Verb Noun Collocations in Arabic and Their Patterns in Lexicography

Hussein Soori[1]

Su'ad Awab[2]

ABSTRACT

This article is part of a study aiming to disambiguate Arabic verb-noun collocation in Machine Translation (MT) and Natural Language Processing (NLP). This article investigates verb noun collocation in the Arabic language from a linguistic point of view. It reviews the main theoretical studies on collocation including early indications, concept, terminology and classification of collocation in the Arabic language. In addition to that, it also reviews contemporary collocation dictionaries and verb-noun collocation patterns and the possibility of using both traditional and current lexicography resources in sense disambiguation for NLP and MT. The final part of this article is dedicated to a suggested eclectic model of verb-noun collocation patterns in Arabic.

Keywords: Arabic collocation patterns; verb noun collocation; collocation in Arabic; lexicography in Arabic

1. First Indications and the Concept of Collocation in Arabic.

Perhaps the first reference to collocation in Arabic literature and thesauruses is in al-Jāḥiẓ's[3] book, *al-Bāyan Wāttābyīn*, 'The Book of Eloquence and Demonstration', (1960: Vol. 1, 20-21). He states that a group of words that go under one part of speech and share, more or less, the same meaning, may not collocate with all words from the same or another part of speech that share the same meaning. However, al-Jāḥiẓ, in this early indication does not give any technical term to this linguistic phenomenon. al-Jāḥiẓ mentions some examples of words that collocate with each other in the Qur'an such as:

- *ṣālat* and *zākat* (prayer and charity)
- *jwū'* and *khāwf* (hunger and fear (of danger))
- *jānnaṯ* and *nar* (heaven and fire (hell))
- *muhajīryīn* and *anṣar* (Muhājirs[4] and Helpers[5])
- *jīn* and *īns* (jinns and humans)

According to 'ābdūl-'zyz (1990:8), the first scholar who named collocation in Arabic was Abu al-Faraj (1966) in the 10th century and he used the term, *mūṣahābaṯ*. However, according to Ghazala (2007), collocation was given an equivalent term in Arabic as *mūtālazīmat lāfẓayyāt* "verbal cooccurrences" for the first time in the history of English-Arabic\ Arabic-English dictionaries by Ba'albaki in *The Dictionary of Linguistic Terms* (1990).

---

1 Hussein Soori, PhD student, The Faculty of Languages and Linguistics, University of Malaya, Kuala Lumpur, soori2003@hotmail.com.
2 Su'ad Awab, Assoc. Prof., The Faculty of Languages and Linguistics, University of Malaya, Kuala Lumpur, suad@um.edu.my.
3 Died between AD 868-869
4 Those who forsook their homes and property in Makkah, in order to assist Prophet Muhammad in his migration to Madinah (from 'Ali, 2005:574).
5 The people of Madinah, who accepted Islam when it was persecuted in Makkah (from 'Ali, 2005:574).

Contemporary Arab scholars use different terms to describe collocation. According to Hassan (1990), some scholars followed Abu-al Faraj in using the term, *mūṣahābaṯ*, while other scholars use the term, *tālazūm*.

This article will follow the term and linguistic classification given to collocation by Ḥāssan's model (1981; 1985; 1990) because it is believed to be the most modern and comprehensive description of this phenomenon in Arabic. Ḥāssan uses the term, *tāḍam* for collocation in his books (Ibid).

Ḥāssan (1981: 296) states that in traditional Arabic grammar, *nāhw* (Arabic syntax) does not deal with spheres below the morphological constructions and not above the single meaningful sentence. In other words, *nāhw* is restricted to grammar, while meaning and style fall under *fiḡh al-lūḡhṯ* (philology). According to him, while *fiḡh al-lūḡhṯ* (philology) has to do with the single word and its relation to other words in a sentence, and its relation to meaning and usage, *'īlm al-bālaghṯ* (rhetoric) has to do with style and meaning and how they relate to each other (Ibid: 310). As far as the meaning is concerned, (Ibid: 1985) *'īlm al-bālaghṯ* is sub-divided into: *'īlm al-bāyan* (diction/ rhetoric), *'īlm al-bādyī* (science of figures of speech) and *'īlm al-mā'any* (semantics) .

'īlm al-mā'any studies the aspects related to the functional meaning of the Arabic sentence. In Arabic, *tāḍam* falls under *'īlm al-mā'any* and is sub- classified into: *tāwarūd* (succession: the possibility of a word/ words to occur after another word) and *tālazūm* (association/ accompaniment).

According to Hassan (2006: 136-7), *tālazūm* and *tāwarūd* are best looked at as two components rather than two sub-categories of *tāḍam*. While *tālazūm* has to do with the functional aspect of *tāḍam*, *tāwarūd* focuses on the lexical aspect of *tāḍam* (Ibid.: 160:164). Ḥāssan argues that *tāḍam* starts with lexicon and ends with function. It starts with giving the lexical meaning of one word to another word and ends up with the other word functioning syntactically as the first word does, including its collocational characteristics. When either or none of these two components work accordingly, we are left with what is called, *al-mūfarāqt al-mū'jāmayyṯ* (incollocability\ anti-collocability) (Ibid: 137).

A few examples of *al-mūfarāqt al-mū'jāmayyṯ* are mentioned by 'Ābdūl-'zyz (1990: 39) as follows:
1. شربت اللحم   (I drank meat)
2. الفكرة قطعت الشجرة (The idea cut the tree)
3. قام الرجل أن خرج * (*The man stood up once he went out)
4. خرج الرجل مدرسة * (*The man went out school)

Where sentences 1 and 2 are examples of lexical incollocability and 3 and 4 are examples of functional incollocability. Obviously, as for sentences 1 and 2 above, one can not drink meat as much as an idea can not cut a tree. As for sentences 3 and 4 above, the verb, (قام) in 3 can not be followed by *māṣdār mwawāl* (interpreted infinitive), and the verb, (خرج) in 4 does not need *māf'wl bīh* (an object) (Ibid. 39-4). Thus, both 3 and 4 above are functionally ill-formed sentences.

Although collocation had not been explained and incorporated in a full linguistic theory by classical lexicographers, some early lexicographers mentioned collocations in their books.

2. Collocation and Lexicography in the Arabic Language

The following is a review of the most important works of early lexicographers and collocation in Arabic, and a discussion of the viability of these works to be used for collocation retrieval in machine translation (MT) and Natural Language Processing (NLP).

2.1 Works of Classical Lexicographers

According to Anees (1984), the idea of compiling dictionaries goes back between the first half of the second and the third Hijri centuries[6]. During that period, linguists started writing small booklets (or *letters*, as they called them) each about a specific topic/ subject, such as al-Asma'iy[7] who wrote *letters* on horses, goats, plants, trees, palm trees, etc. However, alphabetical order of dictionaries did not start until the fourth Hijri century[8] (for a historical review of the beginnings of making dictionaries in the Arabic language, see Anees, 1984: 225-248).

The first to be mentioned here is Adb al-Katib (The Writer's Literature) by Ibn-Qutaybt[9] (1967) who wrote books on many subjects: history, literary criticism, grammar and philology. Ibn-Qutaybt complied this dictionary as a guide for his secretary. The arrangement of Adb al-Katib is not logical. However, according to Heywood (1965), the idea of compiling it is based on a mixture of word- measure and meaning.

After about one and a half century, another lexicographer, named, al-Iskafy[10] (1985), compiled Mabadi' al-Lught (Principles of Language). The arrangement of Mabadi' al-Lught is relatively more logical than Adb al-Katib. It starts with natural phenomena such as time, day and night, and then compiles words for clothes, implements, food drinks and weapons. Following that, come a large section about horses, and a short one about camels and lions and other animals, agricultural implements, trees, plants, trade and illnesses.

Another remarkable work is Fiqh al-lught (Phonology) by Ath-Tha'aliby (1981). According to Haywood (1965), Fiqh al-lught is a store of words and expressions. Ath-Tha'aliby[11] was a contemporary of al-Iskafy. Fiqh al-lught contains synonyms and states the very minor differences in meaning for words that are near synonyms. It contains a large number of small sections that are all grouped under 30 chapters. To be fair to Fiqh al-lught, it is worth mentioning here that it is not only a book about words. It includes, in its own style -along with phonology- debates on philosophy, theology and religion. It also reflects the writers own belief about al-Mu'tazilt sect[12], and some sectarian discussions within the same sect in contrast with other contemporary sects in that era.

Haywood classifies Arab classical books of lexicographers into: general classified vocabulary and specialized vocabulary. As for general classified vocabulary, although arranged according to meaning under subject headings, they cover the same ground as current dictionaries. According to Haywood (1965), those days, there was a stream of these kinds of

---

6 Between AD 719-911.
7 Died AD 831.
8 Between AD 913-1008.
9 Died in AD 889.
10 Died in AD 1030.
11 Died in AD 1008.
12 A Sunni Islamic school of speculative theology that flourished in Basra and Baghdad between 8th and 10th centuries AD (The New Encyclopedia Britannica, 2007).

lexicography books. This stream culminated in the specialized dictionary al-Mukhassas by Ibn-Syidh[13] (1903).

Ibn-Syidh was a very talented lexicographer, grammarian and poet. Although he was blind, he compiled a dictionary of 28 volumes that he called al-Muhkam (2000). He rearranged its contents to form al-Mukhassas (1901), which, according to Haywood (1965), reached the ultimate limit of fullness, exactitude and authoritativeness. In al-Mukhassas, which is considered the biggest-ever Arabic lexicon, Ibn-Syidh achieved the highest level in chaptering, ordering and accommodation.

Emery (1988; 1991) states that lexicographers such as ath-Tha'aliby and Ibn-Syidh were aware of collocation as a phenomenon. According to Emery, although Ibn-Qutaybt and al-Iskafy included a wealth of collocation information in their dictionaries of meaning, the arrangement of those dictionaries was not in alphabetical order but according to meaning, where words were arranged under subject headings. Nevertheless, they still covered the same ground as current alphabetically ordered dictionaries.

Emery concludes that the arrangement of the material in classified dictionaries (arranged according to subject headings) is idiosyncratic and from a modern linguistic point of view unsystematic, because old lexicographers' books necessitate a lot of searching and some classical usages in those books are obsolete or obsolescent.

This opinion can be supported with many examples from the same old dictionaries. By going back to Fiqh al-lught (Ath-Tha'aliby: 1981), in many instances, we noticed a discussion about a linguistic point that is linked next to Islamic sharia, and leads eventually to compare different habits and customs among nations in the east and west, all in one section and, at times, on the same page.

In our opinion, the reason has to do with the fact that the purpose of writing dictionaries during that era was different from our time. Writers during that time were highly influenced by philosophy, logic and religion, and in many instances, they used their books to express their beliefs and advocate their ideas. An example of that can be given from Fiqh al-lught where ath-Tha'aliby (1981) advocates al- al-Mu'tazilt sect and explains the way they differ from other sects. This should not be surprising or even puzzling because old books were written in a different era when the trend was to involve religion, phonology and philosophy in books studying language. It also should not be any more surprising or puzzling than Natural Language Processing (NLP) or multimodal discourse analysis in the 20th and 21st centuries where mathematics, computer science, space, colour, 3 dimension technology, signals, gestures, visual images, video texts and interactive digital sites are incorporated with linguistics for the study of language in computational linguistics, NLP, speech synthesis and MT.

Anees (1984) concludes that looking for sense relations, the evolution and evaluation of the semantics of words and the etymology of words in old dictionaries is futile and the researcher ought to go back to the texts in old Arabic literature books to be able to find about those relations.

---

13 Died in AD 1066.

In a paper dedicated to collocation in Arabic, Emery (1988: 181) concludes that "the fact that equivalent items in different languages have contrasting collocational ranges is not, in itself, problematic. The difficulty here is, primarily, a practical one – the lack of collocational information in lexicographical materials, a lack which is most conspicuous in modern Arabic-English and Arabic-Arabic dictionaries". The present writers believe that after almost 30 years, the above extract from Emery remains –to a large extent- true.

In our opinion, old Arabic lexicography books are futile for writers and translators in modern standard Arabic (MSA). Moreover, collocation pairs are not easy to locate in old Arabic lexicography books and dictionaries because of the subject order system, and the dryness and archaicness of the language used, which is difficult to understand even by many native speakers of Arabic today. Thus, these lexicography books are impractical to be used as sources for researchers working in the fields of NLP and MT, unless multilevel (adverb-noun, verb-noun, adjective-noun, etc.) scanning of collocates is executed manually, which is time and fund consuming, yet not impossible. It is time and fund consuming for the reason that many collocates used in old lexicography and dictionaries are not in use anymore in MSA. However, digitalizing these old lexicography books using Optical Character Recognition (OCR) or other modern technologies can be a useful and practical option for extracting collocation for the purpose of research on old literary Arabic.

For the above-mentioned reasons, the following sections will investigate the works of contemporary lexicographers as a means to find sources that are more practical for researchers in NLP and MT to locate collocates in MSA so to encode them in thesauri, electronic dictionaries or corpora or any other electronic form.

2. 2. Works of Contemporary Lexicographers

In the following sub-sections, two modern collocation dictionaries are reviewed. This includes the arrangement of entries and some contrastive aspects between these two dictionaries.

2.2.1 Al-Hafiz Arabic Collocations Dictionary (2004) (Arabic-English)

According to Hafiz, this dictionary is the first of its kind in Arabic. It does not contain open collocations, for example, بيت +جميل (beautiful + house), اكل +طعام (ate + food). Instead, it contains, restricted collocations in Arabic such as, نشب + قتال (a fight + broke up) and bound collocations such as, حرب + ضروس (fierce + war), as well as short idioms such as, حرب + باردة (cold + war). The dictionary also includes grammatical collocations [14] such as the grammatical collocations of the verb, ضرب, *ḍharaba*:

- حض, نصيب, في الامر بـ سهم   'participated with, contributed'
- عن الامر   'stopped, stepped back'
- إلى ...   'pointed towards'

There are 7756 entries in this dictionary and they are arranged according to the first letter of the first word in the collocational combination. Quoting Hafiz (2004:13):

> "for a speedy and time-economical reference process, the entered collocation should be included both under the core word alphabetically in one entry and under its other collocate(s) in a separate entry. For example, the (verb-noun) collocation اشتدت الحرب

---

[14] A phrase consisting of a dominant word (noun adjective, verb) and a proposition or grammatical structure such as an infinitive or clause (BBI, op.cit) (from Hafiz 2004)

can be found in the entry of the verb اشتد under the (ا) letter section
as well as in the entry of حرب under the (ح) letter section of the
dictionary."

Surprisingly, after trying to look up the collocation, اشتدت الحرب, 'the war has intensified', quoted above, it could not be found in his dictionary neither in the entry of the verb اشتد under the letter (ا) section, nor in the entry of حرب under the (ح) letter section. The dual method mentioned above, and described by Al-Hafiz as, the inclusion of collocation "under the core word alphabetically in one entry and under its other collocate(s) in a separate entry", does not exist. Instead, the numbering system for the entries assigns a different number for every sub-entry. For example, the entry for the word, يوم, *yom*, 'day', has six sub entries that all are given a separate entry number (Ibid: 374) as in:

➤	7749	يوم + ابيض	a good day
➤	7750	يوم + إجازة ـ عطلة	day off
➤	7751	يوم + اسود	a bad, unlucky day
➤	7752	يوم + طويل	a long day
➤	7753	يوم + عمل	working day
➤	7754	يوم + ممل-	boring day
➤	7755	يوم + ما	some, one day,

One more additional entry is given for a slightly different form of the same word, يوماً + ما, 'some/ one day'.

It seems that Hafiz used this method so to save the dictionary readers' time. However, the number of entries in this dictionary does not depend on the number of word entries but rather on the number of combinations of collocates.

We believe that the number of entries in this dictionary could have been more accurate and realistic had Al Hafiz followed the core word system, *i.e.*, if the word, يوم, 'day', with all its collocates, is put under one entry as follows:

➤ 7749	يوم	day
●	يوم + ابيض	a good day
●	يوم + إجازة ـ عطلة	day off
●	يوم + اسود	a bad, unlucky day
●	يوم +طويل	a long day
●	يوم +عمل	working day
●	يوم + ممل	boring day
●	يوم +ما	some, one day

Hafiz's (2004) dictionary is following a modern system and can be used as a source for MT encoding purpose, but it is not sufficiently rich. However, it is the first of its kind in Modern standard Arabic collocation dictionaries.

2.2.2 Dar El-Ilm's Dictionary of Collocations (2007) by Ghazala (English-Arabic)
This dictionary includes 12,000 entries, under which there are 120,000 combinations of English collocations translated into 150,000 combinations of Arabic collocations. According to Ghazala (2007), this number of collocations is unprecedented in one volume for both Arabic and English.

In comparison with Hafiz (2004), this dictionary follows a more logical and space saving entry-system. While Hafiz's dictionary lists the meaning of collocates, Ghazala's includes

meaning or usage, structure and situation. For instance, 'Day' in this dictionary has 27 sub-entries. These are followed by a new entry for 'daybreak', 'daydream', 'daylight', and 'days' as in:

> ➢ *day*   يوم

• to celebrate a ~;	يحتفل بـ ~
• to look forward to ~;	يتطلع إلى ~
• a blissful ~;	يوم سعيد/ مبهج/ مفرح
• a busy ~;	~ مليء بالعمل
• a chilly ~;	يوم بارد نوعاً ما
• a cool ~;	~ قليل البرودة؛~بارد قليلاً
• a holy ~;	~ مقدس؛ ~ مبارك؛ عيد (ديني)
• an opening ~;	~ الإفتتاح
• a ~ breaks;	يبداء يوم جديد، يشرق نهار جديد
• ~ dawns	يطلع ~؛ تشرق شمس فجر جديد
• a ~ comes;	ياتي/ يحل ~
• a ~ drags on;	يتجرجر/ يتململ ~

The sub-entries for every core-word in this dictionary, as may be noticed in the entry for *day* above, are arranged according to the following order:

- Verb-Noun collocations
- Adjective-noun collocations
- Noun -verb collocations
- Of-genitive/ prepositional collocations
- Miscellaneous
- Idiomatic collocations.

This dictionary is rich with a long lists of collocates in every entry and the examples provided in every sub-entry. It is practical and time saving because the order shown above enables both the reader and researcher to find collocations according to their pattern. It follows a more logical and space saving entry-system. For these reasons, in my opinion, this dictionary can serve as a good source for researchers in NLP to be encoded in MT systems.

Nevertheless, it is evident that the number of entries in this dictionary is small in comparison with collocation dictionaries in other languages such as German, English or French. For this reason, it is hoped that further works are to follow in the near future with a bigger number of entries.

3. Verb-Noun Collocation Patterns in Arabic

The following sections include a comparative study between the different English-Arabic collocation patterns. Idioms are outside the sphere of this study. Following this, we propose an eclectic model of verb-noun collocation patterns.

3.1 The Nijmegen Arabic- Dutch- Arabic Dictionary Project

The idea of this project came out first in 1992 in a report by a committee of experts in bilingual education. This report was entitled, 'Cedars in the Backyard', and came up with some suggestions to improve the education of minorities (Turks and Moroccans students) in the Netherlands. The project was initiated because of the idea that immigrant students need a curriculum in their own language, in line with current thinking about the importance of the first language acquisition in the acquisition of the second language. Accordingly, there was

the need for teaching materials in Arabic including Arabic bilingual dictionaries. In 1995, a commission financed by the Dutch and Flemish governments called, CLVV, decided to initiate a project that had to lead to two sets of bilingual dictionaries: a smaller one of about 15,000 lemmas and a larger one of about 40,000 lemmas (Hoogland & Versteegh 1999). The smaller project was entrusted to Mark van Mal from the University of Leuven in Belgium, and the larger project was given to a team from the University of Nijmegen in the Netherland. These team members were Prof. Manfred Woidich, Jan Hoogland and Kees Versteegh.

Although this project is concerned with Arabic/ Dutch/ Arabic and not English/ Arabic lexicography, it is included here because it is considered one of the very few, if not the only, serious lexicography projects, at the university level, in the 20th century for Arabic bilingual dictionaries. Having Arabic as one of its languages, this project is comprehensive in terms of the methods used into the making of Arabic bilingual dictionaries and the results it came out with. One of the methods was to include a considerable number of expressions, examples and collocations in word entries. The Arabic part included three levels of semantic units, lexical, examples and idiom units.

In 1993, Jan Hoogland conducted a study about the frequency of collocation examples in modern standard Arabic dictionaries. He concluded that learners of Arabic have very few materials to consult in order to find collocations in bilingual and monolingual Arabic dictionaries (Hoogland & Versteegh: 1999). Hoogland (1993), in his classification of collocation, used three models: Benson (1989), Benson *et al.* (1986) and Emery (1991). These collocation pattern classifications are reviewed in the next subsections.

3.2 Benson and Benson & Ilson (The BBI combinatory Dictionary of English)
The authors of this dictionary make a distinction between grammatical and lexical collocation. According to them, grammatical collocation is a phrase consisting of a dominant word (noun, adjective, and verb) and a preposition or grammatical structure, such as an infinitive or a clause. According to them, grammatical collocations fall into the following combinations: noun + preposition, noun + to-infinitive, noun + that-clause, preposition + noun, adjective + preposition + predicate adjective + to + infinitive, adjective + that-clause. Within the same category, the authors of this dictionary also included 19 verb patterns; these are the different syntactic combinations of the verb with different elements (components) in a sentence. For more verb pattern classifications see: Hornby (1975), Hornby and Cowie (1980) and Quirk *et el* (1985).

On the other hand, the authors of this dictionary looked at lexical collocations as seven main collocations that do not contain prepositions, infinitives, or clauses, but contain nouns, adjectives, verbs and adverbs.

In this study, we are more concerned with the lexical collocation category out of which three involve noun and verb collocations as follows:

A. *CA verb (usually transitive) + noun/pronoun (or prepositional phrase)*

These collocations are called CA collocations because they contain verbs denoting creation, e.g., *compose music*, or activation, e.g., *set an alarm*, or both, e.g., *impose an embargo*. According to the BBI authors, CA collocations are extremely important for polysemous nouns. For example, the entry for the noun 'line' has the following collocations: draw a line (on paper); form a line (= 'line up'); drop *smb*. a line (='write *smb*. a letter').

B. *EN verb + noun*

These collocations are called EN verb collocations because they contain verbs denoting eradication and/ or nullification, e.g., *lift a blockade* and *scrub (cancel) a mission.*

C. *noun + verb*

Collocations of this type consist of a noun and verb; the verb names an action characteristic of the person or thing designated by the noun, e.g., *adjectives modify* and *boxers box*.

This study focuses primarily on A and B above because in most cases C doesn't pose a problem for English Arabic translation because it requires word for word translation.

3.3 Emery 1991

Peter Emery classifies collocation into four categories:

A. Open collocations

In this category, both elements (for example, the verb and the object) are freely combinable and each element is used in a common literal sense as in:

- بدءات/ انتهت + الحرب/ المعركة   Badāt/ intahat + al-ḥarb/ al-maʿraṯ (The war/ battle + began/ ended).

B. Restricted collocations

These are combinations of two or more words used in one of their regular, non-idiomatic meanings, following certain structural patterns. According to Aisenstadt (1979), these combinations are restricted in their commutability, not only by grammatical and semantic valency but also, by usage.

In a restricted collocation, only one of the elements may be either literal or figurative. For example, in *explode + a myth/ a belief* the verb is figurative, while in *clench + one's teeth/ fists*, it is literal. An example of the verb-noun (as an object) collocation is given by Emery as follows:

- خاض+ المعركة/ المفاوضات   khāḍa + al maʿrakaṯ/ al mufawaḍat (He rushed into/ embarked on + battle/ negotiations)

C. Bound collocations

According to Emery, Cowie (1981) describe this category of collocations as a bridge between collocations and idioms. This category exhibits unique contextual determination, *i.e.*, one of the elements is uniquely selective of the other where the selecting element is typically specialized (non-literal) in meaning. Emery describes this type as a composite unit. According to Emery, this type of collocation is uncommon in English as in, *foot + the bill* and *curry + flavour*, but the derivational richness in Arabic permits a particular root/ pattern combination to be embarked for specific collocates as in:

- حرب + ضروس   ḥarb + ḍarws (fierce/ murderous + war)
- جيش + جرار   jaysy + jarrār (huge/ tremendous + army),

where the adjective collocates uniquely with a specific noun and the selecting item is not invariably figurative. Another verb/ noun (body part) examples are given by Emery as follows:

- اطرق+الرأس   atraqa al ras (he bowed + his head)

- شمر عن + ساعده syammara 'n + sa'idehi (he bared + his forearm)

Bond collocations readily lend themselves to idiomatic use where *syammara 'n + sa'idehi* , more commonly, carry the meaning of 'get down to work'. However, they are not considered as idioms.

D. Idioms

In this type, the constituent elements of idioms are opaque; in other words, they are used in specialized sense, together forming a single semantic unit as in:

- الحرب الباردة , al ḥarb al baridaṯ (the Cold War)
- حرب الكواكب , ḥarb al kawakib (Star Wars)

3.4 Hoogland 1993

Hoogland mentions eleven types of collocation of which only three contain verb-noun collocations as follows:

A. *verb + noun*

In this type, the noun is always the subject, but in some other cases, the verb can be passive. Hoogland mentions a few examples as in: اشتدت الأزمة (the crises has intensified), تضاءل الأمل (hope faded) and انخفض السعر (the price dropped).

B. *verb + noun*

This is the second noun + verb collocation pattern in Hoogland's collocation classification, where the noun is the object. A few examples are mentioned by Hoogland as in, أحرز تقدما (he made progress) and منح جائزة (he awarded a prize).

C. *verb + preposition + noun*

This is the third collocation pattern in Hoogland's collocation classification model, where the noun is the indirect object after a preposition that comes with the verb. Some examples are mentioned by Hoogland as in, اعتذر عن مضايقة (apologized for bothering) and اختار بين بدائل (chose between alternatives).

3.5 Al-Hafiz 2004

Unlike Ghazala, Al-Hafiz only mentions the Arabic collocation patterns without providing the reader with the English equivalents to these patterns. He classifies collocation patterns, or as he calls them 'parts of speech', into twelve types of which only one focuses on verb-noun collocation as follows:

A. Collocation pattern 1: verb + noun

Al-Hafiz classifies the noun in this pattern into three subcategories:

    A1. Noun as a subject as in, هدأ الموج (the waves hushed)
    A2. Noun as an object as in, ضرب الخيمة (he stood the tent) and
    A3. Noun as a state as in, استشاط غضباً. (he flared up with anger).

Al-Hafiz states that this collocation pattern composes a large section of Arabic collocations where almost every single verb in Arabic has its own numerous noun collocates.

B. Collocation pattern 2: verb + prepositional noun phrase (noun as indirect object), as in إستقال من العمل (he resigned from work).
C. Collocation pattern 3: verb + prepositional noun phrase (phrase acts as an adverb) as in ضرب بشدة (he hit hard).

   D. Collocation patter 4: verb + noun phrase (noun in the form of adverbial condition) as in اتصل هاتفياً (he called telephonically).

3. 6 Ghazala 2006
Ghazala classifies collocation patterns into twelve types of which there are three that include verb-noun collocations (patterns 2, 8 and 9.4):

   A. Collocation pattern 2:

This pattern includes verb + noun collocation such as, يحضر محاضرة (attend a lecture),يبذل جهداً (exert an effort), يسن قانوناً (pass a law), يدير شركة (run a company), يلقن درساً (teach a lesson), ينتهز (pay attention), يعير الانتباه (pay attention), يحرز انتصاراً (score a victory), يكسب الثقة (win confidence), يستل سيفاً (draw a sward) الفرصة(seize the opportunity) and يستل سيفاً (draw a sward).

This collocation pattern is the primary in Arabic because such verb-noun collocation patterns are very common in Arabic. According to Ghazala, the central point in this pattern is to match a certain verb with a certain noun in Arabic.

The Arabic grammatical structure corresponding to the English structure is mostly a verb + noun or verb + object. However, some exceptions are mentioned by Ghazala as follow:

   • *Noun + verb:(i.e., verb + preposition + noun (i.e. prepositional phrase))*
   • pay a visit: يقوم بزيارة/ يزور
   • recognize a state: يعترف بدولة
   • shake hand : *verb* only (zero collocation because no noun is needed in Arabic) يصافح/ يتصافح

where in يزور /يقوم بزيارة, (pays a visit) and يعترف بدولة, (recognizes a state) the grammatical construction in Arabic is verb + preposition + noun (*i.e.* prepositional phrase), but in /يصافح يتصافح, (shakes hand) the collocation disappears in Arabic because the English verb and noun are translated into one word in Arabic.

   B. Collocation pattern 8:

This pattern is made of noun + verb collocations, but it is restricted to names of sounds, for example: نباح الكلاب (dogs bark) and مواء القطط (cats mew), رنين الاجراس (bells ring), دوي النحل (bees buzz), (dogs bark). According to Ghazala, the grammatical structure for this collocation pattern in Arabic is noun + noun (i.e. genitive), but a noun + verb can be possible in cases as in, سمعت الكلاب تنبح ليلة امس (I heard the dogs bark last night). This collocation pattern was mentioned earlier by Benson (1989), Benson et al. (1986) under C category in section 3.2 above. This collocation pattern is not included in this study because it requires word for word translation and doesn't lead to ambiguity.

   C. Collocation pattern 9.4

Collocation pattern 9 is made of four sub-patterns of which the fourth 9.4 sub-pattern includes verb + preposition as in, يشتاق إلى/ لـ (long for) and ينتظر (فلاناً) (wait for somebody).

4. Discussion and Eclectic Verb-Noun Collocation Model
Based on the five different collocations patterns discussed above in section 3, some concluding remarks can be made. The Nijmegen Arabic- Dutch- Arabic Dictionary Project is not included in this discussion because it is concerned with Arabic/ Dutch/ Arabic and not English/ Arabic lexicography.

Benson, Benson and Ilson try to come up with two separate collocation patterns: one grammatical (or as they call it, syntactic pattern) and the other lexical –both are based on the idea of parts of speech. Their grammatical collocation model demonstrates the different possibilities of nouns and adjectives with prepositions, to + infinites and that-clauses. They also include verb patterns within their grammatical collocation patterns model. Their lexical collocation pattern is more interesting in the sense that they try to give semantic labeling to the verbs according to the verbs semantic features. This semantic labeling can be used in MT engines as selectional restrictions for the polysemy of verb in sentences that are syntactically different. For other works concerned with verb semantic features, see Quirk *et al.*'s (1985) verb complementation (verb classification) and verb situation type. The verb-noun collocations in this model are focused on CA verb (usually transitive) + noun/pronoun (or prepositional phrase) and eradication and/ or nullification (EN) verbs.

Peter Emery's classification according to the relationship between collocates, *i.e.*, relational collocation patterns, is useful in drawing a clear-cut line between open, restricted and bound collocations, and idioms.

The next three collocation patterns, Hoogland, Al-Hafiz and Ghazala's, focus on Arabic collocation patterns:

Three verb-noun patterns are found in Hoogland's model. According to Hoogland, his "classification system for Arabic collocation, which aims at serving the learners of Arabic, should include both syntactic information and information based on parts of speech". Hoogland's model contains three interesting verb-noun collocation patterns for this study: noun as a subject, noun as an object and noun as indirect object after a preposition that comes with the verb. Obviously, Hoogland's aim to serve dictionary consultants is different from the computational aim of this study. For this reason, we may notice that the number of patterns he mentions is rather limited to only three patters.

Six verb-noun patterns are found in Al-Hafiz's model (patterns A, B, C and D. (pattern A has three sub-patterns). Patterns A1, A2 and B match those three found in Hoogland's model. Al-Hafiz's model demonstrates three more sub-patterns in Arabic than those found in the other two models; *verb + noun* (*noun as a state*), *verb + prepositional noun phrase* (*phrase acts as an adverb*) and *verb + noun phrase* (*noun in the form of adverbial condition*).

Three verb-noun patterns are found in Ghazala's model (patterns 2, 8 and 9.4), of which 9.4 matches Hoogland's pattern C and Al-Hafiz's pattern B. It is noticed that Ghazala's model does not include a basic pattern: verb + noun (the noun being the subject.) In spite of the limited number of verb- noun patterns mentioned in Ghazala's model, an electronic version of Ghazala's dictionary can be a very good source for NLP and MT because of its richness and arrangement of entries.

In an attempt to come up with an eclectic model of verb-noun collocation patterns in Arabic, we merged the three above verb-noun collocation models into one model. This model is illustrated in Table 1.

Table 1 Eclectic Noun-Verb Collocation Patterns in Arabic

1. verb + noun (*the noun being the subject*):	
تضاءل الأمل -	Hope faded.
تصعّدت الأوضاع -	The situation has escalated.
2. verb + noun (*the noun being the object*):	
يدير شركة -	runs a company
اعار الإنتباه -	paid attention
3. verb + preposition + noun (*noun as state*):	
إستشاط غضبا -	flared up with anger
4. verb + prepositional + noun/ phrase (*the noun/ phrase being the indirect object after a preposition that comes with the verb*):	
يشتاق إلى/ ل -	long for *something/ somebody*
إستقال من العمل -	resigned from work
5. verb + prepositional noun phrase (*the phrase acts as an adverb*):	
ضرب بشدة -	(he) hit hard (zero preposition in English)
6. verb + noun phrase (*noun in the form of adverbial condition*):	
اتصل هاتفيا -	(he) called telephonically

5. Conclusion

Classical Arabic lexicography books are futile for writers and translators in modern standard Arabic (MSA). Moreover, collocation pairs are not easy to locate in old Arabic lexicography books and dictionaries because of the subject order system, and the dryness and archaicness of the language used, which is difficult to understand even by many native speakers of Arabic today. Thus, these lexicography books are impractical to be used as sources for researchers working in the fields of NLP and MT, unless multilevel (adverb-noun, verb-noun, adjective-noun, etc.) scanning of collocates is executed manually, which is time and fund consuming, yet not impossible. It is time and fund consuming mainly because many collocates used in old lexicography books and dictionaries are not in use anymore in MSA. However, digitalizing these old lexicography books using Optical Character Recognition (OCR) or other modern technologies can be useful and practical for extracting collocation for the purpose of research on old literary Arabic. Considering the importance of Arabic as one of the main languages in the world, the number of collocation dictionaries in Arabic is limited in comparison with other main language. It is believed that more collocation dictionaries are needed to fill the demand in MSA for both, computational and non-computational purposes. Many scholars mention different collocation patterns in Arabic language. In an attempt to come up with a unified model in this article, the verb patterns in Hoogland (1993), Hafiz (2004) and Ghazala (2006) are merged into an eclectic model made of six patterns, out of which the first two are the most common verb noun collocation patterns in Arabic.

REFERENCES

'abdul-'zyz, Muhammad Hasan. (1990). *al-musahabat fi atta'bir al-lughawi* (in Arabic). Cairo: Dar al-Fikr al-'araby.

'abdul-'zyz, Muhammad Hasan. (1990). *al-musahabat fi atta'bir al-lughawi* (in Arabic). Cairo: Dar al-Fikr al-'araby.

Aisenstadt, E. (1979). Collocability restrictions in dictionaries in dictionaries and their users. *Exeter Linguistic Studies Exeter*, *4*, 71-74.

Ali, A. Y. (2005). The meaning of the Holy Qur'an: *Complete translation with selected notes*. Kuala Lumpur, Islamic Book Trust.

al-Iskafy, Muhammad ibn 'Abd Illah. (1985). *kitab mabadi' al-lught ma'a sharh abyat mabadi' al-lught* (in Arabic). Dar al-Kutub al-'Ilmiyat.

Anees, Ibrahim. (1984). *Dilalat al-alfaz* (in Arabic). Maktabat al-Anjlw al-Misriyyat.

Ath-Thaʻaliby, A. (1981). *Fiqh al-lughṯ* (in Arabic). Dar al-Kutub al-ʻIlmiyaṯ.

Baalbaki, R. M. (1990). *Dictionary of Linguistic Terms: English-Arabic with sixteen Arabic glossaries*. Dar El-Ilm Lilmalayin.

Benson, M. (1989). The structure of the collocational dictionary. *International Journal of Lexicography*, *2*(1), 1-14.

Benson, M., & Benson, E. Ilson. R. (1986). *The BBI Dictionary of English Word Combinations.* Amsterdam and Philadelphia: John Benjamins.

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied linguistics*, *2*(3), 223-235.

Emery, P. (1991). Collocation in modern standard Arabic. *Journal of Arabic Linguistics*, *23*(1), 56-65.

Emery, P. (1991). Collocation in modern standard Arabic. *Journal of Arabic Linguistics*, *23*(1), 56-65.

Emery, P. G. (1988). Collocation, a problem in Arabic-English translation. *Quinquereme-new studies in modern languages*. *11*(2), 178-184.

Ghazala, H. (2006). *Dar el-Ilm's dictionary of collocations: a comprehensive English-Arabic dictionary of accuracy of word combination and usage*. Beirut: Dar El Ilm Lilmalayin.

Hafiz Al-Tahir A. (2004). *Al-hafiz Arabic collocations dictionary*. Beirut: Librairie du Liban Publishers.

Hassan, Tammam. (1981). *al-ausuwl* (in Arabic). Casa Blanca: Dar al-Thakafaṯ.

Hassan, Tammam. (1985). *al-lughaṯ al-ʻarabiyyaṯ: maʻnaha wa mabnaha* (in Arabic). Cairo: al-Hayaṯ al-Misriyyaṯ al-ʻAammaṯ Lil Kitab.

Hassan, Tammam. (1990). *Manahij al-baḥt fi al-lughaṯ* (in Arabic). Cairo: Maktabaṯ al-Anjlaw al-Misriyaṯ.

Hassan, Tammam. (2005). *al-Khulasaṯ an nahawiyaṯ* (in Arabic). Cairo: ʻAlam al-Kutub.

Hassan, Tammam. (2006). *Maqalat fi al-lughaṯ wal adab* (in Arabic), (Vols. 1-2). Cairo: ʻAlam al-Kutub.

Haywood, John A. (1965). *Arabic lexicography*. Leiden: E. J. Brill.

Hoogland, J. (1993). Collocation in Arabic (MSA) and the treatment of collocations in Arabic dictionaries. In *Proceedings of the Colloquium on Arabic Lexicology and Lexicography*. Budapest: Eôtvôs Lorànd University and Csoma de Kôrôs Society (pp. 75-93).

Hoogland, J., & Versteegh, K. (1999). The Nijmegen Arabic/Dutch Dictionary Project. Retrieved from
http://wba.ruhosting.nl/Content1/1.2.3%20Aim_methods.htm#biblio

Hornby, A. S. (1975). *A guide to patterns and usage in English*. London: Oxford University Press.

Hornby, A. S., & Cowie, A. P. (1980). *Oxford advanced learner's dictionary of current English* (Revised and reset of the third edition 1974). New Delhi: R. Dayal, Oxford university press.

Ibn Qutaybṯ, ʻAbd Allah ibn Muslim. (1967). *Adb al-katib* (in Arabic). Dar Sadir.

Ibn-Syidh, Abu l-Hasan `Ali. (1903). Kitab al-Mukhassas (in Arabic). Al-Matba al-Kubra al-Amiriya.

Ibn-Syidh, Abu l-Hasan `Ali. (2000). Kitab Muhkam (in Arabic). Dar Al-Kotub Al-Ilmia.

Quirk, R., & Greenbaum, S. G. Leech, J. Svartvik (1985). *A comprehensive grammar of the English language*. Harlow: Longman.