FOLIA 351

Annales Universitatis Paedagogicae Cracoviensis

Studia Linguistica 17 (2022)
ISSN 2083-1765
DOI 10.24917/20831765.17.20

Jarosław Wiliński
ORCID 0000-0002-3136-6529
Uniwersytet Przyrodniczo-Humanistyczny w Siedlcach, Polska

What-clefts with adjectives in English: A corpus-based analysis

Keywords: what-clefts, construction grammar, frame semantics, COCA, corpus-based analysis

Słowa kluczowe: zdania rozszczepione, gramatyka konstrukcyjna, semantyka ramowa, korpus, analiza korpusowa

Introduction

The past few decades have witnessed a striking proliferation of publications dealing with different aspects of ordinary *what*-clefts or pseudo-clefts in English (Prince 1978; Delahunty 1984; Declerck 1988; Collins 1991; Hedberg 1990; Huddleston and Pullum 2002; Kim 2007; to name a few). Some researchers (Kim 2007; Bierwiaczonek 2016: 168–169) have focused on the structural properties of different types of clefts, whereas others (Prince 1978, Collins 1991) have drawn a comparison of *what*-clefts with other types of clefts, such as *it*-clefts or reversed *wh*-clefts. Many research studies (Declerck 1984; Collins 1991, 2006; Biber et al. 1999; Herriman 2003, 2004; Deroey 2011) have also examined semantic, pragmatic, and/or discourse-functional properties of *what*-clefts. However, very little attention has hitherto been paid to the *What be ADJ be*-construction, a specific variant of the *what*-cleft construction in English. To the best of the author's knowledge, previous studies have failed to explore the nature of this construction: some researchers have solely mentioned this pattern in passing or have given one or two examples of its usage (e.g., Deroey 2011; Gast and Levshina 2014). To date, no single study has sought to examine this construction at length and determine its different linguistic features, not to mention the quantification of adjectives in this construction and the identification of the most frequent patterns of its usage.

Thus, because of these research gaps, this paper attempts to achieve two goals: first, to identify structural, semantic, pragmatic, distributional, and discourse-functional properties of this construction, and second, to determine those adjectives that are strongly attracted to the pattern in question. More specifically, on the basis of the data retrieved from the Corpus of Contemporary American English (COCA), the paper aims to indicate that the construction in question has a specific structure

and meaning, serves different functions in discourse, is distributed differently across various registers in COCA, and occurs with specific categories of adjectives evoking different semantic frames.

The rest of the paper is divided into five sections. Section 2 briefly discusses the theory and methodology applied in the corpus-based investigation. Section 3 elucidates data sources and their nature, data retrieval and evaluation, and the tools and procedures used for their extraction and quantification. Section 4 considers structural, semantic, distributional and discourse-functional properties of the *What be ADJ be*-construction. Section 5 presents and evaluates the results of the quantitative study of adjectives occurring in this construction. Finally, section 6 provides concluding remarks.

Theory and methodology

This paper uses construction grammar and the theory of frame semantics to determine the structural and semantic properties of the construction at issue. Construction grammar (for example, as defined by Goldberg 2006, 2013, Hoffmann and Trousdale 2013: 1–2; Hoffmann 2017) assumes that grammar consists of constructions, form-meaning pairings, at different levels of complexity and schematicity, from morphemes through words to the most schematic syntactic patterns and associated rules of their semantic, pragmatic and discourse-functional interpretation. Thus, for example, the *what*-cleft (e.g., *What is important is…*) is a construction since it has a complex form (e.g., [*What be ADJ be…*]) and a specific evaluative and highlighting function/meaning (the sense 'the thing that has a major effect on someone or something is…') that are conventionally associated with each other. All grammatical units can be stored and represented as constructions on the condition that they occur with sufficient frequency, which in turn is an essential prerequisite for their entrenchment in a speaker/hearer's grammatical system (Croft and Cruse 2004: 292–293).

Frame semantics (Fillmore 2006) in turn is a semantic theory that emphasizes a close connection between lexical semantics and encyclopedic knowledge. It is based on the assumption that we cannot apprehend the meaning of a linguistic unit without access to all the background knowledge that is connected with that unit. For example, we would not be able to comprehend the verb *to lecture* without familiarity with the situation of SPEAKING ON TOPIC, which involves three core elements, such as a speaker, audience and topic, and several other peripheral elements, such as a place, occasion, medium, and manner. In this study, this theory is used to define the semantics of the construction under study and the meanings of adjectives occurring with this construction. The names of all semantic frames and their modified descriptions were taken from the FrameNet lexical database (Fillmore and Baker 2010; Boas 2017; see The FrameNet project in data sources).

In addition, the paper makes use of corpus-based methodology. The quantitative method referred to as *the attraction-reliance measure* (Schmid 2000) is used to gauge the reciprocal association between adjectives and the *what*-cleft construction in American English. In other words, it is supposed to aid in identifying adjectives

that are more strongly attracted to this construction than others and those whose occurrence in this construction is more significant than their use in other contexts in the corpus. *Attraction* is expected to measure the degree to which the *What be ADJ be* construction attracts a particular adjective, while *reliance* is intended to quantify the degree to which an adjective appears in the construction in question versus other patterns or contexts in the corpus (cf. Wiliński 2021). Although the method requires quantification and statistical evaluation, the quantitative findings are assessed qualitatively and subjectively. In this case, adjectives are classified into semantic groups, while their specific meanings in the construction are defined relative to the semantic frames they evoke.

Corpus, data retrieval, and statistical evaluation

Data were obtained from an earlier version of the well-balanced Corpus of Contemporary American English (COCA). This version covers the years between 1990 and 2017 and contains more than 560 million words. It consists of the following registers and genres: spoken (transcripts from various TV and radio programs), fiction (short stories and plays), popular magazines (various magazines covering a wide range of domains such as news, health, sports, religion, or finance), newspapers (ten American newspapers: USA Today, San Francisco Chronicle, New York Times, etc.), and academic texts (approximately 100 various peer-reviewed journals). The most recent update was made in March 2020. Hence, COCA presently includes more than one billion words of text from eight genres: spoken, fiction, popular magazines, newspapers, academic texts, TV and Movies subtitles, blogs, and other web pages. The illustrative sentences examined in section 4 were taken from the updated version of COCA.

The procedure followed in this study involved taking several steps. First, the observed frequencies of adjectives from the corpus were retrieved by a search engine. More specifically, the concordancer installed in COCA searched all the texts for the adjectives occurring in the *What be ADJ be*-pattern. To extract all the occurrences, thirty-two different 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 wildcard (*What is * ADJ * * is*) was used to extract the combination *What is so interesting to me is* (…), while the wildcard (*What 's * ADJ is*) to find *What 's less apparent is* (…). The retrieval was restricted to ten words in the pattern: thus, it was possible to retrieve relatively long combinations such as *What was so strange to me about that was* (…), *What would be new of course is* (…), or *What has been most striking is* (…). However, due to its limitations, the program did not allow for extracting combinations consisting of more than ten words.

Then, all specific uses of adjectives were manually examined to determine true combinations, i.e. the combinations that accurately reflect grammatical patterns of usage indicated by wildcards. All false hits (i.e., the occurrences which did not correspond with the pattern searched by wildcards) were discarded from further analysis. The observed frequencies of occurrence (e.g., a: the frequency of the adjective *important* in the *what*-cleft construction and x: the total frequency of all

adjectives in the construction), as shown in Table 1 below, were calculated manually by examining concordance lines, whereas the total frequencies of adjectives in COCA (e.g., e: the total frequency of the adjective *important*) were computed automatically by the concordance program.

After the estimation of the observed frequencies, the numbers (a, x, and e) in Table 1 below were entered into an Excel worksheet and were used for calculating Schmid's measures of attraction and reliance (cf. Wiliński 2020). *Attraction* was calculated by dividing the raw frequency of an adjective in the *what*-cleft construction by the total frequency of all adjectives in this construction, whereas *reliance* was computed by dividing the frequency of occurrence of an adjective in the construction in question by its frequency of occurrence in COCA (cf. Schmid 2000: 54). The results of these calculations were expressed as percentages by multiplying the raw frequency of a particular adjective in the *what*-cleft construction in each case by one hundred. The percentage was regarded as an indication of attraction or reliance: the higher the percentage, the stronger the attraction to, and reliance on, the *what*-cleft construction (cf. Wiliński 2021).

Table 1. Cooccurrence table for a statistical analysis

adjective	a	x	e	attraction	reliance
important	1,179	8,728	200,570	13.51%	0.59%

Note: a = The frequency of the adjective *important* in the *what*-cleft construction; x = The total frequency of all adjectives in the construction; e = The total frequency of the adjective *important* in COCA

As can be noticed in Table 1, the scores of attraction and reliance for the adjective *important* are 13.51% and 0.59%, respectively. This means that the adjective occurs in 13.51% of the uses of adjectives in the *what*-cleft construction: in other words, *important* is a highly significant lexeme, which is very strongly associated with this construction. However, solely 0.59% of the occurrences of the same adjective are found in this construction, which means that the adjective relies on other constructions in a proportion of 99.41%. In other words, it occurs more frequently in other contexts or patterns. These quantitative findings were sorted according to the measure of attraction and then evaluated qualitatively and subjectively.

What-clefts with adjectives

Properties

Although much attention has been paid to ordinary *wh*-clefts, *what*-clefts with adjectives have been largely neglected. Thus, this section aims to examine one specific variant of the ordinary *wh*-clefts, referred to as the *What be ADJ be*-construction, the use of which can be exemplified by the following sentences extracted from COCA:

(1) [*What 's so interesting about this company Mossack Fonseca*] *is* ***the kind of menu of options they provide*** (SPOK: Fresh Air).

(2) [*What has been most striking*] *is* ***that my patients make careful and deliberate decisions*** (NEWS: Washington Post).

(3) [*What will be important to us*] *is* ***to gauge audience response from Off the Edge and ask,*** […] (NEWS: Atlanta Journal Constitution).

(4) *Overall,* [*what is fascinating*] *is* ***how the past and present websites illustrate and record the progression of this election*** (BLOG: Mediabistro.com).

(5) [*What was so strange to me about them*] *was that* ***they literally identify with front-line foot soldiers and pilots, and*** […] (SPOK: NPR_Morning).

As can be observed above, these examples are divided into two main parts: one is backgrounded and placed in a subordinate (dependent) clause (Biber et al. 1999) or a fused relative construction (cf. Huddleston and Pullum 2002); the other is foregrounded (focused on) and functions as the complement of the verb *be.* The specially foregrounded element is in bold in the examples above, and the dependent *what*-clause is placed in square brackets. This kind of information-packaging construction (see Huddleston and Pullum 2002: 67; Bierwiaczonek 2016: 159, for this term) contains one fixed lexical item (*what*) and at least four flexible slots that can be filled by different forms of the verb *be,* adjective phrases, the copula *be*, and the complement of the copula *be*:

Table 2. Properties of the *What be ADJ be*-construction

form	What + backgrounded element (be AdjP)	copula *be*	foregrounded element: noun phrase/ ing-clause/to-infinitive clause/ that-clause/wh-clause
meaning	The thing that (or that which) + is (stative and ascriptive) + adj (property)	state	value/states/activities/achievements/ accomplishments/
IP function	presupposition/topic/given evaluative judgment	specifying	focus/comment /new

The syntactic and semantic properties of this construction can be represented structurally and schematically in the form of Table 2, where *what* is followed by the backgrounded component (*be* + an adjective phrase), the copula *be* in the specifying sense, and the foregrounded component (a noun phrase, an infinitive clause, an ing-clause, a finite *that*-clause, or a *wh*-clause). In rare cases, the focused component can be a prepositional phrase (as in *What was important was for Susan to tell what happened*, or in *What is unusual is for one of those former lawmakers to keep his campaign fund open*). Occasionally, some optional elements can be inserted in the *wh*-clause (e.g., adverbials: *overall* or *however*, as in 4 and 6). The foregrounded element cannot be an adverb phrase or an adjective phrase. However, as the corpus data shows, it is possible to use an adverbial clause: e.g., as in *But what is interesting is if he was negotiating with Mousavi, it would not be as tough.*

As with ordinary *wh*-clefts, the foregrounded element of the *What be ADJ be*-construction is located at the end, in accordance with the topic-comment structure and the information principles (IP) of end-weight and end-focus (Quirk et al. 1985: 863, Biber et al. 1999: 896–898, Huddleston and Pullum 2002: 1368–1372). The primary communicative purpose of this construction is to signal explicitly what is considered background information and what is the prime focus of attention:

(1) [*What was new about the book*], *however, was* ***that it spoke about Americans through first-hand experience*** (ACAD: Scandinavian Studies).

(2) [*What was so unique about the show*] *was* ***that it was from an African-American perspective*** (MAG: Essence).

The *what*-clauses in (6) and (7) begin with a topic (an established idea) and given (background) information by carrying an explicit reference to the preceding context: *the book* and *the show*, respectively. They end with new information by commenting on the topic. In both cases, the main communicative idea (the new information) is provided at the end of the sentence, in agreement with the principles of end-weight and end-focus: i.e., long and complex clauses tend to be used in the final position to process new information more easily and efficiently.

As for the meaning and function of the *What be ADJ be*-construction, in the *what*-clauses above, *what* means *the thing (s) that, be* designates a state, and an adjective phrase denotes a property. The whole *what*-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 certain established idea or entity by providing the evaluative comments in the form of various adjectives, frequently followed by prepositional phrases (e.g., *so strange to me about that*, as in 5). The verb *be* in a fused relative clause seems to be used in the ascriptive sense: e.g., in (7), it ascribes a property of being unique to an abstract referent indicated by *what* (i.e. *the thing that*). By contrast, the copula *be,* connecting the background element with the foregrounded one, denotes a state and appears to be used in the specifying sense: it defines a variable and specifies its value (cf. Huddleston and Pullum 2002: 1421, Gast and Levshina 2014). For example, in (6), it serves to specify, or identify, a variable in an open proposition or presupposition (*What was new about the book was* x, x = *that it spoke about Americans through first-hand experience*). Finally, the clauses following the copula *be* designate situations: states or occurrences (activities, accomplishments, or achievements), and they fulfil an explanatory function.

The Polish counterpart of pseudo-clefts with adjectives

The construction in question has its counterpart in Polish. However, the Polish construction does not possess all structural characteristics of English pseudo-cleft sentences, and it seems to be restricted mainly to the variant with a noun phrase used as the foregrounded element. The Polish counterpart can be represented structurally in the following way: [*To, co* Rel_{pron} *be* adj, *to* (V_{cop}) Compl_{NP}], where *to* 'this' is a demonstrative pronoun, *co* 'what' is a relative pronoun, V_{cop} is a copular verb, and Compl_{NP} is a noun complement. A typical instantiation of this pattern in Polish is the sentence (*To*)*, co jest ważne dla mnie, to miłość* 'What is important for me is love'. In this structure, the demonstrative pronouns *to* form a specific template which is superimposed on the sentence: [*To, co…, to…*]. Thus, the sentence is divided into two parts, i.e. the backgrounded element and the foregrounded element. The demonstrative pronouns *to* are used cataphorically to indicate the foregrounded element, which is placed at the end of the sentence, whereas the relative pronoun *co* introduces the background information. The first pronoun *to*, which is optional in

many contexts, is followed by the relative pronoun *co,* the verb *be* and an adjective phrase. The second demonstrative *to* functions solely as a pointer, an indicator that makes the foregrounded element emphatic. It can be followed by the copula *be*, but it more frequently stands on its own and replaces the copula.

Variants of the What be ADJ be-construction

The corpus data reveal that the construction has four major variants (*What be bare-ADJ be NP, What be bare-ADJ be to-INF, What be bare-ADJ be that-clause,* and *What be bare-ADJ be wh-clause*), each of which tends to occur with specific types of adjectives. For example, the pattern with that-clauses exhibits a preference for adjectives denoting OBVIOUSNESS (e.g., *clear, obvious, apparent,* or *evident*) and CERTAINTY (*certain, sure, undeniable,* or *indisputable*), while the pattern with to-infinitive clauses displays a strong tendency to combine with adjectives evoking the STIMULUS FOCUS frame (e.g., *amazing, astonishing, extraordinary,* or *surprising*), the IMPORTANCE frame (*important* and *critical*), the NECESSITY frame (*necessary* and essential), and the semantic frames of DIFFICULTY, UNIQUENESS, or MORALITY EVALUATION (such as *difficult, unusual,* or *wrong*). The pattern with NPs in turn tends to colligate with adjectives invoking the IMPORTANCE frame (*important, significant, crucial,* and *critical*), the NECESSITY frame (*necessary* and *essential*), the STIMULUS FOCUS frame (*interesting, extraordinary, astonishing, surprising, amazing,* etc.), the OBVIOUSNESS frame (*clear, obvious, apparent,* etc.), and the DIFFICULTY (*difficult* and *hard*), and morality evaluation (*bad, good,* and *wrong*). Lastly, the pattern with *wh*-clauses commonly occurs with adjectives evoking the following semantic frames: OBVIOUSNESS (*unclear* and *apparent*), IMPORTANCE or RELEVANCE (*important, crucial, critical,* and *relevant*), and STIMULUS FOCUS (*amazing, unusual, curious, funny, nice, impressive, shocking, frightening, disturbing, fascinating, surprising*, etc.).

In addition to these four patterns, the study identified several other sub-patterns imposing constraints upon the adjectives with which they colligate: *What be bare-ADJ be ing-clause, What be ADJ in NP be NP/that-clause/wh-clause*, *What be ADJ with NP be NP/that-clause/wh-clause*, *What be ADJ for NP be NP/that-clause/wh-clause*, *What be ADJ to-infinitive be NP/that-clause/wh-clause, What be ADJ about NP be NP/that-clause/wh-clause,* and *What be ADJ to NP be NP/that-clause/wh-clause*. The first sub-pattern is mainly restricted to the adjectives *important* and *difficult*, the second one to *important, new, unique, striking, essential, amazing*, or *interesting,* the third one to *wrong, different, incompatible,* or *important,* and the fourth one to *important, interesting, critical, great, vital, necessary, unique, good, clear, characteristic, exciting, easy, frustrating, problematic, best, better*, or *expensive*. The pattern with an adjective followed by a to-infinitive clause mainly occurs with *interesting, important, easy, difficult, hard*, and *likely*. The sub-pattern with *about* commonly occurs with *interesting, different, unique, good, important, special, great, remarkable, distinctive, significant, new, best,* and many others, while the last one with *interesting, important, amazing, clear, fascinating, surprising, frustrating, remarkable*, or *shocking.*

Distribution across different registers

Regarding distributional properties of the *What be ADJ be*-construction, Table 3 presents the frequency distribution of the most significant adjectives occurring in one specific instantiation of this construction, i.e. the *What is ADJ is*-pattern. The reason why this sub-pattern was selected for distributional analysis is that it is the most basic and common variant of occurrence of the construction under study in COCA.

Table 3. Distribution of the *What is ADJ is*-pattern

What is ADJ is-pattern	all	spoken	fiction	magazine	newspaper	academic
clear	236	51	3	50	56	76
important	207	75	12	25	26	69
certain	99	11	8	27	22	31
interesting	85	47	1	10	1	26
new	82	21	4	17	18	22
surprising	70	13	1	18	13	25
true	47	26	5	8	5	3
remarkable	33	3	-	8	3	19
significant	28	3	1	4	2	18
unusual	28	8	2	6	7	5

A cursory examination of frequency data in Table 3 reveals that the sub-pattern under study is relatively infrequent in fiction; however, it predominates in academic and spoken registers and tends to occur commonly in written texts, such as magazines and newspapers. These findings are partly congruent with the results of the studies into *wh*-clefts conducted by Collins (1991) and Biber et al. (1999). Collins observed that ordinary *wh*-clefts are commonly used in speech and written texts, while Biber et al. found that they are most frequent in conversation and less common in fiction, newspapers, and academic prose.

It is clear from the table that there is great variation in the use of the adjectives across different registers. *Clear*, *certain surprising*, *remarkable*, and *significant* are much more common in academic discourse than they are in other registers, while *important*, *interesting*, *true*, and *unusual* are more frequent in spoken discourse. *New* in turn has a high frequency not only in academic and spoken corpora, but also in magazines and newspapers. The close association between the adjectives and these registers may have something to do with the low communicative significance (cf. Prince 1978, Collins 1991) or the low information content (Biber et al. 1999) that is commonly found in the *what*-clause. In other words, speakers and writers generally tend to use *what*-clauses with adjectives as a 'springboard starting an utterance' (Biber et al. 1999), as a 'framework of evaluation' (Gast and Levshina 2014), or as a 'please-pay-attention message' (Miller and Weinert 1998: 124), thus contributing little information to the propositional content expressed in the sentence. In other words, such *wh*-clauses with adjectives mainly serve an evaluative and highlighting

function and direct readers or hearers' attention to the specific and key information which is clarified in the foregrounded component.

Quantitative findings and their discussion

The quantitative data retrieved from the corpus proved to contain 629 types of adjectives, out of which 337 occurred only once in the *What be ADJ be*-construction. However, this section will solely discuss and interpret the findings for the 50 most strongly attracted adjectives of the pattern, due to the limited amount of space here reserved. Table 4 presents the quantitative results of the measures of attraction and reliance for the 30 most significant adjectives, including the raw frequency of adjectives in the construction in question, the total frequency of all adjectives in the construction, and the total frequency of these adjectives in the whole corpus. The results for the next twenty significant adjectives are provided in Table 5 at the end of this section, since it would be difficult to present all the findings for the 50 adjectives in Table 4 on one single page. They are evaluated qualitatively and quantitatively in the appropriate paragraphs along with the results displayed in Table 4.

These results in both tables are sorted according to the measure of attraction and evidently imply that there are adjectives that are strongly attracted to the *what*-cleft construction. Even a glance at the findings already reveals that the top of Table 4 includes fairly common adjectives, such as *important*, *interesting*, *clear*, *new*, and *different.* The reason why these adjectives occupy the top positions in the ranking list is that the total frequency of these adjectives in COCA overall is expected to influence the likelihood of their occurrence in the construction in question. For example, *important* (attraction score 13.51%) and *interesting* (attraction score 12.95%) attained much higher scores for attraction than *clear* (attraction score 6.40%) and *new* (attraction score 3.14%), since they occurred much more frequently in the *what*-cleft than *clear* and *new*, as rendered in Table 4. By contrast, the list for reliance comprises much higher scores for less frequent adjectives occurring in the construction, such as *interesting* (reliance score 2.00%), *surprising* (reliance score 1.33%), *remarkable* (reliance score 1.13%), and *fascinating* (reliance score 1.26%), since the statistical formula applied to the estimation of reliance takes the total frequency of each adjective in COCA into consideration. In other words, though *important* occurs much more frequently in the *what*-cleft than *interesting*, the latter obtains a much higher score for reliance because its total frequency of occurrence in COCA is much lower (56485 occurrences). Consequently, the reliance of *interesting* on the *what*-cleft is greater (2.00%).

Table 4. The 30 most strongly attracted adjectives of the *What be ADJ be*-construction

a = The frequency of an adjective (e.g. *important*) in the *what*-cleft construction; **x** = The total frequency of all adjectives in the pattern; **e** = The total frequency of an adjective (e.g. *important*) in corpora						
rank	**adjective**	**a**	**x**	**e**	**attraction**	**reliance**
1.	important	1,179	8,728	200,570	13.51%	0.59%
2.	interesting	1,130	8,728	56,485	12.95%	2.00%

3.	clear	559	8,728	111,637	6.40%	0.50%
4.	new	274	8,728	875,730	3.14%	0.03%
5.	different	237	8,728	236,402	2.72%	0.10%
6.	surprising	222	8,728	16,753	2.54%	1.33%
7.	good	196	8,728	500,960	2.25%	0.04%
8.	great	186	8,728	259,611	2.13%	0.07%
9.	remarkable	182	8,728	16,130	2.09%	1.13%
10.	amazing	178	8,728	25,936	2.04%	0.69%
11.	certain	171	8,728	91,490	1.96%	0.19%
12.	true	151	8,728	111,513	1.73%	0.14%
13.	signficant	127	8,728	86,284	1.46%	0.15%
14.	wrong	127	8,728	82,200	1.46%	0.15%
15.	fascinating	121	8,728	9,600	1.39%	1.26%
16.	unusual	112	8,728	23,034	1.28%	0.49%
17.	funny	111	8,728	29,199	1.27%	0.38%
18.	nice	71	8,728	62,472	0.81%	0.11%
19.	worse	68	8,728	38,616	0.78%	0.18%
20.	disturbing	63	8,728	7,343	0.72%	0.86%
21.	striking	60	8,728	7,385	0.69%	0.81%
22.	difficult	57	8,728	87,596	0.65%	0.07%
23.	exciting	57	8,728	15,196	0.65%	0.38%
24.	unique	56	8,728	31,552	0.64%	0.18%
25.	sad	53	8,728	22,160	0.61%	0.24%
26.	cool	51	8,728	36,015	0.58%	0.14%
27.	likely	50	8,728	90,413	0.57%	0.06%
28.	best	45	8,728	179,702	0.52%	0.03%
29.	essential	45	8,728	27,361	0.52%	0.16%
30.	shocking	45	8,728	5,953	0.52%	0.76%

The adjectives ranked in Tables 4 and 5 can be classified into several semantic groups. The first category is constituted by adjectives evoking the IMPORTANCE frame. In this frame, the importance of a certain entity or situation is assessed by a speaker, as shown in the example (3) above. *Important*, the most significant lexeme of this group, is followed by *significant*, *crucial* and *critical* in ranks 13, 38, and 45. The quantitative findings reveal that *important* amounts to 13.51% of the occurrences of the pattern under study in COCA, and that 0.59% of occurrences of the same adjective can be observed in that specific construction. By comparison, the two other adjectives are much less relevant slot fillers for the pattern (attraction scores: 1.46%, 0.45%, and 0.39%) and hinge on this pattern to a lesser degree (reliance scores: 0.15%, 0.20%, and 0.07%).

The second group consists of adjectives invoking the STIMULUS_FOCUS frame, a frame in which a certain stimulus (an object or event) brings about a particular feeling, emotion or experience in a speaker, as in the examples (1), (2), (4) and (5) above. *Interesting*, ranked second, is the most significant lexeme of this set. It is attracted to the pattern in a proportion of 12.95%, and relies on the pattern in a proportion of 2.00%. It is accompanied by a range of adjectives indicating that someone judges a certain state of affairs to be *surprising*, *remarkable*, *amazing*, *fascinating*, *funny*, *nice*, *disturbing*, *striking*, *exciting*, *weird*, *sad*, *cool*, *shocking*, *strange*, *extraordinary*, *impressive*, *frustrating*, *intriguing*, *ironic*, *astonishing*, or *scary.*

Some adjectives mentioned above also seem to instantiate other semantic frames. For example, *remarkable* in rank 9, along with *different*, *unusual*, *unique* and *special*, can be interpreted with reference to the UNIQUENESS frame, a semantic frame in which a certain state of affairs (an entity or event) is evaluated with regard to whether it exhibits a unique characteristic that distinguishes it from other entities or events, as in (7). *Amazing*, *astonishing*, *extraordinary*, and *cool* can in turn be relativized to the DESIRABILITY frame, in which a certain evaluee (an entity or situation) is judged for its quality (good or bad) relative to other instances of its type, as in *What is good is that there is a commitment*. *Bad, good, great, best, worse,* and *unfortunate* also reflect this frame, as in *What is unfortunate is that many of the other elements of the profession are still missing. Bad* and *good* (like *wrong* in rank 14) can be also used to judge the morality and rightness of somebody's behaviour, thereby evoking the MORALITY_EVALUATION frame, as in *What was bad was that everybody thought I was drinking again.*

The next significant group comprises adjectives pertaining to the OBVIOUSNESS frame. This frame describes a certain phenomenon (an entity or fact) with respect to the degree of likelihood, given some evidence or circumstances, as in *What is clear is that the pope could have done more*. Its main lexeme, *clear* (attraction score 6.40%), in rank 3, is followed by *evident*, *obvious*, and *unclear* in ranks 36, 37, and 45, respectively. Apart from these adjectives, the list contains two adjectives, the meanings of which can be understood with respect to the NECESSITY frame, in which a certain requirement must be fulfilled as a prerequisite for obtaining or occurring, as in *What is necessary is a significant increase in the cost*. *Essential*, occupying the twenty-ninth position, precedes *necessary*, ranked thirty-first. Both achieved similar scores for attraction (0.52% and 0.50%) and reliance (0.16% and 0.08%).

Among the most significant lexemes in the ranking list, there are also adjectives, such as *new*, *true*, *certain*, *likely*, *relevant*, and *real.* The first adjective evokes the FAMILIARITY frame, a scene in which a certain entity (a concrete or abstract entity) is assessed in terms of how familiar or recognizable it is to a cognizer, as in (6). *True* is used in the pattern in the situations where a particular piece of information is evaluated for its accuracy, as in *What is true is that Hemon is a world class talent.* Thus, its meaning can be interpreted relative to the CORRECTNESS frame. The next two adjectives instantiate the LIKELIHOOD frame, a frame in which the likelihood of a hypothetical event is assessed by a speaker, as in *What 's certain is that there will be a lot more Americans. Relevant* pertains to the RELEVANCE frame, in which a certain phenomenon is directly connected with a mental endeavor (some cognitive action,

such as discussion or decision) of a cognizer, as in *What is relevant to this discussion is the context in which these desperate, violent acts occur.* Finally, *real* is used in the pattern to declare the existence of an entity (abstract or concrete), as in *What is real is that vast debt exists and money does not.* Hence, it invokes the EXISTENCE frame.

Table 5. The next 20 strongly attracted adjectives

rank	adjectives	a	x	e	attraction	reliance
31.	necessary	44	8,728	55,725	0.50%	0.08%
32.	relevant	44	8,728	19,851	0.50%	0.22%
33.	strange	44	8,728	30,285	0.50%	0.15%
34.	extraordinary	43	8,728	16,802	0.49%	0.26%
35.	real	43	8,728	173,713	0.49%	0.02%
36.	evident	40	8,728	11,555	0.46%	0.35%
37.	obvious	40	8,728	28,637	0.46%	0.14%
38.	crucial	39	8,728	19,576	0.45%	0.20%
39.	impressive	39	8,728	13,026	0.45%	0.30%
40.	weird	38	8,728	12,966	0.44%	0.29%
41.	frustrating	37	8,728	4,956	0.42%	0.75%
42.	intriguing	36	8,728	5,039	0.41%	0.71%
43.	bad	35	8,728	122,517	0.40%	0.03%
44.	critical	34	8,728	52,225	0.39%	0.07%
45.	unclear	33	8,728	7,971	0.38%	0.41%
46.	astonishing	32	8,728	4,132	0.37%	0.77%
47.	ironic	32	8,728	5,339	0.37%	0.60%
48.	scary	32	8,728	9,238	0.37%	0.35%
49.	special	31	8,728	107,140	0.36%	0.03%
50.	unfortunate	31	8,728	6,911	0.36%	0.45%

Concluding remarks

This article has sought to identify different properties of the *What be ADJ be*-construction and the most common adjectives occurring in this pattern. The results of the corpus-based study have revealed that the construction in question has many different variants, is relatively infrequent in fiction, tends to be used very frequently in academic and spoken registers, and exhibits a strong tendency to occur with particular categories of adjectives activating the following semantic frames: IMPORTANCE, STIMULUS FOCUS, NECESSITY, RELEVANCE, UNIQUENESS, DESIRABILITY, OBVIOUSNESS, LIKELIHOOD, or FAMILIARITY.

The frequent occurrence of the adjectives evoking these frames is connected with semantic and discourse-functional properties of the *What be ADJ be*-construction, though their high frequencies may in time also enhance, reinforce, and affect those properties. The pattern under consideration can be treated as a discourse-functional

construction that reveals a strong preference for adjectives expressing a speaker's/writer's evaluative judgment about the importance, necessity, relevance, uniqueness, desirability, obviousness, likelihood, or familiarity, of a certain entity, fact, statement, or state of affairs. In addition, it is commonly used with adjectives denoting specific feelings, emotions or experiences to highlight that a certain stimulus (an entity or event) brings about such states in a speaker's mind. Thus, the construction serves an evaluative and highlighting function and directs readers'/hearers' attention to new pieces of information introduced at the end of a sentence.

After a summary of the findings, several limitations of this study need to be acknowledged. First, it was impossible to present and assess the quantitative findings for all the adjectives found in the construction. Second, the current study was unable to discuss sub-patterns and their instantiations in much detail. Thirdly, the study was confined to the examples found in American English. Lastly, the current study was unable to establish which adjectives are strongly associated with, or loosely related to, different sub-patterns. Future research might, therefore, concentrate on comparing and contrasting adjectives occurring in different variants of the construction. Such a quantitative study would uncover minor distributional differences in the use of adjectives in different sub-patterns, and it could allow us to explain the existence and degree of semantic differences in their usage. Moreover, given the limitation of the current study to COCA and the potential existence of subtle variations in the use of adjectives in the construction, it is recommended that future research explore the occurrence of this construction in the British National Corpus across various kinds of both written and spoken registers.

References

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

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

Boas H., 2017, *Computational resources: FrameNet and Constructicon*, [in:] *The Cambridge handbook of cognitive linguistics*, ed. B. Dancygier, Cambridge: Cambridge University Press, p. 549–573.

Collins P., 1991, *Cleft and pseudo-cleft constructions in English*, London and New York: Routledge.

Collins P., 2006, *It-clefts and wh-clefts: Prosody and pragmatics*, „Journal of Pragmatics" 38, p. 1706–1720.

Croft W., Cruse D.A., 2004, *Cognitive linguistics*, Cambridge: Cambridge University Press.

Declerck R., 1984, *The pragmatics of it-clefts and wh-clefts*, „Lingua" 64, p. 251–289.

Declerck R., 1988, *Studies on copular sentences: Clefts and pseudo-clefts*, Leuven: Leuven University Press.

Delahunty G., 1984, *The analysis of English cleft sentences*, „Linguistic Analysis" 13, p. 63–113.

Deroey K.L.B., 2011, *What they highlight is...: The discourse functions of basic wh-clefts in lectures*, „Journal of English for Academic Purposes" 11(2), p. 112–124.

Fillmore C.J., 2006, *Frame semantics, Encyclopedia of language and linguistics*, ed. K. Brown, Amsterdam: Elsevier, p. 613–620.

Fillmore C.J., Baker C., 2010, *A frames approach to semantic analysis*, [in:] *The Oxford handbook of linguistic analysis*, eds. B. Heine, H. Narrog, Oxford: Oxford University Press, p. 791–816.

Gast V., Levshina N., 2014, *Motivating w(h)-clefts in English and German: A hypothesis-driven parallel corpus study*, [in:] *Frequency, forms and functions of cleft constructions in Romance and Germanic: Contrastive, corpus-based studies*, ed. A.-M. De Cesare, Berlin: De Gruyter, p. 377–414.

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

Goldberg A., 2013*, Constructionist approaches*, [in:] *The Oxford handbook of construction grammar*, eds. T. Hoffmann, G. Trousdale, Oxford: Oxford University Press, p. 15–31.

Hedberg N., 1990, *Discourse pragmatics and cleft sentences in English*. PhD dissertation, University of Minnesota.

Herriman J., 2003, *Negotiating identity: the interpersonal functions of wh-clefts in English*, „Functions of Language" 10(1), p. 1–30.

Herriman J., 2004, *Identifying relations: the semantic functions of wh-clefts in English*, „Text" 24(4), p. 447–469.

Hoffmann Th., 2017, *Construction grammars*, [in:] *The Cambridge handbook of cognitive linguistics*, ed. B. Dancygier, Cambridge: Cambridge University Press, p. 310–329.

Hoffmann Th., Trousdale G., 2013, *Construction Grammar: Introduction*, [in:] *The Oxford handbook of construction grammar*, eds. T. Hoffmann, G. Trousdale, Oxford: Oxford University Press, p. 1–14.

Huddleston R., Pullum G., 2002, *The Cambridge grammar of the English language*, Cambridge: Cambridge University Press.

Kim J., 2007, *English cleft constructions: Corpus findings and theoretical implications.* „Studies in Generative Grammar" 17, p. 99–122.

Miller J., Weinert R., 1998, *Spontaneous spoken language: syntax and discourse*, Oxford: Clarendon Press.

Prince E., 1978, *A comparison of wh-clefts and it-clefts in discourse*, „Linguistic Society of America" 54(4), p. 883–906.

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

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

Wiliński J., 2020, *The preposition under and its noun collocates in the under-NOUN pattern: A quantitative corpus-based investigation*, „Crossroads: A Journal of English Studies" 31, p. 23–43.

Wiliński J., 2021, *Nouns in the be of N-construction: A corpus-based investigation*, „Studies in Logic, Grammar and Rhetoric" 66(3), p. 747–768.

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/
Mediabistro. Available from http://www.mediabistro.com

What-clefts with adjectives in English: A corpus-based analysis

Abstract

This paper aims to investigate the *what*-cleft construction with adjectives and establish its structural, semantic, and distributional features by adopting frame semantics and usage-based construction grammar, exploiting the data from the Corpus of Contemporary American English (COCA), and applying quantitative corpus-based methodology. To this end, the author extracts the occurrences of the *What be ADJ be*-construction from a large corpus of naturally-occurring data, determines its structural, semantic, distributional, and discourse-functional properties, and identifies adjectives that are strongly associated with the construction in question. The paper makes a significant contribution to a growing body of literature on the *what*-cleft construction by conducting a qualitative and quantitative analysis of one of its variants, a grammatical pattern with adjectives that has not been hitherto investigated in much detail.

Zdania rozszczepione z przymiotnikami w języku angielskim

Streszczenie

W niniejszym artykule autor stawia sobie za cel zbadanie konstrukcji *what*-cleft z przymiotnikami i ustalenie jej cech strukturalnych, semantycznych i dystrybucyjnych poprzez zastosowanie semantyki ramowej i gramatyki konstrukcji, wykorzystanie danych z korpusu (COCA) i użycie ilościowej metodologii korpusowej. W tym celu autor wydobywa wystąpienia konstrukcji z dużego korpusu naturalnie występujących danych, określa jej strukturalne, semantyczne, dystrybucyjne i dyskursowo-funkcjonalne właściwości oraz identyfikuje przymiotniki, które są silnie związane z omawianą konstrukcją. Artykuł wnosi znaczący wkład do rosnącej literatury na temat konstrukcji *what*-cleft poprzez jakościową i ilościową analizę jednego z jej wariantów, konstrukcji z przymiotnikami, który nie był do tej pory szczegółowo badany.