The Grove. Working Papers on English Studies 27 (2020): 83-102.
DOI: 10.17561/grove.v27.a6

A STUDY OF DIRECT SPEECH COMPLEMENTATION WITH EMBEDDING VERBS: THE COLLOSTRUCTIONAL ANALYSIS

UN ESTUDIO DE LA SUBORDINACIÓN DE ESTILO DIRECTO CON VERBOS COMPLETIVOS: EL ANÁLISIS COLOSTRUCCIONAL

Aroa Orrequia-Barea
Universidad de Jaén

orrequia@ujaen.es

Abstract

Non-relational verbs, as opposed to relational ones, cannot replace their complement clause with a complex nominal, meaning that they do not denote a proposition, as the Relational Analysis states. However, direct speech seems to be a proper replacement for the complement clause in the non-relational verb cases. This paper deals with the analysis of some of the most representative taxonomies of embedding verbs using the *British National Corpus,* to check whether they can occur with direct speech complements; the *collostructional analysis*, which is a technique of statistical significance; and the programming language *R* to do it in a computational and automatic way. Thus, the collostructional method will measure the strength between the embedding verbs and their corresponding complement clauses in the direct speech form.

Keywords*:* collostructional analysis; complement subordination; corpus linguistics; direct speech; embedded complements; non-relational verbs.

Resumen

Los verbos no relacionales, al contrario que los relacionales, no pueden sustituir sus cláusulas subordinadas por un nominal complejo, es decir, estas cláusulas no denotan una proposición como establece el Análisis Relacional. Sin embargo, el estilo directo parece que reemplaza de manera correcta a las cláusulas subordinadas en el caso de los verbos no relacionales. El presente artículo realiza un análisis de algunas de las taxonomías más representativas de los verbos completivos usando el *British National Corpus*, para comprobar si estos verbos ocurren con claúsulas en estilo directo; el análisis colostruccional, que es una técnica de significatividad estadística; y el lenguaje de programación R para

hacerlo de manera automática. De este modo, el análisis colostruccional medirá la fuerza entre los verbos completivos y sus cláusulas subordinadas en la forma de estilo directo.

Palabras clave: análisis colostruccional: subordinación completiva; lingüística de corpus; estilo directo; verbos completivos; verbos no relacionales.

1. Introduction

In English, there are some embedding verbs which behave differently from the typical transitive ones both semantically and syntactically. A *prima facie* problem is that these verbs cannot replace their complement clauses with complex nominals because doing so either makes the sentence ungrammatical or does not paraphrase the full meaning of the complement clause. Therefore, this group of verbs cannot be analysed as the transitive one since they do not express the same type of relation (Moltmann 82–5; Pietroski 217–32; Rosefeldt 302). For this reason, we will refer to them as *non-relational* verbs (Moltmann).

Nevertheless, there seems to be a grammatical construction that works as a proper substitution for the complement clauses of non-relational verbs, that is, direct speech (Orrequia-Barea 251). The main aim of this paper is to statistically demonstrate that there is a group of embedding verbs, non-relational ones, that attracts the construction of direct speech as their complement clause. This feature can be added to a list of characteristics that split embedding verbs into two groups differentiated by their singular behaviour in semantic and syntactic terms. To prove this hypothesis, we have explored whether embedding verbs extracted from two selected taxonomies in English (Hooper and Cattell) occur with direct speech complements in the *British National Corpus* (BNC henceforth). Preliminary results indicate that this set of verbs mostly appear with direct speech complements in the corpus. However, to prove it statistically, a corpus linguistics technique of statistical significance is used: the *collostructional analysis* (Gries and Steefanowitsh) which measures the attraction strength between the direct speech complements and the embedding verbs (Schmid and Küchenoff) to shed some light on the semantic differences between two apparently synonymous constructions, that of embedding verbs taking *that*-clauses as complements.

The article is organised as follows. Section 2 offers a review of the literature, focusing on the taxonomies of the studied verbs and the collostructional analysis. Section 3 describes the methodology used to analyse the embedding verbs of Section 2 and their clausal complements. In Section 4, we analyse the results obtained from the application of the collostructional analysis. Section 5 is a

discussion of the results from the previous section. Finally, Section 6 offers a summary and some conclusions.

2. Literature Review

Traditionally, transitive verbs are analysed using the Relational Analysis, which has two main tenets (Moltmann 79):

1) The complement clause denotes a proposition, and for this reason, it can be replaced by complex nominals such as *the proposition that S* or *the fact that S* without a change in meaning. Hence, examples (1) and (2) are synonymous.

    1. John regretted/forgot the fact that he had destroyed the files
    2. John regretted/forgot that he had destroyed the files.

2) The verb expresses a relation between the subject and the complement clause. In example (3) what made Mary very upset is the thing that John accepted.

    3. John accepted something that made Mary very upset.

Unlike most embedding verbs in English, there are some verbs that do not fulfil these two requirements of the Relational Analysis. On the one hand, their complement clause does not denote a proposition and, as a consequence, they cannot be replaced by complex nominals, as can be seen in the ungrammatical (4). This is known as the *Substitution Problem* (Moltmann 82).

    4. #John thinks/hopes/argues/comments/says the proposition that the earth is round.

On the other hand, these verbs do not express a relation between the subject and the complement clause of the verb. This phenomenon is called *Objectivization Effect* (Moltmann 86). The difference between examples (5) and (3), repeated here as (6), is that in the former what made Mary very upset was the fact that John said something, whereas in the latter, it is the thing itself what upset Mary, not the fact of accepting it.

    5. John said something that made Mary very upset.
    6. John accepted something that made Mary very upset.

Taking into account the inability of these verbs to be analysed by the Relational Analysis, they are called *non-relational* verbs, as opposed to those that do establish a relation between the subject and the complement clause, the *relational* ones. This is further reflected in the fact that there are several grammatical phenomena that take place in non-relational verb environments, whereas they are

not allowed in relational ones. Some of these constructions are: root transformations, parenthetical constructions, raising passive, 'so' anaphor, complementizer omission or adjunct extraction, as can be seen in examples (7)-(11) respectively (Orrequia-Barea).

7. John says he doesn't want to meet Mary, but Mary says that meet her he will.
8. Mary is not at home today, she believes.
9. He said/thinks so.
10. They said (that) they were too busy.
11. How$_i$ did Mary say that he was doing t_i?

It seems that the correlation between these phenomena, the direct speech complementation and the semantics of these verbs reveal a pattern of coincidence that is too systematic to be neglected.

2.1. Verb taxonomies

To gather as many instances of embedding verbs in the English language as possible, two taxonomies were chosen to analyse their members: Hooper's (92) classification and Cattell's taxonomy (77). These two compilations include a large number of embedding verbs, that is, verbs taking *that*-clause complements.

Firstly, Hooper's (92) classification will be analysed. The criterion used to divide these verbs is the ability of some of them to change the prominence of the information from the main clause, where it is by default, to complement clauses. Therefore, some subordinated clauses can be asserted as main clauses usually are. Thus, when embedded clauses work as the focus of the sentence, the main clause is reduced in content and meaning to a second place in the sentence. As a consequence, not only do main clauses allow for grammatical constructions such as root phenomena (adjunct extraction, topicalization, anteposition), raising passive, complementizer omission or parentheticals, but also complement clauses do so. This classification also includes a previous division made by Kiparsky and Kiparsky (143) according to the concept of *factivity*. Table 1 shows the verbs included.[1]

[1] Verbs with really low frequency in the corpus have been excluded from the analysis.

	Assertive verbs	Non-assertive verbs
Non-factive verbs	*Acknowledge, admit, affirm, allege, answer, argue, assert, assure, certify, charge, claim, contend, declare, divulge, emphasise, explain, grant, guarantee, hint, hypothesize, imply, indicate, insist, intimate, maintain, mention, point out, predict, prophesy, postulate, remark, reply, report, say, state, suggest, swear, testify, theorize, verify, vow, write, agree, be afraid, be certain, be sure, be clear, be obvious, be evident, calculate, decide, deduce, estimate, hope, presume, surmise, suspect, appear, believe, expect, figure, guess, imagine, seem, suppose, think*	*Be likely, be possible, be probable, be conceivable, Be unlikely, be impossible, be improbable, be inconceivable, doubt, deny*
Factive verbs	*Discover, find out, know, learn, note, notice, observe, perceive, realise, recall, remember, reveal, see*	*Amuse, be exciting, be interesting, be odd, be relevant, be sorry, be strange, bother, care, forget, make sense, regret, resent, suffice*

Table 1. Hooper's (92) classification of embedding verbs

As far as non-factive verbs are concerned, there are two types: assertive and non-assertive verbs. The first group includes all the verbs that always allow the assertion of their complement, that is, verbs whose complement clauses can work as the focus of the sentence, as examples (12) and (13) illustrate (from Hooper 92).

12. Many of the applicants are women, it seems
13. He wants to hire a woman, he says.

In examples (12) and (13), the main clause is now relegated to a second place at the end of the sentence since the prominence of the information is now placed on the embedded clause. Parenthetical constructions are the structural reflection of the fact that the complement clause is semantically more relevant than the main clause.

Conversely, non-assertive verbs do not allow the assertion of their complements in any case, therefore the main clause always bears the prominence of the information as examples (14) and (15) demonstrates (from Hooper 113).[2]

14. *Many of the applicants are women, it's likely.
15. *He wants to hire a woman, it's possible.

Regarding factive verbs, we can also find two different groups which behave differently: *semifactive* verbs, which are assertive, and *genuine* or *pure* factive, which are classified as non-assertive (Hooper 114). On the one hand, the latter always presupposes the truth of the complement clause, therefore, they cannot allow assertion. In example (16) the speaker must presuppose that he has not told the truth (from Hooper 115).

16. It is possible that I will {regret/forget} later that I have not told the truth.

On the other hand, semifactive verbs are considered to have a special status in semantic terms. They are included in the factive verb group because they also imply the presupposition of the truth of their complement. However, they only do so in at least one reading, that is, the presupposition is cancellable in certain contexts, for example, under the negation. As a consequence, this characteristic of presupposition is not constant as it is in the case of the genuine factive verbs. Compare the following examples (from Kartunnen 343):

17. John regretted that he had no money.
18. John didn't regret that he had no money.
19. John remembered to lock his door.
20. John didn't remember to lock his door.

In examples (17) and (18), despite the negation, the presupposition in both sentences is 'John had no money'. Regarding sentence (20), the presupposition is that 'John didn't lock the door', as the verb *remember* allows the negation to reach the infinitive clause. However, there is no way in which this presupposition remains the same for example (19) since John actually locked the door. Unlike (17) and (18), in sentences in which there is a semifactive verb like (19) and (20), the presupposition can be cancellable because of the presence of the negation. It is worth mentioning that due to this cancellability, this group also shares one of the main characteristics of the assertive verbs, that is, they allow for parenthetical constructions.

21. Meg is pregnant, I found out.
22. You've painted your house, I see.

[2] Further classifications inside each group have been ignored for expository purposes.

In the present paper only assertive verbs and semifactives have been taken into account, since they are the only ones which fall under our non-relational verb terminology, that is, they do not fulfil the criteria of the Relational Analysis. The pure factive and non-assertive verbs are considered relational verbs, as they behave like run-of-the-mill transitive verbs.

The second taxonomy to be analysed is Cattell's (77). According to Stalnaker (701), assertion and presupposition are understood as opposed terms. Presupposition is defined as the set of cases that interlocutors share in a conversation, that is, the *common ground* (CG henceforth), whereas assertion is what contributes to transforming that set. There is a set of theories that claim that certain lexico-semantic properties of some embedding verbs are able to make their subordinated clauses change the context as main clauses usually do. Stalnaker (704) defines the clause in terms of their potential of changing the context. For example, to define subordinated clauses, it is considered to what extent the propositional information introduced by the verb belongs to the set of cases shared between the interlocutors, that is, the CG. In a conversation, the CG shared by the interlocutors is what determines whether presuppositions of a sentence are satisfied (Peters 122).

Cattell (66) represents the first application of these ideas to the complement subordination phenomenon. As this scholar states, the classification criterion is whether the information conveyed by the complement clause belongs to the CG or not. There are two main groups: *stance* verbs, which contribute to change the CG, either including new propositions (*volunteered stance* verbs) or questioning and reinforcing the presence of propositions already included (*response stance* verbs); and *nonstance* verbs, which do not add anything to the CG. According to Cattell's classification (77), the subject of an embedding verb can:

> a) not mean to alter the CG or set of propositions that are taken for granted in the pertinent context: *nonstance* verbs
>
> 23. John commented that Madrid is the capital of Spain.
>
> b) try to contradict or reinforce the truth of the propositions taken for granted in the CG: *response stance* verbs.
>
> 24. John accepted/denied that Madrid is the capital of Spain.
>
> c) try to introduce a new true proposition in the CG: *volunteered stance* verbs.
>
> 25. John claimed that Madrid is the capital of Spain.

Table 2 shows the verbs included in these three groups:

Stance verbs		Nonstance verbs
Volunteered stance verbs	**Response stance verbs**	*(be) aware, (be) certain, comment, convey, convince, detail, doubt, emphasise, forget, mention, notice, point out, realise, recall, recognise, regret, remember, remind*
Allege, assert, assume, believe, claim, conclude, conjecture, consider, decide, declare, deem, envisage, estimate, except, fancy, feel, figure, imagine, intimate, judge, maintain, propose, reckon, report, say, state, suggest, suppose, suspect, tell, think	*Accept, admit, agree, confirm, deny, verify*	

Table 2. Cattell's classification of embedding verbs (77)

In his theory, Cattell focuses on the root phenomenon of adjunct extraction and the different interpretation some verbs have regarding this phenomenon. When making a question using an adjunct like *why* in a complex sentence like (26), this adjunct should apply to any of the clauses of that sentence, that is, it should have at least the two interpretations in (a) and (b) (adapted from Cattell 61).

26. Why do the they think (that) Sue killed Harry?
a) Why_i do they t_i think [that Sue killed Harry]?
b) Why_i do they think [that Sue killed Harry t_i]?

Interpretation (a) asks why they consider that Sue killed Harry, whereas (b) looks for Sue's reasons to kill him. Regarding the classification presented above, *volunteered stance* verbs are the ones allowing this double interpretation. However, not all the verbs are ambiguous between these two interpretations. Consider (27) and (28).

27. Why do they deny that she killed him?
28. Why did Richard comment that Sue killed Harry?

As there are two clauses, we expect to have two different interpretations, one raised out of each clause. Nevertheless, in (27) and (28) the adjunct is only extracted by the main clause and the subordinated one blocks such interpretation. According to the author, complement clauses of *response stance* and *nonstance* verbs behave like "islands" and they can only be interpreted as in the reading of (a).

In this paper, only *volunteered stance* verbs have been selected for our study. Nonetheless, since there are some verbs from the *response stance* group, such as *admit* or *agree*, which are included in Hooper's (92), they have also been taken into account.

2.2. Collostructional Analysis

The collostructional analysis (Stefanowitsch and Gries 214) is a statistical technique frequently used in Corpus Linguistics. This approach is based on Construction Grammar (Lackoff, Goldberg) which sets the construction as the basic unit of linguistic organisation, hence the study of the lexicon and the grammar is not seen as something completely different. Stefanowitsch and Gries define the term *construction* as "any linguistic expression … that is directly associated with a particular meaning or function, and whose form or meaning cannot be compositionally derived" (212). The main aim of the collostructional analysis is to provide an objective method to determine "the degree to which particular slots in grammatical structure prefer, or are restricted to, a particular set or semantic class of lexical items" (Stefanowitsch and Gries 211). This method can be applied to any linguistic unit, such as morphemes, compounds, multi-word expressions, phrasal verbs and even more abstract ones like tense, aspect or mood.

According to these scholars, the collostructional analysis "investigates which lexemes are strongly attracted or repelled by a particular slot in the construction (i.e., occur more frequently or less frequently than expected)" (214). It measures the collostruction strength, that is, the "attraction" or "repulsion" of a *collexeme* and a *collostruct*. The *collexeme* is the word that is attracted to a particular construction, in this particular case, the collexemes are the embedding verbs selected from each taxonomy. Conversely, the *collostruct* is the construction which is associated with a particular lexeme, namely the direct speech construction, in this case. The combination of these two elements is known as *collostruction* (Stefanowitsch and Gries 215).

Collostructional analysis uses Fisher's exact test as the statistical measure because it "neither makes any distributional assumptions, nor does it require any particular sample size" (Stefanowitsch and Gries 218). Then, to calculate the collostruction strength four frequencies are needed: the frequency of the collexeme in the construction, the frequency of the collexeme in all other constructions, the frequency of the construction with lexemes other than the collexeme and the frequency of all other constructions with lexemes other than the collexeme. Using this information, the Fisher's exact test provides the probability of this distribution. The main disadvantage of this statistical test is the numerous

calculations that need to be made but, fortunately, it can be calculated by using computational programs.

3. Methodology

Direct speech seems to be a proper replacement for the complement clause of non-relational verbs. In fact, the search for most of the selected verbs from the taxonomies with the direct speech returned at least one hit (Orrequia-Barea). However, the mere presence of one occurrence of this construction in the verb environment is not relevant enough since we want to measure the degree to which embedding verbs prefer the direct speech complements. To do this, two main steps have been followed: firstly, to compile a subcorpus of the selected embedding verbs with the direct speech complementation from the BNC; and secondly, to apply the collostructional method using the programming language R.

3.1. Data collection

The first step was to obtain a subcorpus of examples from the BNC. For this reason, we searched for the embedding verbs included in the taxonomies in 2.1 followed by the direct speech complementation to explore if there is empirical evidence of this pattern.

The *British National Corpus* was chosen since it is classified as a reference corpus because of its representativeness of the most important grammatical constructions, vocabulary and varieties of the English language. Additionally, it is considered a balanced corpus and its validity has been proved due to the number of studies that have been carried out based on its examples. The BNC is a monolingual corpus which contains samples of the late-20th century British English and it consists of 98,000,000 words (10% of spoken English and 90% of written language).

The BNC can be accessed from different websites; however, we chose the one from Lancaster University[3] because it seemed to be the most adequate for our purposes. On the one hand, this interface allows to download the samples, even in KWIC (*Key Word in Context*) format; on the other hand, it allows the use of quotations marks, something essential in this study to look for concordances including direct speech. Although we downloaded every instance of each verb with apparent direct speech constructions, a vast number of examples also turned up, which had to be manually discarded since quotation marks were used for other purposes.

[3] http://www.natcorp.ox.ac.uk/. Free registration is necessary.

The queries were restricted to the past tense of the verbs for two main reasons: firstly, because most of the texts compiled in the corpus are written, where narratives tend to be used, which means a frequent use of verbs in the past tense for obvious reasons; secondly, there are a lot of verbs which coincide in form in the past tense and the past participle form, thus returning more hits than any other tense. When looking for the samples, the search engine of the corpus requires a special syntax for specific constructions as is the direct speech due to the use of quotation marks. The syntax used to query this pattern is shown in (29):

29. <verb> \, "[4]

In the interface, the quotation marks are straight, then there is no distinction between the opening and closing ones. However, this webpage allows the user to choose between open or close quotation marks in the option *frequency breakdown.* In this tab, the type of quotation mark can be selected and the interface shows the number of occurrences and percentage corresponding to each option. For example, the form [*said* \, *"*] returned 10,781 hits with the closing quotation marks, which means 98.65% out of the total. However, there were only 148 occurrences with opening quotation marks, meaning 1.35% of the total. Obviously, the first option was chosen as we are looking for samples in which the embedding verb takes the direct speech as its complement, rather than embedding verbs which are placed at the end of a quote. It is not surprising that there were so many hits of the verb *say* as it is typically used in introducing direct speech in narratives. The same procedure was followed for each of the verbs of the taxonomies.[5] Once we searched for each verb, we downloaded the information to create our subcorpus of embedding verbs with direct speech complements.

3.2. Data analysis procedure

As previously mentioned, the main disadvantage of Fisher's exact test was the summations of all the probabilities, which means a tiresome and time-consuming task. This is the reason why we performed the analysis using the programming language R, implemented in RStudio, a desktop application and a free environment to easily manage R. There is a script to compute the collostructional approach and make it in a computational and automatic way (Flach).[6] The package contains the functions to do Simple, Distinctive and Co-Varying Collexeme Analysis (Stefanowitsch and Gries 100). We used the Simple Collexeme analysis,

[4] Angle brackets are used to mark the elements that must be introduced by the user, but they should not be included in the real query.
[5] The verb *tell* requires a special syntax to include the presence of the dative in the query: [told *\, "]. The asterisk means that any word can appear in that slot. Using the same pattern as the rest of the verb, the corpus only returned passive voice examples.
[6] For more information: https://sfla.ch/collostructions/

called *collex*, in which the information required is a data frame with three columns containing the word, the construction frequency and the corpus frequency, respectively. Apart from that, the size of the corpus and the association measure, in this case Fisher's exact test, are also required. Additionally, it is needed to be specified if the output will be ordered in descending or ascending order. There are other options that can be customized, such as the number of decimals. Although the package includes datasets to explore the different methods, the user needs to upload their own data extracted from a corpus to see the results. When the analysis is finished, the output is displayed in a table that can be downloaded. In the subsequent sections, we will show and comment the most relevant results.

4. Results

The output displayed on the screen after the computational analysis shows interesting and valuable information about the embedding verbs and the collostruction strength they have regarding the direct speech construction. For example, there are two columns which are really interesting for our study. First, the ASSOC column refers to the collostruction strength and it gives two parameters: "attr", which stands for attraction, and "rep" which is displayed when the collexeme and the collostruct are repelled. This measure is based on the difference between the observed frequency in the corpus of each verb with direct speech complementation and the expected frequency of the collexeme in the construction. Secondly, the significance of the association is also shown in the output using asterisks for the different levels of representativeness or "ns" to indicate that it is "non-significant". The more asterisks used, up to five, the more significance. The output can be seen in figure 1.

In our analysis, we have explored 74 verbs selected from the previously mentioned taxonomies of embedding verbs. Only 38 verbs were found with the direct speech complementation in the BNC, from which 28 are analysed as significant according to the Fisher's exact test, which means 74% of the verbs. Apart from that, there are 8 verbs which are considered to be attracted to the construction but their association measure is not considered statistically significant because the occurrences in the corpus are very few. The remaining two are neither attracted to the construction nor significant. The verbs which are attracted can be seen in figure 1.

	COLLEX	CORP.FREQ	OBS	EXP	ASSOC	COLL.STR.FYE	SIGNIF
1	reply	5063	217	0.4	attr	Inf	*****
2	say	195278	7193	16.3	attr	Inf	*****
3	tell	35375	182	3.0	attr	250.56538	*****
4	write	9703	125	0.8	attr	221.82615	*****
5	think	53567	164	4.5	attr	189.55867	*****
6	answer	3614	88	0.3	attr	180.86852	*****
7	remark	1641	69	0.1	attr	158.62021	*****
8	comment	1917	45	0.2	attr	92.23929	*****
9	explain	6762	47	0.6	attr	71.45498	*****
10	declare	4399	38	0.4	attr	61.51640	*****
11	conclude	3085	23	0.3	attr	36.12174	*****
12	state	5127	25	0.4	attr	34.62670	*****
13	observe	4997	19	0.4	attr	24.49723	*****
14	admit	5569	18	0.5	attr	22.00666	*****
15	recall	1795	12	0.1	attr	18.65343	*****
16	suggest	10793	19	0.9	attr	18.33381	*****
17	insist	3230	9	0.3	attr	10.79565	*****
18	argue	6370	8	0.5	attr	7.00851	*****
19	note	6070	6	0.5	attr	4.81887	****
20	reveal	5243	4	0.4	attr	2.96769	**
21	divulge	35	1	0.0	attr	2.53512	**
22	point out	3378	3	0.3	attr	2.51949	**
23	conjecture	40	1	0.0	attr	2.47722	**
24	claim	8223	4	0.7	attr	2.27038	**
25	allege	3051	2	0.3	attr	1.56235	*
26	agree	14350	4	1.2	attr	1.47450	*
27	figure	418	1	0.0	attr	1.46493	*
28	maintain	4052	2	0.3	attr	1.33943	*
29	realise	4711	2	0.4	attr	1.22396	ns
30	remember	5011	2	0.4	attr	1.17732	ns
31	report	11918	3	1.0	attr	1.10081	ns
32	predict	1812	1	0.2	attr	0.85282	ns
33	acknowledge	2066	1	0.2	attr	0.80032	ns
34	confirm	4861	1	0.4	attr	0.47694	ns
35	feel	26035	3	2.2	attr	0.43186	ns
36	believe	7964	1	0.7	attr	0.31377	ns

Figure 1. Distribution of the verbs that attract the direct speech

As expected, the verbs with the highest significance are those related to speech acts. Speech act verbs are defined as those referring to "any type of verbal behaviour or to the much smaller subset of verbs expressing specific speaker's attitude" (Proost 912). The terms 'illocutionary verbs', 'verbs of communication' or 'verbs of saying' are also used to name this type of verbs. Some of the verbs with the highest frequency belonging to this group are: *reply*, *say* and *tell*

(250.56).[7] Surprisingly, in the first two cases, the significance is so high that the output does not even provide an exact number, but the abbreviation "inf" which stands for "infinite". More verbs related to speech acts are also found in different positions of the results: *answer* (180.86) in the sixth position followed by *remark* (158.62), *comment* (92.23), *explain* (71.45), *declare* (61.51) or *conclude* (36.12), among others. However, the verb found in the fourth position is *write* (221.82). It may not be fully considered a speech act verb but taking into account the previous definition, it also makes reference to a verbal behaviour, in this case in the written form.

Unexpectedly, although the analysed construction is one related to the discourse, as is direct speech, not all the verbs that have a great significance are speech act verbs. For example, according to figure 1, in the fifth position we find the verb *think* (189.55), considered a propositional attitude verb or traditional *verba sentiendi*, which makes reference to a mental state and not to a speech act. Similarly, *observe* (24.49), which is more related to the senses, occupies the thirteenth position in figure 1.

Additionally, some speech act verbs which returned hits with direct speech, such as *report* (1.1) or *confirm* (0.47), are not considered significant by the collostructional method since the difference between the observed frequency and the expected one is not remarkable enough. Both verbs are classified as having an attraction with the direct speech ("attr") but due to their low frequency in the corpus with this construction, it is not statistically representative.

Nevertheless, there are other verbs, such as *decide* (0.18) or *consider* (0.15), which can also be considered speech act verbs but the occurrences in the corpus are so few, particularly only one, that are classified as non-significant and verbs that repelled the direct speech construction.

To return to the taxonomies abovementioned, semifactive verbs, as previously stated, share some characteristics with assertive verbs, being the parenthetical reading the most important one. However, in this paper, we have found out that there is another feature that these verbs also share with the assertive group, that is, taking direct speech as complement clause. Actually, we did not expect to find occurrences of these verbs with the direct speech construction, as most of them do not refer to speech acts. Contrariwise, most of them returned hits with the direct speech as a complement clause. For instance, *observe* (24.49), *recall* (18.65), *note* (4.8) or *reveal* (2.96) are among the first twenty verbs, as can

[7] Results of the association measure used, in this case Fisher's exact test, which shows the probability of distribution by means of calculating the single and joint frequencies of the construction and the collexeme in the corpus.

be seen in figure 1. Additionally, *remember* (1.17) occupies the thirtieth position but is considered to attract the construction, although non-significantly.

5. Discussion

The collostructional analysis has shown that the embedding verbs are generally attracted to the direct speech complementation. It is not surprising that the top positions of the results are mainly taken by so-called speech act verbs. In semantic terms, direct speech denotes utterances and they cannot be paraphrased by complex nominals. This is one of the main reasons why non-relational verbs cannot replace their complement clause by referential expressions without yielding ungrammaticality or a change in meaning (García Núñez and Orrequia-Barea 171). However, if direct speech complementation denotes utterances, how then is it possible for a verb like *think* to take this type of complement clause? As previously stated, *think* is considered a mental state verb. However, the direct speech complement clause "can report on the possibly unexpressed utterance-based content of a mental state the speaker manages to describe the relevant eventuality by giving the content of the utterance that, to her knowledge, best sums it up" (García Núñez and Orrequia-Barea 173). This contrast can be seen in examples (30) and (31), whereas the former has an embedding verb denoting an utterance, the latter does not.

> 30. In a cool voice she answered, "I thought you might be late" (JXS 64).[8]
>
> 31. Catriona took a look at the brown liquid and thought, "I'm nineteen years old, I've failed all my O-grades, and I'm a disgrace to my family" (BN1 390).

However, one of the most revealing findings in this paper is the fact that semifactive verbs allow direct speech complementation. As reviewed above, semifactive verbs cannot be included in the factive verb groups since the presupposition of the complement can be cancellable in some contexts. For this reason, in these contexts, the complement can be asserted in the same way as it happens with assertive verbs. The verbs belonging in semifactive verb group apparently do not make any reference to a verbal behaviour, hence these verbs are not classified as speech act verbs. However, taking into account the results, we decided to look them up in a dictionary to see whether there was an explanation in terms of lexicography. We used the Oxford English Dictionary (OED) to look

[8] Data cited in this work have been extracted from the British National Corpus, distributed by the University of Oxford on behalf of the BNC Consortium. All rights in the texts cited are reserved.

them up ("note", "observe", "recall", "remember", "reveal"). The point becomes clear as soon as we look carefully at the entry of each verb, as can be seen in (32).

32. *Note*
(i) To notice, observe.
(ii) To mention
Observe
(i) To take notice of, be conscious of; to notice, perceive.
(ii) To say by way of comment; to remark or mention in speech or writing.
Recall
(i) To bring back to the mind.
(ii) To recount (a circumstance, event, etc.) from one's memory.
Remember
(i) To think of, recall the memory of (a person) with some kind of feeling or intention.
(ii) To mention.
Reveal
(i) To make manifest or apparent; to demonstrate, exhibit.
(ii) To disclose or make known (something previously unknown or kept secret) in speech or writing; to divulge, declare publicly or openly.

As stated by the dictionary entries included in (32), all semifactive verbs are ambiguous between a relational reading (i) and a non-relational one (ii). The definitions of the verbs in (i) make reference to the most widely used meaning of these verbs, that is, the relational one. Conversely, the meanings in (ii) are non-relational and in semantic terms, these readings can even be included in the set of speech act verbs as all of them make reference to a verbal behaviour. Actually, the meanings in (ii) are synonyms with verbs such as *say* or *mention,* which is the most used verb in defining them. Taking into account the definitions in (ii), it is not surprising then that direct speech complementation is allowed in these verbs environments since all of them are related to the discourse, as the following examples illustrate.

33. When he reflected with sadness on the unhappiness of his marriage towards the end of 1929, he noted, "Friendship with Lewis compensates for much" (A7C 494).
34. Hellman cautiously observed, "Does this mean you don't like them?" (AP0 619).
35. Crawford later recalled, "I didn't enjoy my last three years at school" (HRF 74).
36. So I felt really good about it until I remembered, "Damn, I was only plugged into a Zoom" (C9K 274).

37. Rosemary agreed, and revealed, "If the rent on this place wasn't paid until the end of the year, I'd be in trouble myself" (JY1 219).

Generally speaking, it seems that embedding verbs which actually take direct speech complement clauses are those which somehow make reference to discourse. This reference does not necessarily mean that the speaker uses the verbatim utterance but a potential utterance the speaker thinks is the best summary of the subject's attitudes and mental state.

6. Conclusions

Collostructional analysis seems to be a useful method to measure the association between the collexemes and the collostruction, that is, between the embedding verbs and the direct speech construction. Although the mere existence of a hit in the corpus can be taken as a kind of evidence of this phenomenon, the collostructional approach indeed provides more insight in the construction since it establishes whether that presence in the corpus is actually significant or not.

According to the results, embedding verbs are attracted to direct speech complementation since 74% of the verbs allow for this type of complement clause. Collostructional analysis determines that in 28 verbs out of 38 the results are significant, whereas there are other 8 that, although attracted, are not considered statistically significant. That means a total of 36 verbs out of 38 which are attracted to the direct speech complementation. Therefore, it can be stated that the direct speech is a proper replacement for complement clauses in non-relational verbs.

As the discussion revealed, direct speech complements denote utterances and this is the reason why these complements cannot be properly paraphrased by complex nominals of the type *the proposition that S*. The main difficulty arises with mental state verbs, such as *think*, in which there is no a particular utterance to report on. However, it is then the speaker the one who reports on a potential utterance that best describes the subject's mental state. These results open the possibility for the collostructional approach to be a valid method to explore the semantic-syntactic interface.

The main problem this study has faced is the lack of evidence in the BNC of some verbs. Therefore, there are some verbs which have been identified as "repelled" by the collostructional analysis but which in reality can take direct speech complementation, as is the case of verbs like *affirm, guess, imply* or *indicate*, among others. This can be seen as an inconsistency of the theory. However, we can confirm that direct speech complementation is one more defining characteristic of the group of non-relational verbs. Although in this paper we have only focused on verbs which actually take direct speech complements, in

previous research (Orrequia-Barea), we queried all the members of the taxonomies to prove that direct speech is a construction that cuts off the embedding verbs. As discussed, there is a lack of evidence in the corpus of some non-relational verbs. However, no hits were retrieved in the case of so-called relational verbs with direct speech complementation. We believe that further research into this topic can search for the same non-relational verbs in other corpora, such as the *International Corpus of English*, which is not only restricted to British English but collects samples of English worldwide.

WORKS CITED

Cattell, Ray. "On the Source of Interrogative Adverbs". *Language* 54 (1978): 61–77. DOI: https://doi.org/10.2307/412999

Flach, Susanne. *Collostructions: An R implementation for the Family of Collostructional Methods*, 2017. Package version v.0.1.0. Web. 15 Nov. 2020. <https://sfla.ch/collostructions/>

García Núñez, José María and Aroa Orrequia-Barea. "Non-relational Embedding Verbs: Quotes and Reports". *Complutense Journal of English Studies* 28 (2020): 163–75. DOI: https://doi.org/10.5209/cjes.65989

Goldberg, Adele E. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press, 1995.

Hooper, Joan. "On Assertive Predicates". *Syntax and Semantics* 4 (1975): 91–124. DOI: https://doi.org/10.1163/9789004368828_005

Karttunen, Lauri. "Some Observations on Factivity". *Papers in Linguistics* 4 (1971): 55–69. DOI: https://doi.org/10.1080/08351817109370248

Kiparsky, Paul and Carol Kiparsky. "Fact". *Progress in Linguistics*. Ed. Manfred Bierwisch and Karl Heidolph. The Hague: Mouton, 1970. 143–73.

Lakoff, George. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press, 1987. DOI: https://doi.org/10.7208/chicago/9780226471013.001.0001

Moltmann, Friederike. "Propositional Attitudes without Propositions". *Synthese* 135.1 (2003): 77–118. DOI: https://doi.org/10.1023/A:1022945009188

"Note." *OED Online,* Oxford University Press, Sept. 2020. Web. 15 Sept. 2020. <www.oed.com>.

Orrequia-Barea, Aroa. *La Semántica de los Verbos Proposicionales del Inglés Actual*. Diss. University of Cádiz, 2017.

---. "La Relación entre la Discursividad y la No-Relacionalidad: Un Análisis Basado en Corpus". *Pragmalingüística* 26 (2018): 246–64. DOI: https://doi.org/10.25267/Pragmalinguistica.2018.i26.12

"Observe." *OED Online,* Oxford University Press, Sept. 2020. Web. 15 Sept. 2020. <www.oed.com>.

Peters, Stanley. "Presuppositions and Conversation". *Texas Linguistic Forum* 2 (1975): 122–33.

Pietroski, Paul. *Events and Semantic Architecture.* Oxford: Oxford University Press, 2005. DOI: https://doi.org/10.1093/acprof:oso/9780199244300.001.0001

Proost, Kristel. "Speech Act Verbs". *Encyclopaedia of Language and Linguistics, Volume 2*. Ed. Keith Brown. Oxford: Elsevier, 2006. 912–17. DOI: https://doi.org/10.1016/B0-08-044854-2/01075-0

"Recall" *OED Online,* Oxford University Press, Sept. 2020. Web. 15 Sept. 2020. <www.oed.com>.

"Remember" *OED Online,* Oxford University Press, Sept. 2020. Web. 15 Sept. 2020. <www.oed.com>.

"Reveal" *OED Online,* Oxford University Press, Sept. 2020. Web. 15 Sept. 2020. <www.oed.com>.

Rosefeldt, Tobias. "That-Clauses and Non-Nominal Quantification". *Philosophical Studies* 137 (2008): 301–33. DOI: https://doi.org/10.1007/s11098-006-0002-8

Schmid, Hans-Jörg and Helmut Küchenhoff. "Collostructional Analysis and other Ways of Measuring Lexicogrammatical Attraction. Theoretical Premises, Practical Problems and Cognitive Underpinnings". *Cognitive Linguistics* 24.3 (2013): 532–77. DOI: https://doi.org/10.1515/cog-2013-0018

Stalnaker, Robert. "Common Ground". *Linguistics and Philosophy* 25 (2002): 701–21. DOI: https://doi.org/10.1023/A:1020867916902

Stefanowitsch, Anatol and Stefan T. Gries. "Collostructions: Investigating the Interaction of Words and Constructions". *International Journal of Corpus Linguistics* 8.2 (2003): 209–43. DOI: https://doi.org/10.1075/ijcl.8.2.03ste

The British National Corpus, version 3 (BNC XML Edition), 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Web. 13 Nov. 2020. <http//www.natcopr.ox.ac.uk/>.

Received October 21, 2020

Revised version accepted November 15, 2020