-816. 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.

Hands, P. (ed.) 2017. *Collins Cobuild English Grammar*. Glasgow: HarperCollins.

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.

Hoffmann, Th. 2022. *Construction grammar: the structure of English*. Cambridge: Cambridge University Press.

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.

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

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

Schmid, H.-J. 2000. *English abstract nouns as conceptual shells. From corpus to cognition.* Berlin, New York: Mouton de Gruyter.

Schmid, H.-J. 2016. *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Berlin and Boston: Mouton De Gruyter.

Szcześniak, K. 2014. *The meaning of constructions. The cognitive denial of the lexicon-syntax division.* Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Tummers, J., K. Heylen and D. Geeraerts 2005. Usage-based approaches in cognitive linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1(2): 225-261.

Van Linden, A. 2012. *Modal adjectives: English deontic and evaluative constructions in synchrony and diachrony*. Berlin and Boston: Walter de Gruyter.
Wiliński, J. 2018a. Adjectives in extraposed constructions with *that*-clauses: A quantitative corpus-driven analysis. *Brno Studies in English* 44(1): 83-102.
Wiliński, J. 2018b. Distinctive adjectives in two variants of the it-extraposition construction: a quantitative corpus-based investigation. *Linguistica Silesiana* 39: 21-45.

Data sources

The Corpus of Contemporary American English (COCA). Available from https://www.englishcorpora.org/coca/
The FrameNet project. Available from https://framenet.icsi.berkeley.edu/fndrupal/

Appendix 1

Table 7. The next 30 strongly attracted participles

rank	participles	a	x	y	attraction	reliance
31.	acknowledged	197	25820	975	0.76%	20.21%
32.	emphasized	196	25820	479	0.76%	40.92%
33.	established	195	25820	676	0.76%	28.85%
34.	rumored	181	25820	181	0.70%	100.00%
35.	felt	187	25820	1184	0.72%	15.79%
36.	anticipated	180	25820	350	0.70%	51.43%
37.	stated	177	25820	1535	0.69%	11.53%
38.	proposed	170	25820	708	0.66%	24.01%
39.	demonstrated	164	25820	1726	0.64%	9.50%
40.	documented	142	25820	312	0.55%	45.51%
41.	claimed	132	25820	1031	0.51%	12.80%
42.	speculated	121	25820	379	0.47%	31.93%
43.	proven	116	25820	636	0.45%	18.24%
44.	stressed	107	25820	297	0.41%	36.03%
45.	mentioned	92	25820	788	0.36%	11.68%
46.	presumed	92	25820	155	0.36%	59.35%
47.	learned	90	25820	3347	0.35%	2.69%
48.	alleged	89	25820	277	0.34%	32.13%
49.	realized	75	25820	2089	0.29%	3.59%
50.	recalled	73	25820	165	0.28%	44.24%
51.	confirmed	72	25820	868	0.28%	8.29%
52.	feared	72	25820	256	0.28%	28.13%
53.	inferred	65	25820	87	0.25%	74.71%
54.	asserted	64	25820	300	0.25%	21.33%
55.	projected	64	25820	181	0.25%	35.36%
56.	admitted	62	25820	765	0.24%	8.10%
57.	suspected	61	25820	403	0.24%	15.14%
58.	supposed	60	25820	130	0.23%	46.15%
59.	disclosed	58	25820	316	0.22%	18.35%
60.	held	58	25820	818	0.22%	7.09%