

Using Sketch Engine to Investigate Synonymous Verbs *Complete* and *Finish* in English

Xu Yuting

Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangzhou, China.

Abstract: Objective: Synonymy is an important yet intricate linguistic feature in the field of lexical semantics. **Method:** Using the 100 million-word British National Corpus (BNC) as data and the software Sketch Engine (SkE) as analyzing tool, this study examines the usage differences between *complete* and *finish*, two synonymous verbs that are often misused for their complex semantic and syntactic usage patterns. In addition to examining the collocates of the verbs, the study also investigated the syntactic patterns that the verbs typically occupy in the sentence structure and their functional implications. **Result:** The data analysis yields an informative description of the internal semantic structure of the synonym set. The results also show the need for the corpus approach to go beyond collocational analysis in the study of synonymous verbs. The limitations of using Sketch Engine to extract and disambiguate synonyms are also addressed. **Conclusion:** This paper ends by discussing the pedagogical implications that this research may have when the results are introduced into the classroom teaching.

Keywords: *Synonymy, Semantic and syntactic usage patterns, Collocational analysis.*

Article Received: 25 Feb. 2020

Revised: 20 March 2020

Accepted: 27 March 2020

Introduction

Synonym is an important yet complicate area in the field of lexical semantics. Synonyms are not completely interchangeable; rather, they differ in shades of meaning and vary in their connotations, implications, and register. There is a considerable number of synonymous words in any language among which English is particular rich in them.

This enables English speakers “to convey meanings more precisely and effectively for the right audience and context” [16]; they also constitute a thorny area for EFL (English as a Foreign Language) learners because of their subtle nuances and variations in meaning and usage. An important aspect of English linguistics is to find the proper measures of automatically identifying and extracting synonyms [8] and of distinguishing one word

from its synonyms or near-synonyms [13], [3], [11], [25], [4], [15]. Although these two areas of researching synonymous words are equally important, the main orientation of this study is methodological in that we will discover what the relative strengths and weaknesses of using Sketch Engine to research synonyms and what their relative scope of applicability is.

This paper is structured as follows. Sections 2 will give an overview of related work by introducing corpus studies of lexical semantics and then discussing corpus-based automatic extraction and identification of synonymous words. Section 3 will present corpus data and tools used in this study. The results of this study are presented and discussed in Section 4. Section 5 will summarize the major findings and

implications for future research and teaching.

Related Work

Corpus Studies of Lexical Semantics

In the field of lexical semantics, there are a number of closely related key issues such as “How do we know what words mean? What evidence do we have? Is this evidence observable and objective? How can large text collections (corpora) be used to study what words mean?” [24]. For centuries, researchers, language teachers and dictionary makers have used both their intuitions and the attested uses of words from printed books. However, it is only since the mid-1980s that corpus methods have been able to provide evidence about word meaning by searching across large text collections.

The pioneering figures in using corpus evidence to study meanings of words and phrases are called neo-Firthian corpus linguists. John Sinclair might be the first one to bring Firth’s ideas together with a corpus linguistic methodology [23]. Other important neo-Firthians include Michael Hoey, Susan Hunston, Bill Louw, Michael Stubbs, Wolfgang Teubert and Elena Tognini-Bonelli [17].

At the core of corpus linguistics is searching for the units of meaning. Words and lemmas have been viewed as the main unit of meaning for hundreds of years. However, Sinclair and his associates provide a considerable amount of evidence that units of meaning are phraseological units rather than single words. Inspired by “you shall know a word by the company it keeps”, the maxim in Firth’s work in 1957, Sinclair paid much attention to the context in which a word is used. He firmly believes in the principle of “trust the text” [20] and claims that “the language looks rather different when you look at a lot of it at once” [18].

Reading concordance and calculating collocates from corpus are two important ways to study a lexical item in its context

used by Sinclair, hence his well-cited book is entitled as *Corpus, Concordance, Collocation* [18]. For Sinclair [18], “A concordance is a collection of the occurrences of a word-form, each in its own textual environment. In its simplest form, it is an index. Each word-form is indexed, and a reference is given to the place of each occurrence in a text”.

Closely related to concordance is the notion of collocation. Firth [7] defines collocations of a given word as “statements of the habitual and customary places of that word”. Nevertheless, Firth’s research on collocation is largely intuition-based, which is in sharp contrast with most corpus linguists’ belief that the only way to reliably identify the collocates of a given word is to study patterns of co-occurrence in a corpus [4]. For example, Hunston [6] argues, “Collocation may be observed informally in any instance of language, but it is more reliable to measure it statistically, and for this a corpus is essential.”

Sinclair and associates’ early work from 1970 (reprinted in 2004) is considered a methodological elaboration on the concordance. A collocation is a co-occurrence pattern that exists between two items that frequently occur in proximity to one another-but not necessarily adjacently or, indeed, in any fixed order. Closely related to collocation is the notion of node and collocates. A node is an item whose total pattern of co-occurrence with other words is under examination; a collocate is any one of the items which appears with the node within a specified span [20].

In order to test whether two words are significant collocates, four pieces of data are required: the length of the text in which the words appear, the number of times they both appear in the text, and the number of times they appear together [22]. The optimal span is 4:4, as demonstrated in Sinclair’s [18] definition of collocation, “collocation is the co-occurrence of two or more words within a short space of each other in a text.

The usual measure of proximity is a maximum of four words intervening". On the basis of Sinclair's work, Hoey [5] defines collocation as "a psychological association between words (rather than lemmas) up to four words apart and is evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution".

In sum, Sinclair and his associates have shown that lexical items tend to occur in particular linguistic contexts, e.g. they tend to co-occur or collocate with certain other words, phrases, and/or grammatical structures, and these distributional tendencies help define their meanings. Sinclair's pioneering work has shaped contemporary research on lexical semantics, leading to experimental and corpus approaches to the synonymous words.

Corpus Approaches to Synonyms

Boosted by the computer era and the central ideas of corpus linguistics, the past decades have witnessed significant advances in the studies on synonym. Based on the Brown Corpus, Miller & Charles [9] found that the more two words are judged to be substitutable in the same linguistic context (i.e. the same location in a sentence), the more synonymous they are in meaning.

Employing a "lexical substitutability" test in a corpus study of the near-synonyms *ask for*, *request* and *demand*, Church *et al.* [5] produced the same finding: the substitutability of lexical items in the same linguistic context constitutes a good indicator of their semantic similarity. Gries [11] quantifies the similarity between English adjectives ending in *-ic* or *-ical* (like *economic* and *economical*) on the basis of the overlap between their collocations. Gilquin [9] investigates the difference between the English causative verbs *get* and *have*, Glynn [10] compares intra- and extralinguistic factors in the contexts of *hassle*, *bother* and *annoy*, and Gries & Otani [12] study the synonyms *big*, *great* and *large* and their antonyms *little*, *small* and *tiny*.

Other sets of synonyms that have attracted attention include *strong* and *powerful* [5], *absolutely*, *completely* and *entirely* [17], *big*, *large* and *great* [3], *quake* and *quiver* [2], *principal*, *primary*, *chief*, *main* and *major* [15], and *actually*, *genuinely*, *really*, and *truly* [16], *raise* and *increase* [4].

One corpus-based approach to synonym is sometimes called corpus-based behavioral profile (BP) study. Generally, a BP study uses corpus data to examine the distributional patterns of lexical items, such as the linguistic contexts a word is typically used in and the words it usually collocates with, so as to identify its unique semantic and usage patterns [4]. In recent years, the BP approach has been developed by Gries and associates [6], [11], [12], which is more sophisticated in examining both adjectives and verbs.

Although BP approach proves to be effective in studying synonyms, it might be complex for pedagogical purpose and thus it may have limited scope of application. This study, based on corpus tool Sketch Engine, aims to introduce a simple method that can be widely used by researchers, language teachers and even EFL learners.

Method

Corpus Data: BNC

The British National Corpus (BNC) is a 100-million-word collection of samples of written and spoken language from a wide range of British English from the later part of the 20th century, both spoken and written [1].

The written part of BNC (90%) includes, for example, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. The spoken part (10%) consists of orthographic transcriptions of unscripted informal conversations and spoken language collected in different contexts, ranging from

formal business or government meetings to radio shows and phone-ins.

BNC is, by nature, monolingual, synchronic, general and sample-based, in that it deals with modern British English, it covers British English of the late twentieth century, it includes many different styles and varieties instead of being limited to any particular subject field, genre or register, and that it contains many samples which allows for a wider coverage of texts within the 100 million limit.

The corpus is encoded according to the Guidelines of the Text Encoding Initiative (TEI) to represent both the output from CLAWS (automatic part-of-speech tagger) and a variety of other structural properties of texts (e.g. headings, paragraphs, lists etc.). Full classification, contextual and bibliographic information is also included with each text in the form of a TEI-conformant header.

Corpus Tool and Analysis Procedure

The Sketch Engine is a leading corpus tool, widely used in lexicography, language teaching, translation and the like [14]. Sketch

Engine is the ultimate tool to explore how language works. Its algorithms analyze authentic texts of billions of words (text corpora) to identify instantly what is typical in language and what is rare, unusual or emerging usage. It is also designed for text analysis or text mining applications. Sketch Engine has a number of core functions: Thesaurus, Wordlist, Concordance, Word Sketches and Word Sketch Difference, which will be elaborated in the following sections.

Thesaurus

In Sketch Engine, the automatic identification and summary of synonymy is achieved by Thesaurus. The synonym list may contain words which should not be included; this is a result of automatic processing. Sketch Engine cannot compare the meanings directly but determine the collocates according to a large computation. If two words share the same collocates, they will be in each other's synonym list in Thesaurus even though they are dissimilar in meaning. To avoid this, we can use a large corpus to search a word with a relatively higher frequency.

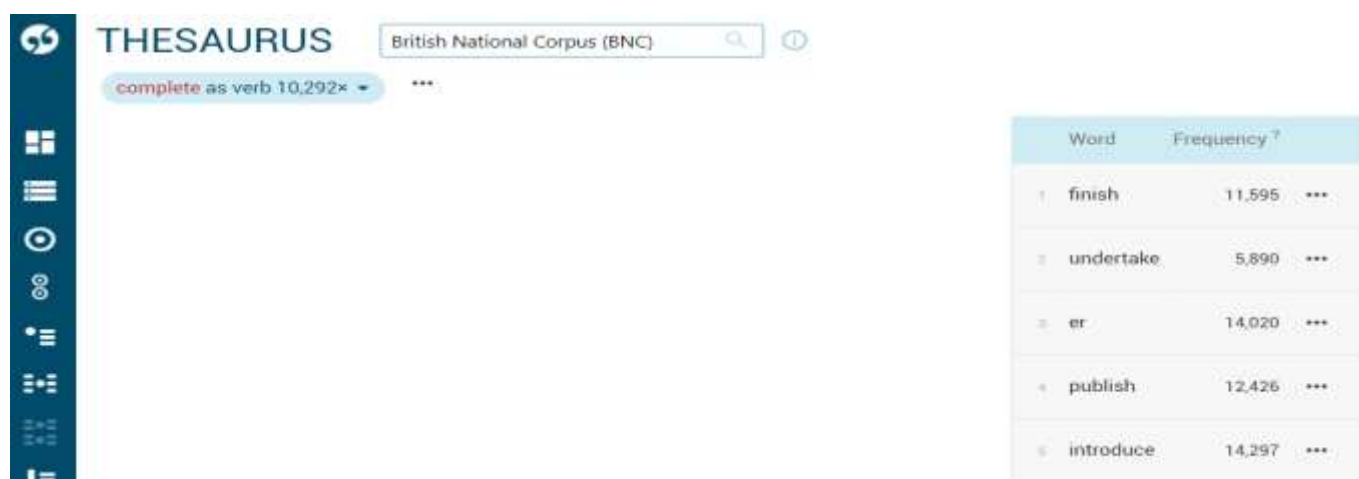


Figure 1: Top 5 synonyms of *complete* in BNC

Figure 1 demonstrates the top 5 words in the synonym list generated by Thesaurus. Obviously, only the first word *finish* has similarity in meaning with the search word *complete* while the others do not.

The word *complete* and *finish* have many collocates in common, among which the “work” in object position (*complete the work* and *finish the work*) is a piece of evidence.

Concordance

The basic method in Sketch Engine to generate concordance is to use the basic search function. However, users may want more control over the searched item. By clicking on “ADVANCED”, they can see the

options in Figure 2, and can specify the query type: simple, lemma, phrase, word, character or CQL as well as the part of speech of a lemma: noun, verb, adjective, adverb, pronoun or any (with an option to match for case).

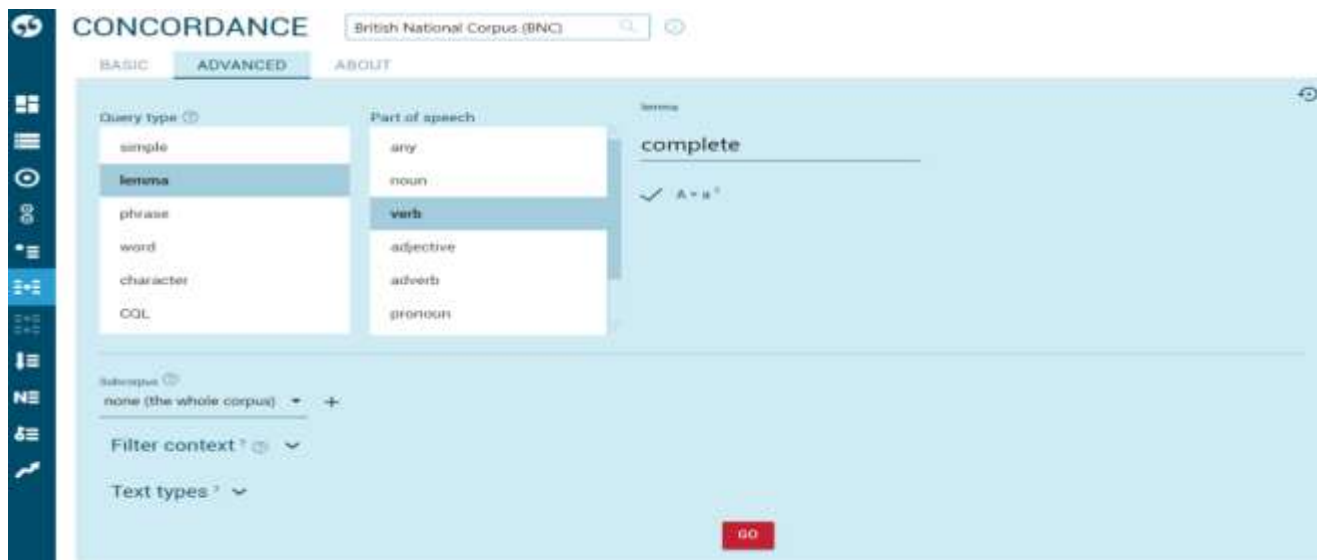


Figure 2: The advanced concordance of *complete*

If users click the button “GO”, the software

will generate the concordance lines as Figure 3.



Figure 3: Search hits for the verb *complete* in BNC

Collocations

Closely related to Concordance is Collocates. In the concordancing interface like Figure 3,

if we click the “Collocations” menu, a new interface will jump up, as demonstrated in Figure 4 below:

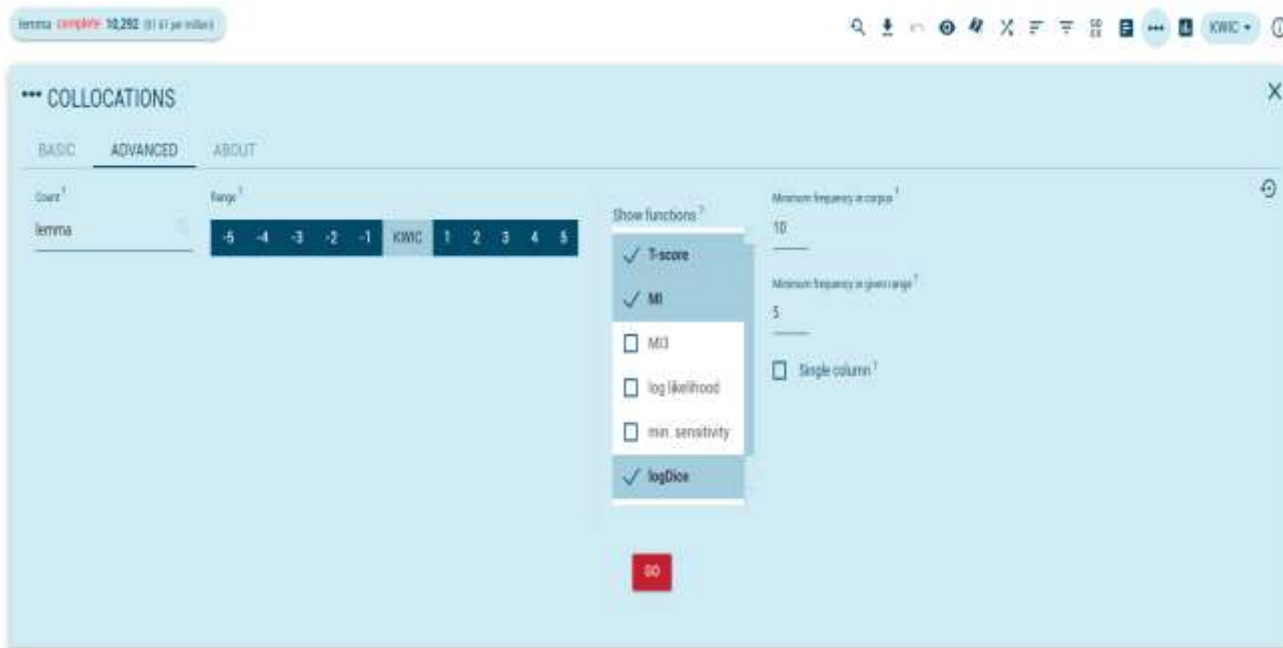


Figure 4: Collocation candidates

In “Count” we should choose an attribute that will be used. Since some collocates of *complete* may have different forms (for example, *work* and *works*). Lemma will count different word forms as one. In “Count”, we choose lemma. The span (the number of words left and right of the search word) is (-5, 5), the minimum frequency of each collocate being set 10 and minimum frequency in a given range (in our study -5, 5) 5. Of seven measures to calculate the strength of collocation (T-score, MI, MI3, log likelihood, min. sensitivity, and logDice), I choose the default one logDice which is considered more reliable than the frequently used MI (mutual information) measure.

Word Sketch

A word sketch processes the word’s collocates and other words in its surroundings. It is a summary of the word’s grammatical and collocational behavior. The results are organized into grammatical relations, such as the words that serve as an object of the verb, a subject of the verb and words that modify the verb etc. Figure 5 shows part of the word sketch for the verb *complete* in BNC. In the second column, a lot of words like *questionnaire*, *form*, *course*, *task*, *project* are classified in the category of *object*, i.e., they are used as the object of *complete*.

modifiers of "complete"	objects of "complete"	subjects of "complete"	"complete" and/or ...	prepositional phrases
successfully successfully completed	questionnaire form	end to be completed by the end of	return complete and return the	"complete" by ...
satisfactorily been satisfactorily completed	course	student	submit	"complete" in ...
recently recently completed	task	july	sign	"complete" with ...
duly duly completed	project	candidate	send	"complete" on ...
nearly nearly completed	work	mission	pass	"complete" at ...
correctly	hat-trick	december	start	"complete" within ...
	coupon complete the coupon	participant	open	"complete" before ...
		june	take	"complete" for ...

Figure 5: Word sketch for the verb *complete* in BNC

Word Sketch Difference

The word sketch difference is designed for making comparisons by contrasting collocations. This function is probably the most straightforward when researching synonymous words. Three options are available: lemma-compares the use of two

different lemmas via their collocations; word forms-compares the use of two different word forms of the same lemma via their collocates; subcorpora: compares the use of the same lemma in two different subcorpora of the same corpus via their collocates. Figure 6 presents the interface of Word Sketch Difference.



Figure 6: The interface of Word Sketch Difference

If we click the button “GO”, the software will generate a summary of the two synonymous verbs *complete* and *finish* according to the collocates arranged by the grammatical categories.

Results and Analysis

Frequencies of *complete* and *finish*

Concordance helps the researchers to compare the frequencies of the synonymous verbs in BNC. As is shown in Table 1, the synonymous words *complete* and *finish* are

quite close in term of frequencies. The two verbs under study are defined in Oxford English Dictionary as follows:

Complete: *transitive*. To bring to an end, finish (an action, performance, work, a distance, period of time, etc.)

Finish: *transitive*. To bring to an end; to come to the end of, go through the last period or stage of. Often with gerund (formerly with infinitive) as object: To ‘make an end of’, cease (doing something). †Also, *rarely*, To put an end to, cause to cease.

Table 1: Frequencies of *complete* and *finish* in BNC

Lemma	Frequency	Frequency per million
Complete	10292	91.61
Finish	11595	103.21

The results provide substantial evidence that *complete* and *finish* are two verbs that are quite comparable in BNC.

Collocates of *Complete* and *Finish*

Studying collocation is one of the major branches in corpus linguistics. As Firth [7]

puts it, “You shall know a lot about a word from the company it keeps.” Collocation is thus “a way of understanding meanings and associations between words” [3]. The software generates a list of the collocates for the searched words. We download the first 1000 rows of the collocates of *complete* and *finish* in

BNC.

The dominate content collocates of the two verbs are nouns. We summarize the top 50 nominal collocates of each verbs and the results are demonstrated as Table 2 and Table 3.

Table 2: The top 50 collocates of *complete* in BNC

Rank	Collocates	Freq	logDice	Rank	Collocates	Freq	logDice
1	questionnaire	174	8.90197	26	operation	71	6.52581
2	task	196	8.16369	27	withdrawal	33	6.43187
3	form	432	7.79109	28	mission	34	6.41504
4	project	197	7.74323	29	journey	41	6.41174
5	hat-trick	63	7.59709	30	tour	44	6.37073
6	coupon	50	7.21432	31	preparation	36	6.34417
7	transaction	65	7.18819	32	hour	89	6.20449
8	return	185	7.1318	33	sale	65	6.20414
9	course	279	7.11556	34	acquisition	30	6.18814
10	student	128	7.04068	35	purchase	39	6.18499
11	assignment	47	7.00396	36	education	66	6.16185
12	application	103	7.00099	37	thesis	26	6.14489
13	programme	123	6.8729	38	job	89	6.12988
14	training	95	6.80498	39	contract	58	6.11936
15	schedule	49	6.78588	40	construction	34	6.07659
16	picture	89	6.77904	41	assessment	36	6.06026
17	survey	67	6.74903	42	exercise	46	6.03729
18	process	127	6.73876	43	week	112	6.03164
19	sentence	63	6.68101	44	deal	75	6.01659
20	work	461	6.66664	45	review	43	6.00617
21	month	149	6.63145	46	apprenticeship	21	5.98864
22	circuit	40	6.62073	47	marathon	21	5.97539
23	cycle	43	6.59743	48	section	58	5.91983
24	degree	67	6.57756	49	research	60	5.89082
25	study	142	6.55928	50	plan	80	5.88867

As is shown in Table 2, the dominant content collocates of *complete* can be grouped into four categories:

- Learning-related activity: course, student, assignment, picture, sentence, degree, study, education, thesis, exercise, apprenticeship, section, research;
- Abstract activities with temporal

continuity: project, programme, process, circuit, cycle, operation, withdrawal, mission, journey, tour, preparation;

- Items that needs further information: questionnaire, form, coupon, application, survey, contract, assessment;
- Business and economic terms: transaction, sale, purchase, deal;

Table 3: The top 50 collocates of *finish* in BNC

Rank	Collocates	Freq	Log Dice	Rank	Collocates	Freq	Log Dice
1	meal	82	7.21808	26	hour	66	5.72638
2	sentence	84	7.00551	27	week	91	5.69896
3	race	80	6.87607	28	film	40	5.6818
4	tea	70	6.86704	29	course	105	5.67745
5	coffee	63	6.85858	30	task	36	5.63733
6	job	149	6.82887	31	bottle	26	5.61464
7	drink	78	6.68108	32	wine	26	5.53726
8	runner-up	37	6.64	33	conversation	25	5.52719
9	breakfast	45	6.5362	34	training	41	5.52669
10	lunch	39	6.23939	35	goal	30	5.47304
11	work	343	6.22749	36	school	74	5.43125
12	minute	85	6.2	37	letter	42	5.42667
13	career	45	6.15479	38	month	66	5.419
14	winner	35	6.11	39	afternoon	26	5.40711
15	tour	38	6.05517	40	night	61	5.38086
16	time	375	6.01899	41	today	46	5.36574
17	dinner	33	5.93791	42	song	22	5.32878
18	game	55	5.85669	43	tonight	22	5.32364
19	story	43	5.84542	44	painting	22	5.3009
20	championship	26	5.82038	45	year	206	5.29293
21	round	74	5.78024	46	evening	30	5.28301
22	match	43	5.74766	47	supper	15	5.24522
23	morning	52	5.73849	48	war	38	5.22899
24	book	78	5.73033	49	beer	17	5.21508
25	day	163	5.72715	50	yesterday	34	5.17392

As is shown in Table 3, the dominant content collocates of complete can be grouped into three categories:

- Food, drinks and meals: food, tea, coffee, drink, breakfast, lunch, dinner, bottle, wine, supper, beer;
- Specific time: minute, time, day, hour, week, month, night, today, tonight, year, evening, yesterday;
- Sports activity: race, runner-up, winner, game, championship, round, match;

It can be evidently observed that the verb *complete* tend to occur with nouns that refers to complex and abstract activities like learning, business, etc. In contrast, *finish* is

more frequently used in referring to the ending of specific action such as eating, drinking, reading, writing etc. More often, the action object *-ing* is omitted and the verb *finish* occur with the patient of the action verb. For example, *finish (drinking) coffee*, *finish (eating) breakfast*, *finish (reading) the book* etc.

Syntactic patterns of *complete* and *finish*

The syntactic patterns of the two verbs are based on the Word Sketch function of Sketch Engine as shown in Table 4. In order to make a fine-grained comparison of syntactic patterns of the two verbs, I summarized the 27 patterns of *complete* and 30 patterns of *finish* in Table 4 and Table 5.

Table 4: The syntactic behavior of verb *complete* in BNC

Category	Freq	Example
object	6848	ICL completed their donation of computer equipment by giving another two p.c.s
subject	1702	The University has completed its own survey (through consultants) of the backlog of repair
modifier	1462	She has recently completed the Science Foundation Course of the Open University
+ by	375	Almost every transaction can now be completed by the use of a transfer deed.
+ in	305	The information elements of the project cannot be completed in small amounts of time
and/or	271	Please complete and return before December.
+ pron	172	The head of the household is required by law to complete it
+ with	119	In 1987/88, 80% of all modules were completed with full success
+ on	89	The work was completed on schedule.
+ at	65	They are being technically completed at the moment.
+ within	53	The transaction must be completed within the current financial year
+ wh-words	53	I last week completed what I believe to be one of my best achievements.
+ before	42	This period must be completed before the application is made
+ for	32	First orders have just been completed for Air New Zealand
+ during	30	Three house conversions were started and completed during the year
+ to	22	The hospital wing has been completed to the highest standards.
+ as	21	Yet in the race to complete as many transactions as possible
+ after	17	The module was expected to be completed after 31 July 1991
+ without	17	The first part of the mowing was completed without any breakage.
+ over	16	James completed over sixty miles to raise funds.
+ until	13	The third round was not completed until early yesterday morning.
+ under	9	Projects are completed under the supervision of experienced consultants.
+ between	7	Buildings were completed between 1986 and 1989.
+ if	7	This form is to be completed if the transfer of a single entry is required.
pron subject	7	Few of them complete a life-cycle compatible with cycles of plants
complete + of	6	The building project for the centre was completed ahead of schedule
particle + object	5	Everything you want to complete your evening out at the theatre

Table 5: The syntactic behavior of verb *finish* in BNC

Categories	Freq	Example
object	3825	She wouldn't finish that sentence
pron subjects	2521	Let us finish by reminding ourselves of the traditional skills
modifier	2211	you're nearly finished with your script
subject	2005	When the song finished we were both silent.
-ing object	701	Just read on for us James please when you've finished yawning.
pron objects	509	You actually stopped her finishing it.

+ particles	428	Who would finish up as victor ludorum?
+ with	389	You have completely finished with a file or device for this program.
particles + object	357	I've got to finish off this term's work
+ in	321	Come on and we'll finish in time .
and/or	239	We can continue and try and finish the business
+ at	182	It's easier to get if an afternoon meeting can finish at a reasonable time
+ on	104	She finished on a slightly bitter note .
+ wh-words	83	The drafter's job is not quite finished when the terms are drawn up .
+ by	72	It must be finished by Friday
+ of	36	Northampton's new Visitor Centre, the start and finish of our tour .
+ adj and noun	32	Swindon finished fifth Leicester sixth... on equal points
+ as	31	It was finished as a political force
+ for	26	He was finished for the day .
+ before	17	And these, too, he hastened to finish before September .
+ to	16	It's finished to a very high standard and the hardware is equally attractive
+ until	13	Thus it did not finish until two o'clock in the morning
+ within	12	The instruction cache unit will be finished within a week
+ after	11	Of the 102 starters, 81 finished after six days of hard driving .
+ till	8	The monument won't be finished till the millennium
+ under	7	Whether lambs from the first cross to be finished under hill conditions .
+ from	6	I think that's 26 minutes from start to finish from the bare canvas' .
+ over	6	Did he not get it finished over the weekend or something?
+ about	5	We were finished about five minutes before
+ among	5	She and Genesis could well finish among the leaders .

It has to be noted that although the syntactic patterns of the two verbs are similar in many ways; there are also some differences in terms of specific collocational patterns. For example, *-ing* object only collocates with *finish* and the prepositional word *during* is only used with *complete*. In order to investigate more about the differences, we will use Word Sketch Difference of the software. The results will be presented in the following section.

Direct Comparison of Lexical and Grammatical collocates

The Word Sketch Difference function of the software allows users to visually compare and contrast synonymous words according to their salient collocational context. Figure 7 is part of the result when clicking “GO” in Figure 6. In the figure, the greener a word is, the more closely it relates to *complete*. The redder a word is, the more closely it relates to *finish*.



Figure 7: Comparison of *complete* and *finish* in terms of collocational patterns

Obviously, despite the fact that the two verbs *complete* and *finish* have a lot of syntactic patterns in common, the collocates in each pattern differ considerably. For example, in “and/or” pattern, *return* frequently collocates with *complete* but never used with *finish*. On the other hand, *complete* occurs 66 times with *finish* but only 5 times with *finish*, in which big discrepancy exists.

In the object pattern, the collocation tokens for *complete* are 6848 and only 3825 for *finish*, which indicates that there are more words used as object of *complete*. Words like *questionnaire*, *form*, *hat-trick*, *coupon*, *transaction*, *cycle*, *circuit* and *operation* only collocate with *complete* instead of *finish*. Words that refer to the food and drink like *sandwich*, *cake*, *supper*, *dinner*, *coffee*, *tea* only collocate with *finish* but are never used with *complete*.

Discussion

In the preceding sections, we have introduced how to use the corpus tool Sketch Engine to research the two synonymous verbs *complete* and *finish* in English. Each function has both advantages and disadvantages. Concordance not only enables us to look at the linguistic context of the words under investigation but also presents their respective frequency as in Table 1, which helps to judge the comparability of the words. Concordance can also make the invisible patterns visible as wisely pointed out by Tognini-Bonelli [17].

“In an individual text, we can observe neither repeated syntagmatic relations nor any paradigmatic relations at all, but it is precisely these two things which concordance make visible”. Because it gives access to many important language patterns in texts, the concordance is considered “at the center of corpus linguistics” [18].

Given the fact that the concordance of both *complete* and *finish* consist of nearly more than 10 thousands of concordance lines, it would be more valuable to generate a list of collocates which tend to occur near or next to

the verbs under investigation. Collocation plays a central role in the research of synonyms, as strongly articulated by Gries [11]: the meaning of words can be defined “in terms of their significant collocates”. Word Sketch enriches the traditional study of collocation by providing syntactic patterns and the collocates, as demonstrated in Figure 5 and Table 4 & 5. An apparent limitation is that when the words under investigation has more than one sense under a certain part of speech, the list generated by Sketch Engine would include the syntactic patterns that is irrelevant with the target meaning.

For example, one meaning of the verb *finish* under investigation is “a decorative texture or appearance of a surface (or the substance that gives it that appearance)”. In prepositional phrases summarized by Word Sketch function of the software, some phrases occur when the verb is used with the meaning of decoration and perfection.

- The process can be carried out during the normal dyeing and **finishing of fabrics** (BNC)
- Just the three parts of the project, he wrote: **the finishing of the glass**, the showing of the glass, and the end. (BNC)

These concordance lines constitute noisy information for the precise comparison of *complete* and *finish* in the study.

Word Sketch Difference may be the most straightforward method to discriminate between the synonymous words. Nevertheless, Word Sketch Difference alone is insufficient to demonstrate the semantic and syntactic features of the words under investigation.

The summary list like Figure 7 is incomplete in that many important collocates may be missing. For example, it can be observed in Table 4 & 5 that the verb *finish* tends to collocate with more precise temporal adverbials than *complete*. The noun *clock* is an important collocate of *finish*, which is not demonstrated by the summary list of Word Sketch Difference.

In addition to what has mentioned above, using Word Sketch Difference will make the users fail to take a whole look at the collocates and syntactic patterns of synonyms as a whole. For example, we may lose the opportunity to observe and classify the nominal or adjective collocates of the words under investigation, which will help us to find the semantic prosody of the target item.

In a word, while using Word Sketch Difference alone can give the researchers a quick look at the apparent differences between synonyms in the light of both collocations and syntactic patterns, it would be rewarding to examine synonymous words by using other core functions of Sketch Engine.

It has to be pointed out the software Sketch Engine itself has some limitations. One apparent limitation is its automatic extraction of synonyms or similar words. In Figure 1, some synonymous words seem to have little similarity with the searched item.

Conclusion

Despite its intricacy, researching synonym is a crucial task in lexical semantics. This paper has introduced the leading corpus Sketch Engine and its advantages in investigating synonymous words. Different functions of Sketch Engine can make different contributions to discriminate between *complete* and *finish*.

This study also has a lot of pedagogical implications. It has been noticed that students tended to confine their use of *complete* into a limited scope such as *complete this dialogue/this form*. In contrast, they tend to overuse *finish* because it tends to occur with more specific activities in our life such as food, drink, meal and some work in everyday life, which more frequently occurs in usage patterns.

Studies in language acquisition show that people memorize not only words in isolation, but also, to a large extent, groups (or chunks) of words. These chunks are viewed as building blocks of language. Thus, if EFL

teachers aim to help their students to achieve a great amount proficiency and fluency, they may hope to use examples extracted from corpus as in Table 4 & 5.

Due to the fact that English is so rich in synonymous words that the it is impossible for the teachers to teach all of them to the students. It might be more promising to teach the students how to use the Sketch Engine to investigate the synonymous words that are to their interests. This paper might provide a clear instruction to such investigation.

Acknowledgement

This work is sponsored by the Postgraduate Research Innovation Project of Guangdong University of Foreign Studies “A Corpus-based Comparative Study of English and Chinese Human Metaphor in Economic Discourse” (Project No.: 19GWCXXM-03).

References

1. Aston G, Burnard L (1998) The BNC Handbook: Exploring the British National Corpus with SARA. Edinburgh University Press.
2. Atkins B & Levin B (1995) Building on a corpus: A linguistic and lexicographical look at some near-synonyms. *International Journal of Lexicography*, 8(2):85-114.
3. Biber D, Conrad S & Reppen R (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511804489>
4. Chunyu H & Bei Y (2015) Using Sketch Engine to Investigate Synonymous Verbs. *International Journal of English Linguistics*, 5(2):29-41. <http://dx.doi.org/10.5539/ijel.v5n4p29>
5. Church KW, Gale W, Hanks P, Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical*

- Acquisition: Exploiting On-line Resources to Build a Lexicon (p. 115-164). Hillsdale, NJ: Lawrence Erlbaum.
6. Divjak D & Gries, S. Th (2006) Ways of trying in Russian: Clustering behavioral profiles. *Journal of Corpus Linguistics and Linguistic Theory*, 2(1):23-60.
 7. Firth JR (1957) A synopsis of linguistic theory 1930-1955. In *Philological Society (Eds.), Studies in Linguistic Analysis* (p. 1-32). Oxford, UK: Blackwell.
 8. Geeraerts D, Grondelaers S, Speelman D (1999) *Convergentie en Divergentie in de Nederlandse Woordenschat*. Amsterdam, Netherlands: Meertens Institute.
 9. Gilquin G (2003) Causative 'get' and 'have': So close, so different. *Journal of English Linguistics*, 31(2):125-148.
 10. Glynn D (2007) *Mapping meaning. Towards a usage-based methodology in Cognitive Semantics*. (Unpublished doctoral dissertation). University of Leuven, Leuven, Belgium.
 11. Gries S. Th. (2001) A corpus-linguistic analysis of -ic and -ical adjectives. *ICAME Journal*, 25:65-108.
 12. Gries S. Th. & Otani N (2010) Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, 34:121-150.
 13. Hanks P (1996) Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, 1(1):75-98.
 14. Kilgarriff A, Baisa V, Busta J, Jakubicek M, Kovar V, Michelfeit J, Rychly P, Suchomel V (2014) The Sketch Engine: Ten years on. *Lexicography*, 1(1):7-36.
 15. Liu D (2010) Is it a chief, main, major, primary, or principal concern? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics*, 15(1):56-87.
 16. Liu D, & Espino M (2012) Actually, Genuinely, Really, and Truly: A corpus-based Behavioral Profile study of near-synonymous adverbs. *International Journal of Corpus Linguistics*, 17(2): 198-228.
 17. Partington A (1998) *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam, Netherlands: John Benjamins.
<http://dx.doi.org/10.1075/scl.2>
 18. Sinclair J (1991) *Corpus, Concordance, Collocation*. Oxford University Press.
 19. Sinclair J (1992) The automatic analysis of corpora. In J. Svartvik (Ed.), *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991* (pp. 379-397). Berlin and New York: Mouton de Gruyter.
 20. Sinclair J (2004a) *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
 21. Sinclair J (2004b) Intuition and annotation: the discussion continues. In K. Aijmer & Altenberg (Eds.), *Advances in Corpus Linguistics* (p. 39-60). Amsterdam: Rodopi.
 22. Sinclair J, Jones S, Daley R, Krishnamurthy R (2004) *English Collocational Studies: The OSTI Report*. London: Continuum.
 23. Stubbs M (1996) *Text and Corpus Analysis: Computer Assisted Studies of Language and Culture*. Oxford: Blackwell.
 24. Stubbs M (2001) *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

25. Xiao R & McEnery T (2006) Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103-129.
26. Yves Peirsman, Y, Geeraerts D, Speelman, D (2015) The corpus-based identification of cross-lectal synonyms in pluricentric languages. *International Journal of Corpus Linguistics*, 20(1): 54-80.