V-collocates with *will* and *be going to*:
A Corpus-based Analysis

Jarry Chia-Wei Chuang
Department of English
National Chengchi University
Taipei, Taiwan

cwchuang.academia@gmail.com

ORCID: 0000-0002-3029-4463

Abstract

A corpus-based study is conducted to investigate verbs collocating with present-tensed *will* and *be going to*. Data from Corpus of Contemporary American English (COCA) are addressed, especially for lexical verb-collocates (V-collocates). Data-driven results show that lexical V-collocates of *will* and *be going to* are sensitive to semantic and pragmatic functions. MI scores manifest that verbs with deficient meanings but higher functionality can be commonly applied to *will* contexts. V-collocates in *will* contexts have a wider acceptance of register, despite a potential stylistic preference to formal use; in contrast, V-collocates in *be going to* reveal its lower formality but higher frequency in informal spoken context. Additionally, semantic expansion with metaphorical and hilarious use is interestingly found in informal *be going to* contexts, especially on words involved with serious legal or security issues. In general, meaning shifts within *will* and *be going to* are greatly influenced by and pragmatically derived from context, which is coordinated with the monosemous account by Nicolle (1998a). The present findings offer new information about the distinction of English future event markers.

Keywords: verb collocates (V-collocates), will, be going to, corpus-based, COCA

1   Introduction

1.1   Overview

As primary English future event expressions, *will* and *be going to* have been long explored for their intricacies in syntactic, semantic, and pragmatic aspects (Berglund, 1997, 2005; Haegeman, 1989; Nicolle, 1998a; Szmrecsanyi, 2003). Researchers have demonstrated their efforts in the distinction with different methods and from theoretical aspects, while the relation between *will* and *be going to* in actual use remains complicated.

1.2   Research Difficulties

The difficulties in distinguishing them can be derived from (1) their intimate associations with tense, aspect, and mood/modality (TAM), (2) their sensitivity to syntactic and semantic conditions, and (3) disambiguation from their semantics or pragmatics. As Haegeman (1989) noted: "to analyze future markers in English, we address not only tense and aspectual auxiliaries but also the relation between surface representations of tense and truth conditions." In addition, by examining syntactic environments, Szmrecsanyi (2003) further indicates that English future event markers are sensitive to negation, subordinate contexts, IF-clause, and even sentence length. As for their disambiguation, a monosemous account was proposed by Nicolle (1998a), in which the varieties of *will* and *be going to* interpretations and their function to express

future events are pragmatically derived. Multiple factors have been widely discussed, suggesting that follow-up studies should better consider those intervenient elements at different linguistic levels.

1.3 Research Gap

Previous discussions mainly focused on the complementizer layer and inflectional layer (CP & IP), while attention was seldom paid to the lexical layer (vP & VP). As the meaning core of subject predicates, sentential expressions, and utterances, lexical verbs (V_{LEX}) (with lower functionality but rich semantic knowledge), in comparison with *will* and *be going to*, should provide vital information. Verbs collocating with *will* and *be going to* are thus believed to play a significant role in explaining their meanings as well as marking the differences in use.

1.4 Purpose of Study

The present study investigates present-tensed *will* and *be going to*, with an innovative point of view. We explore the pattern at the lexical layers, by scrutinizing their verb collocates (V-collocates). Research Data is retrieved from the Corpus of Contemporary American English (COCA). V-collocate differences are expected to show different semantic and pragmatic functions of the two markers. It is worth noting that considering the sensitivity to TAM as well as the complex interaction with tense and aspectual auxiliaries. So, in the paper, we mainly discuss lexical V-collocates (*will* / *be going to* + V_{LEX}).

2 Syntax

Some of the previous studies on *will* and *be going to* attach importance to their syntactic environments. To observe their similarity and difference in syntactic patterns, they may show implications on the distinction. Those attempts done previously can be summarized and classified in the 3-layered structure of syntax.

Rizzi (1997) categorizes the syntactic structure into three layers, including the complementizer layer (CP), inflectional layer (IP), and lexical layer (vP & VP). Associating syntactic layers with previous discussions, we find that studies on the CP and IP levels were mostly explored previously, while the lexical layer in *will* and *be going to* contexts was seldom under inspection.
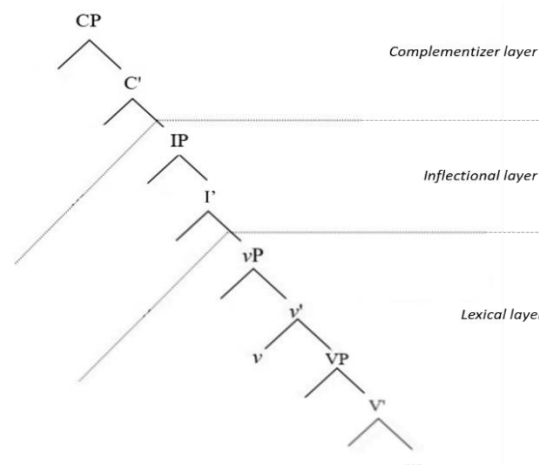

Figure 1. The three-layered structure of English syntax (Rizzi, 1997)

2.1 Complementizer layer (CP)

The complementizer layer is generally headed by a functional word or morpheme (e.g., *for*, *since*, *when*, *if*, *as*, etc.). Type of sentence is also determined here. CP-level discussions about *will*-and-*be-going-to* distinction include their sensitivity to subordinate contexts and IF-clause.

In Szmrecsanyi's (2003) study, *be going to* presents a high-frequency pattern in syntactic dependent contexts (i.e., subordinate contexts) but a relatively lower frequency in independent contexts (i.e., the main clauses of sentences); on the other hand, reservations are held for *will* contexts, since *will* seems not to demonstrate an obvious tendency towards syntactic independence within sentences. In general, subordinate contexts have perceptible influences on the occurrence of *be going to*. Moreover, Yeh (2021) reexamined subordinate structures in both contexts, using data from the British National Corpus (BNC). It is indicated that in the BNC data, *will* is much more frequent in independent contexts. Note that Szmrecsanyi's (2003) data sources embraced British English (BE) and American English (AE), while Yeh (2021) only adopted a BE-based corpus. This infers the use of *will* and *be going to* may have dialectal variations and cross-cultural bias, in which AE and BE speakers show different degrees of preference.

Szmrecsanyi (2003) also pointed out that *will* and *be going to* have extraordinary sensitivity in if-subclause (IF-Sub) structure, among subordinate

constructions: *will* is more dominant in main clauses of IF-Sub clauses; *be going to* is frequent in IF-Sub clauses. IF-Subs usually lead conditional sentences, where their truth values are opaque or unsure. Conditional sentences often coordinate with mood and modality to avoid over-affirmation. Divergent distributions in IF-Subs confirm that *will* and *be going to* are tied to TAM.

2.2 Inflectional layer (IP)

The inflectional layer is often headed by functional heads. According to the split-IP hypothesis (Pollock, 1989), IP can be divided into several sub-layers, some of which may be subtly different across languages. Verb can move from lower phrases to TP, which is called verb raising. In English syntax, one of the sub-phrases in the split IP is Negation Phrase (NegP).

In previous studies, negation is reported to be influentially associated with the distribution of *will* and *be going to* (Szmrecsanyi, 2003) in the IP level: *won't* frequently occur in negated contexts, while data excluding *won't* illustrates that *will* is hardly negated. As for *be going to*, data of its negation does not reveal an explicit preference. This phenomenon should be noted by the analogical pressures in AE, in which AE speakers prefer contracted forms (cf. Hofland & Johansson, 1982; Hundt, 1997; Szmrecsanyi, 2003); thus, the frequency of contracted *won't* in AE can be higher than that of *not going to*. To study *will* and *be going to*, cross-dialectal difference should be particularly noted.

2.3 Lexical layer (vP & VP)

The lexical layer is normally headed by the verb, in which the argument structure and theta roles work. In the syntax of *will* and *be going to*, the main IPs are occupied by *will* and *be going*. V_{LEX} serves as the head of vP, as demonstrated in (1). The syntax of *will* is less complicated than *be going to*, in that V_{LEX} in the latter context is located at the vP of another lower-layered CP.

(1) Syntax of present-tensed *will* and *be going to*

a. *will* + V_{LEX}:
[$_{TP}$ *will* [$_{vp}$ V_{LEX} …

b. *be going to* + V_{LEX} :
[$_{TP}$ *is/am/are* [$_{AgrP}$ *going* [$_{CP}$ [$_{TP}$ *to* [$_{vP}$ V_{LEX} …

Literature about *will* and *be going to* mostly discuss their behaviors at CP and IP layers, hardly coping with the syntax at the lexical layer (vP & VP); however, V_{LEX} following them can provide vital information in the distinction, since V_{LEX} can not only be C-selected but also be S-selected by *will* and *be going to*.

It would be argued here that V_{LEX} is worth noting and might be more truthful to capture the pattern of *will* and *be going to*, since it is dominated at the lower layer.

3 Semantic-Pragmatic Interactions

Aside from the syntactic distributions, the semantics and pragmatics of *will* and *be going to* should also be examined and well defined first. They seem to be polysemous in surface, in which *will* is more than expressing volition and *be going to* is more than signifying movement. Checking the evolution of their meanings, we might find a successive process of meaning expansion and they might not be polysemous despite multiple meanings that have been derived in both contexts.

As the semantic model of English modal auxiliaries proposed by Klinge (1993), Nicolle (1997) argued that *will* and *be going to* should be analyzed as monosemous, under the Relevance Theory. Even though users usually think of more than one possibility that *will* and *be going to* can derive, this does not make their polysemy. Instead, such expressive diversity of *will* and *be going to* is considered to be pragmatically derived (Nicolle, 1998a). The apparent polysemy is caused by their context-sensitive interpretations, highly associated with grammaticalization (i.e., shifting from heavy verbs to light verbs/modal auxiliaries) (Nicolle, 1998b). Grammaticalized *will* and *be going to* make themselves seemingly polysemous. This can be attributed to semantic retention. Original senses are kept but only triggered in certain contexts, in which lexical meanings and grammaticalized senses co-exist. So, *will* and *be going to* offer several potential interpretations in contexts and those should be interpreted as sense derivation at pragmatic levels.

3.1 *will*

Commonly known as modal auxiliaries in English, *will* derives several contextual meanings, appearing to state (1) possibility with present or future time references, (2) habitual expressions, (3) willingness

to make performance or take action, and so on. Previous analyses have drawn the monosemous analysis and attributed such polysemous interpretations to semantic retention; a single sense is actually taken. Before grammaticalized, *will* represents its old (lexical) meaning to express one's volition. But, inferential meanings become accessible from the relevance-theoretic perspective. Interpretations through three major functions of English modal auxiliaries (e.g., epistemic, deontic, and dynamic) hence expand its semantically underdefined sense, leading to various contextual meanings.

3.2 *be going to*

In use, *be going to* reflects activities in the upcoming future or inevitability (Coates, 1983; Leech, 2014; Palmer, 2014). Two different meanings are usually considered: (1) prior intention, and (2) current activities (Nicolle, 1998a). Distinct default senses challenge the monosemous account for *be going to,* while it is solved by semantic retention as well. The default meaning is found to be kept and proved evident in native children processing their L1 (English) (Ziegeler, 1996, 1997). Polysemous interpretations are in fact by-products of grammaticalization. V-collocates thereby have the possibility to trigger such contextually polysemous meanings in *be going to* conditions. Brisard (2001) further notes that *be going to* can draw a "paradoxical but pragmatically plausible" account for the near future, in which events remain unknown in the present timing usually, before declared by the speaker.

3.3 Importance of the Monosemy

In general, the semantics of *will* and *be going to* can be treated as monosemy, though. Monosemy should be noted before the following analysis of corpus data. As *will* and *be going to* are monosemous and are semantically based on a single sense, V-collocates or the whole VP can take the floor to greatly influence and even determine the sentence and contextual meanings. Then, V-collocates gain the importance to be checked and discussed in the following sections.

4   Method

The present analysis relies on corpus data from native users' examples, which is believed to better connect close-to-nature patterns to the distinction between *will* and *be going to*.

4.1 Data Source

Research data under close scrutiny are consulted from the Corpus of Contemporary American English (COCA) (Davies, 2008), one of the most representative corpora of American English (AE). It comprises over 1 billion words and a wide array of sentences in genres with various formality, from spoken text to academic context. COCA also provides instant operation of frequency and MI value.

4.2   Procedure & Data Analysis

To explore V-collocates with *will* and *be going to*, we will first categorize types of V-collocates, since they can potentially be attached by lexical verbs or tense-aspectual auxiliaries. Modified data have been under operation and checked by two well-trained research fellows. Two V-collocates with English future markers are examined through frequency ($Min_{freq} = 20$) and MI scores ($Min_{MI} = 3$).

5   Results & Discussion

V-collocates may include lexical verbs and auxiliary verbs (e.g., progressive auxiliary, passive auxiliary, perfect auxiliary, etc.). Raw data consulted from COCA should be under re-analysis to scout out lexical V-collocates. V-collocates can further be explored via frequency and MI scores.

5.1   Type of V-Collocates

Verbs attached to *will* and *be going to* are miscellaneous. Note that BE-verbs should be cautiously examined, for they can serve as (1) intensive verbs proceeded by subject predicates like adjectives or nouns, (2) progressive auxiliaries [PROG] followed by Ving, and (3) passive auxiliaries [PASS] to express passive voice. Besides, *have* also needs close examination, since it can be a lexical verb as well as the perfect auxiliary [PREF]. So, the data presented below has been amended based on raw frequency excerpted from COCA. Four types of V-collocates dominating most of the uses are shown as follows:

	will	*'ll*	**Total**	**Occurrence**
X + V$_{LEX}$	1,480,274	835,879	2,316,153	90.81%
X + be Ving	39,687	29,190	68,877	2.70%
X + be Vpp	140,797	15,515	156,312	6.13%
X + have Vpp	8,200	1,105	9,305	0.36%
Total	1,480,274	835,879	2,316,153	100.00%

Note: X represents two different forms of WILL, *will* or *'ll*.

Table 1: Type of V-collocates with *will*

	is	*'s*	*am*	*'m*	*are*	*'re*	**Total**	**Occurrence**
Y + V$_{LEX}$	78,467	91,977	7,403	65,106	56,010	126,957	182,967	88.63%
Y + be Ving	1,823	1,689	174	1,192	2,512	4,431	11,821	5.73%
Y + be Vpp	4,046	2,346	57	462	2,917	1,745	11,573	5.61%
Y + have Vpp	14	14	-	6	19	34	87	0.04%
Total	84,350	96,026	7,634	66,766	61,458	133,167	206,448	100.00%

Note: Y stands for different forms of BE going to, in which BE can be *is* (*'s*), *am* (*'m*), or *are* (*'re*).

Table 2: Type of V-collocates with *be going to*

In terms of frequency, corpus data illustrates *will* is more prevailing than *be going to*. V-collocates of present-tensed *will* and *be going to* in **Table 1&2** reveal that they are regularly attached to lexical verbs (N$_{1, LEX}$=2,316,153; N$_{2, LEX}$=182,967). In addition, BE-verbs collocating with two verbs demonstrate subtly different patterns, especially when they are progressive auxiliaries and passive auxiliaries. Collective frequencies of passive and progressive use in *will* context display the frequency of passive use (be+Vpp) (N$_{1, be+Vpp}$=140,797) is divergently higher than progressive use (be+Ving) (N$_{1, be+Ving}$=39,687) in *will*, while progressive use (N$_1$=29,190) overwhelms passive use (n=15,515) in cliticized *'ll* conditions. Progressive-passive biased use is not obviously found in *be going to* context (N$_{2, PROG}$=11,821; N$_{2, PASS}$=11,573).

5.2 Frequency of V-collocates

Results of V-collocates by frequency manifest a wide variety of actions that seem not to be semantically associated with *will* and *be going to*. In **Table 3**, high-frequency verbs collocating with present-tensed *will* and *be going to* are displayed. Note that in present-tensed *will* and *be going to* contexts, frequency of *be* is counted, as it serves as an intensive verb to bring the subject predicates out. Yet, V-collocates presented as be+Ving and be+Vpp have been excluded.

will$_{[PRES]}$	*be, have, take, make, get, do, continue, come, help, go, see, find, give, need, become, tell, say, happen, work, keep*
be$_{[PRES]}$ going to	*be, have, get, do, take, go, make, happen, see, come, need, give, say, try, tell, talk, start, put, find, look*

Table 3: V-collocates with present-tensed *will* and *be going to* by frequency

The patterns of V-collocates are analogous to each other. First, *be*, as a lexical verb, is attached to both most frequently. Secondly, verbs carrying abundant meanings (e.g., *have*, *get*, *make*, *take*, *do*, *go*, etc.) are frequently attached to both future markers. This could be taken as part of the evidence

that *will* and *be going to* share similar functions in identifying the high probability of the occurrence.

5.3 MI Scores & V-collocates

As frequency patterns between *will* and *be going to* provide limited information on the distinguishment, the following section will analyze MI scores of V-collocates. Four dominant genres are listed in **Table 4&5**, in which the majority of the register distribution can be found: (1) spoken, (2) magazine, (3) newspaper, and (4) academic contexts.

5.3.1 *will*

Formality will also be counted based on the sum of V-collocates in most of the formal contexts (i.e., magazine, newspaper, and academic contexts). To avoid selection bias, the ratios of occurrence in all the selected contexts are calculated, with the time of use in all documented contexts as the denominator. Results of mutual information reveal a variety of differences in V-collocates. It is found that V-collocates have intimate interaction with semantic and pragmatic functions, including meanings of lexical verbs, genres, and formality.

V-collocate	MI Score	All	Informal		Formal						Formality (M+N+A)/All	
			Spoken		Magazine		Newspaper		Academic			
		N	N	%	N	%	N	%	N	%	N	%
continue	7.42	25,413	4,348	17.1%	3,075	12.1%	4,306	16.9%	2,832	11.1%	10,213	40%
prevail	7.14	827	214	25.9%	98	11.9%	97	11.7%	53	6.4%	248	30%
enable	7.05	2,152	115	5.3%	509	23.7%	234	10.9%	586	27.2%	1,329	62%
depend	6.99	2,493	209	8.4%	425	17.0%	379	15.2%	642	25.8%	1,446	58%
remain	6.70	8,038	779	9.7%	1,160	14.4%	1,635	20.3%	1,059	13.2%	3,854	48%
require	6.48	5,799	271	4.7%	1,042	18.0%	839	14.5%	1,624	28.0%	3,505	60%
receive	6.29	5,240	255	4.9%	582	11.1%	1,309	25.0%	536	10.2%	2,427	46%
allow	6.27	8,386	809	9.6%	1,390	16.6%	1,323	15.8%	1,144	13.6%	3,857	46%
emerge	6.14	1,075	90	8.4%	243	22.6%	191	17.8%	197	18.3%	631	59%
happen	6.04	11,608	2,359	20.3%	1,113	9.6%	1,295	11.2%	425	3.7%	2,833	24%
affect	5.97	3,007	397	13.2%	456	15.2%	494	16.4%	511	17.0%	1,461	49%
begin	5.92	6,892	882	12.8%	948	13.8%	1,829	26.5%	469	6.8%	3,246	47%

Table 4: Lexical verbs collocating with present-tensed *will* by MI score.

V-collocates with *will* have an inclination toward verbs like (1) intensive verbs, (2) prepositional (or intransitive) verbs, and (3) verbs proceeded by non-finite clauses. Such verbs require subject-predicates or object-predicates to make the VPs semantically complete. For example, *remain* serves as an intensive verb and requires a subject-predicate, such as NP or AP, to complete the meaning of the whole VP. Besides, intransitive verbs or prepositional verbs also usually attach to *will*, including *begin*, *continue*, *depend*, *emerge*, *happen*, etc. They can either exist independently in intransitive use or look for an infinitive to complete the meaning of the whole VP. What's more, verbs like *allow*, *enable*, and *require* are commonly followed by a non-finite clause to complement the

expressions. See also *affect*, as in (2a) and (2b). *Affect* is a semantically fuzzy word like *influence* or *impact*. It infers that the prediction is uncertain to some extent. If there is a high probability to occur, speakers should know more details and would rather take more informative verbs to elaborate on how the situation can be influenced. In sum, V-collocates in *will* contexts seem to be semantically defective and possess higher functionality.

(2a) Your vote will **affect** the future and be recorded in eternity. (BLOG, 2012)
(2b) […], which I also don't think will **affect** many people. (NEWS, 1990)

Pragmatically speaking, V-collocates with *will* appears more often in formal contexts. Checking the tendency of formality, we find most of the verbs present a relative increase in frequency, transitioning from spoken context to academic contexts. For example, *enable* and *depend* only show 5.3% and 8.4% occurrence in spoken contexts, but their occurrences obviously surge to 27.2% and 25.8% in academic contexts. Similar patterns can be found in other V-collocates with *will*. Most of the cumulative occurrence rates (including use in magazines, newspapers, and academic contexts) reach 40% and more. Yet, the formality distributions mostly hover around 40-60%. It reveals *will* has a wider acceptance of different registers to some extent (cf. formality value of *be going to* in **Table 5**), in spite of the potential preference towards formal contexts.

5.3.2 *be going to*

V-collocate	MI Score	All	Informal		Formal							Formality (M+N+A)/All	
			Spoken		Magazine		Newspaper		Academic				
		N	N	%	N	%	N	%	N	%		N	%
happen	7.01	17,571	7,568	43.1%	989	5.6%	1,621	9.2%	186	1.1%		2,796	16%
die	6.06	5,619	1,230	21.9%	429	7.6%	359	6.4%	73	1.3%		861	15%
explode	5.95	335	58	17.3%	23	6.9%	13	3.9%	2	0.6%		38	11%
kill	5.55	4,756	1,034	21.7%	194	4.1%	243	5.1%	39	0.8%		476	10%
jail	5.50	1,403	328	23.4%	60	4.3%	104	7.4%	16	1.1%		180	13%
lose	5.38	3,347	1,320	39.4%	220	6.6%	337	10.1%	27	0.8%		584	17%
win	5.35	4,885	2,095	42.9%	318	6.5%	617	12.6%	24	0.5%		959	20%
marry	5.34	841	164	19.5%	52	6.2%	38	4.5%	2	0.2%		92	11%
be	5.26	156,090	70,649	45.3%	8,903	5.7%	15,274	9.8%	1,751	1.1%		25,928	17%
retire	5.20	315	135	42.9%	22	7.0%	63	20.0%	1	0.3%		86	27%
solve	5.20	897	445	49.6%	62	6.9%	99	11.0%	13	1.4%		174	19%
continue	5.04	3,702	2,310	62.4%	151	4.1%	366	9.9%	46	1.2%		563	15%

Table 5: Lexical verbs collocating with present-tensed *be going to* by MI score.

As shown in Table 5, *die*, *explode*, *kill*, and *jail* are verbs involved with serious legal or security issues. Normally, speakers would not address such verbs for sensitivity, especially when things have been known to certainly happen. Referring to the original sense of *be going to*, we find its V-collocates in part related to the meanings: (1) events in the near future and (2) inevitability. The distributions are attributed to the latter one. While it is confirmed to have occurrence and display the certainty, those verbs can be linked to the use of *be going to* (i.e., pure future and inevitability) much better than *will*. The rules can likewise explain other V-collocates like *lose*, *win*, *marry*, and *retire*. Those verbs are used when the happening of the action or the appearance of status is for sure for speakers.

Furthermore, an intriguing use with semantic expansion is frequently found in *be going to* context. For example, *explode* and *kill* are commonly associated with legal or security issues. As they are thrown into the spoken context, the original meaning of *explode* and *kill* are weakened; yet, metaphorical use and humorous expressions take place to expand their uses. Look at the examples below.

(3a) Bob is going to *explode*. (SPOK, 2014)
(3b) I believe you, Joe. […] You're not going to *kill* me? (SPOK, 2017)

In (3a), it would be generally reckoned that Bob is going to be angry, instead of breaking up into pieces. The speaker describes that Bob's emotions would erupt angrily, dangerous as a bomb. As for

(3b), where a question is hilariously asked. By observing the context, it can be easy to understand that the speaker did not really confirm whether his friend was going to kill him, but made fun of his friends and try to tell his friend not to be angry. The speaker covertly manipulates the cooperative principle, flouting the maxims. In fact, using *kill* to play the boundary is a typical example, in that many examples do not derive *kill* from "cause to die" (cf., Fodor, 1970; Wierzbicka, 1975).

Though semantic expansion seemingly takes place, such processes are pragmatically derived. In (3a) and (3b), word senses seem to be humorously or metaphorically applied to informal contexts. However, it should be noted such meanings originally occur in specific contexts. Expansion of those word senses takes place, as the meanings have been conventionalized under valid inference for a long time. So, (3a) and (3b) should be known as the transition for their pragmatics to semantics.

With respect to genres and formality, V-collocates with *be going to* have a preference for informal registers. Generally, they share a high frequency in spoken contexts rather than formal registers. Most of the formality distributions are lower than 20%, which implies that *be going to* may be commonly used in informal contexts or daily use, in comparison to *will*.

5.3.3 Comparison between *will & be going to*: *continue & happen*

Comparing *will* and *be going to*, we further found *continue* and *happen* cooccur in the top10 V-collocates lists by MI score, with high MI scores in both. For their high accessibility, this section is proposed to comparatively discuss their behaviors in both contexts. They are especially noticeable in frequency as well as MI score (see **Table 6**).

V_{LEX}	*will*	*be going to*
continue	7.42 (25,413)	5.04 (3,702)
happen	6.04 (11,608)	7.01 (17,571)

Note: Numbers in the table refer to MI Score (Frequency).

Table 6: Representation of *continue* and *happen* by MI score and frequency

Considering their MI scores and frequencies, we found *continue* is preferred in *will* contexts than *be going to* contexts, with higher MI scores and frequencies (MI: 7.42 > 5.04; Freq: 25,413 > 3,702). In contrast, *happen* has been used more in *be going to* contexts (MI: 6.04 < 7.01; Freq: 11,608 < 17,571). The numbers of *happen* is less divergent, which needs further inspection. Following the previous section, we summarized their distribution of formality to check the preferences (**Table 7**).

V_{LEX}	*will*		*be going to*	
	Informal (SPOK)	Formality	Informal (SPOK)	Formality
continue	4,348 / 17.1%	40%	2,310 / 62.4%	15%
happen	2,359 / 20.3%	24%	7,568 / 43.1%	16%

Table 7: Formality preference of *continue* and *happen* in *will* and *be going to* contexts.

V-collocates of *continue* and *happen* confirm that *will* is less preferred in informal contexts, while *be going to* bears lower formality and is used more frequently in informal contexts. Generally speaking, V-collocates with *will* and *be going to* can be sensitive to formality.

6   Conclusion

The study employs a corpus-based approach to scrutinize V-collocates with *will* and *be going to* in the present tense. Results show their V-collocates are highly associated with their semantics and pragmatics. First, MI scores manifest V-collocates with *will* are verbs with deficient meanings but higher functionality like (1) intensive verbs, (2) prepositional (or intransitive) verbs, or (3) verbs proceeded by non-finite clauses. Secondly, their V-collocates are sensitive to genres and formality. V-collocates of *will* has a wider acceptance of register than those of *be going to*; on the other hand, V-collocates with *be going to* display lower formality and higher frequencies in informal spoken context. Third, associating verb behaviors and contexts of use, we find that semantic expansion with metaphorical or hilarious use is frequently found in informal *be going to* contexts, in that figurative and analogical meanings are potentially derived from

their denotations and pragmatically manipulated to be semantic expansion. This is coordinated with the monosemous account drawn by Nicolle (1998a). By and large, the present findings reveal that V-collocates with present-tensed *will* and *be going to* are sensitive to their distinction.

Acknowledgments

References

Berglund, Y. (1997). Future in present-day English: Corpus-based evidence on the rivalry of expressions. *ICAME journal, 21*, 7-20.

Berglund, Y. (2005). *Expressions of future in present-day English: A corpus-based approach.* Acta Universitatis Upsaliensis,

Brisard, F. (2001). Be going to: An exercise in grounding. *Journal of Linguistics, 37*(2), 251-285.

Coates, J. (1983). *The semantics of the modal auxiliaries*: Routledge.

Davies, M. (2008). The corpus of contemporary American English (COCA). In.

Fodor, J. A. (1970). Three reasons for not deriving" kill" from" cause to die". *Linguistic Inquiry, 1*(4), 429-438.

Haegeman, L. (1989). Be going to and will: a pragmatic account. *Journal of Linguistics, 25*(2), 291-317.

Hofland, K., & Johansson, S. (1982). *Word frequencies in british and american english*: Norwegian computing centre for the Humanities.

Hundt, M. (1997). Has British English been catching up with American English over the past thirty years. *Corpus-based studies in English. Amsterdam: Rodopi*.

Klinge, A. (1993). The English modal auxiliaries: from lexical semantics to utterance interpretation1. *Journal of Linguistics, 29*(2), 315-357.

Leech, G. N. (2014). *Meaning and the English verb*: Routledge.

Nicolle, S. (1997). A relevance-theoretic account of be going to. *Journal of Linguistics, 33*(2), 355-377.

Nicolle, S. (1998a). Be going to and will: a monosemous account. *English Language & Linguistics, 2*(2), 223-244.

Nicolle, S. (1998b). A relevance theory perspective on grammaticalization.

Palmer, F. R. (2014). *Modality and the English modals*: Routledge.

Pollock, J.-Y. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry, 20*(3), 365-424.

Rizzi, L. (1997). The fine structure of the left periphery. In *Elements of grammar* (pp. 281-337): Springer.

Szmrecsanyi, B. (2003). Be going to versus will/shall: Does syntax matter? *Journal of English Linguistics, 31*(4), 295-323.

Wierzbicka, A. (1975). Why" kill" does not mean" cause to die": the semantics of action sentences. *Foundations of language, 13*(4), 491-528.

Yeh, T.-F. (2021). A corpus-based investigation of semantic and syntactic differences between the two major future tense constructions. *Concentric, 47*(1), 34-60.

Ziegeler, D. (1996). A synchronic perspective on the grammaticalisation of WILL in hypothetical predicates. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language", 20*(2), 411-442.

Ziegeler, D. (1997). Retention in ontogenetic and diachronic grammaticalization.