Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-05566-5 OPEN

Arabic translations of the English adjective 'necessary': a corpus-driven lexical study

Rukayah Alhedayani[1]✉ & Ghuzayyil Mohammed Al-Otaibi[1]

Modal adjectives of non-epistemic necessity are very common in language corpora. However, such adjectives are expected to behave differently in context, and thus differences between them should be highlighted in dictionaries. Nevertheless, there are a few studies that have examined modal adjectives with respect to their associated constructions and meanings in English. More importantly, studies on equivalent Arabic modal adjectives are scarce. Hence, the present study is quantitative and corpus-driven utilizing monolingual (i.e., the arTenTen18 and the enTenTen18) and parallel (i.e., Open Parallel Corpus or OPUS for short) corpora. Further, it is based on construction grammar and frame semantics to explore Arabic and English words of necessity. Using distinctive collexeme analysis, covarying-collexeme analysis, and LogDice as an association measure, the results reveal that the English *necessary* has various senses but occurs more significantly in predicative and extradosed constructions. Further, the Arabic words *ḍarūrī* and *lāzim* are more associated with *necessary* as translations, and each is commonly used in a unique construction to evoke certain semantic frames. More specifically, the word ḍarūrī associates with the extraposed construction to evoke semantic frames of cognitive processes, while the word *lāzim* is attracted to attributive constructions with nouns denoting procedures and intentional acts. They are considered to be cognitive synonyms, and hence they do not alter the truth value of the expressed proposition, and they have at least one sense in common. Based on such results, it is recommended that lexicographers exploit monolingual and parallel corpora for the purpose of creating more accurate dictionaries.

[1] King Saud University, Riyadh, Saudi Arabia. ✉email: ralhedayani@ksu.edu.sa

Introduction

The availability of electronic corpora made it possible for corpus linguists to identify mutual association between lexemes (i.e., a lexeme is an abstract unit of meaning from which other related forms are derived) and certain constructions (e.g., Deshors, 2017; Kim and Davies, 2016; Stefanowitsch (2014); Wiliński, 2017, 2018, 2020, 2021) in natural discourse. For example, Deshors (2017) examined verbs employed in the construction of the progressive in different World Englishes, whereas Stefanowitsch (2014) and Kim and Davies (2016) listed collexemes (i.e., a collexeme is a word filling a specific slot in a grammatical construction) of verbs used in the into-causative construction, and Wiliński (2019, 2021) outlined the nouns occupying the slot in *be of N*-pattern and pairs of adjectives and verbs used in extraposition constructions. However, words associated with attributive, predicative, and cleft constructions and those of extraposition have received little or no attention from previous studies (cf. Collins, 1991; Wiliński, 2019). Some of the words that frequently occupy a slot in such constructions are modal adjectives of necessity. As noted by Hilpert (2014), such adjectives are very common in corpora as the British National Corpus (BNC), but each adjective behaves differently in context (Wiliński, 2021). Jäger (2018) and Wiliński (2021) reported findings on English words of non-epistemic modality including adjectives of necessity in it-extraposed constructions which leaves room for similar studies in Arabic. In addition, studies that investigated modality in Arabic mainly focused on how English modal verbs are translated into Arabic (AL-Khazraji (2023); El-Hassan, 1990; Zraigy et al., 2019) regardless of respective constructions in which they occur. Others considered verbal indicators of modality. For example, Mochón (2022) discussed two mood markers on the verb and explained their function in expressing modality. Hence, the present study begins with the English adjective *necessary* and investigates its senses, construction, and its translations in Arabic.

Modality encompasses meanings of possibility, probability, necessity, volition, and obligation (Depraetere and Reed, 2006, p. 269), and it is not only expressed through modal verbs (e.g., *can*, *could*, *must*, etc.), lexical verbs (e.g., *demand*, *require*, etc.), nouns (e.g., *belief*, *certainty*, etc.) but also through adjectives (e.g., *necessary*, *possible*, etc.) and adverbs (e.g., *certainly*, *probably*). In Arabic, modality can be expressed using verbs like *yajib*, adjectives like *lāzim*, nouns like *ḍarūrah*, and particles like *qad*, in addition to mood marking on the verb (Ryding, 2005).

Conducting collexeme analysis, Wang, Jiangping (2022) reported that *necessary* is one of the 10 attracted adjectives in it-extraposed constructions. Similarly, Hilpert (2014) stated that *necessary* is one of the frequently occurring adjectives suggesting importance that are attracted to *It is ADJ to V* constructions. Additionally, Wiliński (2019) used co-varying collexeme analysis and found that *necessary* is one of the frequent 11 adjectives occupying a slot in it-extraposed constructions. Nevertheless, much of corpus investigation considered modal verbs such as *can* (Szymański, 2021), *must*, *will*, *can*, *should*, *may* (Akinseye, 2020), and *must*, *have to*, *have got to*, *got to* and *need to* (Penry Williams and Korhonen, 2020), and modal adverbs such as *certainly* (Maíz and Arús, 2008) and *surely* (Downing, 2001).

There are a few studies exploring modal adjectives in it-extraposed constructions (Wiliński, 2019). However, except for Wiliński's (2019) study, such corpus studies focused on the diachronic evolution of the word (Maíz and Arús, 2008), examined the sociolinguistic aspect of the modal words (Penry Williams and Korhonen, 2020), and analysed modal adjectives using Talmy's concept of force dynamics (Salama, 2021). Hence, more studies are needed on English and Arabic modal adjectives of non-epistemic necessity especially with focus on nouns and verbs attracted to such adjectives in attributive and predicative positions and extraposed and cleft constructions. It has been postulated that though languages vary, findings about one construction in one language can be generalized to similar constructions in other languages because cognitive processes are similar (Boas, 2010a; Goldberg, 2013). Therefore, based on the above-advanced argument, the present study aims to answer the following:

(i) Which Arabic lexical units are commonly associated with the Arabic word of necessity (i.e., *ḍarūrī*)?
(ii) What are the different senses of the English word *necessary*?
(iii) Which constructions are commonly associated with the English *necessary*?
(iv) Which Arabic adjectives are commonly used as prototypical translation candidates for the English word *necessary*?
(v) Which constructions are commonly associated with each Arabic word of necessity?
(vi) What are the distinctive collexemes commonly attracted to each Arabic modal adjective?
(vii) What are the semantic frames of significantly attracted lexemes in each construction?

The point of departure for this study is the translation found in the Cambridge English-Arabic Dictionary (https://dictionary.cambridge.org/dictionary/english-arabic/) for the word *necessary*. The only equivalent that is provided for this word is the word *ḍarūrī*. However, the adjective *necessary* occurs in the enTenTen18 (Suchomel, 2020) 3,635,220 times with a normalized frequency of 140.76 per million words. On the other hand, the word *ḍarūrī* as an adjective occurs 79,901 times in the arTenTen (Arts et al., 2014) with a normalized frequency of 14.96 per million words. This large discrepancy in the occurrence of the two adjectives is not the only motivation for this study. An investigation of the translation equivalents of *necessary* in the OPUS corpus (Tiedemann, 2012) reveals a more convincing purpose of this study. The words *ḍarūrī*, *asāsī*, *muhim*, and *lāzim* are all adjectives that are used in the OPUS as translational equivalents of the English adjective necessary. Although the study reveals that not all these adjectives exhibit a statistically significant use as equivalents of *necessary*, it is justifiable to investigate the statistically significant ones. The Cambridge Dictionary (https://dictionary.cambridge.org/) asserts that corpora are used in the process of choosing equivalents. However, the corpus investigation presented here shows that it falls short in the case of the word *necessary*.

The present study proceeds following a number of steps. First, the Arabic lexical units related to the field of necessity are identified independently from the English word *necessary*. Second, the English word *necessary* is investigated also independently in order to identify its different senses and the constructions with which these senses collocate. The third step is to identify the translations of the word *necessary* in Arabic and determine which ones are used significantly as translations of *necessary* and to decide which Arabic words from the first step are constantly eliminated as translations of the word *necessary* which can have important implications from a translational perspective. The fourth and final step is to highlight the constructions in which the Arabic words are frequently used and align these constructions with the ones identified for *necessary* in the second step.

The paper is structured as follows: the following section dwells on four constructions in English and Arabic and discusses previous research conducted on English only as no similar studies were found for Arabic. Section 3 elaborates on the methodology adopted in this study. It presents a brief explanation of the statistical measures employed, and the data that is used in this study.

Section 4 discusses the results pertaining to the senses of the English *necessary*, the constructions it occupies, the Arabic words for necessity, and the translational equivalents of the word *necessary*, and section 5 dwells on the results of comparing the use of two Arabic equivalents of *necessary*. Finally, section 6 presents the conclusions with suggestions for further study.

Constructions hosting adjectives in Arabic and English

There are two types of adjectives in English and Arabic with respect to their positions and functions. The first is attributive adjectives, whereas the second is predicate adjectives. An attributive adjective is part of the noun phrase whether it occurs before (i.e., in a prenominal position) or after the noun (i.e., in a postnominal position) it modifies (Alotaibi, 2022). It is called so because it gives information about the qualities or attributes of the noun. The other type of adjectives is known as predicative, and it functions to give information about the subject of the sentence. In English, most adjectives can be used attributively or predicatively. Some adjectives starting with the letter a- (a-adjectives) are never used attributively. On the other hand, adjectives such as *principal*, *chief*, *main*, and *little* occur in the attributive position only. In attributive positions, the adjective occurs before the noun it modifies and after the verb if it is used in predicative positions. Hence, attributive adjectives function to modify, whereas others are for predication (Quirk and Greenbaum, 1973). Examples (1) and (2) below are illustrative of attributive adjectives in prenominal (i.e., prepositive adjectives) and postnominal (also known as postpositive adjectives) positions, respectively (Quirk et al., 1985). On the other hand, in Example (3) below, the adjective (responsible) is used predicatively because it occurs after the linking verb. Predicative adjectives can stand alone or become part of an adjective complement as in (3).

(1) We should find a responsible person.
(2) We should find him responsible.
(3) James is responsible of what has happened.

In Modern Standard Arabic, both types of adjectives are found. A predicative adjective occurs after the linking verb or the subject. It can occur without a verb if the sentence is in the present (Example 4 below). Otherwise, it needs a verb (Example 5). Semantically, there is no difference between adjectives in attributive or predicate positions. Nevertheless, functionally, an attributive adjective modifies the preceding (Example 6) or the following noun (Example 7), whereas a predicate adjective modifies the subject (Alotaibi, 2022).

(4) Al-baytu wāsi'un
DEF-house spacious
The house is spacious.
(5) Al-baytu kāna wāsi'an
DEF-house COP spacious
The house was spacious.
(6) Jā'a al-rajulu al-ṭayyib
come.PST.3SG.M DEF-man DEF-good
The good man came.
(7) Jā'a al-rajulu al-karīma-tu akhlāqa-hu
come.PST.3SG.M DEF-man DEF-good-F moral.PL-3SG.M
The man with good morals came.

The *it BE ADJ* construction is another similar grammatical pattern. In English, it consists of a projected and a projecting clause. The latter is realized by an anticipatory *it*, a copular *BE*, and an adjective, whereas the former can be filled with an *-ing* clause, an infinitive clause, or a *that* clause. The resultant clause is called hypotactic clause complex, which is commonly known as *it*-extraposition (Quirk et al., 1985). Quirk et al. (1985) noted that an *it*-extraposed construction is the result of a syntactic process that involves shifting one clause (i.e., *-ing* clause as in Example 10, an infinitive clause as in Example 9, or a *that* clause as in Example 8 below) from being in subject position to occupy that of the predicate position. The subject position, however, is filled with anticipatory *it*. In terms of structure, the *it*-extraposed construction with the infinitive verb is made of three fixed lexical units realized as (it is […] to […]) where the empty slots are filled with verbs and adjectives.

(8) It is surprising that you traveled to America.
(9) It will be difficult to take the vaccine again.
(10) It is fun playing video games with your friends.

Extra-position can be expressed using a number of constructions in Arabic. For example, personal pronouns can be used to link the subject and predicate in an equational sentence (Khan, 1984). Also, demonstrative pronouns and interrogative pronouns can be placed at the beginning of a sentence similar to English *This is what I like, What I like is this*. Most importantly for our purposes is extra-position using a preposition. The examples in (11) to (13) below illustrate this phenomenon.

(11) Yajibu an ta-dhhaba ilay-hā
Must that VOC.SG.M-go to her
You must go to her.
(12) Min al-wājibi ʿalay-ka an ta-dhhaba ilay-hā.
From DEF-obligation on you that VOC.SG.M-go to her
You must go to her.
(13) Min al-wājibi ʿalay-ka al-dhahāb ilay-hā
From DEF-obligation on you DEF-going to her
You must go to her.

Extra-position using prepositions, as exemplified in (12) and (13) above, consists of the extra-posed phrase *min*+adjective followed by either a verbal clause (example 12) or a verbal noun (example 13; Khan, 1984). As a construction, extraposed constructions are more common in corpora than non-extraposed constructions (Kaltenböck, 2005). They serve various discourse functions as they allow speakers to express their opinions as commonly accepted facts and help users convey new and given information. They also aid listeners in processing information (Quirk et al., 1985).

One more construction that hosts adjectives of necessity is known as the pseudo-cleft construction or the WH-cleft, where there is a separation between given and new information (Quirk and Greenbaum, 1973). According to Lambrecht (2001), a cleft construction is a complex one consisting of a matrix clause headed by a copular verb and a relative clause "whose relativized argument is co-indexed with the predicative argument of the copula" (p.4). Cleft constructions function to emphasise a piece of information encoded in an element that is moved from its normal position to give it *focus* (Lambrecht, 2001). In English, the construction is made of two clauses in which the first preceding the copula is a relative clause or a WH-clause, whereas the second is the complement of the copular verb and termed the head of the cleft (Delin, 1990). Example 14 is illustrative.

(14) What Tom wants to buy is a blue sweater.
Compared to *it*-cleft constructions, instead of having the pronoun it as the subject, the nominalized relative clause acts like the subject. The dummy *do* can be used to help the subject and the verb to be used in different halves (Example 15).
(15) What Jane is *doing* is taking her son to school.

The copular verb is followed by a noun, a gerund, or an infinitive verb with or without *to*. The head of the cleft can be

introduced by *that* or different WH-words. As maintained by Lambrecht (2001), the order of the clauses in a cleft construction can be reversed (Bara, 2010). In general, the cleft construction is of three parts: (a) the cleft constituent, which is the focus of the construction, (b) the copula (be), and (3) the cleft clause, which contains the remaining parts of the construction (Calude, 2008).

In Arabic, clauses in cleft constructions are introduced by relative pronouns such as *huwa*, *alladhī*, *hiyā*, *allatī*, *humā*, *alladhān*, *allatān*, *humm*, *alladhīn*, *hunn*, *al-lawāti*, *al-lātī*, and *mā*. The following sentence is an example.

(16) alladhī ḍaraba ʾaḥmad huwa ʿalī
Who.3SG.M hit.SG.M.PST Ahmed he.SG.M Ali
Who hit Ahmed is Ali.

Arabic allows for focus on time, but this is not possible in English. Arabic Time words are embedded in the two clauses before relative pronouns. The following are examples.

(17) *When he ate was two hours ago.
(18) al-ʾusbūʿ alladhī sāfar fīh huwa al-ʾusbūʿ al-māḍī
DEF-week that.SG.M travel.PST Prep he.SG.M DEF-week DEF-past
The week he traveled was last week.

Previous research on extraposed and nonextraposed constructions

Generally speaking, extraposition has been the focus of a few research papers (cf. Kaatari, 2010; Kaltenböck, 2005; Hilpert, 2014; Quirk et al., 1985). Some compared it with other grammatical patterns such as clefts (Calude, 2008) or the non-extraposed predicative construction (Hilpert, 2014). Some explored its discourse functions (Hewings and Hewings, 2002), whereas others examined the association of deontic or epistemic adjectives with extraposition (Biber et al., 1999; Van Linden, 2012). As for constructions with predicative and attributive adjectives, a few studies focused on the semantics of predicative adjectives (Chierchia, 1985), the distribution of predicative and attributive adjectives in spoken English (Englebretson, 1997), etc.

On the other hand, there are a few studies that address the grammatical features and functions (Hobi, 2011; Alotaibi, 2022) of equivalent constructions in Arabic or examined them from syntactic and semantic perspectives (Al-Sharifi and Sadler, 2009). Thus, more studies are needed to explore collocational preferences of such constructions in Arabic in relation to deontic adjectives. This section, reviews studies on English constructions that have a scope similar to that of the present study.

Hilpert (2014) used collexeme analysis and more specifically covarying-collexeme analysis to find out words that are attracted to the adjective and the verb slots in it-extraposed and nonextraposed constructions. Then, he categorized resultant adjectives using Fillmore's Frame Semantics. Utilizing the BNC, Hilpert (2014) found that *necessary* is one of the nine strongly attracted adjectives in the *it's ADJ to V* construction besides others suggesting IMPORTANCE such as *essential*. Other strongly attracted lexemes are adjectives denoting POSSIBILITY (e.g., possible, impossible) and ADVISABILITY (e.g., advisable, better). Such adjectives happened to be among the most frequent adjectives in the corpus. The researcher then compared the extraposed construction with that of the non-extraposed predicative adjective. Results showed that *necessary* was not one of the top adjectives attracted to the non-extraposed construction. The construction attracts adjectives of truth values (e.g., right, true) and other adjectives (e.g., dead, successful) that referentially describe nominals. Hence, such adjectives are not associated with the extraposed construction and cannot describe the anticipatory *it*. Results of covarying-collexeme analysis of the *it*-extraposed construction revealed that the construction favored combinations (e.g., easy to see, true to say, interesting to note, etc.) that are used to introduce a new piece of information. In other words, they do not carry any focal information in themselves.

In another study, Hilpert (2013) referred to collostructional change where a change in meaning, form, function, distribution, and frequency affects the construction. Mair (2006) reported that the change affects a node in a network of constructions, and it proceeds to affect similar nodes in other constructions and eventually groups of constructions. One example of collostructional change is when one variant of a construction becomes more frequent than the other as revealed by frequency measures and inferential statistics. With some time, the prevailing variant became the prototype of the construction. This phenomenon became known as *layering* only if the two constructions express the same grammatical meaning (Hilpert, 2013; Hopper, 1991). This can be traced through using diachronic corpora or dictionaries. For instance, Mair (2006) noted that in British and American English the get-passive became more frequent than other variants expressing the passive, as it is commonly used by young speakers (i.e., English speakers of the age group 14 to 24 years old) and revealed by scores of relative frequencies. However, such changes are not meant to fill functional gaps (Traugott (2003b)), and such constructions tend to cluster especially in reference to the grammatical meaning of deontic modality (Hilpert, 2013). In some cases, as part of host-class expansion, some constructions become very productive to accommodate more elements of one or more syntactic categories (Hilpert, 2013) as in the use of adverbial or prepositional phrases functioning as complements of the copula in *it*-clefts (e.g., It was *in Manchester* that she met her brother; Patten, 2010). As part of change, other constructions tend to appear in specific genres or develop to be used commonly in the formal register (Mair, 2006).

Another relevant study is by Wiliński (2019) who based his study on Usage-Based Construction Grammar (Goldberg, 1995, 2006, 2013) and Filmore's Frame Semantics (1982) and utilized co-varying collexeme analysis (Stefanowitsch and Gries, 2005) to determine strongly attracted and loosely associated combinations of verbs and adjectives in extraposition constructions. He compared the construction to the non-extraposed one that starts with an infinitive clause (i.e., *To find a job here is impossible* compared to *It is impossible to have a job here*). Using the Corpus of Contemporary American English (COCA) and applying Fisher exact test, results indicated that combinations evoking semantic frames of IMPORTANCE and its relationship with frames of BEOMING AWARE, REMEMBERING INFORMATION, STATEMENT, (e.g., important to recognize, important to remember) dominate the extraposed construction. Additionally, co-varying collexeme analysis showed some of the repelled combinations such as *important to imagine*, *difficult to have*, etc. The researcher reported that such combinations are expected to be repelled since the construction shows preference for verbs introducing new information.

As shown above, the two studies by Hilpert (2014) and Wiliński (2019) concluded that the extraposed construction mainly attracts words belonging to the semantic frame of IMPORTANCE such as *necessary*. Further, collexemes that are strongly associated with *necessary* are those used to introduce information such as *note*. Since there are no similar studies on Arabic adjectives, this study aims at bridging this gap. The following section discusses the theoretical and the methodological perspectives this paper adopts.

Theoretical and methodological background

Theoretically, the paper is based on Goldberg's (1995, 2006, 2013) Usage-Based Construction Grammar and Frame Semantics (Fillmore, 1982; Fillmore and Atkins, 1992; Fillmore and Baker 2010). This section reviews both theories and presents the statistical measures used in this study. It also discusses the sources used for data collection.

Construction grammar. Proponents of Construction Grammar postulate no separation between lexicon and grammar. Thus, constructions found at various levels of abstraction from the English plural morpheme up to grammatical constructions such as extraposed constructions, idioms, etc. are mainly pairings of meaning and form (Goldberg (2009)). Semantics is directly linked to the surface form (Goldberg, 2002). Different constructions are placed on a continuum of lexicon and syntax (Fillmore et al., 1988; Goldberg, 2003), but they are linked together to form a network (Fillmore et al., 1988; Goldberg, 1995). Further, lexemes prefer certain constructions, and the opposite is true. Therefore, this theory places a lot of emphasis on the frequency of usage or occurrences of items in constructions and thus the term "usage-based." Collocational analysis (i.e., identifying collocates or words accompanying other words) is central to any constructional study. Nevertheless, using corpus linguistics, it has been proven that relative frequency counts, not raw frequency counts, of frequently occurring items in a construction can provide an answer to which lexemes are significantly associated with which constructions or unlikely to appear in other constructions (Goldberg (2009)). A lexeme that is attracted to a construction is called a collostruct, and a construction that is associated with a lexeme is called a collostruction. Collostructional analysis is an extension of the traditional collocational analysis, and thus the term collostruction is based on construction and collocation (Stefanowitsch, 2013). Collostructional analysis of constructions stems from the fact that the meaning of a construction tends to harmonize with the meaning of the individual linguistic units that are typically used in it. This is known as the Semantic Coherence Principle (Goldberg, 1995). For instance, a verb is used appropriately in a construction if its arguments correspond with the roles specified by a certain construction. Hence, one of the basic tenets of Construction Grammar is to analyse mutual associations between constructions and component elements for the purpose of understanding the meaning of the construction. This distinguishes collostruction studies from collocation-based studies because the former are based on quantification and employ inferential statistics besides the fact that they are theory-driven focusing on the interface between grammar and lexis (Stefanowitsch, 2013). More notably, collostructional analysis is a general term for various types of analyses. The types that will be explained below will be used in the study.

(i) Distinctive collexeme analysis (Gries and Stefanowitsch, 2004). This is used to contrast between two constructions that are considered to be similar or variants of the same construction through identifying words typically associated with each respective construction. For example, one can identify typical verbs associated with active or passive constructions. Also, this analysis measures the strength of attraction or repulsion between lexemes and two construction variants with similar meanings. Hence, lexemes are considered "distinctive" to some constructions. One important measure used for such associations is Fisher exact test which assumes no normal distribution of data or certain sample size. Therefore, this test works well with items of low frequency (Zipf, 1965). As a measure of collostructional strength, resultant p-values associated with the test can be higher to indicate a stronger mutual attraction.

(ii) *Covarying-collexeme analysis* (Gries and Stefanowitsch, 2004b; Stefanowitsch and Gries, 2005). This is useful in identifying mutual dependencies of two lexemes occupying two different slots in a certain construction. For example, in *it*-extraposed constructions followed by infinitive clauses, one slot is occupied by an ADJ and another by a verb. Covarying-collexeme analysis can yield important information on ADJ-Verb combinations that are significantly used in the construction. This helps one draw conclusions about semantic frames commonly associated with the construction.

(iii) Another association measure that can be exploited is LogDice whose score ranges between a fixed maximum of 14 and 0. The given score does not depend on the total size of the corpus. This makes LogDice beneficial in comparing between different corpora (Rychly (2008)).

Frame semantics. Charles Fillmore's Frame Semantics Theory is a cognitive theory that links meanings of words to the syntactic context in which they occur (Atkins et al., 2003: 254). The theory of Frame Semantics is based on the assumption that every experience is associated with a meaningful context, and one's ability to remember such experiences stems from the fact that humans possess mental schemes that give meaning and order to surrounding objects, events, and relations. A frame is a representation of an event or situation in which each has a number of elements known as frame elements. For Fillmore (1982, 1985b), grammatical meaning, though abstract, is realized through semantic frames, as it adheres to the speech situation it represents (Bybee, 2013). Based on this assumption, Filmore argued that words are learned in their meaningful contexts, and such contexts aid in comprehension when some experiences are evoked, and this eventually enhances learning of words. Fillmore further identified a number of frames in which each represents an experience with its related entities and events (Fillmore, 1976). Fillmore acknowledged how Case Grammar (i.e., semantic roles representing situation types; Fillmore, 1968) is relevant to Frame Semantics (Fillmore, 1982) and how *frame*, as a cognitive device, plays a critical role in Frame Semantics compared to semantic roles which can be defined in relation to a semantic frame (Fillmore, 2008).

As part of Fillmore's (1967) work on the grammar of some synonymous verbs (i.e., sick vs. ill) and other related verbs (e.g., hit vs. break), he referred to the fact that some verbs can be used in attributive as well as predicative positions, but some are only used predicatively. This knowledge pertaining to the use and grammar of synonymous verbs never appears in standard dictionaries. Fillmore emphasizes that the meaning of verbs is better understood in light of both semantics and syntax. Fillmore's (1968) contribution to argument structure and Case Grammar (i.e., where formal pattens have been assigned semantic roles) paved the way for the emergence of Frame Semantics and Construction Grammar and their various applications (e.g., Hilpert, 2014; Wiliński, 2019) in the investigation of similar constructions. Fillmore emphasized the universality of Frame Semantics and its usability for every language through utilizing a number of semantic frames in place of semantic roles (Fillmore, 2006). As for Construction Grammar, the goal of introducing the theory is to give an account of idiomatic structures besides regular syntactic ones. Constructions are treated as words that have different forms and meanings and serve various functions. Examining the *let alone* construction, Fillmore et al. (2014) notes

that it is very productive, and it can contribute various pragmatic and semantic meanings besides its syntactic information.

We used FrameNet (LU index | fndrupal (berkeley.edu)), an English database, that includes annotated examples of how a certain lexical unit (LU) is used in different contexts. The premise of proposing such a theory is that LUs are best defined through semantic frames (Ruppenhofer et al., 2016), and LUs are grouped together if they evoke the same frame. The Project is meant to organize the English lexicon (Fillmore, 1982). Special attention is given to frame-evoking words which can be verbs, nouns, adjectives, or adverbs. Following a FrameNet analysis, one can identify the specific frame that a LU belongs to after reading the description of the frame and the situation it represents along with its typical adjuncts, arguments, and participant roles (i.e., frame elements). For example, through the semantic frame of DOCUMENTS, one can define nouns like *license*, *diploma*, *contract* and adjectives such as *contractual*. Words belonging to the same frame are listed together (Fillmore et al., 2003). Frames are connected to one another forming a network in the semantic space. LUs can be described hierarchically in terms of super- and sub- frames. The total number of frames is 1200, and the website includes about 200,000 annotated sentences (Ruppenhofer et al., 2016).

Since a similar project is not available in Arabic, the English Frame Net was used in the study. Frames are sometimes modified in order to be adapted to the specific meanings of Arabic collexemes. There are a few initiatives on Arabic frames (cf. Abdul-Baquee and Atwell, 2009; Alhedayani 2016; Al-Qahtani 2005; Gargett and Leung, 2020), albeit they are less developed. Hence, the English FrameNet based on the British National Corpus (BNC) was used as some proved its usability for Arabic (cf. Abdul-Baquee and Atwell, 2009; Alhedayani, 2016). Yet, the researcher may face a few problems when they translate a collexeme including the fact that some words are homonymous (i.e., words sharing the same form but give two distinct meanings in two different sentences), polysemous (i.e., a word that gives multiple, similar meanings; Hilpert, 2014), or used metaphorically (Abdul-Baquee and Atwell, 2009). For instance, it is very common that *hard* is used to mean DIFFICULT more than SOLID. Hilpert (2014) reported that polysemous words are at a disadvantage in collexeme analysis because manual coding is not always easy. Hence, we selected the common sense of the lexeme after checking concordances. For example, the collexeme (*murāja'ah*), that commonly occurs with (*min al-ḍarūrī*) in extraposed constructions, can mean to be sent to a physician or a specialist (i.e., *murāja'at al-ṭabīb*) as part of the semantic frame of SENDING or to review (i.e., *murāja'at al-qanūn*) motivated by the ASSESSING frame. Thus, both frames were considered. Further, the meaning of the lexeme can be affected by its inflections (i.e., morphemes attached to words to convey information on tense, gender, number, voice, etc.). For instance, if *bayān*, as in *Al-bayān al-lāzim li-mulābasāt al-ḥādith* ("the necessary statement of the circumstances of the accident"), is pluralized (*al-bayanāt al-lāzimah*), it will be related to the semantic frame of INFORMATION. However, if it is in the singular form, it will be associated with the STATEMENT frame. Yet, the plural form of the lexeme is more common than the singular one. As reported by Fillmore and Atkins (2012), FrameNet accounts for the meanings of polysemous words. On the other hand, dictionaries may list only familiar meanings of words and treat various meanings of a polysemous word as having different senses.

Data. We drew on a variety of resources to obtain the data for this study. Arabic and English words of necessity have been established using Almaany Arabic-Arabic dictionary and the Thesaurus Function in SketchEngine. The monolingual Arabic and English data were obtained from the comparable corpora arTenTen18 (Arts et al., 2014) and the enTenTen18 (Suchomel, 2020), respectively. The TenTen corpus family includes several text corpora that are created from the web and that are prepared according to the same criteria (Jakubíček et al., 2013). The enTenTen18 corpus contains 21.9 billion words that are part-of-speech tagged (i.e., marked for part of speech) and lemmatized (Suchomel, 2020). The arTenTen18 (Arts et al., 2014) is also part-of-speech tagged and lemmatized, and it contains 4.6 billion words. Also, a parallel corpus is used to investigate the equivalents of the word *necessary* in Arabic. There are two parallel corpora available at the SketchEngine web application: the United Nations Parallel Corpus (UNPC, Lison and Tiedemann, 2016) and the Open Parallel Corpus (OPUS; Tiedemann, 2012). We used the OPUS parallel corpus as it is more general in its genre (i.e., different types of writing identified by certain conventions) coverage than the UNPC. The OPUS parallel corpora (Tiedemann, 2012) provide text corpora that are aligned on the sentence level to allow searching for translations. It contains 290.9 million aligned tokens (i.e., the total number of words in a corpus). The downside of this corpus is that it is not lemmatized nor is it tagged for part of speech. It is only tagged by shallow tagging that distinguishes numerals and punctuations from content and other words.

To overcome the shortcomings of the OPUS corpus, search terms are chosen in a way that allows for the extraction of all inflections of a certain lemma. For example, the search terms should reflect all inflections of adjectives in Arabic: number, gender, case, and state. Another measure that was taken to manipulate the Arabic results is that they are downloaded as .csv files and searched through using a spread sheet editor.

Results and discussion

This Results section is divided into two sub-sections. In the first sub-section, the results are given in the order reflecting the procedure adopted for data collection and data analysis. Hence, the first sub-section reports the results found on the different senses of the English and Arabic words of necessity and the constructions hosting various English words of necessity. In addition, the translation equivalents for the English word *necessary* using a parallel corpus. Then, a covarying collexeme analysis was conducted to reveal results on the collocates of the various translation equivalents based on data drawn from a monolingual corpus. After that, a distinctive collexeme analysis was performed using a parallel corpus to decide on stronger translation candidates for *necessary*. The Arabic monolingual corpus was employed again to identify collocates of nouns for the Arabic translation equivalents in the attributive position to decide on to what extent such Arabic equivalents are similar in meaning. A distinctive collexeme analysis was utilized one more time to find out the constructions that are typically associated with each Arabic translation equivalent. On the other hand, the second sub-section mainly focusses on outlining results pertaining to the semantic frames of collocates of nouns accompanying each translation equivalent using parallel and monolingual corpora. Because of the scarcity of similar research in Arabic, the results are compared to those of English studies.

Being "necessary" in Arabic. In order to identify the lexical units related to the field of necessity in Arabic, we looked up the word *ḍarūrī* in Almaany dictionary [1] which yields the words *ḍarūrah*, *ḥājah*, *lāzim*, *muhim*, *asāsī*, *wājib*, *mu'akkad*. We then used the Thesaurus function in SketchEngine for the corpus arTenTen

which revealed that that the word *ḍarūrī* is similar in its meaning to the words *tālī*, *muhim*, *mumkin*, *lāzim*, *khāṣ*, *muḥaddad*, *asāsī*. Both sets of words include the words *lāzim*, *asāsī*, *muhim*. Using the same function in SketchEngine, we identified the synonyms for each of these words. These words are listed in Table 1.

Looking at the synonyms for each of these three words, their subtle differences arise. The word has *ḍarūrī* and *muhim* listed as its synonyms but also words like *possible*, *consequence*, and *appropriate*, which indicates that this word activated a semantic frame related to resulting. On the other hand, the word *asāsī* evokes frames of priority and primal in addition to importance as it is found to be a synonym of *muhim* and *ra'eesi*, but it also evokes frames of specificity as words such as *muḥaddad* and *muʿayyin*, which are found to be synonyms of it. The synonyms for the word *muhim*, however, do not include any of the other words *asāsī*, *lāzim*, and *ḍarūrī* but includes words of unrelated meanings.

Different senses of the English word 'necessary'. The Thesaurus in SketchEngine enTenTen18 reveals the synonymous words of the English word *necessary* which are *essential appropriate, useful, important, sufficient*. The subtle differences of these synonyms can indicate the different senses the word *necessary* has. To unravel these subtle differences, different constructions typically hosting the word *necessary* are investigated next. This part of the investigation is based on the assumption that different senses of a particular word tend to be attracted to different constructions (Fillmore and Atkins, 1992; Gilquin, 2010; Gilquin, 2013). The word *necessary* was found in the literature to occur in a number of constructions, such as the what-cleft construction (Wiliński, 2022; Jäger, 2018), *it*'cleft construction (Jäger, 2018), attributive construction (Hilpert, 2014; Jäger, 2018), and predicative construction (Hilpert, 2014; Jäger, 2018). An extra construction was also recognized in the data is *NP makes it necessary to V*. Table 2 shows the frequency of each construction in the enTenTen18, along with the frequency of the construction when hosting the synonyms of the word 'necessity'.

We preformed collexeme analysis which showed that the word *necessary* collogates with these constructions significantly. However, when a distinctive collexesme analysis was performed on each of the words with *necessary*, the results showed that *necessary* collogates with all constructions significantly except for the attributive construction which collogates significantly more with all other words than with the word *necessary*. An exception for this result is the word *important*. The following is the results table for this word.

Table 3 shows that the word *important* collogates with the attributive use more than *necessary* similar to all other words investigated here. Unlike other words, the word *important* collogates more with the construction *What is important is* more than the word *necessary*.

Translational equivalents of the word "necessary.". The Cambridge Dictionary provides one Arabic translation for the adjective *necessary*, i.e., *ḍarūrī*. However, when we used the OPUS English-Arabic parallel corpus (Tiedemann, 2012) to find out translations of the word *necessary* in Arabic, twelve translation candidates in different parts of speech were found. These candidates are listed in Table 4 along with their frequencies.

Table 1 Synonyms for Arabic words of 'necessity' as suggested by the Thesaurus Function in SketchEngine.

Arabic words of necessity	Synonyms
lāzim	ḍarūrī, mumkin, muhim, tālī, khāṣ, mulāʾim
ḍarūrī	lāzim, muhim, mumkin, tālī, khāṣ, muḥaddad, asāsī
muhim	tālī, akthar, mumkin, awwal, murtabiṭ, mukhtalif, khāṣ
asāsī	ra'eesi, muḥaddad, hām, muʿayyin, tālī, muhim

Table 2 The frequency of the constructions hosting the synonyms of the English 'necessary' in the enTenTen18 and the frequency of each construction in the Corpus.

	frequency	necessary	essential	appropriate	useful	important	sufficient
Adj_N	1,660,243,545	877,421	1,049,593	1,559,430	837,156	4,443,511	597,877
N_be_Adj	32,387,753	327,188	324,785	81,549	120,618	498,363	75,413
It_be_Adj_to	4,236,482	176,471	55,271	19,017	31,361	680,997	10,446
What_be_Adj_is	90,080	947	798	67	93	14,151	11
NP_makes_it_Adj_to_V	152,269	2536	218	113	30	331	4
NP_made_Adj	712,511	1903	372	804	362	2936	265

Table 3 The results of the distinctive collexeme analysis of the constructions collogating significantly with the word *necessary*.

COLLEX	O.CXN1	E.CXN1	O.CXN2	E.CXN2	ASSOC	COLL.STR.LOGL	SIGNIF	SHARED
N_be_Adj	327188	162891.5	498363	662659.5	necessary	200880.22046	*****	Y
NP_makes_it_Adj_to_V	2536	565.7	331	2301.3	necessary	6329.26264	*****	Y
NP_made_Adj	1903	954.8	2936	3884.2	necessary	982.37214	*****	Y
It_be_Adj_to	176471	169189.1	680997	688278.9	necessary	440.67889	*****	Y
What_be_Adj_is	947	2979	14151	12119	important	2220.26495	*****	Y
Adj_N	877421	1049885.9	4443511	4271046.1	important	135847.58052	*****	Y

O.CXN observed construction, *E.CXN* expected construction, *ASSOC* association, *COLL.STR.LOGL* collexeme loglikelyhood, *SIGNIF* significance.

Table 4 shows that there are four adjectival forms, the words *muhim*, *ḍarūrī*, *asāsī*, and *lāzim*. To determine the level of attraction or repulsion between the collocates of the four adjectives, we performed a covarying collexeme analysis (Gries and Stefanowitsch, 2004) which was done using the R collostructions package (Flach, 2021). The statistical measure used in this analysis is the Fisher-Yates Exact Test (Pedersen, 1996) because it works well with low frequencies. These are listed in Table 5.

Table 5 lists the frequencies of both slots of the construction N ADJ in the translated corpus in addition to the observed and expected frequencies. The next column shows the association between the two slots. The last three columns show the collocation strength, the Fisher-Yates Exact Test results, and the significance level of the association, respectively. One can observe that some words are attracted significantly to a construction. For example, the word amr "subject matter" is attracted to the adjectives *ḍarūrī* and *muhim* but not to *lāzim*. Also, the word sharṭ "condition" is attracted to *ḍarūrī* and *asāsī* but is repelled against the adjective *lāzim*. The attracted/repelled words reveal that there are subtle differences between these synonyms. The word amr "subject matter," for example, collocates with muhim, while the word sharṭ "condition" collocates with *asāsī* "essential." The fact that both collocate with ḍarūrī indicates that this word evokes frames associated with both importance and essentiality.

Finding four adjectival translation candidates for the word *necessary* does not mean that they all should be included in the dictionary because translations of lexical items are based on context. Therefore, we performed a distinctive collexeme analysis (Gries and Stefanowitsch, 2004) using the "collostructions" R package (Flach, 2021) in order to identify the ones that can be considered stronger candidates out of all twelve translation candidates. This measure illustrates whether these words associate with being translations of *necessary* or other constructions in the translation corpus. Table 6 shows the results obtained from this calculation.

Table 6 shows that the word ḍarūrī as an adjective is used as a translation of the word *necessary* rather than other words in the source corpus. This explains why this word appears in the Cambridge Dictionary as a translation of the word *necessary*. However, a similar result was found for the three parts of speech of the word *lāzim*. The other translation candidates did not associate with being translations of the word *necessary* including the nominal form *ḍarūrah*. This finding preliminarily suggests that the word *lāzim* is a stronger candidate as a translation for *necessary* than the word *ḍarūrī*, but further investigation is needed.

Hence, because the two adjectives *asāsī* and *muhim* significantly associate with being translations of other words in the parallel corpus, we eliminated them as possible translations for the word *necessary*. We limit our search to the adjectival forms of the words *ḍarūrī* and *lāzim*. We wanted to observe how these two adjectives behave in context. Therefore, we turned to an Arabic monolingual corpus, the arTenTen18 corpus, to investigate the

Table 4 Translations of the word *necessary* in the OPUS corpus.

translation	POS	overall frequency	frequency as a translation of "necessary"
ḍarūrī	adjective	68739	37,369
ḍarūrah	Noun	78900	8045
lāzim	Adjective	124583	46,331
lazam	Verb	37282	10,356
lazūm	Noun	3247	1046
yahtāj	verb	4448	227
hājah	Noun	98741	2777
iqtiḍā'	Noun	38313	6587
lā bud	Verb	44562	1648
asāsī	Adjective	51409	310
yaqtaḍī	Verb	18942	1411
muhim	Adjective	35498	83

Table 5 Covarying collexeme analysis using the Fisher-Yates Exact Test on different collocates of the four synonyms.

SLOT1	SLOT2	fS1	fS2	OBS	EXP	ASSOC	COLL.STR.FYE	P.FYE
al-tadābīr	ḍarūrī	65	154	65	21.5	attr	36.07294	8.45E-37
amr	ḍarūrī	37	154	36	12.3	attr	17.17214	6.73E-18
ijrā'āt	lāzim	81	305	81	53.1	attr	16.7011	1.99E-17
al-mawārid	lāzim	60	305	55	39.4	attr	6.14014	7.24E-07
al-tashīlāt	lāzim	26	305	26	17.1	attr	4.92897	1.18E-05
al-ẓurūf	lāzim	22	305	22	14.4	attr	4.14716	7.13E-05
al-irādah	lāzim	21	305	21	13.8	attr	3.95309	0.00011141
al-tartībāt	ḍarūrī	8	154	8	2.6	attr	3.89344	0.00012781
al-taghyīrāt	ḍarūrī	7	154	7	2.3	attr	3.39989	0.0003982
sharṭ	ḍarūrī	16	154	12	5.3	attr	3.22071	0.00060157
al-khuṭuwāt	lāzim	79	305	64	51.8	attr	3.07719	0.00083716
al-wasā'il	ḍarūrī	6	154	6	2	attr	2.90834	0.001235
ṣammām	muhim	1	2	1	0	attr	2.36642	0.0043011
sharṭ	asāsī	16	4	2	0.1	attr	2.19329	0.0064078
iltizāmāt	asāsī	1	4	1	0	attr	2.06539	0.0086022
wasīṭ	asāsī	1	4	1	0	attr	2.06539	0.0086022
al-qadr	lāzim	10	305	10	6.6	attr	1.85397	0.013997
al-da'm	lāzim	9	305	9	5.9	attr	1.66629	0.021563
al-taswiyyāt	lāzim	5	305	5	3.3	attr	0.92071	0.12003
al-musā'adāt	lāzim	5	305	5	3.3	attr	0.92071	0.12003
amr	muhim	37	2	1	0.2	attr	0.8154	0.15297
al-ḍaght	lāzim	4	305	4	2.6	attr	0.73557	0.18384
'ilāj	lāzim	1	305	1	0.7	attr	0.18315	0.65591
al-khuṭuwāt	ḍarūrī	79	154	15	26.2	rep	2.71607	0.0019228
sharṭ	lāzim	16	305	2	10.5	rep	4.8884	1.29E-05
al-mawārid	ḍarūrī	60	154	5	19.9	rep	5.70817	1.96E-06

Table 6 The results of distinctive collexeme analysis performed on frequencies of translation equivalents of *necessary* found in the OPUS corpus.

COLLEX	O.CXN1	E.CXN1	O.CXN2	E.CXN2	ASSOC	COLL.STR.LOGL	SIGNIF	SHARED
ḍarūrī	37369	13208.6	31370	55530.4	necessary	49391.98675	*****	Y
lāzim	46331	23939.4	78252	100643.6	necessary	29004.97621	*****	Y
lazam	10356	7164	26926	30118	necessary	1715.83849	*****	Y
lazūm	1046	623.9	2201	2623.1	necessary	310.56516	*****	Y
iqtiḍā'	6587	7362.1	31726	30950.9	other	110.69693	*****	Y
yahtāj	227	854.7	4221	3593.3	other	761.50111	*****	Y
yaqtaḍī	1411	3639.8	17531	15302.2	other	2147.71447	*****	Y
ḍarūrah	8045	15161.1	70855	63738.9	other	5412.50962	*****	Y
lā bud	1648	8562.9	42914	35999.1	other	10190.14234	*****	Y
muhim	83	6821.2	35415	28676.8	other	14722.58281	*****	Y
asāsī	310	9878.6	51099	41530.4	other	20086.21889	*****	Y
hājah	2777	18973.7	95964	79767.3	other	28017.16009	*****	Y

Table 7 Distinctive collexeme analysis of the two adjectival translations of *necessary*.

COLLEX	O.CXN1	E.CXN1	O.CXN2	E.CXN2	ASSOC	COLL.STR.LOGL	SIGNIF	SHARED
ḍarūrī	106790	34087.8	237766	310468.2	min.Adj	158407.7519	*****	Y
lāzim	20250	57131.3	557228	520346.7	N.Adj	44165.26408	*****	Y

attributive use of these two adjectives in Arabic. We chose the attributive use over the predicative one because this can help us find out the nominal collocates for them. In addition, the predicative use is different in Arabic and English. In English, the adjective is always preceded by a verb, but in Arabic it is preceded by a verb in the past tense but not in the present tense. We focused on the construction N Adj in order to detect subtle semantic relations. Overlap of collocations for the two words reveals that they share conceptual content and small divergence in collocations indicates their different modes of construal, which can suggest that they are near-synonyms (e.g., words sharing almost the same meaning).

The first test aims to find out how the two words are distributed in relation to the three main constructions that the word *necessary* colligates (i.e., typically occurs in) with. A distinctive collexeme analysis (Gries and Stefanowitsch, 2004) was done using the R collostructions package (Flach, 2021) on the occurrences of these words in the arTenTen18 corpus. The results show that the word ḍarūrī associates with the extraposed construction, while the word lāzim associates with the construction N Adj (See Table 7). This finding has been highlighted by Hilpert (2014) and Wiliński (2019) who found that words suggesting IMPORTANCE such as the English *necessary* are associated with the extraposed construction. Further, examples of non-extraposition are very rare in corpora as emphasized by Biber et al. (1999). Hence, modal adjectives were not found to be associated with non-extraposed constructions. This also highlights that similar constructions in English and Arabic might host similar words in the two languages, which confirms findings by Boas (2010a) and Goldberg (2013).

The two adjectives are used in different constructions that the word *necessary* is used in. Both the English *necessary* and the Arabic translation *ḍarūrī*, which is the one found in the Cambridge Dictionary, associate with the extraposed construction. However, because the adjective *lāzim* is significantly attracted to be a translation of the word *necessary* and is used in a different construction than *ḍarūrī*, it has to be viewed as an important contributor to the translation of the word *necessary* and in turn included in an Arabic-English bilingual dictionary.

As a second step, we investigated the frames that are used with the two words. First, we answer the question *What frames are used with each word when they are used attributively?* Second, we answer the question *What frames are associated with each word when they are used in their preferred construction: attributively for the word* lāzim *and in extraposed construction for the word* ḍarūrī? We leave the second question for the next section. As for the first one, we used semantic frames to identify the different frames used with each adjective. Nouns modified by both adjectives were assigned frames based on the sentences they appear in. The resulting frames are listed in Appendix A. The table in Appendix A shows that a number of frames are shared by both adjectives; for example, the frames of CURE (e.g.,'ilāj "treatment," dawā "remedy"), TOPIC (e.g., amr "topic or subject matter"), and NEEDING (e.g., ḥājah "need," iḥtiyāj "needs"). However, the word lāzim is used for wider semantic contexts than ḍarūrī in the attributive construction. This is understandable as the word lāzim associates with this construction more than ḍarūrī. For example, the word lāzim is used for frames as those of ATTENTION (e.g., ihtimām "attention"), TIME SPAN (e.g., zaman "period," muddah "period"), AWARENESS (e.g., ma'rifah "knowledge"), and SUPPLY (e.g., wuqūd "fuel").

The next section answers the final question of this paper and discusses the frames that are associated with the word *lāzim* in the attributive position and the ones associated with the word *ḍarūrī* in the extraposed position.

Frames associated with the two Arabic constructions. Since *lāzim* and *ḍarūrī* are associated with different constructions, one expects that nouns and verbs associated with each construction belong to different semantic categories which suggests that each construction conveys multiple meanings. In this section, we restrict our search for collexemes in Sketch Engine to nouns occurring before *lāzim* in attributive constructions and verbs and nouns following min *al-ḍarūrī* in extra-position.

Starting with *lāzim* in attributive positions, it occurs with 93 nouns (e.g., al-ijrā'āt "procedures," al-tashilāt "facilities," al-ẓurūf "circumstances") whose LogDice score is above 5.37. On the other hand, *ḍarūrī* in the extraposed position commonly occurs with 50 nouns (e.g., ijrā "procedure," al-tadhkīr "reminding") and

verbs (e.g., nushīr "refer") whose Logdice is above 4.5. Each word was assigned a frame, and the resulting frames are listed in Appendix A.

The results showed the interdependence between the adjective *lāzim* and some nouns indicating procedures, measures, treatment, reports, money, skills, support, equipment, condition, etc. Such words evoke the semantic frames of MEANS (e.g., ijrāʾ "procedure," wasīlah "means," ālyyah "mechanism"), INTENTIONAL ACTS (e.g., tadbīr "measure," khuṭwah "step," iḥtiyaṭ "measures"), CURE (e.g., ʿilāj "treatment," dawā "remedy"), STATEMENT (e.g., maḥḍar "report"), MONEY (e.g., māl "money," siyūlah "liquidity," mablagh "amount of money"), EXPERTISE (e.g., mahārah "skill," khibrah "experience"), SUPPORTING (e.g., da'm "support"), and GIZMO (e.g., mu'iddat "equipment," adāt "tool"), INSPECTING (e.g., ikhtibār "exam," taḥlīl "test," faḥṣ "examination"), respectively. Further, the LogDice score of collocates showed that words suggesting procedures, resources, facilities, circumstances, states of an entity, etc. evoking frames of MEANS (e.g., al-ijrāʾāt "procedures"), SOURCE OF GETTING (e.g., al-mawārid "source"), DESIRING (e.g., al-irādah "will"), INTENTIONAL ACTS (e.g., al-khuṭuwāt "steps"), QUANTIFIED MASS (e.g., al-qadr "amount"), CURE (e.g., ʿilāj "treatment"), ASSISTANCE (e.g., al-musāʿadāt "help"), etc. are strongly attracted to the attributive construction of the adjective *lāzim*. However, words suggesting conditions evoking the semantic frame of TERMS OF AGREEMENT is not commonly used in this construction with *lāzim*. To illustrate uses of *lāzim*, consider, for example, sentence (14) in which an Agent uses a Means, which can be an action, to achieve a Purpose.

(19) Wa yāḥtāj al-ashkhāṣ alladhīna yatʿarraḍ-ūn li-daraj-āt al-ḥarārah al-murtafiʿah aw yaʿmal-ūn taḥt al-shams mubāsharah, ilā ittikhādh al-ijrāʾ-āt al-lāzimah lil-wiqāyah min ḥudūth al-jafāf.
And need DEF-people who expose-PL.M to degree-PL.F Def-temperature DEF-high or work- PL.M below DEF-Sun directly to adopt DEF-procedure-PL.F DEF-necessary to preventing from occurrence DEF-dehydration
People (agents) who are exposed to high temperatures or work under direct sunlight need to take the necessary measures (means) to prevent dehydration (purpose).
In another commonly motived semantic frame of intentional acts, acts (e.g., deeds, measures, activities, etc.) are done by sentient beings. Sentence (15) shows how steps are described as necessary intentional acts to be executed by a group of agents.

(20) ʿIlman bi-anna al-hayʾah qad waffara-t irshād-āt mufassalah l-ijamīʿ al-khuṭuw-āt al-lāzimah li-lḥuṣūl ʿalā al-khidmah ʿalā bawwābat al-hayʾah al-iliktrūnīyah
Noting Prep that DEF-authority has provide-F instruction-PL.F detailed Prep all DEF-step-PL.F DEF-necessary for obtaining Prep DEF-service Prep portal DEF-authority DEF-electronic
Noting that the Authority (agents) has provided detailed instructions for all necessary steps (intentional acts) to obtain the service on the Authority's electronic portal.
On the other hand, *ḍarūrī* in extraposed position commonly attract verbs and verbal nouns denoting doing, finding, knowing, consultation, attention, emphasizing, reviewing, etc. Attracted nouns and verbs are related to the semantic frames of INTENTIONAL AFFECT (e.g., al-qiyām "do," 32.ittikhādh "do"), ACHIEVING FIRST (e.g.,ījād "find"), DISCUSSION (e.g., istishārah "consultation"), ATTENTION (e.g., intibāh "attention"), FAMILIARITY (e.g., al-taʿarruf "know," ma'rifah "know"), EMPHASIZING (e.g., taslīṭ al-ḍaw' "focus," al-ta'kīd "emphasize"), ASSESSING (e.g., murāja'ah "review or send," iʿādat al-naẓar "review"), INGESTION (e.g., shurb "drink,"tanāwul "eat"), COGITATION (e.g., al-tafkīr "think," mura'āt "consider," al-akhdh biʿayn al-iʿtibār "consider"). The following sentences illustrate how the Cognizer thinks about a Topic over a period of time if associated nouns evoke the COGITATION frame with *ḍarūrī* in extraposed position.

(21) Yu-ʿadd al-tafkīr al-ījābī hāmman min ajl al-najāḥ, fa-min ajl taḥqīq ahdāfi-ka min al-ḍarūrī al-tafkīr bi-ṭarīqah tatamāshā maʿ tilka al-ahdāf
Consider-PRS.PASS DEF-thinking DEF-positive important Prep for DEF-success Prep Prep for achieving goal.PL-VOC Prep DEF-necessary DEF-thinking Prep way go with those DEF-goal.PL
Positive thinking is important for success. In order to achieve your (Cognizer) goals, it is necessary to think in a way (Topic) that is consistent with those goals.

(22) Hal min al-ḍarūrī murāʿāt al-maṭālib al-akhlaqī-yah fī al-mumāras-ah al-siyāsī-yah?
Q Prep DEF-necessary consider DEF-demand.PL DEF-moral-ADJ Prep DEF-practice-F DEF-politics-ADJ
Is it necessary to take into account moral demands (Topic) in political practice?

Such results have been reported by Hilpert (2014) and Wiliński (2019) with regard to the English *necessary*. They found that the extraposed construction is employed to introduce new information and evoke the semantic frame of BECOMING AWARE and REMEMBERING INFORMATION that are similar to COGITATION. Hence, verbs occurring in the construction evoke frames pertaining to cognitive processes. As a result, as noted by Kaltenböck (2005), extraposed constructions allow language users to present their personal views as generally accepted facts and thus the association with semantic frames of EMPHASIZING, ATTENTION, and COGITATION.

The frames discussed so far show that *lāzim* and *ḍarūrī* share some conceptual content because there are frames that are shared by both in the monolingual corpus. Next, we turn our attention to the parallel corpus in order to check whether the use of *ḍarūrī* and *lāzim* in the translated corpus matches their use in the monolingual corpus. Using Frame semantics and following the same procedure, nouns modified by the adjectives are assigned frames; and the frames evoked by them are listed in Appendix B.

The table shows that *ḍarūrī* is used in varied semantic contexts than *lāzim* in the translated corpus. It also shows that *lāzim* is restricted to a few frames that are also shared by *ḍarūrī*. Nouns modified by *lāzim* do not evoke frames not evoked by nouns that collocate with *ḍarūrī*. This is different from the findings we obtained from the monolingual corpus.

The context of a lexical item reveals key facts about its syntactic and semantic properties (Sinclair, 1991; Biber et al., 1998), and the results of this section have shown that *lāzim* and *ḍarūrī* share many semantic properties, but their syntactic ones are different. Therefore, both words complement each other as translations of the word *necessary* that can be found in both the attributive position, similar to *lāzim*, and the extra-posed one, similar to *ḍarūrī*. However, in the parallel corpus, this complementarity between the two adjectives is not observed, which suggests that parallel corpora are not enough resources when creating dictionaries. Monolingual corpora need to be consulted for the aim of creating more accurate dictionaries.

Conclusions

This is a corpus-driven rather than a corpus-exemplified study (Tummers et al., 2005) in which we explored large corpora to

support or falsify our claim. The starting point for this study was an exploration of the various senses embedded in the Arabic and English words of *necessity*. Then, the Arabic translations of the word *necessary* have been examined using Cambridge Dictionary. It provides one translation: *ḍarūrī*. We have shown in this study that both adjectives (*ḍarūrī* and *lāzim*) are used as translations of *necessary* in the OPUS parallel corpus. We have also shown that their use in the monolingual corpus arTenTen18 indicates that they share the sense of NECESSITY that indicates non-epistemic modality. This finding supports Héja's (2010) and Hilpert's (2013) assertion that using parallel corpora facilitates the creation of more accurate bilingual dictionaries. However, the use of these two adjectives in the parallel corpus diverges from their use in the monolingual one. The results revealed that the use of the word *lāzim* is restricted in the parallel corpus which suggests that monolingual corpora have to be exploited for the goal of creating more accurate bilingual dictionaries. We advocate the many calls for using corpora in lexicography (Héja, 2010; Wible and Tsao, 2011).

As for the other words used to translate *necessary* into Arabic in the OPUS, *muhim* and *asāsī*, etc., they are used based on context and do not need to be included in the dictionary as their association with *necessary* is weak. However, they do carry some sense of non-epistemic modality and learners of Arabic as well as translators need to put in mind that dictionaries are not the only source of foreign vocabulary. Corpora play an important role as useful tools to be exploited.

Parallel corpora may support the transfer of items from source to target language complementing other sources such as existing bilingual dictionaries and human expert knowledge; they are mainly used for validation and disambiguation using collocational contexts. (Tiedemann, 2014, p. 1433)

From a lexical semantics point of view, the words *ḍarūrī* and *lāzim* are cognitive synonyms and not near-synonyms based on the findings of this study. Cognitive synonyms do not alter the truth value of a proposition (Cruse, 1986), and they have at least one sense in common (Murphy, 2003). Near-synonyms, on the other hand, are context dependant and cannot be linked to each other logically (Murphy, 2003). The adjectives *asāsī* and *muhim* can be viewed as near synonyms of *ḍarūrī*, as they are also used to convey non-epistemic modality. However, they cannot be linked to it logically. The meaning of a linguistic unit "involves both the conceptual content and the construal of that content" (Langacker, 2008, p. 44). Construal, how a conceptual representation is presented through the choice of linguistic units, differs when it comes to how *ḍarūrī* and *lāzim* are used. The adjective *lāzim* is used more frequently in the attributive position and *ḍarūrī* is used in the extraposed one.

Limitations and suggestions for future research. The present corpus-driven study is limited to the use of one parallel corpus (i.e., OPUS) and the investigation of Arabic equivalents for one English deontic modal adjective. Hence, the results cannot be generalized to other English and Arabic modals occupying various constructions. In addition, initiating a similar study on the Arabic equivalents of the English modal verbs of necessity (i.e., must and need) might show which Arabic verbs are commonly associated with English verbs of necessity and how such verbs behave collocationally in context. One may find that verbs that are not associated with the English *necessary* such as *yahtāj* and *lā bud* are more attracted as translations to constructions with *must* and *need*. More importantly, it would be intriguing if future research addresses which collexemes of verbs and nouns typically fill slots in passive and active Arabic constructions and compare them with those of English. Since in Arabic the use of the passive necessitates some ignorance of the doer, conducting the covarying collexeme analysis to identify the collexmes occupying slots in the two constructions in the two languages might reveal contradictory results. Moreover, performing some collocational analysis of synonymous words and identifying the semantic frames for collocates are both recommended for future studies on synonymous words and their equivalents.

Implications of the study. The study has important implications for Arabic lexicographers, translator trainers, and English as a Foreign Language (EFL) instructors. For instance, for lexicographers interested in making bilingual dictionaries, the study clearly shows the importance of checking both bilingual and monolingual corpora to identify translation equivalents for key English words suggesting possibility, probability, necessity, etc. Further, translator trainers should utilize parallel corpora as well as monolingual corpora in their teaching to decide on Arabic equivalents of English words. Corpora can reveal important information on constructions hosting some words and how common is one construction in the language corpus to express a specific meaning compared to its variants. Diachronic corpora might be used to decide on the frequency of one construction and the role of translation, as parallel corpora may contribute to the frequency of one construction. Thus, monolingual corpora are capable of giving more accurate results using quantification and relative frequencies. Additionally, EFL instructors are ought to explain to learners common constructions in English along with commonly used words in them to ensure effective transfer of meaning.

Data availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

Received: 7 February 2025; Accepted: 15 July 2025;
Published online: 18 August 2025

Note

1 A website that searches for meanings of words in different Arabic dictionaries at once. It can be found at almaany.com.

References

Abdul-Baquee S, Atwell ES (2009) Knowledge representation of the Quran through frame semantics: a corpus-based approach. In: Proceedings of the Fifth Corpus Linguistics Conference, 1-12

Alhedayani R (2016) Antonymy in Modern Standard Arabic. Dissertation, Sussex University

Alotaibi YH (2022) Adjectives in Arabic. Intern J Arabic Ling Lit Stud 4:37–52. https://doi.org/10.31559/JALLS2022.4.1.3

Al-Qahtani DM (2005) Semantic valence of Arabic verbs. Librairie du Liban. Publishers, Lebanon

Al-Sharifi B, Sadler L (2009) The adjectival construct in Arabic. In: Proceedings of the LFG09 Conference, p. 26−43

Akinseye TA (2020) Epistemic versus non-epistemic modality: an analysis of English modals in selected English as a second language (ESL) classroom discourses. J Lang Lit Pop Cul 2:67–87

AL-Khazraji NHF (2023) Arabic equivalents of the English modality. Inter J Hum Edu Res 5:399–423. https://doi.org/10.47832/2757-5403.18.22

Arts T, Belinkov Y, Habash N et al. (2014) arTenTen: Arabic corpus and word sketches. J King Saud Uni– Comp Info Sci 26:357–371. https://doi.org/10.1016/j.jksuci.2014.06.009

Atkins S, Fillmore CJ, Johnson CR (2003) Lexicographic relevance: selecting information from corpus evidence. Inter J Lexico 16:251–280. https://doi.org/10.1093/ijl/16.3.251

Bara G (2010) Comparison between It-clefts and Wh-clefts: similarities and differences. Grin Verlag, Germany

Biber D, Conrad S, Reppen R et al. (1998) Corpus linguistics: investigating language structure and use. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511804489

Biber D, Johansson S, Leech GN et al. (1999) Longman grammar of spoken and written English. Pearson Education, Harlow
Calude AS (2008) Clefting and extraposition in English. ICAME J 32:7–34
Boas HC (2010a) Contrastive studies in construction grammar. John Benjamins, Amsterdam
Bybee JL (2013) Usage-based theory and exemplar representations of constructions. In: Hoffmann T, Trousdale G (eds) The Oxford handbook of construction grammar. Oxford, Oxford University Press, p. 49–69
Chierchia G (1985) Formal semantics and the grammar of predication. Ling Inq 16:417–443
Collins P (1991) Cleft and pseudo-cleft constructions in English. Routledge, London
Cruse DA (1986) Lexical semantics. Cambridge University Press, Cambridge
Delin JL (1990) Cleft constructions in discourse. Dissertation, University of Edinburgh
Depraetere I, Reed S (2006) Mood and modality in English. In Aarts B, McMahon A, Hinrichs L (eds) The handbook of English linguistics, 2nd edn. Blackwell Publishing, Oxford, p. 269−290. https://doi.org/10.1002/9780470753002.ch12
Deshors SC (2017) Zooming in on verbs in the progressive: a collostructional and correspondence analysis approach. J Eng Ling 45:260–290. https://doi.org/10.1177/0075424217717589
Downing A (2001) 'Surely you knew!': surely as a marker of evidentiality and stance. Func Lang 8:251–282. https://doi.org/10.1075/fol.8.2.05dow
El-Hassan S (1990) Modality in English and Standard Arabic: paraphrase and equivalence. J King Saud Uni 2:149–166
Englebretson R (1997) Genre and grammar: predicative and attributive adjectives in spoken English. Ann Meet Berk Ling Soc 23:411–421. https://doi.org/10.3765/bls.v23i1.1272
Fillmore CJ (1967) The grammar of hitting and breaking. In: Jacobs R, Rosenbaum P (eds) Studies in English Transformational Grammar, p. 9−29. https://kb.osu.edu/server/api/core/bitstreams/6f246a7c-e36d-52fe-86ee-7f92df1ce276/content
Fillmore CJ (1968) The case for case. In: Bach E, Harms R (eds) Universals in linguistic theory. Holt Rinehart and Winston, New York, p. 1–88
Fillmore CJ (1976) Frame semantics and the nature of language. Ann New York Acad Sci 280(1):20–32. https://doi.org/10.1111/j.1749-6632.1976.tb25467.x
Fillmore CJ (1982) Frame semantics. In: The Linguistic Society of Korea (ed) Linguistics in the morning calm. Hanshin Publishing Company, Seoul, p. 111−137
Fillmore CJ (1985b) Frames and the semantics of understanding. Quad di Semant 6(2):222–254
Fillmore CJ (2006) Frame semantics. In: Geeraerts D (ed) Cognitive linguistics: basic readings. Mouton de Gruyter, Berlin/New York, p. 373−400
Fillmore CJ (2008) Merging frames. In: Favretti RR (ed) Frames, corpora and knowledge representation. Bononia University Press, Bologna, p. 1−12
Fillmore CJ, Atkins BT (1992) Toward a Frame-Based lexicon: The semantics of RISK and its neighbors. In: Lehrer A, Kittay EF (eds) Frames, fields and contrasts: new essays in semantic and lexical organization. Lawrence Erlbaum Associates, Hillsdale, p. 75-102
Fillmore CJ, Atkins BT (2012) Toward a frame-based lexicon: the semantics of RISK and its neighbors. In: Lehrer A, Kittay E (eds) Frames, fields, and contrasts. Routledge, London, p. 75−102
Fillmore CJ, Baker C (2010) A frames approach to semantic analysis. In: Heine B, Narrog H (eds) The Oxford handbook of linguistic analysis. Oxford University Press, Oxford, p. 313−340. https://doi.org/10.1093/oxfordhb/9780199544004.013.0013
Fillmore CJ, Petruck MR, Ruppenhofer J, Wright A (2003) FrameNet in action: the case of attaching. Inter J Lexico 16(3):297–332. https://doi.org/10.1093/ijl/16.3.297
Fillmore CJ, Kay P, O'Connor MC, (1988) Regularity and idiomaticity in grammatical constructions: the case of let alone. Lang 64(3):501–538. https://doi.org/10.2307/414531
Fillmore CJ, Kay P, O'connor MC (2014) Regularity and idiomaticity in grammatical constructions: the case of let alone. Lang 64(3):501–538
Flach S (2021) Collostructions: an R implementation for the family of collostructional methods. Package version v.0.2.0. https://sfla.ch/collostructions/. Accessed 8 August 2023
Gargett A, Leung T (2020) Building the Emirati Arabic Framenet. In: Proceedings of the International FrameNet Workshop 2020: Towards à Global, Multilingual FrameNet, p. 70−76
Gilquin G (2010) Corpus, cognition and causative constructions. John Benjamins, Amsterdam
Gilquin G (2013) Making sense of collostructional analysis: on the interplay between verb senses and constructions. Cons frames 5(2):119–142
Goldberg AE (2002) Surface generalizations: an alternative to alternations. Cogn Ling 13:327–356
Goldberg AE (2003) Constructions: a new theoretical approach to language. Trends Cogn Sci 7(5):219–224
Goldberg AE (2009) Constructions at work: the nature of generalization in language. Oxford University Press, Oxford
Goldberg A (2013) Constructionist approaches to language. In: Hoffmann T, Trousdale G (eds) Handbook of construction grammar. Oxford University Press, Oxford, p. 15-48. https://adele.princeton.edu/wp-content/uploads/sites/277/2015/01/13constructionist-approaches-Oxfordfinal-edits.pdf
Goldberg AE (1995) Constructions: a construction grammar approach to argument structure. University of Chicago Press, Chicago
Gries STH, Stefanowitsch A (2004) Extending collostructional analysis: a corpus-based perspective on 'alternations.'. Intern J Corp Ling 9:97–129. https://doi.org/10.1075/ijcl.9.1.06gri
Gries STH, Stefanowitsch A (2004b) Co-varying collexemes in the into-causative. In: Achard M & Kemmer S (eds) Language, culture, mind. The University of Chicago Press, Chicago, p. 225−236
Héja E (2010) Dictionary building based on parallel corpora and word alignment. In: Dykstra A, Schoonheim T (eds) Proceedings of the 14th EURALEX International Congress, Fryske Akademy, Leeuwarden/Ljouwert, p.341–352 https://www.euralex.org/elx_proceedings/Euralex2010/015_Euralex_2010_1_HEJA_Dictionary%20Building%20based%20on%20Parallel%20Corpora%20and%20Word%20Alignment.pdf
Hewings M, Hewings A (2002) "It is interesting to note that…": a comparative study of anticipatory 'it' in student and published writing. Eng Spec Purp 21:367–383. https://doi.org/10.1016/S0889-4906(01)00016-3
Hilpert M (2013) Corpus-based approaches to constructional change. In: Hoffmann T, Trousdale G (eds) The Oxford handbook of construction grammar. Oxford University Press, Oxford, p. 334–345
Hilpert M (2014) Collostructional analysis: measuring associations between constructions and lexical elements. In Glynn D, Robinson JA (eds) Corpus methods for semantics. John Benjamins, Amsterdam, p. 391−404
Hilpert M (2014) Construction grammar and its application to English. Edinburgh University Press, Edinburgh
Hobi E (2011) A contrastive study of attributive adjectives in English and Arabic. J Al-Ma'moon Uni Col 17:265–274
Hopper PJ (1991) On some principles of grammaticalization. In: Traugott EC, Heine B (eds) Approaches to grammaticalization, 1st Vol. John Benjamins, Amsterdam, p. 17–35
Jäger VC (2018) Expressions of non-epistemic modality in American English: a corpus-based study on variation and change in the 20th century. Dissertation, Johannes Gutenberg-Universityät
Jakubíček M, Kilgarriff A, Kovář V et al. (2013) The TenTen corpus family. In: The 7th International Corpus Linguistics Conference CL, 125−127
Kaatari H (2010) Complementation of adjectives: a corpus-based study of adjectival complementation by that- and to-clauses. MA thesis, Uppsala University
Kaltenböck G (2005) It-extraposition in English: a functional view. Inter J Corp Ling 10:119–159. https://doi.org/10.1075/ijcl.10.2.02kal
Khan G (1984) Extraposition and pronominal agreement in Semitic languages. University of London School of Oriental and African Studies, London
Kim JB, Davies MA (2016) The into-causative construction in English: a construction-based perspective. Eng Lang Ling 20:55–83. https://doi.org/10.1017/S1360674315000271
Lambrecht K (2001) A framework for the analysis of cleft constructions. Ling 39(3):463–516
Maíz C, Arús J (2008) Towards a diachronic study of modal adverbs: a case study of "certainly.". Philo 6:17–25
Mair C (2006) twentieth-century English: history, variation and standardization. Cambridge University Press, Cambridge
Mochón EV (2022) Modality in Arabic: the multiple functions of the (non)-indicative markers –ūn and b-. Lang 7(4):272. https://doi.org/10.3390/languages7040272
Murphy M (2003) Semantic relations and the lexicon: antonymy, synonymy, and other paradigms. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511486494
Langacker RW (2008) Cognitive grammar: a basic introduction. Oxford University Press, Oxford
Lison P, Tiedemann J (2016) Opensubtitles2016: extracting large parallel corpora from movie and TV subtitles. In: Proceedings of the 10th International Conference on Language Resources and Evaluation, p. 923−929
Patten A (2010) Grammaticalization and the *it*-cleft construction. In: Traugott EC, Trousdale G (eds) Gradience, gradualness and grammaticalization. John Benjamins, Amsterdam, p. 221–243
Pedersen T (1996) Fishing for exactness. In: Proceedings of the South-Central SAS Users Group Conference, p. 188−200
Penry Williams C, Korhonen M (2020) A sociolinguistic perspective on the (quasi-) modals of obligation and necessity in Australian English. Eng World-Wide. J Var Eng 41:267–294. https://doi.org/10.1075/eww.00051.pen
Quirk RS, Greenbaum S (1973) A universal grammar of English. Longman, London

Quirk R, Greenbaum S, Leech G et al. (1985) A comprehensive grammar of the English language. Longman, London
Ruppenhofer J, Ellsworth M, Schwarzer-Petruck M et al. (2016) FrameNet II: extended theory and practice. International Computer Science Institute, Berkeley
Rychly P (2008) A lexicographer-friendly association score. In: Sojka P, Horák A (eds) Proceedings of recent advances in Slavonic natural language processing, Raslan 2008, December 2008. Masaryk University, Brno, p. 14–17
Ryding K (2005) A reference grammar of modern Standard Arabic. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511486975
Salama AHY (2021) The force dynamics of adjectival deontic modality in the mediatised register of the fatwa: a corpus cognitive–semantic analysis. Corp 16:1–30. https://doi.org/10.3366/cor.2021.0207
Sinclair J (1991) Corpus, concordance, collocation. Oxford University Press, Oxford
Stefanowitsch A (2013) Collostructional analysis. In: Hoffmann T, Trousdale G (eds) The Oxford handbook of construction grammar. Oxford, Oxford University Press, p. 215–229
Stefanowitsch A (2014) Collostructional analysis: a case study of the English into-causative. In: Herbst T, Schmid HJ, Susen Faulhaber S (eds) Constructions collocations patterns. De Gruyter, Berlin, p. 217−238. https://doi.org/10.1515/9783110356854.217
Stefanowitsch A, Gries STH (2005) Covarying collexemes. Corp Ling Ling Theor 1:1–43. https://doi.org/10.1515/cllt.2005.1.1.1
Suchomel V (2020) Better web corpora for corpus linguistics and NLP. Dissertation, Masaryk University
Szymański L (2021) A corpus-based analysis of the modal can with the perfect in American English from a synchronic and diachronic perspective. Stud. Neo 93:287–313. https://doi.org/10.1080/00393274.2020.1846613
Tiedemann J (2012) Parallel data, tools and interfaces in OPUS. In: Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC2012), p. 2214−2218
Tiedemann J (2014) Supplementary volume dictionaries: an international encyclopaedia of lexicography: supplementary volume: recent developments with focus on electronic and computational lexicography. De Gruyter, Berlin
Traugott EC (2003b) From subjectification to intersubjectification. In: Hickey R (ed) Motives for language change. Cambridge University Press, Cambridge, p. 124–139
Tummers J, Heylen K, Geeraerts D (2005) Usage-based approaches in cognitive linguistics: a technical state of the art. Corp Ling Ling Theo 1:225–261. https://doi.org/10.1515/cllt.2005.1.2.225
Van Linden A (2012) Modal adjectives: English deontic and evaluative constructions in diachrony and synchrony. Walter de Gruyter, Berlin. https://doi.org/10.1515/9783110252941
Wang J, Jiangping Z (2022) A corpus-based study of semantic categorizations of attracted adjectives to the it BE ADJ clause construction. SAGE Open 12(2):21582440221091242. https://doi.org/10.1177/21582440221091242
Wible D, Tsao NL (2011) Towards a new generation of corpus-derived lexical resources for language learning. In: Meunier F, De Cock S, Gilquin G, Paquot M (eds) A taste for corpora. John Benjamins Publishing Company, Amsterdam, p. 237 – 256. https://doi.org/10.1075/scl.45.14wib
Wiliński J (2019) Co-varying verbs and adjectives of it-extraposed constructions with to-infinitive clauses in academic discourse: a quantitative corpus-driven study. Ling Prag 29:7–28. https://doi.org/10.14712/18059635.2019.1.1
Wiliński J(2021) Nouns in the Be of N-Construction: a corpus-based investigation. Stud Log Gram Rhet 66:747–768. https://doi.org/10.2478/slgr-2021-0045
Wiliński J (2022) What-clefts with adjectives in English: a corpus-based analysis. Annales Univers Paedag Cracoviensis Stud Ling 17:258–272
Zipf GK (1965) The psychobiology of language: an introduction to dynamic philology. MIT Press, Oxford
Zraigy AS, Saleem HS, Omar OD (2019) Meaning of English Modals in Arabic. Alfarahidi Let J 37:516–539

Acknowledgements

This research received grant no. (478/2024) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

Author contributions

Rukayah Alhedayani is the first and the corresponding author of the article. She designed the research, collected, analysed and interpreted the data, as well as drafted and revised the Introduction, the Review of Literature, the Data Analysis Section, the Results and the Discussion Section, and the References. Ghuzayyil Mohammed Al-Otaibi is the second author of the article. She participated in collecting, analysing, and interpreting the data besides writing, drafting, and revising the Introduction, the Review of Literature, the Results, the Limitations, Implications, and Suggestions for Further Research Sections besides the References.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-05566-5.

Correspondence and requests for materials should be addressed to Rukayah Alhedayani.