A Quantitative Corpus-Driven Approach to Disambiguation of Synonymous Grammatical Constructions

Viktoriia Zhukovska[0000-0002-4622-4435]

Zhytomyr Ivan Franko State University, 40, Velyka Berdychivska Str., 10008
Zhytomyr, Ukraine
victoriazhukovska@gmail.com

Abstract. Adopting usage-based construction grammar and quantitative corpus linguistic research approaches, this study attempts to empirically analyze the distributional properties of the two English synonymous grammatical constructions, i.e. detached augmented and unaugmented Participle I clauses with the explicit subject. Despite the extensive research on the morphological and functional features of the syntactic patterns under study, semantic and pragmatic dissimilarities between the two alternative syntactic structures thus far have not been studied, at least in a way of the quantitative examination of the subject slot of these constructions. Applying the methods of simple collexeme analysis and distinctive collexeme analysis to the linguistic data retrieved from the BNC-BYU corpus, the study explores the semantics of the investigated pair of constructions by identifying semantic frames instantiated by their significantly attracted noun collexemes. The distributional data obtained prove that the analyzed synonymous constructions are pragmatically distinct and semantically determined.

Keywords: usage-based construction grammar, quantitative corpus linguistics, detached participial constructions with explicit subject, collexeme analysis, distinctive collexeme analysis.

1   Introduction

The English detached present participle clauses with the explicit subject considered in this study are illustrated by the examples collected from the BNC-BYU corpus [1]:

1. *He clutched at a rail and held on,* **[[NP *heart*] [XP *thumping*]], [[NP *the blood*] [XP *pounding in his ears*]], [[NP *his mind*] [XP *wailing for mercy*]]** [BNC-BYU, B1X];
2. *I stood up, holding on to the back of my chair,* **[[NP *my heart*] [XP *beating like a hammer*]]** [BNC-BYU, FPU];
3. *The experiment was repeated many times,* **[with [NP *the bats*] [XP *taking turns to be the starved victim*]]** [BNC-BYU, ARR];
4. *The plain is like a field of poppies,* **[with [NP *the flowers*] [XP *growing most thickly near the river*]]** [BNC-BYU, FAJ].

The syntactic patterns under scrutiny are part of a minimally bi-clausal structure comprising a matrix clause and a Participle I clause with its own explicit subject. The participial clause is detached from the matrix by a punctuation mark, primarily by a comma. The participial clauses with an explicit subject are secondary predication patterns of binary structure [NP XP], where (NP) is a secondary subject, different from the subject of the matrix clause, and (XP) is a secondary predicate. The patterns have fixed subject and predicate slots, and their syntactic form can be represented schematically as [with/with-less[Subj COMMON NOUN][V PARTICIPLE I]].

In recent years, there has been a growing interest in the study of qualitative and quantitative features of the participial clauses with the explicit subject [2; 3; 4; 5; 6; 7, to name but a few]. While this paper draws on the insights of the existing studies, it differs from these works in the aspect that it focuses only on the detached Participle I clauses with the explicit subject that have been traditionally viewed as part of a family of absolute constructions [3; 4], small clauses [8], or non-finite/verbless adjunct clauses [9]. In our research, we subsume the syntactic patterns exemplified in (1) – (4) under the term detached constructions with the explicit subject, following B. Combettes [10] and S. Thompson [11]. The detached constructions with the explicit subject embrace non-finite secondary predication structures where a secondary subject is expressed by a noun group (NP) and a secondary predicate is represented by a non-finite verb form (Participle I, Participle II, to-Infinitive) or a nonverbal part of speech (NP, PP, AdvP, AP). Though such structures are detached from the matrix clause by a punctuation mark, they are logically and semantically related to it.

The unaugmented (*with-less* [5]) and augmented (*with* [5]) Participle I clauses with the explicit subject are considered synonymous or alternative [7; 12; 13]. With regard to the Principle of No Synonymy [14, p.67], according to which dissimilarities in form between constructions always entail semantic and pragmatic dissimilarities, we hypothesize that the analyzed syntactically distinct (with or without the augmentor *with*) constructions are semantically motivated and perform different pragmatic functions. We assume the investigated patterns represent different constructions instantiating a symbolic (structural) configuration, namely a (complex) sign in which a particular form is paired with a particular function [15; 14; 16; 17; 18].

To validate the stated hypothesis, we carry out an empirical analysis of the English augmented and unaugmented present participle clauses with the explicit subject from a usage-based construction grammar perspective utilizing the methods of quantitative corpus linguistics. To the author's knowledge, the distributional properties of the clauses under study have not been yet considered with the use of the corpus linguistic method of collostructional analysis to examine the interrelation between the syntactic patterns and lexemes that fill them. Moreover, the existing research adopting the collostructional analysis mainly focuses on the relationships between verbs and constructions [19], but since the most conspicuous feature of the analyzed clauses is the presence of the explicit subject different from the matrix subject, this study aims to examine how nouns in the subject slot shape the meaning of the two synonymous constructions. In addition, it determines the underlying semantic factors that motivate the distribution of nouns in the investigated syntactic structures and thereby defines semantic and pragmatic contrast between them.

2   Theoretical and Methodological Assumptions

In our study, we adopt the terminology and theoretical assumptions of construction grammar. A constructional approach to grammar rests on the premise that grammar should be described as a structured inventory of form-meaning pairings, collectively referred to as constructions: Fillmore (1988); Goldberg (1995, 2006); Croft (2008); Hilpert (2019). Knowledge of language is understood as knowledge of a network of constructions [18, p. 2].

These basic principles are shared by all construction grammarians, though it is necessary to note that construction grammar is not a monolithic theory, but rather a family of approaches, each of which has its distinctive postulations [20]. For instance, usage-based construction grammar recognizes frequencies of usage or occurrence of a grammatical unit as a reflecting factor that influences the representation of grammatical units in the minds of speakers. Thus, the frequency of occurrence of a construction determines the degree of its entrenchment in the speech community [21]. Usage-based construction grammar views formally similar or even identical constructions to be different sub-constructions or even constructions, if they have different communicative functions.

To investigate the distributional dissimilarities between the augmented and unaugmented present participial clauses with the explicit subject empirically and reasonably objectively we adopt the collostructional analysis, which is one of the most widely applied methods of quantitative corpus linguistics. The description of the method and its application to study various linguistic phenomena are well-documented and illustrated in the papers of its developers A. Stefanowitsch and St. Th. Gries [22; 23; 24; 25].

The collostructional analysis (the word "collostruction" is a combination of two words "collocation" and "construction") is a set of methods aimed at investigating the relationships between the words and the grammatical structures they occur in [19, p. 290]. This method comprises three major variants (the simple collexeme analysis, the distinctive collexeme analysis, and the co-varying collexeme analysis), each of which concentrates on the particular issue concerning the relationships between lexemes and grammatical constructions. The simple collexeme analysis investigates how strongly lexemes are attracted to a particular slot in a construction [19, p. 291]. The distinctive collexeme analysis contrasts two or more functionally similar constructions regarding the lexical items that occur in them [19, p. 296]. The co-varying collexeme analysis explores interdependencies between lexemes in two different slots of the same construction [19, p. 300]. The lexemes most attracted by a particular slot of the construction are collexemes of that construction; whereas, a construct associated with a particular lexeme is called a collostruct; the combination of a collexeme and a collostruct is called a collostruction.

The output of the collostructional analysis provides important insights into the semantics of the construction by identifying semantic classes of the significantly attracted collexemes. The method is based on the principle of semantic compatibility, according to which "a word may occur in a construction if it is semantically compatible

with the meaning of the construction (or, more precisely, with the meaning assigned by the construction to the particular slot in which the word appears)" [22, p. 213].

Following the observation by A. Stefanowitsch, St. Th. Gries, [23, p. 34] and D. Schönefeld [26, p. 26] that the lexemes appearing in different slots of the given construction display semantic coherence grounded on the world knowledge as organized in frames, the semantics of the analyzed participial clauses is examined in terms of semantic frames instantiated by their collexemes. The theory of semantic frames was worked out by Ch. J. Fillmore and his colleagues within Frame Semantics [27; 28; 13]. The underlying assumption of this linguistic approach is that the meaning of a word is best comprehended on the basis of a semantic frame, that provides important background knowledge of different types of events, relations, or entities and the participants in it [29]. Semantic frames are retrieved from the FrameNet project [29], an English lexical database where the meanings of words are defined based on semantic frames that inform their meaning [13]. Each lexical description in FrameNet connects a particular lexical unit to the semantic frame that it evokes. The analysis of semantics of the investigated constructions in terms of semantics frames provided by the FrameNet project has not yet been applied to study nouns occurring in their subject slots.

3   Corpus, Data and Statistical Procedure

In this article, we focus on the distributional properties of the detached Participle I with the explicit subject constructions augmented by *with* or unaugmented (*with-less*): [*with/with-less*[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]]. We analyze only those instances where the investigated patterns are separated from the rest of a sentence with a punctuation mark (a comma, a dash or a bracket). Such formal detachment automatically sifted off the cases of prepositional phrases introduced by the comitative preposition *with*.

The analysis has been undertaken in five stages comprising the following procedures:

- extracting [*with/with-less*[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] constructions from the well-balanced British National Corpus (BNC-BYU). BNC-BYU's search engine was used to retrieve the occurrences of the investigated constructions in all positional variants to the matrix clause (initial, medial and final). Each concordance line was manually examined to exclude all false hits. The observed frequency of the constructions was then calculated;
- identifying the noun slot of each construction, selecting all common noun lemmas and calculating the observed frequency of each noun lemma in the constructions under study;
- statistical evaluation of the collostructional strength between noun lexemes and the subject slot of each construction by using the simple collexeme analysis;
- statistical evaluation of the nouns attracted to one or another construction by utilizing the distinctive collexeme analysis;
- semantic analysis of the data generated by the simple collexeme and the distinctive collexeme analyses in terms of the semantic frames as presented in FrameNet [29].

The most significant collexemes are considered to display the typical semantic properties of the two analyzed constructions.

The sample consists of 2950 examples of the [$_{with}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction and 1535 examples of the [$_{with\text{-}less}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction, with 1445 and 236 different noun lexemes correspondingly.

This paper adopts the simple collexeme and the distinctive collexeme analyses to determine lexemes that are attracted to the [Subj] slot of each of the analyzed constructions. This approach is employed following A. Stefanowitsch [19, p. 297], who points out that a study directed to uncover both the similarities and the dissimilarities between the constructions should combine the distinctive collexeme analysis with separate collexeme analyses for each construction. The simple collexeme analysis generates a ranking of collexemes according to the degree of association strength to the construction and establishes a distinction between lexemes that are significantly more frequent than expected (significantly attracted collexemes) and significantly less frequent (significantly repelled collexemes). The distinctive collexeme analysis investigates semantic properties of pairs of structurally or functionally (semantically) similar grammatical constructions by establishing the lexemes that appear significantly more frequently with one construction than the other. It identifies distributional dissimilarities between the nearly equivalent constructions by comparing the frequency of lexemes that fill one constructional slot to their frequencies in a corresponding slot of the other construction in a pair. The procedure of calculating the association strength is the same as for the simple collexeme analysis, the only exception is that the second construction occupies the place of the corpus.

To calculate the association strength of each collexeme, four types of frequencies of the lexical unit (L) and the construction (C) in the corpus are retrieved [22, p. 218]: 1) the frequency of the lexeme (L) in the construction (C); 2) the frequency of the lexeme under study (L) in all other constructions in the corpus; 3) the frequency the investigated construction (C) with other lexemes; 4) the frequency of all other constructions with other lexemes in the corpus. These data are cross-tabulated in a 4×4 table for each lexeme occurring in the analyzed construction. The Fisher-Exact test or the Fisher-Yeats test is adopted as a default association measure since it produces precise results even on small sample sizes and fits the distributional reality of linguistic data.

The simple collexeme and distinctive collexeme analyses are performed using a script written by St. Th. Gries for freely available R statistical software environment downloaded from [30].

4   Results

The sample obtained from the BNC-BYU corpus contains 2950 occurrences of the [$_{with}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction and 1535 occurrences of the [$_{with\text{-}less}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction. The observed frequencies highlight a noticeable quantitative discrepancy between the patterns under study. The former significantly outnumbers the latter and is approximately twice as frequent in the

corpus. The data show that the [$_{with}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction co-occurs with 1445 types of nouns, out of which 965 types are used merely once with this pattern. These account for 66.78% of the total number of items in the construction. By contrast, the [$_{with\text{-}less}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction co-occurs with 236 types of nouns, out of which 126 types occur once in it. These items make up 53.40% of all types of nouns in the pattern. It is apparent that items with low frequency are rather loosely associated with the analyzed constructions, and the rest of the items are obviously more strongly attracted to one of the compared constructions. At the same time, hapax legomena, i.e. items occurring only once in the given construction, signify the potential productivity of this construction. The potential productivity ratio is calculated by dividing the number of hapaxes of a construction by the overall token frequency of this construction [31, p. 128]. The bigger the ratio, the higher the potential productivity of the syntactic pattern and a greater number of new types will be produced on the basis of a constructional schema [ibid.]. The estimated productivity ratio of the [$_{with}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction is 0.67 and of the [$_{with\text{-}less}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction is 0.53. Though both indexes are rather high, still the construction introduced by the augmentor *with* appears to be more productive.

The application of the collexeme and distinctive collexeme analyses allow us to corroborate the existence and degree of the semantic difference between the constructions under investigations and determine the semantic restrictions they impose on the nouns filling their subject slots. This aim is achieved by determining the nouns which are highly distinctive for one of the analyzed constructions.

4.1 Collexeme Analysis of the [$_{with\text{-}less}$[Subj$_{\text{COMMOM NOUN}}$][V$_{\text{PARTICIPLE I}}$]] Construction

The results of the collexeme analysis show that out of 236 different nouns in the construction 107 items reveal a significant attraction to the construction (the coll. strength > 1,30103 = p <0,05). But the lower the p-value the greater is the probability that the observed frequency lexemes distribution is not random and the greater is the strength of attraction between the lexeme and the construction. The data suggest that 60 nouns reach high significance (coll. strength > 3 = p <0.001). Table 1 presents the top 20 attracted collexemes ranked according to the value of coll. strength.

The significantly attracted collexemes of the subject slot of the [$_{with\text{-}less}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction are revealing of the pattern's semantics. The analysis of the attracted nouns in terms of semantic frames is expected to shed further light on the semantic and functional specification of the analyzed syntactic pattern.

The largest group of the most strongly attracted nouns of the [$_{with\text{-}less}$[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction includes nouns evoking the BODY_PARTS frame and contains somatisms, i.e. nouns naming the whole body (*body*, *whole being*), external parts of the human body (*head*, *face*, and their constituent parts *eyes*, *lips*, *cheeks*, etc.), internal organs (*heart*, *stomach*, etc.), limbs (*hands*, *legs*, *arms*, etc.), interior elements of the anatomy (the elements of the systems that do not represent parts of the body but are part of it (*voice*, *mind*, *gaze*, *smile*, *breath*, etc.)), emotions and feelings (*spirits*, *senses*).

Table 1. The top 20 significantly attracted collexemes of the
[*with-less* [Subj_{COMMON NOUN}][V_{PARTICIPLE I}]] construction

№	Nouns	Coll. strength	№	Nouns	Coll.strength
1.	eyes	Inf	11.	chest	34.342606
2.	hand(s)	137.473224	12.	gaze	32.006045
3.	tears	78.747349	13.	weather	30.256535
4.	face	63.857815	14.	mind	27.646698
5.	mouth	56.463974	15.	thing(s)	26.037315
6.	heart	49.836669	16.	teeth	24.711382
7.	lips	47.913696	17.	smile	21.995194
8.	voice	46.859533	18.	chin	21.539576
9.	arms	39.742112	19.	head	21.45395
10.	fingers	36.871427	20.	body	20.764775

The second set is constituted by the nouns *weather*, *sun*, *wind*, whose meaning can be interpreted regarding the WEATHER frame. This frame indicates ambient conditions of temperature, precipitation, windiness, and sunniness pertaining at a certain place and time. The noticeable occurrence of the noun *weather* (rank 13) consistently followed by the participle *permitting* suggests that the expression *weather permitting* is a lexically filled construction.

The third group of strongly attracted nouns of the construction under study comprises nouns such as *crew*, *conglomerate* whose meaning can be interpreted with reference to the AGGREGATE frame that contains nouns denoting aggregates of individuals.

The next category is constituted by the noun *father* representing the KINSHIP frame. This frame contains words that denote kinship relations.

Among the most strongly associated nouns of the investigated construction, there are lexemes whose meaning can be interpreted regarding the PURPOSE, REASON, EMPHASING, INCLUSION, and SIMILARITY frames. The PURPOSE frame evoked by the nouns *object*, *purpose*, *intention* describes a state of the world that does not currently hold but the agent wants to achieve. The frame REASON represented by the nouns *reason*, *thing(s)* indicates the eventuality that motivates the agent to perform a particular action, explanation or justification. The noun *emphasis* instantiates the EMPHASING frame, a schematic knowledge structure attributing a degree of importance to an event, state of affairs, an attribute or an entity that has the potential to influence the success of it. The INCLUSION frame is invoked by the noun *exception(s)*. This frame describes a complex entity including components (parts), within this frame the given noun means a part that has been excluded from the total. The meaning of the noun *difference* is understood with respect to the SIMILARITY frame that shows how one entity is the same or different from other entities. Within this frame, the noun *difference* means the state or condition of being different.

4.2 Collexeme analysis of the [*with*[Subj COMMON NOUN][V PARTICIPLE I]] Construction

The same analysis has been undertaken for the [*with*[Subj COMMON NOUN][V PARTICIPLE I]] construction. The output for the collexeme analysis is summarized in Table 2.

Table 2 presents the first 20 most strongly attracted nouns out of 89 items with coll. strength > 3 = p <0.001 ranked according to the value of coll. strength.

It should be noted here, that the [*with*[Subj COMMON NOUN][V PARTICIPLE I]] construction appears to be rather diversified in terms of its semantics. Some nouns show low frequency in the corpus BNC-BYU and as a result they turn to be strongly attracted to the construction. We exclude such items from consideration here and consider them as additional proof to the productivity of the construction that can to co-occur with new not so frequent lexemes.

Table 2. The top 20 significantly attracted collexemes of the
[*with*[Subj COMMON NOUN][V PARTICIPLE I]] construction

№	Nouns	Coll. strength	№	Nouns	Coll. strength
1.	people	25.250147	11.	side	12.131886
2.	proceeds	19.989733	12.	profit	11.171691
3.	student	16.506529	13.	rain	8.746675
4.	poll	16.06899	14.	emphasis	7.739556
5.	wind	15.618497	15.	temperature	7.549623
6.	price	14.626037	16.	unemployment	7.175408
7.	sun	14.624936	17.	train	7.105128
8.	tear	14.042232	18.	winner	6.988417
9.	remainder	13.966305	19.	patient	6.874586
10.	inflation	12.72685	20.	party	6.67096

The results of the collexeme analysis suggest that the most numerous group of the most strongly associated nouns is constituted by the nouns referring to the sphere of commerce (*proceeds*, *prices*, *inflation*, *profits*, *sales*, *cost*, etc.) (21 items). This set of nouns evokes the frame COMMERCE_SCENARIO that instantiates a situation when a buyer and a seller perform the exchange of money and goods.

The next category is the nouns *people*, *woman*, *children*, *individuals*, etc. This set of nouns can be understood with reference to the PEOPLE frame, which concerns individuals that may have such characteristics as age, origin, ethnicity, etc. Directly related to this group is the set of nouns instantiating the frame PEOPLE_BY_VOCATION denoting people of various occupations (*archeologist*, *investigator*, *author*, etc.).

Another group of nouns includes lexemes which denote socio-political realia. The nouns (*minister*, *mayor*, etc.) are understood within the semantic frame LEADERSHIP, and the nouns *election*, *polls*, etc. evoke the semantic frame CHANGE OF LEADERSHIP. The semantic frame PEOPLE ALONG POLITICAL SPECTRUM is evoked by the noun *Republicans*, while the frame LAW_ENFORCEMENT_AGENCY is instantiated by the noun *police*.

The meaning of the nouns *pitch*, *score*, *player*, etc. is understood within the semantic frame COMPETITION. The frame is concerned with the idea that people participate in an organized, rule-governed activity in order to achieve some advantageous outcome.

Another category of strongly attracted lexemes includes nouns such as *team*, *staff*, *company*, etc. as well as nominalized numerals *hundreds*, *thousands*. These nouns denoting groups of individuals instantiate the frame AGGREGATE.

The next group of nouns (*tear*, *hair*, *tail*, etc.) consists of lexemes reflecting the BODY_PARTS frame.

The set of nouns *weather*, *sun*, *temperature*, *snow* and *rain* evokes two semantic frames WEATHER and PRECIPITATION. The meaning of the nouns *river*, *precipices*, etc. can be interpreted regarding the NATURAL FEATURES frame. The noun *winter* instantiates the semantic frame CALENDRIC UNIT.

Another category of nouns (*train*, *jetlag*) instantiates the semantic frame VEHICLE that concerns the vehicles that people use for transportation. The meaning of the nouns *wires*, *program*, etc. is comprehended regarding the frame ARTIFACT, where an artifact is deliberately made or modified by an intelligent entity.

The nouns *emphasis*, *attention*, and *violence* evoke the semantic frames EMPHASING, ATTENTION and VIOLENCE correspondingly.

One more peculiar feature of the [*with*[Subj$_{COMMON\ NOUN}$][V$_{PARTICIPLE\ I}$]] construction is that the inclusion into a frame is not exclusive, a noun may be attested to more than one frame because it is used in more than one way. For instance, nouns *student, pupil, teacher, tutor*, etc. evoke the frame PEOPLE_BY_VOCATION, but at the same time, they instantiate the semantic frame EDUCATION_TEACHING that contains words signifying teaching and participants in teaching. Another example the nouns evoking the semantic frame COMPETITION (*player*, *commentator*) and the frame LEADERSHIP (*minister*, *mayor*) can be attributed to the frame PEOPLE_BY_VOCATION.

4.3 Distinctive Collexeme Analysis of the Investigated Constructions

The next step of the research is to identify what lexemes are distinctive for each of the constructions in question. With this in mind, we apply the distinctive collexeme analysis to distinguish between the [*with-less*[Subj$_{COMMON\ NOUN}$][V$_{PARTICIPLE\ I}$]] and the [*with*[Subj$_{COMMON\ NOUN}$][V$_{PARTICIPLE\ I}$]] constructions. As A. Stefanowitsch [2013: 297] states, the distinctive collexeme analysis highlights the differences between the constructions and hides their similarities. The lexemes that are significantly associated with both constructions but used significantly more frequently in one of them will be regarded as distinctive for that construction. The total number of common lexemes between the constructions in question is 116. These nouns constitute 8.03% from the total number of different noun lexemes in the [*with*[Subj$_{COMMON\ NOUN}$][V$_{PARTICIPLE\ I}$]] construction and 49.15% in the [*with-less*[Subj$_{COMMON\ NOUN}$][V$_{PARTICIPLE\ I}$]] construction.

The results from Table 3 reveal that the two synonymous constructions do indeed possess distinctive collexemes and exhibit distinct preferences for the nouns in their subject slots. The augmented [*with*[Subj$_{COMMON\ NOUN}$][V$_{PARTICIPLE\ I}$]] construction is strongly associated with three nouns that evoke one semantic frame PEOPLE (*people*, *man*, *woman*) while the distinctive collexemes of the unaugmented [*with-less*[Subj$_{COMMON\ NOUN}$][V$_{PARTICIPLE\ I}$]] construction include 23 nouns of the semantic frames

BODY_PART (20 items), WEATHER (one item), KINSHIP (one item) and PURPOSE (one item).

Table 3. The top distinctive collexemes of the [*with*[Subj_{COMMON NOUN}][V_{PARTICIPLE I}]] and [*with-less*[Subj_{COMMON NOUN}][V_{PARTICIPLE I}]] constructions

words	obs. freq.1	obs. freq.2	exp. freq.1	exp. freq.2	pref. occur	coll. strength
people	7	68	25.86381	49.13619	with	6.415589
woman	1	20	7.241867	13.758133	with	2.783807
man	1	16	5.862464	11.137536	with	2.129318
eye	367	12	130.698452	248.301548	with-less	163.220082
hand	114	11	43.10635	81.89365	with-less	40.553806
face	56	2	20.001346	37.998654	with-less	23.312826
heart	39	3	14.483733	27.516267	with-less	14.63404
mind	29	1	10.345524	19.654476	with-less	12.180775
arm	35	5	13.794032	26.205968	with-less	11.345943
tear	43	13	19.311645	36.688355	with-less	10.041268
finger	27	3	10.345524	19.654476	with-less	9.459867
body	26	3	10.000673	18.999327	with-less	9.037761
weather	21	1	7.586718	14.413282	with-less	8.577909

Thus the distinctive collexeme analysis reveals semantic dissimilarities between the two constructions with approximately the same syntactic structure.

5   Discussion

The output of the collexeme analysis, carried out separately for each of the investigated constructions, proves the existence of noun lexemes which are most significantly attracted to the subject slot of a particular construction.

The results of the simple collexeme analysis reveal that the most numerous group of nouns in the [*with-less*[Subj_{COMMON NOUN}][V_{PARTICIPLE I}]] construction evokes the BODY_PART semantic frame. They include 46 items (76.67% of the significantly attracted nouns), with the leading lexeme *eye(s)* demonstrating an exceptionally significant degree of attraction to the construction (coll. strength is Inf).

The analyzed construction is predominantly used in fiction (the distribution of the constructions under study in BNC-BYU registers is given in the appendix Figure 1) where it serves as a means of effective packing descriptive information and giving additional details to the situation in the matrix clause. In case with body part nouns (somatisms) and nouns of kinship in the subject slot only a part of the matrix situation is specified, that is the participant expressed by the subject of the matrix clause acting as AGENT/EXPERIENCER. The referent of the subject of the [*with-less*[Subj_{COMMON NOUN}][V_{PARTICIPLE I}]] construction in extralinguistic context is directly connected with the referent of the matrix subject. Part of a human (animal) body is naturally connected with the whole body as a holistic discrete unit, while family members are connected by family ties.

The most significantly attracted and at the same time distinctive collexemes of the specified construction are the nouns *eye*, *hand* and *face*. Their prevailing use in the construction can be attributed to the specificity of fiction as a genre as well as to the peculiar nature of somatisms. The body part nouns give information about the objects they nominate as well as render information about the emotional and psychological state of an individual. Mimics, gestures, poses, and facial expressions may convey the inner state of a person. Somatisms differ in their potential to indirectly render information about a person's inner state. For instance, among non-verbal signs, facial expressions play a vital role in social interaction [32, p. 3454]. In communication, people tend to focus their attention on the faces of their interlocutors and most of all on the eyes. Therefore, faces and eyes are one of the most important means of expressing the diversity of human emotions and serve as a primary source of information about people's feelings, and this is reflected in language.

The construction under analysis instantiates a set of fully and partially lexically specified constructions. Nouns *thing(s)*, *object*, *purpose*, *intention*, etc. constitute a class of general nouns [33] or shell nouns [34]. They are considered "an open-ended functionally-defined class of abstract nouns that have, to a varying degree, the potential for being used as conceptual shells for complex, proposition-like pieces of information" [34, p. 4] and hold or encapsulate information realized elsewhere in the context. In Schmidt's classification [34, p. 88], the analyzed nouns are included into the group of factual shell nouns that describe facts or states of affairs. Functioning as subjects of the unaugmented [*with-less*[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction, these nouns summarize the matrix clause performing the support function [cf. 35, p. 99]. These constructions do not contribute to the factual meaning but add supplementary information (comments, specification, explanation, etc.). Typically intensified by the adjectives *main*, *major*, *only*, *other* and always followed by the participle *being*, these nouns appear to represent completely or partially lexically specified instantiations of the patterns: *(other) things being equal*, *the object being to V*, *the purpose being to V*, *the intention being to V*, etc. These constructions are used in written narrative texts other than fiction (newspaper, magazine, commerce, academic, non-academic).

The results of the distinctive collexeme analysis support the findings of the simple collexeme analysis. The distinctive collexemes include 23 nouns of the semantic frames BODY_PART, WEATHER, KINSHIP and PURPOSE. The analysis of the distinctive collexemes of the construction indicates a preference for inanimate subject referents (95% of the total analyzed nouns) that do not act intentionally and specify agentless actions. These inanimate subject referents are construed as PATIENTS of a state/process expressed by the present participle, with the AGENT/EXPERIENCER of the action presented by the predicate in the matrix clause. As suggested by their overwhelming quantity and high collocation strength, the subject referents denoting body parts and kinship relations constitute the inalienable property, representing part-whole relations with the matrix subsect referents. Body parts subject referents are construed as PARTITIVE with regard to the matrix subject.

Thus from the results of the distinctive analysis, it becomes evident that the [*with-less*[Subj$_{\text{COMMON NOUN}}$][V$_{\text{PARTICIPLE I}}$]] construction has one prominent and one more peripheral functions. The quantitative data suggest that the depictive function is central. With the majority of collexemes denoting body parts in the [Subj] slot, the construction adds information about physical characteristics and inner state of the matrix subject's

referent. The peripheral function is that of support, realized when the subject of the construction is expressed by a general factual noun.

The augmented [$_{with}$[Subj$_{COMMON NOUN}$][V$_{PARTICIPLE I}$]] construction is predominantly used in newspapers and magazines (Figure 1 in appendix). The collexeme analysis revealed that this construction attracts nouns of heterogeneous semantics. The noun collexemes include items denoting human beings and spheres of their social activity (i.e. commerce, politics, education, sport).

The distinctive collexeme analysis revels that the [$_{with}$[Subj$_{COMMON NOUN}$][V$_{PARTICIPLE I}$]] construction attracts three general nouns *people*, *woman* and *man* which are classified as 'people nouns' [33, p. 99]. The noun *people* is the most frequent and highly significant distinctive collexeme in the construction (rank 1). It is used in two general meanings in the analyzed construction, referring to humans/men/women, children or people of a particular country or social group [33, p. 102]. It occurs with and without determining and modifying elements. Mostly the noun is modified by indefinite quantifying elements (*many*, *more*, *most*, *some*, *few*, *107,000*) which signal that the noun is referred to people in a general way. When used with adjectives *older*, *different*, *deaf*, *happy*, *young*, etc., that do not make the meaning of the word more specific, it still refers to people in general. The generality of the noun *people* accounts for its use in statements about typical actions or processes when it is not important who precisely is involved. The noun *people* is utilized to focus on a situation or action where concrete individuals constitute the background.

The noun *woman* is in the list of collexemes with rank 35 and coll. strength 4.849225, while the noun *man* is also attracted to the construction though the coll. strength is not significant (1.173338). It should be noted here that the number of singular forms of the nouns *woman* and *man* are less frequent than plural uses (4/16 and 4/12 uses respectively). Singular uses of the nouns *woman* and *man* are accounted in fiction where they indicate a specific person without the name of this person being given and specify the sex of a person. These nouns are typically modified by definite or indefinite articles. In this case, the singular noun *woman* or *man* as the referent of the subject of the [$_{with}$[Subj$_{COMMON NOUN}$][V$_{PARTICIPLE I}$]] construction is different from the referent of the matrix subject.

In contrast, the plural forms *women* and *men* are used without determiners and display a greater extent of generality. They are employed to talk about people in an unspecific way. These nouns denote people who are treated as a group, sharing some features in common or the action in which they are involved [33, p. 117]. The plural forms *women* and *men* as subjects of the [$_{with}$[Subj$_{COMMON NOUN}$][V$_{PARTICIPLE I}$]] construction are predominantly used in such BNC-BYU registers as newspapers, popular magazines, academic and non-academic.

The distinctive collexeme analysis of the nouns in the subject slot of the [$_{with}$[Subj$_{COMMON NOUN}$][V$_{PARTICIPLE I}$]] construction reveals a preference for animate subjects that are construed as AGENTS of a process/state expressed by the present participle. The subject of the construction shows no coreference with the subject of the matrix clause, specifying the whole matrix situation. The construction typically performs support function and provides supplementary context to the event presented in the matrix clause. It elaborates on actions and processes promoting the centrality of a human being in general for the message.

The usage of the analyzed constructions strongly depends on the text register. The quantitative analysis of register distribution of the augmented [*with*[Subj_COMMON NOUN][V_PARTICIPLE I]] and unaugmented [*with-less*[Subj_COMMON NOUN][V_PARTICIPLE I]] constructions (Figure1 in appendix) proves that the unaugmented construction utterly prevails in fiction, while the augmented construction is the most numerous in newspapers and magazines. This register distribution specificity directly influences the choice of the subject referents. Newspapers deal with the news that report what happens in the world, discuss global developments as well as problems of individuals with gender issues and social relations among them. Thus nouns referring to people in general moreover in plural form (*people*, *women*, *men*) are very frequent there, they help to concentrate on actions and situations leaving the identity of the people involved anonymous. In fiction much attention is paid to the descriptions of the personages, their outward and inward features. Apart from denoting parts of the body, somatisms represent signs of "non-verbal behavior (movements, postures, facial expression, glances and eye contact, automatic reactions, spatial and touching behavior)" [36, p. 4]. The body language of the literary characters is 'meaningful' in fictional communication as it lends liveliness and authenticity to the action portrayed [36, p. 5]. Hence, nouns referring to parts of the body (*eyes*, *hands*, *face*, etc.) are exceptionally frequent in fiction, especially in prose.

6   Concluding Remarks

The results of the conducted analysis of the English augmented [*with*[Subj_COMMON NOUN][V_PARTICIPLE I]] and unaugmented [*with-less*[Subj_COMMON NOUN][V_PARTICIPLE I]] constructions allow for the following tentative conclusions.

The English augmented and unaugmented present participle with the explicit subject constructions are a very interesting and complex phenomenon, which has already been extensively studied in various linguistic paradigms, and there is still more to be discussed about them. Our focus has been on the underlying semantic relations between two synonymous participial constructions with the explicit subject connected to the matrix sentence with or without the augmentor *with*. To analyze the semantics of the investigated constructions, only the cases with the subject expressed by a common noun have been taken into account.

The underlying assumption of the conducted research has been the following: semantic properties of syntactic constructions can be investigated on the basis of their significant collexemes, i.e. lexemes that occur in a particular slot of the construction more often than expected. The distributional data obtained show that such an analysis provides unexpected and nontrivial findings on the nature of the semantic contrast between   [*with*[Subj_COMMON   NOUN][V_PARTICIPLE   I]]   and   [*with-less*[Subj_COMMON NOUN][V_PARTICIPLE I]] constructions in present-day English.

The data of the simple collexeme and the distinctive collexeme analyses prove that the English augmented [*with*[Subj_COMMON NOUN][V_PARTICIPLE I]] and unaugmented [*with-less*[Subj_COMMON NOUN][V_PARTICIPLE I]] constructions differ in terms of productivity, semantics of nouns in the subject slot, agentivity of the subject's referent, genre distribution, and pragmatic functions.

From the research that has been carried out, it is possible to conclude that the English augmented [*with*[Subj COMMON NOUN][V PARTICIPLE I]] and unaugmented [*with-less*[Subj COMMON NOUN][V PARTICIPLE I]] constructions are pragmatically distinct and semantically substantiated. The constructions display a remarkable consistency in attracting nouns of certain semantic frames to fill their subject slots. The quantitative corpus linguistic methods employed in this study have proved to be beneficial for determining the difference between a pair of synonymous grammatical constructions that differ only in the absence/presence of a particular augmentor.

This study is of preliminary character since the findings are produced on the limited research material. Further more extensive corpus-quantitative research of the English augmented [*with*[Subj COMMON NOUN][V PARTICIPLE I]] and unaugmented [*with-less*[Subj COMMON NOUN][V PARTICIPLE I]] constructions with pronouns and proper nouns in the [Subj] slot would be needed to obtain more reliability and corroborate the data presented.
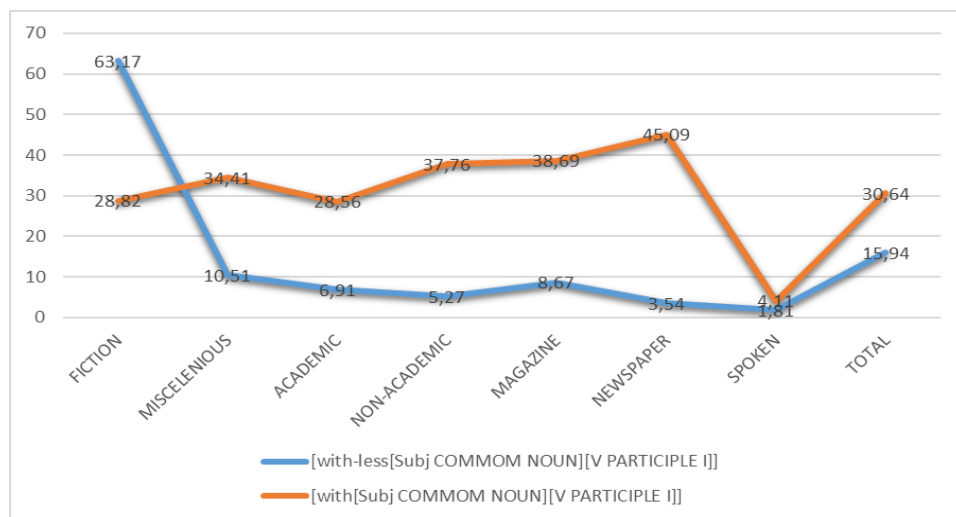
7   APPENDIX

Fig. 1. The register distribution of the [*with*[Subj COMMON NOUN][V PARTICIPLE I]] and [*with-less*[Subj COMMON NOUN][V PARTICIPLE I]] constructions in the BNC-BYU corpus (in frequencies per million)

8   References

1. BNC-BYU. https://www.english-corpora.org/bnc/
2. Beukema, F., Hoekstra, T.: Extractions from with constructions. In: Linguistic Inquiry 15, 689-698 (1984).
3. Stump, G.T.: The semantic variability of absolute constructions Dordrecht: D. Reidel Publishing Co. (1985).
4. Kortmann, B.: Free adjuncts and absolutes in English: Problems of control and interpretation Routledge (1991).

5. Riehemann, S. Z., Bender, E.: Absolute constructions: On the distribution of predicative idioms. In: Bird S., Carnie A., Haugen J., Norquest P (eds.) WCCFL 18 Proceedings, Cascadilla Press, 476-489 (1999).

6. Felser, C., Britain, D.: Deconstructing what with absolutes. In: Atkinson M. (ed.) The Minimalist Muse. Essex Research Reports in Linguistics, vol. 53. 97-134 (2007).

7. van de Pol, N., Hoffmann, Th.: With or without with: The Constructionalization of the with-augmented absolute in English. In: Journal of English Linguistics 44, 318-345 (2016).

8. Fabricius-Hansen, C., Haug, D.T.T.: Co-eventive adjuncts: main issues and clarifications. In: Fabricius-Hansen C., Haug D.T.T. (eds.) Big Events, Small Clauses: The Grammar of Elaboration. Berlin, De Gruyter, 21-54 (2012).

9. Yoo E.-J.: English absolutes, free adjuncts, and *with*: A constructional analysis. In: Korean Society for Language and Information vol. 12(2), 49-75 (2008).

10. Combettes, B.: Les constructions détachées en français, Ophrys (1998).

11. Thompson, S. A.: Grammar and discourse: The English detached participial clause. In: Klein-Andreu, Flora (ed.), Discourse perspectives on syntax. New York: Academic Press, 43-64 (1983).

12. Hasselgård, H.: Possessive absolutes in English and Norwegian. In: Fabricius-Hansen C., Haug D.T.T. (eds.) Big Events, Small Clauses: The Grammar of Elaboration. De Gruyter, 229-258 (2012).

13. Fillmore, C.J., Lee-Goldman, R.R., Rhodes, R.-S.: The FrameNet constructicon. In: Boas H.C., Sag I.A. (eds.) Sign-Based Construction Grammar. CSLI Publications, Stanford, 283-299 (2012).

14. Goldberg, A. E.: Constructions. A construction grammar approach to argument structure. University of Chicago Press (1995).

15. Fillmore, Ch.: The Mechanisms of "Construction Grammar". In: Proceedings of the Fourteenth Annual Meeting of the Berkley Linguistic Society 14, 35-55 (1988).

16. Goldberg, A. E.: Constructions at Work: The Nature of Generalization in Language. Oxford University Press (2006).

17. Croft, W.: Construction Grammar. In: Geeraerts D., Cuyckens H. (eds.) The Oxford Handbook of Cognitive Linguistics, Oxford University Press, 463-508 (2008).

18. Hilpert, M.: Constructional Grammar and its application to English. Edinburgh University Press (2019).

19. Stefanowitsch, A.: Collostructional analysis. In: Hoffman Th., Trousdale G. (eds.) The Oxford Handbook of Construction Grammar. Oxford University Press, 290-307 (2013).

20. Hoffmann, Th., Trousdale, G.: Construction grammar: Introduction. In: Hoffmann T., Trousdale G. (eds.) The Oxford Handbook of Construction Grammar. Oxford University Press, 15-31 (2013).

21. Hilpert, M., Diessel, H.: Entrenchment in Construction Grammar. In: Schmid H.-J. (ed.), Language and the human lifespan series. Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge American Psychological Association; De Gruyter Mouton, 57-74 (2017).

22. Stefanowitsch, A., Gries, St. Th.: Collostructions: Investigating the interaction between words and constructions. In: International Journal of Corpus Linguistics 8(2), 209-243 (2003).

23. Stefanowitsch, A., Gries, St. Th.: Covarying collexemes. In: Corpus Linguistics and Linguistic Theory 1(1), 1-43 (2005).

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

25. Gries, S. Th.: More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013). In: Cognitive Linguistics 26(3), 505-536 (2015).
26. Schönefeld, D.: Things going unnoticed – A usage-based analysis of *go*-constructions. In: Divjak D., Gries S.T. (Eds.) Frequency effects in language representation vol. 2, De Gruyter Mouton, 11-49 (2012).
27. Fillmore, Ch. J.: Frame Semantics. In: The Linguistic Society of Korea (ed.), Linguistics in the Morning Calm, Seoul: Hanshin, 111-138 (1982).
28. Fillmore, Ch. J.: Frames and the semantics of understanding. In: Quaderni di Semantica 6, 222-254 (1985).
29. FrameNet. https://framenet.icsi.berkeley.edu/fndrupal/WhatIsFrameNet.
30. Gries, S.: Th.: Coll. Analysis 3.2a. A program for R for Windows 2.x. http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/.
31. Hilpert, M.: Constructional change in English. Developments in allomorphy, word formation, and syntax. Cambridge: Cambridge University press (2013).
32. Frith, Ch.: Role of facial expressions in social interactions. In: Philosophical Transactions of the Royal Society № 364. 3453–3458 (2009) doi:10.1098/rstb.2009.0142.
33. Mahlberg, M.: English general nouns: a corpus theoretical approach. John Benjamins Publishing Company (2005).
34. Schmid, H.-J.: English Abstract Nouns as Conceptual Shells. From Corpus to Cognition. Mouton de Gruyter (2000).
35. Mahlberg, M.: The textlinguistic dimension of corpus linguistics: The support function of English general nouns and its theoretical implications. In: International Journal of Corpus Linguistics 8:1, 97-108 (2003).
36. Korte, B.: Body Language in Literature. University of Toronto Press (1997).