Оригинални научни рад
811.111`42:378
10.46793/LIPAR90.123S

Katarina B. Subanović[1]
University of Kragujevac
Faculty of Philology and Arts
English Department
ORCID 0009-0005-5144-3586

WHERE THERE'S A *WILL*, THERE'S A *SHALL*: A DISTINCTIVE COLLEXEME ANALYSIS OF *WILL V* AND *SHALL V* IN ACADEMIC DISCOURSE

The present paper deals with the constructions *WILL V* and *SHALL V* from the perspective of distinctive collexeme analysis (DCA), a quantitative corpus-linguistics method developed by Gries and Stefanowitsch (2004a). The data were extracted from the academic subgenre of the *Corpus of Contemporary American English* (COCA). The primary aim of the paper is to identify the semantic categories of verbs that are statistically attracted to each construction and to examine how they relate to the modal meanings of *will* and *shall*. The results show that the two constructions mainly attract the verbs from distinct semantic categories. The *WILL V* construction is predominantly associated with the verbs expressing cause-effect relationships, enablement and necessity, change and development, and research-related actions. Across these categories, *will* conveys predictive meaning, presenting developments, outcomes, and research consequences as expected. On the other hand, *SHALL V* mainly attracts verbs denoting academic commitment, institutional and procedural acts, text structuring and discourse navigation, and acts of definition and delimitation. Within the construction *shall* typically expresses intention, and, in combination with institutional and procedural verbs, it conveys meanings related to norms, obligations, and formal procedures.

Keywords: *WILL V*, *SHALL V*, collostructional analysis, distinctive collexeme analysis, academic discourse

1. INTRODUCTION

Modal verbs form a closed class of auxiliaries in English with clearly defined morphosyntactic properties and an important role in expressing stance, obligation, intention, and prediction (see Erić-Bukarica 2019: 75). *Will* and *shall* can be classified within the group of the primary modal auxiliaries (see Leech 2004: §5)[2] and they are characterized by partially overlapping meanings, different frequencies, and different distributions across registers, both historically and in Present-Day English.

1 katarina.subanovic@filum.kg.ac.rs

2 The consulted publication does not provide page numbers; references are therefore given using a chapter number.

In contemporary usage, *will* is one of the most frequent modal verbs (see Rizvić-Eminović, Šukalić 2019: 357; Sujatna et al. 2019: 169). However, despite its overall high frequency in general English, *will* occurs less often in academic writing than in spoken or journalistic genres, which has been attributed to its weaker hedging potential compared to other epistemic modal verbs (see Rizvić-Eminović, Šukalić 2019: 361–362). When it appears in academic discourse, though, it can appear in research articles, theoretical papers, and review articles (see Karahan 2022). In academic discourse, *will* frequently expresses epistemic meanings such as prediction, inference, and a relatively high degree of authorial commitment. It is also used in metatextual statements and evaluative contexts, for example, when indicating expectations, criticism, or limited certainty in written feedback (see Lee 2019: 38–40).

In contrast, *shall* is one of the least frequent modal verbs in Present-Day English. Large corpus studies consistently show that it accounts for only a small proportion of modal usage, especially outside of formal written contexts (see Rizvić-Eminović, Šukalić 2019: 357–358, 368–369; Sujatna et al. 2019: 169, 171). However, this general rarity stands in sharp opposition to its frequent use in legal discourse where it is the most frequent modal verb which refers to imposing obligations, issuing commands, and formulating rules and prohibitions (see Erić-Bukarica 2020: 80, 81, 85).

These synchronic differences are closely connected to the diachronic development of the two verbs. In Old English, *will* and *shall* occurred in periphrastic constructions and retained clear modal meanings, with *will* expressing volition and intention, and *shall* expressing necessity and obligation. Neither verb functioned as a fully grammaticalized future marker at this stage. Future reference emerged gradually from modal meanings that referred to situations that took place after the moment of speech (Wisher 2008: 131). During Middle English, predictive uses of both verbs increased, with *shall* becoming particularly prominent as a marker of futurity, especially in first- and third-person contexts (*Ibid*. 137–139).

Will and *shall* followed different grammaticalization paths, too, which can be observed in their semantic, morphological, and functional development. When it comes to modality, *shall* extended its use across epistemic, deontic, and dynamic domains, whereas *will* remained more strongly associated with dynamic modality and volitional meanings (Nakayasu 2011: 78). When it comes to the morphological development, in Middle English, for instance, regularized plural forms of *shall* were relatively widespread and appear to have developed under analogical influence of *will*, which indicates close interaction between the two verbs during this period (Gregersen 2022: 26). The awareness of these differences is also reflected in early grammatical descriptions, since grammarians distinguished *shall* and *will* not only according to the grammatical person, but also with regard to modal meaning and syntactic function (Moody 1977: 286).

This overview suggests that *will* and *shall* have attracted scholarly attention for a long time. They have been compared and contrasted from a variety of grammatical and semantic perspectives, both synchronically and diachronically. The present paper approaches these two modal verbs from a constructionist perspective by treating *WILL V* and *SHALL V* as form-meaning pairings and examining their distributional behavior. In particular, the study applies distinctive collexeme analysis which we consider appropriate since the two constructions are functionally comparable and partially interchangeable.

The main goal of the paper is to explore systematic differences between them further, by examining which verbs typically occur within *WILL V* and *SHALL V* collostructions, that is, which collexemes are statistically attracted to or repelled from the said collostructions, and to examine how they relate to the modal meanings of *will* and *shall*. The analysis is conducted on academic discourse because the previous corpus-based studies (e.g. Rizvić-Eminović, Šukalić 2019; Sujatna et al. 2019) have shown that *shall* is generally rare in Present-Day English but occurs with sufficient frequency in academic and formal written registers to allow for meaningful comparison with *will*.

2. THEORETICAL FRAMEWORK AND METHODOLOGY

2.1. COLLOSTRUCTIONAL ANALYSIS AND DISTINCTIVE COLLEXEME ANALYSIS

Distinctive collexeme analysis (DCA) is one of the three methods that constitute collostructional analysis (CA). CA was developed by Stefanowitsch and Gries (2003). It is theoretically grounded in Construction Grammar (CxG) and methodologically rooted in corpus linguistics.

Its connection to CxG lies in the fundamental assumption that the grammar of a language consists of constructions. According to Goldberg (2006: 5), any linguistic pattern is considered a construction if some aspect of its form or function is not strictly predictable from its component parts or other existing constructions. Moreover, even fully predictable patterns are stored as constructions provided that they occur with sufficient frequency. In her most recent work, Goldberg (2019: 7) describes constructions as emergent clusters of lossy memory traces, organized in a high-dimensional conceptual space on the basis of shared formal, functional, and contextual properties. This view entails that constructions themselves encode meaning, rather than serving as mere containers for lexical items, while the newest definition reflects a stronger emphasis on usage, memory, and categorization. Such an understanding is consistent with the principle of semantic compatibility which holds that lexical items may occur in a construction only if their meanings are compatible with

the meaning of the construction, as well as with the principle of semantic coherence, which requires general coherence among all positions within a construction (see Pavlović 2021: 77; Stefanowitsch, Gries 2003: 213; Tošić Lojanica 2021: 106).

Accepting the constructionist view enables the study of grammatical constructions relying on methods traditionally associated with the analysis of the lexicon. This is made possible through corpus linguistics techniques, which form the empirical basis for CA (see Stefanowitsch, Gries 2003: 210–211). CA consists of three quantitative corpus-based methods: simple collexeme analysis (see Stefanowitsch, Gries 2003), distinctive collexeme analysis (see Gries, Stefanowitsch 2004a), and covarying collexeme analysis (see Gries and Stefanowitsch 2004b). All three methods investigate the relationship between constructions and the lexical items that occur within them by measuring the degree of attraction or repulsion[3] between individual lexemes and specific grammatical constructions.

The primary aim of CA is to identify lexemes that are statistically typical, i.e. significantly associated with a given construction or sets of constructions. Unlike traditional collocational analysis which focuses on lexical co-occurrence and relies heavily on raw frequency counts, CA is specifically designed for analyzing lexical-grammatical interface. Raw frequencies alone are insufficient in this context, as they fail to take into account the overall frequency of a lexical item in the corpus. CA therefore determines whether a particular lexical item occurs in a construction more frequently than expected on the basis of chance, relying on relative, rather than absolute frequencies. As a result, the most strongly attracted lexemes are not necessarily the most frequent ones. Lexical items that exhibit a statistically significant attraction to a construction are called *collexemes* of that construction. Conversely, a construction associated with a particular lexeme is called a *collostruct*. The blend of a collexeme and a collostruct is referred to as a *collostruction* (see Stefanowitsch, Gries 2003: 215; Hilpert 2014: 392–393; Pavlović 2021: 77).

For the purposes of the present study, the relevant type of CA is distinctive collexeme analysis. DCA is specifically designed to investigate pairs of semantically similar constructions and the distribution of lexical items across them. It identifies lexemes that show a statistically relevant preference for one construction over the other, thereby making it possible to identify subtle distributional and functional differences between them. While DCA has been frequently applied to cases of grammatical alternation (e.g. the dative alternation), it can also be used to examine other constructional choices provided by grammar (e.g. the two future

3 Attracted collexemes occur more frequently than expected, whereas repelled collexemes occur less frequently than expected. DCA is based on a binary principle – collexemes that are statistically attracted to one construction are automatically repelled by the other. In the results section of the paper, we will present collexemes attracted to each construction, which simultaneously reflects the collexemes repelled by the other.

tense[4] constructions in English). DCA has two main applications. First, it allows for the identification of subtle differences between constructions with similar meanings that can be difficult to capture relying on traditional analytical approaches. Second, it can be used to critically examine the notion of grammatical alternation itself, showing that many alternations are more constrained than previously assumed. In the present paper, DCA is used mainly with the first aim, i.e. to reveal differences between constructions with related meanings (Gries, Stefanowitsch 2004a: 97).

2.2. METHODOLOGY

DCA, like other collostructional methods, is conducted in two main steps: a quantitative step followed by a qualitative one. The quantitative step consists of data extraction and statistical analysis, while the qualitative step involves the interpretation of the statistically obtained data.

The analyzed examples were extracted from the *Corpus of Contemporary American English* (COCA)[5], specifically from the academic genre. The search was conducted using the labels *will V* and *shall V*, which identify occurrences of these modal verbs followed by the bare infinitive. After the initial data extraction, several adjustments were made to the obtained data. The verbs *be* and *become* were removed from the analysis, because when preceded by the analyzed modal verbs, they most frequently occur in constructions such as *WILL BE/BECOME ADJ*. These constructions involve copular verbs in which the adjective functions as a predicative complement. As such, they differ structurally and functionally from the target construction analyzed in the paper, which focuses on modal verbs followed by a lexical verb. In addition, the verb *overcome* was removed from the analysis as well because occurrences with *shall* predominantly appeared as part of the song *We Shall Overcome*, which was repeatedly attested in the corpus. Since the study focused on collexemes occurring in both constructions, *overcome* had to be excluded. Inspection of the data did not reveal any other comparable patterns of this kind.

As expected, the construction *WILL V* yielded a considerably higher number of tokens than the *SHALL V* construction. Due to this discrepancy, the quantitative and qualitative analyses were restricted only to the verbs shared by both constructions. Collexemes unique to either construction are beyond the scope of this paper, which we acknowledge as one of its limitations.

In order to conduct the statistical analysis, frequency data were organized in a 2x2 contingency table:

4 We do not consider futurity as a category of grammatical tense. The term *future-tense constructions* has been directly taken from Gries and Stefanowitsch (2004a: 97).

5 https://www.english-corpora.org/coca/

	construction Y	construction Z	total
verb X	A	B	A + B
all other verbs except from X	C	D	C + D
total	A + C	B + D	A + B + C + D

Table 1: Frequency data needed for the DCA

A refers to the frequency of the verb X in the construction Y. B refers to the frequency of the verb X in the alternating construction Z. C represents the frequency of the construction Y with all verbs except X, while D represents the frequency of the construction Z with all other verbs except X. This type of frequency data is extracted consecutively for each verb occurring in the analyzed constructions (see Pavlović 2021: 87).

The statistical analysis was conducted using *Rstudio* software, relying on the *collostructions* package by Flach (2017). Expected frequencies of individual verbs in the analyzed constructions were calculated using the chi-square test, based on their overall distribution across constructions. Then, Fisher's Exact Test was applied in order to test the null hypothesis, according to which the occurrence of a given verb and the analyzed construction are statistically independent, i.e. the verb does not show any systematic preference for either construction. The resulting *p*-values (*p-exact*) indicate the degree of statistical evidence against the null hypothesis. Smaller *p*-values therefore suggest that the observed distribution of a verb across the two constructions is unlikely to have arisen by chance alone. In the *collostructions* package by Flach (2017) *p*-values are represented using asterisks: five asterisks correspond to $p < .00001$; four asterisks correspond to $p < .0001$; three asterisks correspond to $p < .001$; two asterisks correspond to $p < .01$, whereas one asterisk refers to $p < .05$. If the *p*-value is higher than 0.5, the result is considered statistically non-significant (*ns*). In addition, the analysis provides negative base-10 logarithms of the *p*-values, which are commonly used in collostructional analysis as an index of collostructional strength and as a means of ranking collexemes according to their association with the analyzed collostruction (see Pavlović 2021: 83–85; Tošić Lojanica 2021: 109–110). However, according to Schmid & Küchenhoff (2013), *p*-values primarily reflect the reliability of an observed association, rather than its magnitude or cognitive strength[6]. In the present

6 Small *p*-values obtained through Fisher's Exact Test indicate that a verb shows a statistically reliable attraction to or repulsion from a construction, but they do not in themselves quantify the strength of this association. As Schmid & Küchenhoff (2013: 539) point out, *p*-values primarily express the likelihood that an observed distribution would arise under the null hypothesis, rather than the magnitude of the association itself. Moreover, *p*-values are sensitive to sample size – as corpus size increases, *p*-values tend to decrease, even when relative frequencies remain constant, which means that large corpora have an inherent tendency to reject the null hypothesis (Schmid & Küchenhoff

study, they are therefore used to identify statistically robust patterns of attraction and repulsion, rather than as direct measures of effect size or entrenchment.

After the quantitative analysis, the results were interpreted qualitatively. The qualitative analysis primarily involved classifying the verbs into semantic categories. Given the polysemous nature of verbs, these categories are necessarily flexible because individual verbs could be simultaneously classified within more than one category. This is particularly relevant for the present study since all examples were extracted from the same functional style (academic), where contextual variation is limited and verb meanings could sometimes overlap. The semantic classification of verbs was thus based on their dominant meaning as attested in the analyzed corpus.

3. RESULTS AND DISCUSSION

3.1. COLLEXEMES ATTRACTED TO *WILL V*

Out of 402 shared collexemes between the two analyzed constructions, 184 were attracted to the collostruction *WILL V*, and 52 of these showed statistical relevance. Among them, 14 collexemes showed the strongest statistical evidence of attraction to the construction marked with five asterisks, 5 collexemes showed strong statistical evidence of attraction to the construction marked with four asterisks, 10 collexemes showed moderate statistical evidence of attraction to the construction marked with three asterisks, and 23 collexemes showed weaker, but still statistically significant attraction to the construction, with 12 collexemes marked with two asterisks, and 11 collexemes marked with only one asterisk.

Table 2 shows the top 20 most statistically significant collexemes, followed by their expected and observed frequencies, collostructional strength and p-values:

COLLEX.	Observed freq. *WILL*	Expected freq. *WILL*	Observed freq *SHALL*	Expected freq. *SHALL*	ASSOC.	COLL. STR.	SIGNIF.
help	2362	2262.1	2	101.9	*WILL V*	42.2146	*****
need	2293	2199.9	6	99.1	*WILL V*	35.37889	*****
continue	2832	2742.4	34	123.6	*WILL V*	22.09516	*****

2013: 540–541). For this reason, *p*-values are interpreted here as indicators of statistical reliability, rather than as direct measures of effect size or cognitive entrenchment. In this paper, we did not rely on the Odds Ratio (OR) or its negative logarithm E, which we aknowledge as another limitation. Future research could address this by incorporating these measures to complement p-values and provide a more direct estimate of association strength.

allow	1144	1096.6	2	49.4	*WILL V*	18.9546	*****
increase	938	898.5	1	40.5	*WILL V*	16.44739	*****
require	1624	1570.3	17	70.7	*WILL V*	14.32268	*****
lead	1041	1002.8	7	45.2	*WILL V*	12.08748	*****
provide	1561	1515.7	23	68.3	*WILL V*	10.151	*****
change	465	445.9	1	20.1	*WILL V*	7.60251	*****
affect	515	494.7	2	22.3	*WILL V*	7.4545	*****
occur	508	488	2	22	*WILL V*	7.33099	*****
produce	439	422	2	19	*WILL V*	6.12297	*****
find	1176	1147.3	23	51.7	*WILL V*	5.42751	*****
do	975	949.2	17	42.8	*WILL V*	5.38834	*****
bring	522	505.2	6	22.8	*WILL V*	4.61705	****
work	625	606.7	9	27.3	*WILL V*	4.48316	****
tend	288	276.5	1	12.5	*WILL V*	4.39375	****
serve	500	484.2	6	21.8	*WILL V*	4.29948	****
cause	323	311	2	14	*WILL V*	4.14231	****
enhance	244	234.4	1	10.6	*WILL V*	3.61499	***

Table 2: Verbs most statistically significant in the *WILL V* construction

When it comes to the top 20 most statistically significant collexemes, four semantic groups of verbs were identified. The first semantic group consists of *the verbs of cause-effect relationships*: *help*, *lead*, *cause*, *bring*, *produce*, *affect*. In the following examples, these verbs present the first situation (e.g. *increased access to 3D printing*, *the advent of big data*, *small changes in temperature*, etc.) as the reason why the second situation occurs (e.g. *driving innovation in this field*, *a sizable improvement*, *large changes in density*, etc.).

1) Increased access to 3D printing **will help to drive innovation in this field**, and it is likely that the full potential […]
2) Therefore, we conclude that the advent of big data **will lead to a sizable improvement of our ability to predict complex human traits and diseases**.
3) Small changes in temperature near the critical point -- 31 C -- **will cause large changes in density**, similar to boiling where the liquid changes to a vapor.
4) […] genetic engineering **will bring many significant health and environmental benefits to the world**.
5) In renewable power as in other industries, tough competition **will produce the most cost-effective products.**
6) Given the complex paths by which climate change **will affect human health and migration responses**, effective public health and adaptation strategies must embrace policies […].

Such cause-effect relationships are central to academic argumentation in the analyzed corpus, especially in explanatory contexts and when drawing conclusions. The use of modal verb *will* reinforces this interpretation because the action denoted by the collexeme is not presented as a mere possibility, but rather as a highly reliable consequence of the preceding situation. The predictive use of *will* in such collostructions allows authors to present their claims in a higher degree of certainty, particularly when these claims are supported by evidence, which is a typical feature of academic discourse.

The second semantic group consists of *the verbs of enablement and necessity*: *allow*, *require*, *need*. These verbs are particularly important in academic discourse because they enable authors to describe the conditions under which specific actions are possible or necessary, that is, in the examples below, empirical findings or theoretical considerations are presented as making particular actions feasible (e.g. *will allow more precise forecasting*), or as imposing clear requirements (e.g. *will require transforming the world economy*; *will need to act without a complete model*). The modal verb *will* adds a sense of future expectation and certainty to these statements.

7) […] tracking the dynamics of parameters and their derivatives, **will allow one to more precisely forecast the subsequent disproportions connected with exogenous injections of finances** […]
8) […] avoiding catastrophic impacts of climate change **will require transforming the world economy** at a speed and scale that has no historic precedent.
9) These arguments drive us to the conclusion that every AI system **will need to act without having a complete and correct model of the world**.

The third semantic group consists of *the verbs of change, development, and processes that unfold over time*: *change*, *increase*, *continue*, *occur*, *tend*, *work*, and *enhance*. It is not surprising that the verbs of this type appeared among the most strongly attracted collexemes in the *WILL V* collostruction in the analyzed corpus, because describing trends, tendencies, or ongoing developments are typical of academic discourse. Furthermore, in the examples below, the modal verb *will* is used with a predictive meaning, presenting these developments as expected and well-supported, which allows authors to describe future developments with a relatively high degree of certainty.

10) Absolute risk reduction accounts for the baseline or starting risk but is unstable and **will change** with condition prevalence or baseline risk when applied to other samples and settings.
11) The results predict that the anaerobic core size in each aggregate **will increase** sharply as the network spacing is reduced and the aggregates become tightly packed.
12) If modern fuels are taxed, people **will continue to use traditional fuels**.

13) Most health impacts will be adverse and **will occur** via direct exposures (e.g., heat waves, extreme weather events) and less.
14) In short, over time, a democratic state **will tend to promote the "equality as sameness" principle** through its administrative apparatus.
15) […] concept of nature or physical being that by design excludes mental or experiential being **will work to account for consciousness and its place in nature** […]
16) […] to develop a Silk Road Economic Belt and a twenty-first-century Maritime Silk Road that **will enhance connectivity and cooperation** and provide for integrated and coordinated regional development among the countries involved […]

The final semantic group identified among the top 20 most statistically significant collexemes includes *the verbs of research-related actions*: *find*, *do*, *provide*, *serve*. In the analyzed corpus, they were used to describe different aspects of academic practice. In some cases, they functioned as text structuring devices, as in example 17), and in some other cases, they described the role or contribution of the study, as in examples 18) and 19). On the other hand, there were also cases, as in example 20), where collexemes (e.g. *do* [research]) referred more generally to research activity, rather than to a specific textual function. As in previous semantic groups, the modal verb *will* predominantly expresses prediction.

17) […] You **will find this section** at about 8': 50 " to 12': 00".
18) Both of these studies **will provide strong indications of the potential for efficacy of repellents on sunflowers**.
19) Defining the 'emissions era' as beginning at the present moment **will serve our present purposes**.
20) […] award doctorates, and even in those institutions, it is expected that most academics **will do**, or will recently have done, **research**.

Finally, most of the collexemes attracted to the *WILL V*, but not presented in the table above can be classified within the category of verbs referring to change, development and processes that unfold over time (e.g. *come*, *fall*, *remain*, *develop*, *end*, etc.), and the category of verbs denoting research related actions (e.g. *learn*, *prove*, *perform*, *reveal*, *choose*, *reflect*, *address*, etc.).

3.2. COLLEXEMES ATTRACTED TO *SHALL V*

When it comes to the construction *SHALL V*, 218 collexemes were attracted to it, with 71 collexemes showing statistical significance. Among them, 17 collexemes showed the strongest statistical evidence of attraction to the construction marked with five asterisks, 10 collexemes showed strong statistical evidence of attraction to the construction marked with four asterisks, 9 collexemes showed moderate statistical evidence of attraction to the construction marked with three asterisks, 35 collexemes showed

weaker, but still statistically significant attraction to the construction, with 15 collexemes marked with two asterisks, and 20 collexemes marked with only one asterisk. Table 3 shows the top 20 most statistically significant collexemes:

COLLEX.	Observed freq. *WILL*	Expected freq. *WILL*	Observed freq. *SHALL*	Expected freq. *SHALL*	ASSOC.	COLL. STR.	SIGNIF.
see	931	1298.5	426	58.5	*SHALL V*	244.6298	*****
argue	467	521.5	78	23.5	*SHALL V*	19.64996	*****
return	356	404.8	67	18.2	*SHALL V*	19.34633	*****
call	228	268.9	53	12.1	*SHALL V*	18.88384	*****
inform	80	101.4	26	4.6	*SHALL V*	12.40288	*****
submit	20	33.5	15	1.5	*SHALL V*	11.33951	*****
appoint	12	23	12	1	*SHALL V*	10.17214	*****
consider	316	345.4	45	15.6	*SHALL V*	9.62715	*****
endeavor	15	25.8	12	1.2	*SHALL V*	9.41657	*****
constitute	74	90.9	21	4.1	*SHALL V*	9.25265	*****
issue	38	50.7	15	2.3	*SHALL V*	8.37471	*****
refer	113	129.2	22	5.8	*SHALL V*	7.0822	*****
arrange	10	17.2	8	0.8	*SHALL V*	6.4548	*****
refrain	16	23.9	9	1.1	*SHALL V*	6.25653	*****
try	307	327.3	35	14.7	*SHALL V*	5.59372	*****
rule	22	29.7	9	1.3	*SHALL V*	5.36432	*****
exercise	23	30.6	9	1.4	*SHALL V*	5.23788	*****
cooperate	26	33.5	9	1.5	*SHALL V*	4.88782	****
discuss	436	457.4	42	20.6	*SHALL V*	4.88392	****
include	513	535.9	47	24.1	*SHALL V*	4.86262	****

Table 3: Verbs most statistically significant in the *SHALL V* construction

The first semantic category attracted to the *SHALL V* construction consists of the verbs *argue*, *consider*, *try*, *endeavor*, and *refrain*. These verbs were classified under the label *verbs of academic commitment* because they denote deliberate actions undertaken by the author. In the analyzed data, such verbs are typically used to show what the author intends to do in the study, how they intend to structure the argument, which topics will be addressed, or which issues will be left aside. The meaning of commitment is also reinforced through the meaning of intention expressed by the modal verb *shall*, especially when used with first-person animate subjects:

21) In what follows, I **shall argue that emotional expressions signify emotions** in none of these ways-not by resembling emotions […]

22) Given the large amount of available information, in this paper I **shall consider only that which relates to the years between Cyprian's conversion in about 248** […]
23) In this paper I **shall try to solve the puzzles posed by the triad of metaphysical claims** without availing myself […]
24) […] and in order to facilitate a constructive dialogue in this academic setting, I **shall endeavor to present a Pakistani perspective of the issue** taking into account the genesis of […]
25) […] not help questioning Dr. Morita's classification of shinkeishitsu obsessionality, etc., which **I shall refrain from expounding here**.

The second semantic category attracted to *SHALL V* consists of the *verbs of institutional and procedural acts*: *submit*, *inform*, *cooperate*, *appoint*, *issue*, *exercise*, *constitute*, and *arrange*. Even though they have been identified in academic discourse, they are predominantly associated with the legal subgenre. In the analyzed corpus, they usually referred to formally defined actions assigned to specific institutions or actors, such as submitting information, issuing decisions, appointing officials, exercising authority, etc., all of which are set out in legal or quasi-legal texts. The category label can once again be seen as the result of the interaction between the modal verb *shall* and the semantics of the attracted collexemes. In the examples below, *shall* refers to rules and regulations which is central to legal discourse, and, therefore, naturally co-occurs with verbs that can denote legally defined actions and procedures.

26) A state party intending to establish the outer limits of its continental shelf **shall submit information on such limits**, including scientific and technical data, to the CLCS […]
27) […] establishing a social order in which "justice, social, economic and political, **shall inform all the institutions of national life**."
28) The various attempts by the interested parties **shall cooperate** and exchange information with relevant States, FAO, other international organizations and regional […]
29) […] 3Article II, Section 1 of the Constitution states in part: "Each State **shall appoint**, in such a Manner as the Legislature thereof may direct, **a Number** […]
30) […] is instituted and not dismissed under this chapter, the Patent Trial and Appeal Board **shall issue a final written decision** with respect to the patentability of any patent claim challenged […]
31) Article 85 # The courts of the Hong Kong Special Administrative Region **shall exercise judicial power independently**, free from any interference.
32) The Preparatory Committee shall be composed of mainland members and of Hong Kong members who **shall constitute not less than 50 per cent of its membership**.
33) When a law school submits a completed application for provisional approval, the Consultant **shall arrange for a site evaluation** by a team of qualified and objective persons.

The third semantic category that has been identified consists of the *verbs of text structuring and discourse navigation*: *see*, *return*, *discuss*, *refer*, and *call*. These verbs functioned as metadiscursive devices in the analyzed corpus, but not in the same way. Some of the verbs such as *see* in 34), *return* in 35), and *discuss* in 36) primarily served the text structuring function because they explicitly referred to textual organization. Other verbs, such as *refer* in 37) and *call* in 38) functioned as discourse navigators, which means that they guided the reader in how terms or references should be interpreted. Among these verbs, *see* stands out as a clear outlier, with a considerably higher collexeme strength than the remaining verbs. This can be attributed to its highly conventionalized use in academic discourse (e.g. *as we shall see*), where it functions as a frequent marker of text structuring.

When it comes to the meaning of *shall* within this semantic group, its dominant meaning was the one of intention – it referred to the discourse actions marked by the collexemes as deliberate and authorial decisions about how the text will proceed or how the terms will be used.

34) As we **shall see in the next Section**, the "appropriate scope of governmental power" specifically […]
35) I **shall return to this detail below**, but, on a preliminary assumption, we should […]
36) These are most widely used in teaching, and I **shall discuss them in the section on teaching aids**.
37) […] Instructional Evaluation" or "Teaching Resource Center," but for our purpose we **shall refer to it as the Ministry of Teaching (MOT).**
38) For simplicity we **shall call the region west of the Great Belt " Jutland "** and the region east

The *WILL V* collostruction attracted a similar semantic category of verbs – the *verbs of research-related actions*. However, this category is not as homogenous as it contains verbs which function as text-structuring devices (see *find* in example 17), as well as verbs referring to the role, contribution, or general activity of research (see *provide*, *serve*, *do* in examples 18, 19, and 20).

Many of the verbs outside of the top 20 most statistically significant ones in the *SHALL V* construction can be classified within the already mentioned categories: *the verbs of academic commitment* (e.g. *conclude*, *suggest*, *propose*, *recommend*, *attempt*, *investigate*, *examine*, *obtain*, *establish*, *achieve*, etc.), *the verbs of institutional and procedural acts* (e.g. *execute*, *notify*, *insure*, *preclude*, *notify*, etc.), and *the verbs of text structuring and discourse navigation* (e.g. *specify*, *elaborate*, *review*, *state*, etc.). In addition, another semantic category has been introduced, containing what we have called *the verbs of definition and delimitation* (e.g. *limit*, *confine*, *contain*, *belong*, *include*, *consist*, etc.), which, as the name suggests, were mostly used in the analyzed data to define categories and specify their boundaries.

4. CONCLUSION

This paper has examined two formally identical, but functionally different constructions, *WILL V* and *SHALL V* from DCA perspective in academic discourse. The results show that, in the analyzed corpus, the *WILL V* construction attracts verbs that express cause-effect relationships, enablement and necessity, change and development, and research-related actions. Across these semantic categories, *will* predominantly expresses predictive meaning, presenting outcomes, developments, and research consequences as expected and well-supported.

On the other hand, the *SHALL V* construction mainly attracts verbs of academic commitment, institutional and procedural acts, text structuring and discourse navigation, as well as verbs of definition and delimitation. In this construction, *shall* normally expresses the meaning of intention and, when combined with verbs of institutional and procedural acts, it conveys the meanings related to norms, obligations, and formal procedures.

Taken together, in the analyzed corpus, these two constructions are semantically related, but they differ clearly in the types of verbs they attract. While *will* is primarily associated with predictive meaning and is used to present developments and research outcomes as expected based on the analysis, *shall* is associated with obligation and intention, which supports authorial commitment, text organization, and institutional or procedural acts.

Quantitative and qualitative analyses were restricted to shared collexemes which we consider one of the limitations of the present paper. Future research could expand the analysis by including non-shared collexemes, or systematically including other grammatical parameters (e.g. subject types), or by analyzing the behavior of the two constructions across different academic subgenres.

LITERATURE

Erić-Bukarica 2019: A. Erić-Bukarica, The use of modal verbs in English legal texts and their Serbian equivalents, Novi Sad: *The Journal for Languages and Literatures of the Faculty of Philosophy in Novi Sad*, 9, Novi Sad, 73–96.

Flach 2017: S. Flach, *Collostructions: An R implementation for the family of collostructional methods*, R package version 0.1.0, www.bit.ly/sflach

Goldberg 2006: A. E. Goldber, *Constructions at Work: The Nature of Generalization in Language*, Oxford: Oxford University Press.

Goldberg 2019: A. E. Goldberg, *Explain Me This*: *Creativity, Competition, and the Partial Productivity of Constructions*, Princeton / Oxford: Princeton University Press.

Gregersen 2022: S. Gregersen, Regularized modal verbs in Middle English dialects, Cambridge: *English Language and Linguistics*, 27(1), Cambridge, 23–43.

Gries, Stefanowitsch 2004a: S. T. Gries, A. Stefanowitsch, Extending collostructional analysis: A corpus-based perspective on 'aternations', Amsterdam/Philadelphia: *International Journal of Corpus Linguistics*, (9)1, Amsterdam/Philadelphia, 97–129.

Gries, Stefanowitsch 2004b: S. T. Gries, A. Stefanowitsch, Co-varying collexeme analysis in the into-causative, Stanford: *Language, culture, and mind*, Stanford, 225–236.

Hilpert 2014: M. Hilpert, Collostructional analysis, *in*: D. Glynn, J. A. Robinson (Eds.) *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, 43, Amsterdam/ Philadelphia: Benjamins, 391–404.

Karahan 2022: P. Karahan, Comparative Analysis of Epistemic Modal Verbs in the Three Sub-corpora of Humanities and Social Sciences, *International Journal of Educational Spectrum*, 4(1), Hakan Ulum (Turkey), 43–63.

Lee 2019 K. Y. Lee, Hedging expressions used in academic written feedback: a study on the use of modal verbs, Murcia: *Research in Corpus Linguistics*, 1, Murcia, 33–45.

Leech[3] 2004: G. Leech, *Meaning and the English Verb*, London and New York: Routledge.

Moody 1977: P. A. Moody, *Shall* and *will* in English grammars: A revised history, Amsterdam: *Historiographia Linguistica*, 4(3), Amsterdam, 281–301.

Nakayasu 2011: M. Nakayasu, Towards a pragmatic analysis of modals *shall* and *will* in Chaucer's language, Poznań: *Studia Anglica Posnaniensia*, 46(4), Poznań, 73–96.

Pavlović 2021: V. Pavlović, *Konstrukciona gramatika i kolostrukciona analiza*, Niš: Filozofski fakultet.

Rizvić-Eminović, Šukalić 2019: E. Rizvić-Eminović, Đ. Šukalić, Corpus-based study of the modal verbs in the spoken and academic genres of the Corpus of Contemporary American English, Zenica: *Zbornik radova Islamskog pedagoškog fakulteta Univerziteta u Zenici*, 17, Zenica, 351–375.

Schmid, Küchenhoff 2013: H. J. Schmid, H. Küchenhoff, Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings, Berlin/Boston: *Cognitive Linguistics*, 24(3), Berlin/Boston, 531–577.

Stefanowitsch, Gries 2003: A. Stefanowitsch, S. T. Gries, Collostructions: Investigating the interaction between words and constructions, Amsterdam/ Philadelphia: *International Journal of Corpus Linguistics*, 209–243.

Sujatna et al. 2019: Meita Lukitawati Sujatna, Eva Tuckyta Sari Sujatna, Kasno Pamungkas, Exploring the use of modal auxiliary verbs in Corpus of Contemporary American English (COCA), *in*: R. Widya Setiabudi Sumadinata (Ed.), *Sosiohumaniora: Jurnal Ilmu-ilmu Sosial dan Humaniora* 21(2), Bandung: Faculty of Social and Political Sciences, Universitas Padjadjaran, 166–172.

Tošić Lojanica 2021: T. Tošić Lojanica, Exploring present ability: A collostructional approach, Kragujevac: *Nasleđe*, 48, Kragujevac, 105–115.

Wisher 2008: I. Wisher, *Will* and *shall* as markers of modality and/or futurity in Middle English, Berlin/Boston: *Folia Linguistica Historica*, 29, Berlin/Boston, 125–143.

Катарина Б. Субановић / ГДЕ ЈЕ *WILL*, ТУ ЈЕ И *SHALL*: АНАЛИЗА ДИСТИНКТИВНИХ КОЛЕКСЕМА КОД КОНСТРУКЦИЈА *WILL V* И *SHALL V* У АКАДЕМСКОМ ДИСКУРСУ

Резиме / У раду се анализирају конструкције *WILL V* и *SHALL V* из перспективе анализе дистинктивних колексема. Ова врста анализе представља квантитативну методу у оквиру корпусне лингвистике коју су осмислили Грис и Штефанович (2004а). Анализирана језичка грађа преузета је из *Корпуса савременог енглеског језика* (COCA) и то из академског жанра. Основни циљ рада представља идентификацију семантичких категорија глагола које испитане конструкције статистички привлаче, као и испитивање повезаности између значења модалних глагола и привучених колексема. Резултати показују да ове конструкције привлаче глаголе који припадају различитим семантичким класама када је испитана језичка грађа у питању. Конструкција *WILL V* најчешће привлачи глаголе који упућују на узрочно-последичне везе, затим глаголе који означавају могућност и нужност, глаголе промене стања и развијања радње, као и глаголе који упућују на различите аспекте истраживачких радњи, а модални глагол *will* имао је предиктивно значење у испитаном корпусу. Са друге стране, конструкција *SHALL V* углавном привлачи глаголе који означавају академско обавезивање, глаголе који упућују на различите врсте институционалних и процедуралних радњи, затим, глаголе који упућују структурисање текста и глаголе који означавају дефинисање и разграничавање у оквиру академског дискурса. У оквиру ове конструкције, глагол *shall* најчешће је означавао намеру, док је у комбинацији са колексемама који означавају институционалне и процедуралне радње, имао значење повезано са правним нормама и обавезама и формалним процедурама.

Кључне речи: *WILL V*, *SHALL V*, колострукциона анализа, анализа дистинктивних колексема, академски дискурс

Примљен: 16. 2. 2026.
Прихваћен марта 2026.