A Corpus-Based Study of Near Synonyms: Should and Have To

Sobirova Firuza Urunbaevna

Samarkand state institute of foreign languages Teacher of the department of English philology

Abstract: This study was intended to examine the use of two synonymous modals, should and have to, focusing on their collexemes. Corpus-based approach, which is the study of language based on a collection of electronic texts, is a very effective way to study vocabulary because it provides a resource of natural texts and allows language learners and researchers to investigate actual usages and authentic features of vocabulary. This study utilizes the data from British National Corpus (BNC) and analyzed them following the methods of Gries and Stefanowitsch (2004)'s distinctive collexeme analysis, which enables to differentiate strong and preferable collocates of the two synonymous expressions. The findings indicate that the two modal verbs have distinctive collexemes: should prefers verbs with stative sense and have to prefers verbs with dynamic sense.

Key words: should, have to, synonym, corpus-based study, BNC, obligation force, collexeme.

INTRODUCTION

It is undeniable that knowing a large number of words is an important way to communicate better than knowing only grammar [20]. It is also well-known that English language is an international language which allows people with different native languages to interact with one another. Therefore, equipping with appropriate knowledge of English words is necessary for the global people.

Synonymy is a challenge for foreign learners because they share similar senses but are not the same in all respects and show differences in use. They can be different in registers, genders, or dialects, but it is so difficult to catch the differencesthat most people cannot discern them. Dictionary information on synonyms mostly is not enough for knowing how to use words appropriately in specific contexts, and how the styles differ.

This study aims to compare the use of synonyms, *should* and *have to*, which are known as a modal verb and a semi-modal, respectively, and figure out what exact sense they share and what differences they show in their uses.

The main research objective is to find out what are the important differences in usage of *should* and *have to*. In order to facilitate this goal, one set of questions can help to find the answers. They are concerned to figure out strong collexemes of *should* and *have to*. However, this study aims to set the senses of *should* and *have to* reflected on corpus data.

RESEARCH BACKGROUND

Research Rationale

The study of modal verbs has been one of the most frequently investigated topics in linguistics. It is because their senses are crucial in capturing the speaker's attitudes. Although there are only few modals in English, understanding their meanings is a challenge because they have several meanings and often the same modal verb can be employed to express the same meanings.

Different expressions conveying the same meaning, i.e., synonyms, have been considered to be important data in cognitive semantics. Linguists argue that the similarities and differences can manifest our cognitive structure. This study examines synonymous modal expressions of *should* and *have to* and reveals their synonymous sense and different uses.

The two expressions are conceived as synonymous in many studies because they are commonly used basically to convey the sense of obligation.

(1)   a. You *should* do as he says. [17, 227]

b. All you *have to* do is answer three simple questions, complete the following sentence. (CHO 288, BNC)

Among many obligation modals and semi modals, *should* and *have to* are considered to have weaker strength of obligation than *must.* In many studies (Quirk et al., 1985: 225; Huddleston and Pullum., 2002: 177, 186), *must* is described to have the greatest strength of obligation sense.

(2)   a. You *must* be back by ten o'clock. (Quirk et al., 1985: 225)

 b. Manson will *have to* wait five years for another hearing. (ICE-AUS S2B001208)

The two expressions are also similar in preference. In British English *should* is used as much as *have to*, even in American English *have to* is used much more often. Thus, their frequencies are reported in Table 1.

Table 1. Frequencies of the modal expressions of necessity and obligation

Modals	ICE – AUS	ICE – GB	C - US	Total
Should	1.141 pmw	1.124 pmw	850 pmw	3.115 pmw
Have to	1.311 pmw	1.244 pmw	1.385 pmw	3.940 pmw
Total	2.452 pmw	2.368 pmw	2.235 pmw	7.055 pmw

Table 1 demonstrates the frequencies of modals expressing necessity and obligation senses *should* and *have to* obtained from three corpora - the British component of the International Corpus of English (ICE-GB), the Australian component of the International Corpus of English (ICE-AUS) and a corpus of American English (C-US).

Meanings of Modals

It is a well-known fact that modal verbs, or modals, can be understood in two or more possible senses, thus, they are usually referred as modals having ambiguous senses. English modality includes a range of semantic terms such as possibility, necessity, ability, obligation, permission, and hypotheticality. The common characteristics they share is that all modals imply some kind of non-factuality, that is a situation is introduced not as a straightforward fact. The expressions that help to convey modality in English are modal auxiliaries: *can, may, will, shall, must, should, ought to, need, dare* and *used to* and quasi-modals: *have to, be allowed to, be to, and be supposed to.*

According to Bybee et al (1994) and Coates (1983) modals can express three types of senses: deontic (permission, ability, or obligation (3)), epistemic (it shows speaker's attitude towards the truth of proposition and presents the only possible conclusion or decision based on the evidence available (4)) and dynamic meanings(possible event without obligation (5)).

(3)   If all or part of the guarantee or deposit is lost or taken to pay fines or costs, you *must* pay us the lost amount immediately. (HB5 1539, BNC)

(4)   People tend to think that because many of the problems are global, the answer *must* be global. (ICE-GB W2B-013 46) (Collins 2009:39)

(5)   To get there we *must* negotiate some of the stormiest oceans in the world deadly icebergs and several hundred kilometers of pack ice. (ICE-AUS S2B-035 58) (Collins 2009: 41)

Necessity and obligation are the most common sense of modals. The modals with necessity/obligation sense may be classified into subgroups regarding their modal strength.

Regarding the strength of *must* and *should*, Collins (2009:33) claims that *must, have to, have got to, need, need to, be bound to* and *be to* are related to strong forms and that the medium strength forms involve *should, ought to, be supposed to* and *had better*. Coates (1983:60) claimed that *should* carries the strong obligation meaning while in some cases it can be weaker than *must* and greater than *may*.

Must has been drawn linguists' attention, since among modals of obligation, because it has the strongest sense and other expressions which convey the sense of obligation are always compared their strength of obligation with *must* (Biber et al., 2006:495).

Prior to the analysis, it is important to have a look at previous studies on the senses of *should* and *have to*.

Deontic *should* and *have to* may have subjective and objective meaning. Concerning subjective sense, it shows what the speaker supposes desirable and appropriate (as in (6)). As for objective *should*, it expresses that the desirability or appropriateness of the action does not depend on the speaker's support ((7)).

(6) Well maybe you *should* just let things let him think about what he's doing first. (ICE-AUS S1A-093 214)

(7) Burmese seem to be lactose intolerant and *should* not be given milk. (ICEAUS W2D-019 18)

Examining the works on *should*, *have to* and *must*, it can be inferred that all three modal expressions convey necessity and obligation senses, though they differ in the obligation strength, specifically, *should* and *have to* are considered to have weaker strength of obligation. However, *must* is the strongest obligation modal. Taking into the consideration previous research works, it is relevant to mention that previous researchers compared *must* and *have to*, and *must* and *should*, however, *should* and *have to* were not.

RESEARCH METHODOLOGY

Corpus and data selection

The main data source I used to carry out the Distinctive Collexeme Analysis (DCA) is British National Corpus. The British English (BNC) is composed of 100-million-word collection of samples of written (90%) and spoken language (10%) and it provides data from both parts.

Relating the thesauri, the dictionaries provide with useful grammatical information. Besides the definition, dictionaries provide the etymology, spelling, and the pronunciation as well as it shows synonyms. In addition, it provides a group of words with similar meanings collected together.

RESEARCH METHODS

For the exact purpose of my research, I made a query in the whole BNC including spoken and written corpora. The basic word forms for the query for each lexeme are should_{V} and {HAVE} to.

The results concerning registers were on the basis of the whole number of collocates. However, in order to examine each lexical item based on their different senses, I applied thinning function by random selection to restrict the number of hits into 3000 hits for each word. Thus, taking advantage of search function, the first phase of investigation was a query of the overall frequency of each of two constructions *should* and *have to*. Following this stage, I downloaded required data and made some calculations, in particular, general number of the verbs occurred after *should* and *have to*. The vital thing here is to prepare input data properly, since one can face some issues.

To sum up everything that was stated so far, the DCA can be divided into following steps:

1) Extracting all occurrences of a construction 1 (*have to*) from a corpus (running a concordance list), defining the frequencies of all lexical collocates.

2) Extracting all occurrences of a construction 2 (*should*) from a corpus (running a concordance list), defining the frequencies of all lexical collocates.

3) Applying calculation by Program R that demonstrates which collexemes are attracted and/or repelled by providing a collostructional strength.

4) Conducting qualitative analysis that allows getting information about functions of words or constructions and how to use them in different contexts.

DATA COLLECTION

The extraction of the target lexemes from the data package was obtained with the Key Word In the Context (KWIC) concordance list by making several queries with all the morphological variants of *have to* and *should* including contracted forms ('ve to, 've+participle 2). To provide with more variability, each token was restricted to the tenses (in particular, *have to- has to*, *had to*, *will have to*), different types of sentence, including a declarative, imperative, interrogative and negative sentence. It is relevant that while counting for *have to*, I included some particular uses, such as adverb insertion (e.g., *have* still *to*, *have* only *to*). On the contrary, the following inappropriate contexts were excluded from the analysis. Although such selecting process substantially decreases the number of instances to be analyzed, it is a necessary process for the representative description of variation patterns.

1. Unfinished or elliptic utterances:

   (8) a. I know I don't *have to,* he said calmly. (JY5 1710, BNC)

2. Sentences where conjunctions followed *should* and *have to*.

   (9) a. Ending the scandal of [consultancies and all that] the outside interests   that color, if not compromise, the activities of too many MPs *should,* but won't be, a General Election issue. (K4V 1101, BNC)

3. Sentences where pronouns or people's name followed *should* and *have to*.

   (10) a. She *had to,*Isabel thought. (HH1 6555, BNC)

4. Repetitions and some unclear statements in conversation were excluded.

RESULTS AND DISCUSSIONS

Results

As a first step, we examined overall frequency of *have to* and *should*. In the BNC *have to* occurs 83239 times and *should* 108970 times. Table 2 illustrates the distribution of the two items across spoken and written genres based on the frequency per million words.

Table 2. Distribution of *should* and *have to* across spoken and written genres

Types	Have to		Should	
	No. of hits	Frequency (pmw)	No. of hits	Frequency (pmw)
Spoken	17,147	1647.21	12,277	1179.39
Written	66,092	751.87	96,693	1099.99
Total	83239	846,67	108970	1108,39

Since I applied the thinning function with random selection of 3000 tokens and taking this into consideration, there is a necessity in some calculations, so that the frequencies of collocates correspond to the total frequency in the register. Thus, Table 3 demonstrates the real frequency of *should* and *have to* and the frequency of selected constructions in BNC.

Table 3. List of collexeme frequency

	Have to	Should
Total frequency in the corpus	83.239	108.970
Frequency of selected construction	3000	3000

Based on the selected data, we performed DCA. Table 4 lists strong collexemes of *have to*. The most attracted collexemes to the given constructions are shown in the Table 4.

Table 4. The list of strong collexemes of *have to*

Verb	Collostructional strength
Go	9.437828
Wait	8.760899
Pay	6.35543

In all of the examples below, the collexemes for *have to go*, *wait* and *pay* describe an action that in turn imply changes. These sentences describe what happened within a definite time. So, they can be related to dynamic verbs.

(11) By the time Maggie *had to go* back to London they had never felt closer in warmth, even happiness. (A6N 2249, BNC)

(12) We *had to wait* for an hour for our connection to Frankfurt. (B0U 739, BNC)

(13) This is the community charge which most people *have to pay* in the area where they live. (A0Y 970, BNC)

There are some collexemes which seem to have stative senses such as *keep*, *find, live, rely, stay*, and *share*. However, they are used with sometimes stative and sometimes dynamic.

Even though some collocates show stative senses, most collocates of *have to* can be characterized as dynamic because most of the collexemes (about 78.5 %) have dynamic sense and stative collocates (approximately 21.4 %) comprise just a few.

The so-called dynamic verbs show an action usually physical. In most cases, collexemes of *have to* describe what the subject is doing or has done, such as *go, wait, pay, work, do, make, play, move, run, give, bring*, etc. According to the findings of this work, the collexemes, which are usually attracted to the target word, appear depending on their senses of meaning.

Table 5. The list of strong collexemes of *should*

Verb	Coll. strength
Be	100.5133
Have	65.32857
Like	5.152109
Know	4.748257
Happen	4.72504

From observing of top 5 collocates of *should* from the table, most of the verbs are the verbs that convey the sense of stative verbs. In particular, they depict a state, not the action, for example stative verbs related to thoughts and opinions: *know, think;* feeling and emotions: *like, want, wish;* senses and perceptions: *be;* possession and measurement: *have*. It can be stated that most of the verbs following the modal *should*, convey the meaning related to understanding, discovering, planning, or deciding and verbs that depict states, sense, desires, possession, emotions, or opinion.

(14) a. The doctor may provide a leaflet explaining how to register the death and *should be* able to advise where to do so. (A0Y 63, BNC)

b. Furthermore, we *should have* a shared understanding of the rules of discourse about such subjects. (A1A 745, BNC)

DISCUSSION

In this section, the findings are performed and considered taking into account previous studies and some findings are those which may not have been examined. Thus, it needs to be explained why *should* prefers stative collexemes and *have to* prefers dynamic collexemes.

We can examine whether there is any possibility of grammatical features which can lead the differences. For examples, in comparison with *should*, *have to* shows some unusual feature regarding its collocates. The concordance lines together with collocates can provide useful information about grammatical features.

Firstly, *have to* can be used as the future modal verb (contracted form 'll) that actively demonstrates that it tends to appear in spoken contexts(3345.2pmw)such as conversations.

(15) So, that's what we'll*have to* do then. (KSV 4424, BNC)

Secondly, the negative form of *have to* is found relatively many times. It specifies, unlike *should*, that it appearsto convey the absence of deontic meaning (obligation) as in (16) or implies 'not-required' actions as in (17).

(16) If you *don't have to* bargain, don't bargain. (AYJ 1229, BNC)

(17) Addresses *don't have to* be mentioned, they can easily be looked up in   the electoral roll just from a name. (A16 1858, BNC)

Then, what is the crucial factor for the difference of strong collexemes. Scrutinizing data, we found that the different register preference is the most important reason for the different collexemes. We have seen *have to* is used frequently in spoken and *should* is preferentially used in written register. In spoken register, *have to* is employed to ask, order, or enforce the hearer to do something.

CONCLUSION

Modal and semi-modal expressions of *should* and *have to* have similar meaning of obligation. However, there have not been any studies comparing them explicitly and most studies mention them they have weaker strength of obligation than *must* (Coates 1983:60, Palmer 1990), even though a few studies hint *have to* has stronger obligation sense than *should* (Huddleston and Pullum 2002:177,186).

This study, applying DCA of Gries and Stefanowitsch (2003), compared uses of *should* and *have to* centering on their strong collexemes.

The result showed that *have to* prefers mostly verbs with dynamic senses and *should* prefers verbs with stative senses. Among strong collexemes, *go, wait, pay,work* and *do* are the strongest one for *have to*; *be, have, like, know*, and *happen* are the strongest for *should*. Although there are some collexemes of the two which seem to have the opposite senses, it does not challenge the generalization because the number is just a few and they are also used in terms of both senses.

The register frequencies can be seen to reflect the collexeme behaviors. *Have to* characteristically occurs in the spoken register and *should* in the written register. In spoken register, lots of dialogues are included, where the speaker can ask the hearer to do some actions on the spot. In written register, on the other hand, the hearer, usually the reader, cannot do some actions instantly the speaker asks. Hence the speaker is likely to give readers suggestions or advice which can be actualized or implemented later, rather than directions which he/she needs to comply as soon as he/she reads.

This result can be utilized to explain, if a few, the existing explanation about relative strengths of *have to* and *should*, and the description of their senses. The reason that *have to* has a little stronger sense of obligation than *should* is *have to* is related to the actions of the subject: The speaker asks the subject of the sentence to do an action. Actions by the subject (or, many times, the listener) necessarily accompanies urgency, and they should be performed without delay. This urgency

implies some strength of obligation. On the contrary, *should* is related with the state or situation where the subject is involved: The speaker suggests the subject should be in a certain state or situation. Because the subjects need not show some actions instantly, it does not accompany the sense of urgency. Lack of urgency can give the subject the impression of weak obligation.

In conclusion, collexeme characteristics of the two expressions discovered here can be regarded as reflections of their register frequencies and can explain their relative strengths of obligation.

REFERENCES

1. Baker, P. (2010). *Introduction to Corpora.* A Lecture Note. Ed. Dept. Linguistics and Modern English Language. Lancaster University.

2. Biber, D., Conrad, S &Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

3. Biber**,** D., Johansson, S, Leech, G, Conrad, S &Finegan, E. (2006). *Longman Grammar of Spoken and Written English.* Harlow: Longman.

4. Bybee, J., Pagliuca, W & Perkins, R. (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the languages of the World.* Chicago: Chicago University Press.

5. Coates, J. (1983). *The Semantics of the Modal Auxiliaries.* London/Canberra: Croom Helm.

6. Collins, P.C. (2005). *The Modals and Quasi-modals of Obligations and Necessity in Australian English and other Englishes.* English World-Wide 26:249-273.

7. Collins, P. (2009). *Modals and Quasi-modals in English.* Amsterdam: Rodopi, 33-41.

8. Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective.* Oxford: Oxford University Press.

9. Gries, S. Th &Stefanowitsch, A. (2003). *Collostructions: Investigating the interaction of words and constructions.* International Journal of Corpus Linguistics. 8(2), 209-243. John Benjamins Publishing.

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

11. Gropen, J., Pinker, S, Hollander, M, Goldberg, R & Wilson, R. (1989). *The learnability and acquisition of the dative alternation in English Language*, 65(2), 203—257.

12. Huddleston, R& Pullum, G.K. (2002). *The Cambridge Grammar of the English language.* Cambridge: Cambridge University Press, 177.

13. Leech, G., Hundt, M, Mair, Ch & Smith, N. (2010). *Change in Contemporary English: A Grammatical Study.* Cambridge: Cambridge University Press.

14. Nokkonen, S. (2006). *The semantic variation of NEED TO in four recent British English corpora*. International Journal of Corpus Linguistics, 11, 29-71.

15. Palmer, F. (1990). *Modality and the English Modals*, 2nd edn. London: Longman.

16. Perkins, M. (1983). *Modal Expressions in English.* London: Frances Pinter.

17. Quirk, R., Greenbaum, S, Leech, G, Svartvik, J & Crystal, D. (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

18. Thompson, S & Koide, Y. (1987). *Iconicity and 'indirect objects' in English.* Journal of Pragmatics, 11(3), 399—406.

19. Westney, P. (1995). *Modals and Periphrastics in English: an investigation into the semantic correspondence between certain English modal verbs and their periphrastic equivalents.* Tübingen: Max Niemeyer Verlag, 151.

20. Wilkins, D. (1972). *Linguistics in Language Teaching.* Cambridge, MA: MIT Press.

21. British National Corpus website: www.bncweb.lancs.ac.uk