Dirasat, Human and Social Sciences, Volume 49, No. 5, 2022

Deanship of Scientific Research

Using Technology to Compile an English-Arabic Glossary of the most Frequent Collocations in Ted Talks Parallel Corpus

***Nisreen Issa*[1], *Hanan Al-Jabri*[1]*, *Abdallah Abushmaes*[2]**

[1]Department of English Language and Literature, University of Jordan, Amman, Jordan

[2]Linguistic Research Manager, Mawdoo3, Amman, Jordan

Received: 19/2/2021
Revised: 19/5/2021
Accepted: 13/6/2021
Published: 15/9/2022

* *Corresponding author:*
hanan_aljabri@outlook.com

Citation: Issa, N., Al-Jabri, H., & Abushmaes, A. (2022). Using Technology to Compile an English-Arabic Glossary of the most Frequent Collocations in Ted Talks Parallel Corpus. *Dirasat: Human and Social Sciences*, *49*(5), 446–457. https://doi.org/10.35516/hum.v49i5.2751.

Abstract

Translation has evolved over the years and is now benefiting from the progress made in the field of technology and corpus linguistics. This study aims to use technology, particularly AntConc toolkit, to compile an English-Arabic glossary of the most frequent collocations listed in TED Talks parallel corpus. The glossary contains 1,000 unique English headwords and a total of 3,670 English collocations along with their Arabic translations as produced by Ted Talks translators and the researchers. Headwords in the glossary are alphabetically ordered and each collocation is accompanied with its frequency and its Arabic translation. The paper demonstrates the process through which the bilingual English-Arabic glossary has been compiled and the steps taken to process the collected data.

Keywords: Translation and technology, English-Arabic lexicon, collocations, corpus.

استخدام التكنولوجيا في بناء مسرد للمتلازمات اللفظية الأكثر شيوعاً من اللغة الإنجليزية للعربية

***سرين عيسى*[1]، *حنان الجابري*[1]*، *عبد الله أبو شميس*[2]**

[1]الجامعة الأردنية، عمان ،الأردن.

[2]شركة موضوع، الأردن.

ملخّص

شهدت الترجمة تطوراً كبيراً خلال العقود الماضية، واستفادت كثيراً من التقدم الهائل في مجال التكنولوجيا ولغويات المدوّنات. وتهدف هذه الدراسة إلى استخدام التكنولوجيا، وبالتحديد مجموعة أدوات AntConc لتحليل المدونات، في صناعة مسرد إنجليزي-عربي لأكثر المتلازمات اللفظية شيوعاً في مدونة TED Talks المتوازية. يحتوي المسرد على 1000 كلمة رئيسية فريدة باللغة الإنجليزية، وما مجموعه 3670 متلازمة لفظية باللغة الإنجليزية، مع مقابلاتها العربية التي استخدمها مترجمو Ted Talks وكاتبو هذه الورقة البحثية. وداخل المسرد جرى ترتيب الكلمات الرئيسية الإنجليزية ألفبائياً، وإضافة معلومات حول مقدار شيوعها، وترجماتها العربية. وتوضح هذه الورقة الآلية التي جرى من خلالها صنع هذا المسرد ثنائي اللغة، والخطوات المتخذة لمعالجة البيانات وتنظيمها.

الكلمات الدالة : الترجمة والتكنولوجيا، مسرد، متلازمات لفظية.

1. Introduction

Collocations have been the center of attention of many studies over the past few decades and are still gaining attention by many linguists, EFL students, translation students, and professional translators. This attention is probably derived from the fact that collocations are widely and commonly used by language users in different situations. For foreigner speakers, knowing how to use collocations in their appropriate context is perhaps a clear indication of how well a person is good at that language (Dechert and Lennon, 1989).

Ghazala (1995: 108) defines a collocation as "a combination of two or more words that always occur together consistently in different context in languages." The definition highlights the most important feature of collocations which involves two or more words that are frequently used together and sound natural. Therefore, the nature of the relationship among the words and the frequency of their usage together are essential in recognizing a collocation.

Translating collocations has always been one of the problematic areas that cast a great challenge for many translators because if translated literally, they will sound unnatural. For example, the collocation *strong tea* is translated into شاي ثقيل (heavy tea) in Arabic rather than the literal equivalent شاي قوي (strong tea). Therefore, consulting specialized dictionaries for collocations is important for translators to produce natural and correct translations.

While there are some English specialized dictionaries on collocations, such as *Oxford English Collocation Dictionary*, *BBI Combinatory Dictionary of English*, and the *Dictionary of Selected Collocations*, specialized dictionaries of Arabic collocations have only been created recently under the influence of other dictionaries in different foreign languages although Arab linguists and lexicographers showed an early interest in collocations as a linguistic phenomenon. To the best of the researchers' knowledge, there are three Arabic dictionaries of collocations: *Talal Abu-Ghazaleh Collocations Dictionary* (2012), and *Dar Al-Alam Dictionary of Collocations (* قاموس دار العلم للمتلازمات اللفظية) (2007), and *Al-Hafiz Dictionary of Arabic Collocations* (معجم الحافظ للمتصاحبات العربية) (2004).

However, most of the attempts outlined above were based on limited data in terms of volume and variety of text types. Most importantly, such attempts were largely based on written texts rather than spoken language. The present study aims at using computer software to compile a bilingual glossary of the most frequently used collocations based on Ted Talks parallel corpus which includes 39,288 words making it a reasonable source of the glossary. Furthermore, it includes a huge variety of topics from different specializations produced by different speakers. All English speeches are translated into Arabic by human translators which increases the quality of the results. The researchers will suggest Arabic renditions for any mistranslated or left out collocations.

Many scholars, and since the last century, have focused on the theoretical part of collocations and made very essential contributions to define and categorize collocations, such as Firth (1957), Halliday and Hasan (1976); Sinclair (1991); Wray (2002); Lambert (2004); and Conklin and Schmitt (2012). Other scholars examined the challenges faced by beginner and professional translators in rendering collocations from and into English and Arabic, such as Dweik and Abu Shakra (2011), Nofal (2012), and Bani-Younes (2015), among others.

This research, however, is an empirical research which intends to observe and process a large amount of data, particularly English collocations and their Arabic renditions, using software tools to compile a bilingual glossary of collocations consisting of the most 200 frequent collocations occurring in the spoken corpus using computer software. Although this research is an empirical research, it still contains a theoretical part in which the researchers present the definition and categorization of collocations, and then build on this conceptual data to seek new information from the most frequent patterns observed in Ted Talks parallel corpus along with their Arabic renditions. After compiling the glossary, the researchers aim to reflect on the final product, namely the glossary, analyze the detected patterns of renditions, keep the ones which seem natural, and propose a counter rendition for any mistranslated or left out English collocations.

2. Review of related literature

2.1. Collocations

Collocations are words that enjoy lexical relations which have come to be known from native peoples' frequent use. Collocations as a technical term have only been recognized when Firth (1957) investigated this linguistic phenomenon.

According to Firth (1957:11), a word is known by the company it keeps. Some words naturally collocate with specific words while they fail to collocate with others even if the alternative word is grammatically correct. Halliday (1966) took Firth's work to further steps introducing the term "set" which refers to a group of words that co-occur with one certain word. For example, *fine, gentle, light, slight, soft, heavy* are all words (set) that collocate with the word *rain*.

English collocations have been classified into different categories; the most common two are grammatical collocations and lexical collocations. According to *The BBI Dictionary*, grammatical collocations are those which consist of a dominant word and a grammatical structure. The dominant word or the head word can be a noun, an adjective, or a verb. The Dictionary classifies grammatical collocations into 7 main groups: Noun + preposition; noun + to + infinitive; noun + that clause; preposition + noun combination; adjective + preposition; predicate adjectives + to + infinitive; adjective + that clause. See table 1 below:

Table 1: Types of grammatical collocations in English

Group	Grammatical form	Examples according to the BBI Dictionary
G1	Noun + Preposition	Blockade against
G2	Noun + to + infinitive	A need to do it
G3	Noun + that clause	He took an oath that he would do his duty.
G4	Preposition + noun combination	To somebody's advantage
G5	Adjective + preposition	Angry at everyone
G6	Predicate adjectives + to + infinitive	It was necessary to work.
G7	Adjective + that clause	It was nice that he was able to come home.

Lexical collocations, on the other hand, do not consist of clauses or prepositions. They normally consist of nouns, adjectives, verbs, and adverbs. *The BBI Dictionary* categorizes lexical collocations into 7 groups: verb + noun/pronoun; verb meaning eradication + noun; adjective + noun; noun + verb; noun1 + noun2; adverb + adjective; verb + adverb. See table2 below:

Table 2: Types of lexical collocations in English

Group	Lexical form	Examples according to the BBI Dictionary
G1	Verb + noun/pronoun	Make an impression
G2	Verb meaning essential nullification + noun	Reject an appeal
G3	Adjective + noun	Strong tea
G4	Noun + verb	Blood circulates
G5	Noun1 + Noun2	A pack of dogs
G6	Adverb + adjective	Strictly accurate
G7	Verb + Adverb	Affect deeply

Moreover, collocations can be classified according to the link between their own words into two major groups: strong collocations and weak collocations. According to *Cambridge Dictionary*, a strong collocation is where the link of its words is restricted while a weak collocation is where the link of its words can collocate with other words. In other words, words in weak collocations are more flexible in joining other words. For example, *very hot* is a weak collocation because the adverb *very* collocates with many other words. However, *medical care* is a strong collocation because the headword collocates with a restricted number of alternatives.

The notion of collocations in Arabic is not different from that in English; collocations in Arabic are two or more words that are always associated together, and it is considered incorrect to replace one of these words with another. Many of the modern Arabic literature on collocations and lexical relations are drawn from the modern linguistic efforts of the English linguist Firth.

Similar to English collocations, Arabic collocations can be classified in different ways. For example, Abu Rub (2016: 78) classifies Arabic collocations into two types: open collocations and closed collocations. The first group, open collocations, is similar to weak collocations in the sense that the headword in this group tend to collocate with many words, such as انتصاراً where يحرز انتصاراً/ يسجل انتصاراً collocates with more than one word. Closed collocations, on the other hand, are not as flexible as they tend to collocate with a limited number of words only, such as حرب where حرب ضروس collocates with a very limited set of words.

In terms of structure, Hafiz (2002) classifies Arabic collocations into 12 grammatical patterns outlined in table 3 below:

Table 3: Arabic collocations patterns in terms of grammatical patterns

#	Form	Arabic Example	English equivalent
1	Verb + noun	هدأ الموج	The waves subsided
2	Verb + prepositional noun phrase, the noun is indirect object	استقال من العمل	He resigned from work
3	Verb + prepositional noun phrase, the phrase acts as an adverb	نفّذ بدقّة	He precisely implemented
4	Verb + noun phrase, the noun is in the form of adverbial condition	اتصل هاتفياً	He made a telephone call
5	Verb + conjunction + verb	طار وحلّق	He flew and soared
6	Noun + noun	مسرح الأحداث	The theater of events
7	Noun + conjunction + noun	عزم وإصرار	Intention and insistence
8	Noun + Adjective	قوة عظمى	A supreme power
9	Noun + prepositional noun phrase	غاية في الأدب	Extremely polite
10	Noun + Preposition	مقارنة بـ	In comparison with
11	Adjective + noun	حسن الاخلاق	Having high morals
12	Adjective + adverbial phrase	مستنكر بشدة	Strongly condemns

2.2. Corpus linguistics

According to Crystal (1992), corpus is "a collection of linguistic data, either compiled as written texts or as a transcription of recorded speech." The main purpose of compiling a corpus is to study the language in large amounts of texts or transcriptions in order to examine and analyze how certain words or sounds behave or occur in a language. With the development of technology and the emergence of new devices and applications, the need for linguistic databases has become important for machine education and programming; therefore, computer corpora have emerged and have been widely used to study languages and their notions**.**

Modern dictionaries are corpus-based dictionaries. Dash and Ramamoorthy (2019) noted that language corpus provides an empirical basis in the selection of words and other lexical items. Most big dictionaries in English, like *Oxford English Dictionary* and *Macmillan Dictionary* rely on corpora in their building to guarantee a high level of authenticity in all linguistic information they provide.

Perhaps one of the largest English corpora is The Bank of English. The corpus contains 200 million words of written and spoken English and is still expanding. It is morphologically and syntactically annotated and aims at constantly recording the actual use of English language. The project released a CD-Rom version that gives the user access to 140,000 English collocations and 2,600,000 sentences with these collocations.

Another example of using corpus is the software WordPilot 2000 which is an electronic lexicon developed for English-as-a-second-language learners, researchers and translators. It provides a list of words and phrases from which the user can select and see the meaning of the selected word. The user also can click on the option "Collocation" to view a summary of common collocates of the word selected. WordPilot is based on a corpus of 50,000,000 words.

3. Methodology

3.1. Data collection and software tools

As explained earlier, the glossary is based on Ted Talks parallel corpus. The corpus is open to the public and can be accessed and downloaded by users.

To process and analyze the data, the researchers used the free corpus analyzing toolkit, AntConc3.5.8, which was released in 2019. The software was developed by Laurence Anthony and is available for the public domain online. This tool assists in analyzing the Corpus by extracting the most frequently used collocations along with their Arabic renditions. AntConc has seen a rapid growth in popularity among researchers, teachers, and language learners mainly because it provides simple yet essential tools to perform speedy and accurate analysis for small and mid-sized corpora. Furthermore, it has a freeware license and easy-to-use interface. Finally, according to a major survey conducted by Tribble in 2012 regarding the most used tools by corpus linguists around the globe, three software tools are the most popular nowadays: corpus.byu.edu (a web-based corpus analysis tool 4th generation), WordSmith Tools and AntConc, all of which are easy to use, feature-rich and fast.

3.2. Steps of compiling glossary

The following steps outline the steps carried out by the researchers to collect the data, process the data, and build the bilingual glossary of the frequent collocations used in data. Some steps are accompanied by images which further demonstrate the process as being done.

1- The researchers downloaded Ted Talks Parallel Corpus 2013 available from https://opus.nlpl.eu/TED2013.php.

2- The researchers aligned both English corpus and Arabic corpus so as to be sentence based using Notepad++ and saved the results on an Excel sheet.

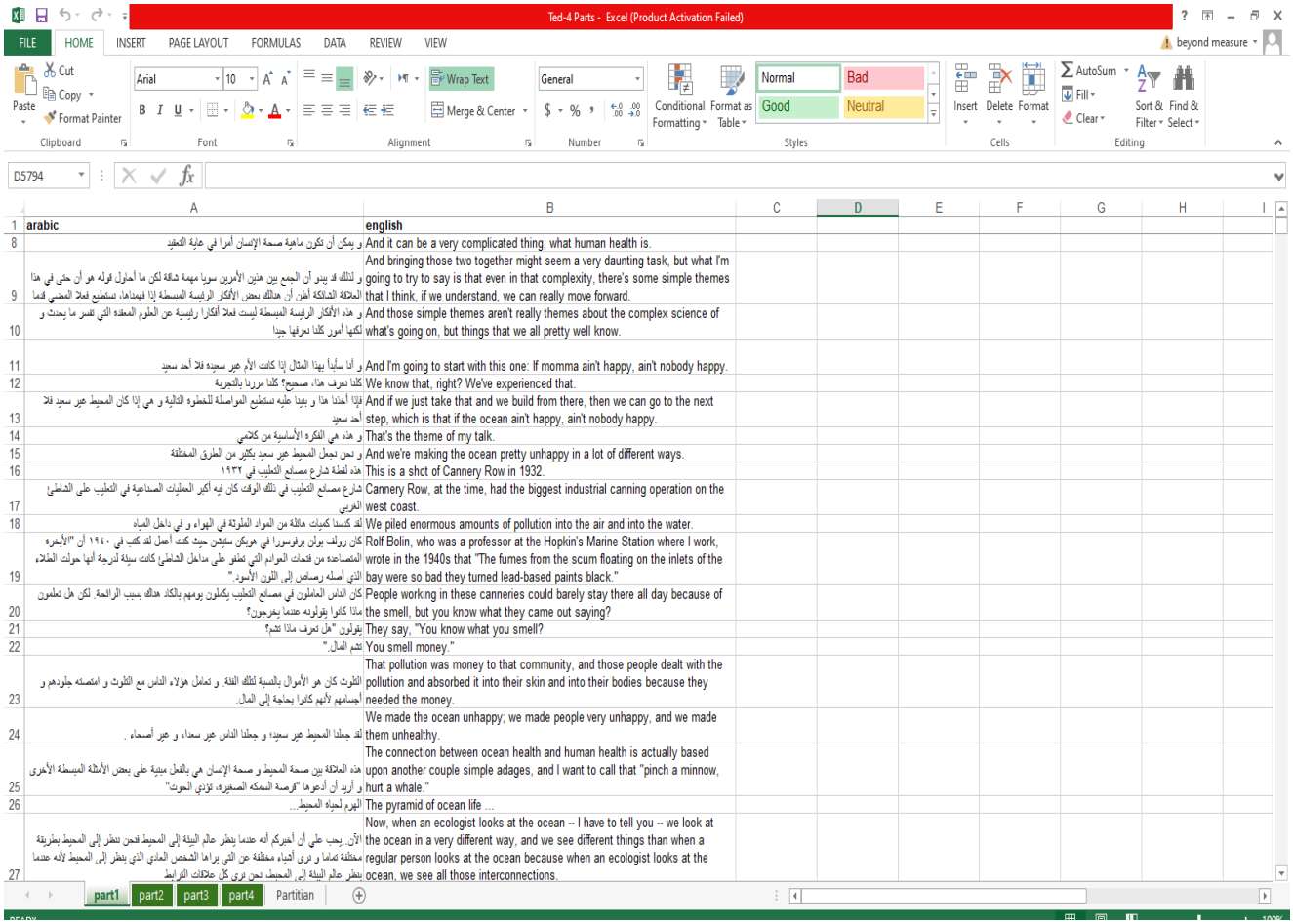


Figure 1: Alignment of data

3- The researchers used AntConc tool to process the resulted data.

4- The researchers applied the feature "word list" available in AntConc to extract a word list from the data. This feature provides word frequencies (i.e. counts) in the corpus and presents them in an ordered list. This helps to quickly identify which word is the most frequent in the corpus. Also, it allows the user to search for a certain word to find its frequency.

The resulted list contained 39,287 words extracted from the software on a notepad file.

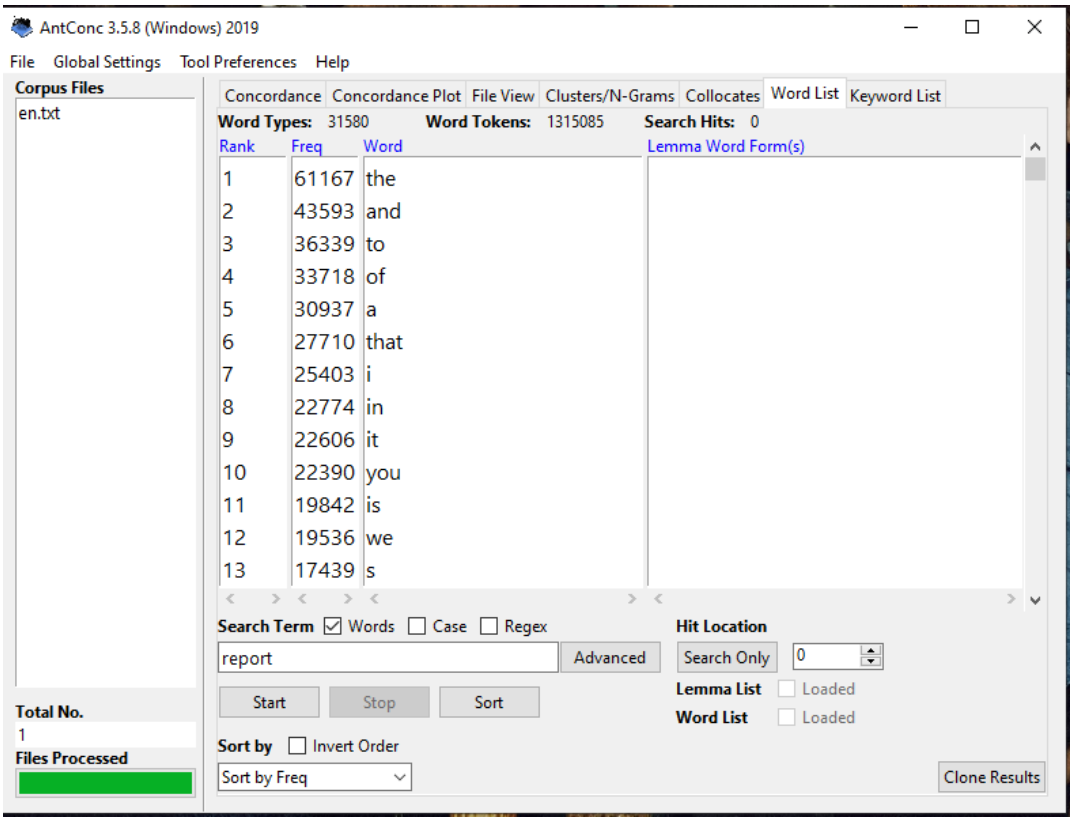
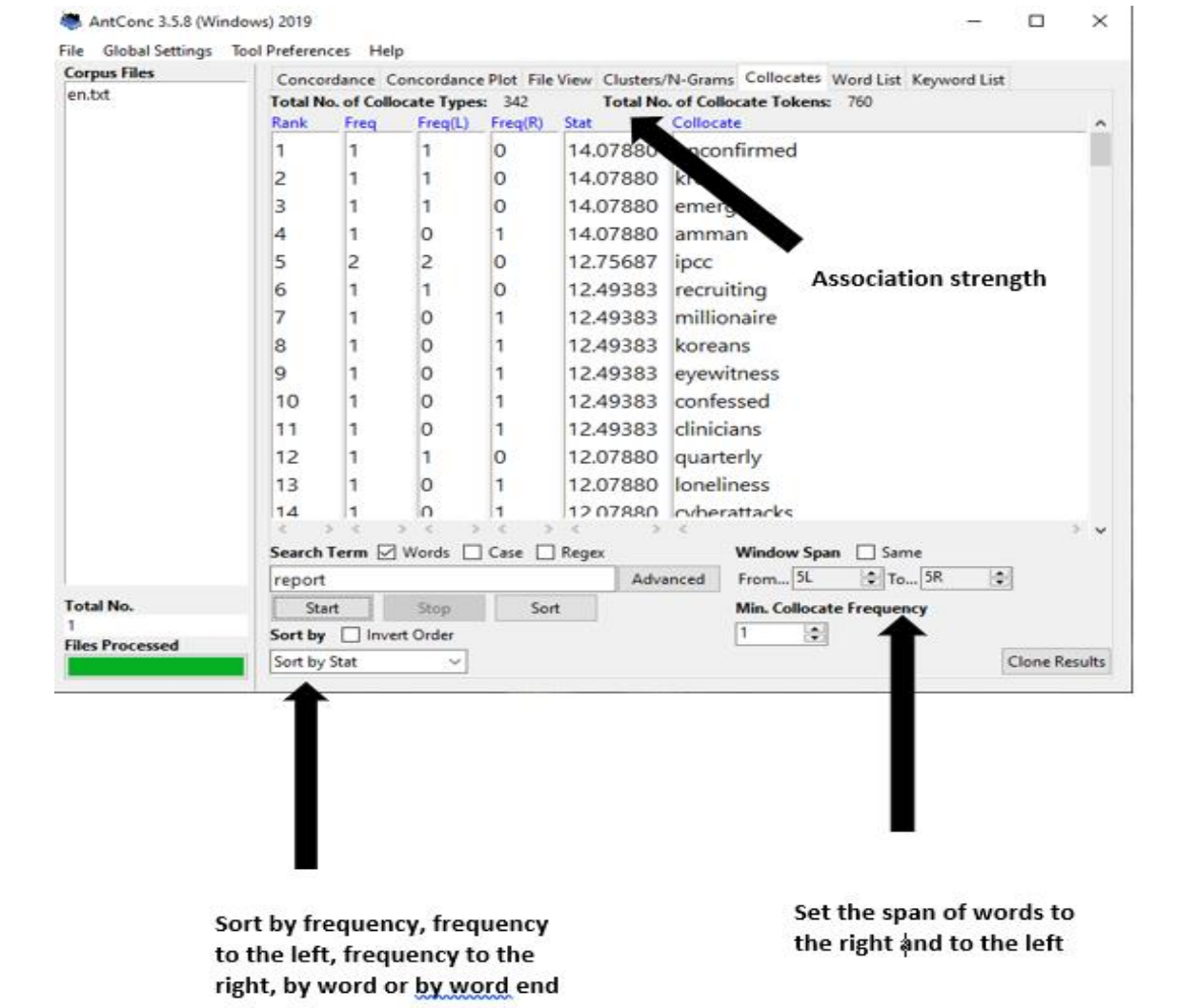


Figure 2: Antconc word list

5- The researchers zoomed in on the first 1,000 unique words and applied the "collocates" feature to each headword. This feature allows the user to search for words that often appear closely together. Different settings in "collocates" were applied to change the span of words to include words on the left of the headword and on the right of the headword. This helped in broadening the span of each headword and in making the glossary as inclusive as possible. For example, when working on the headword *belief*, the researchers managed to locate *strong belief* and *belief system.*



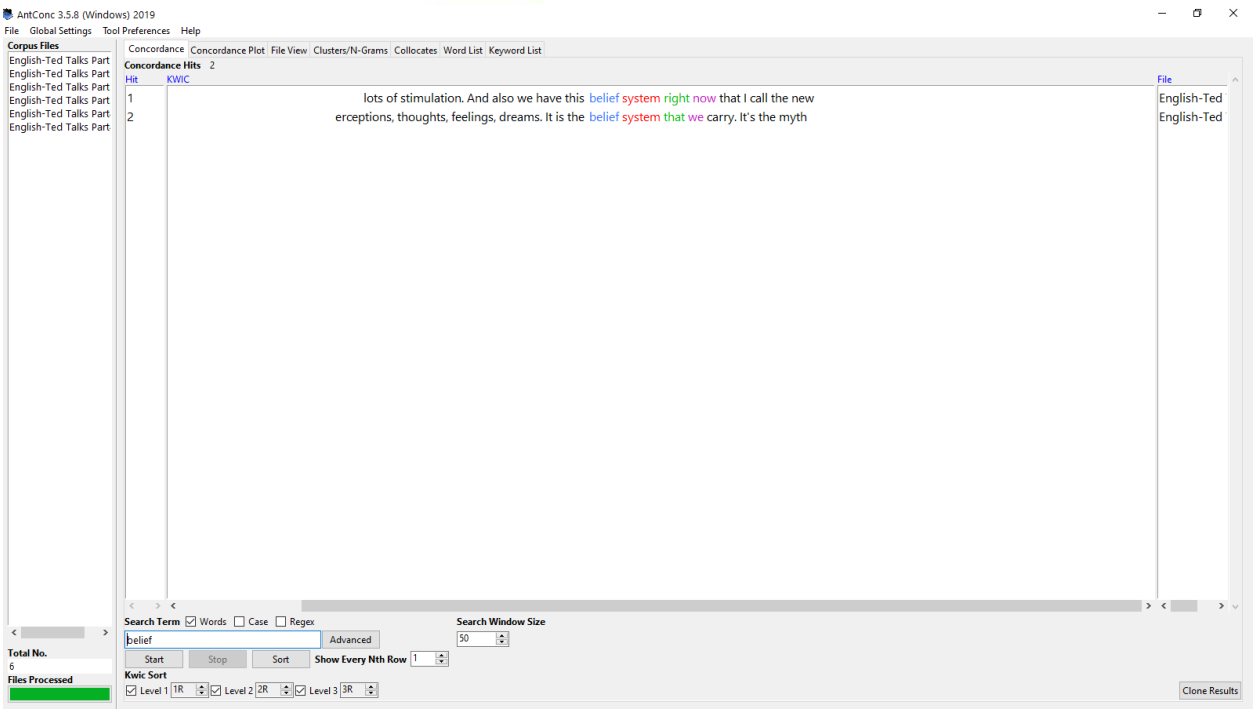


Figure 4: Applying different settings to locate collocations

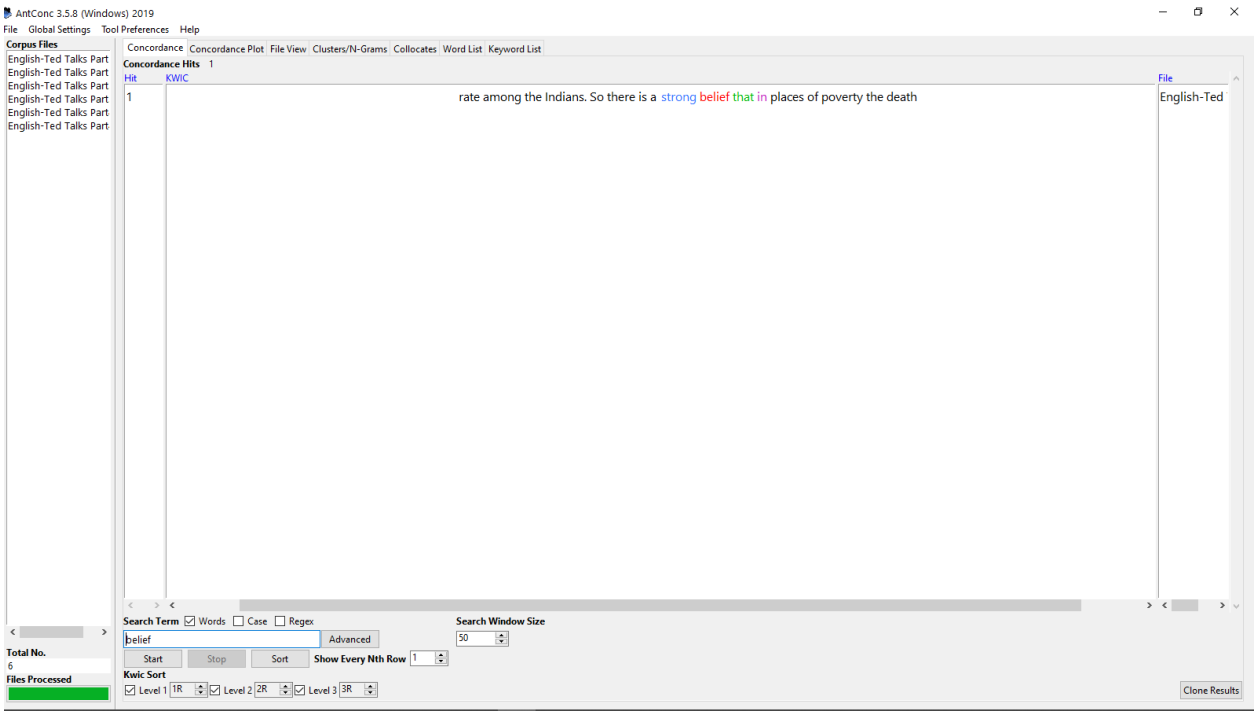
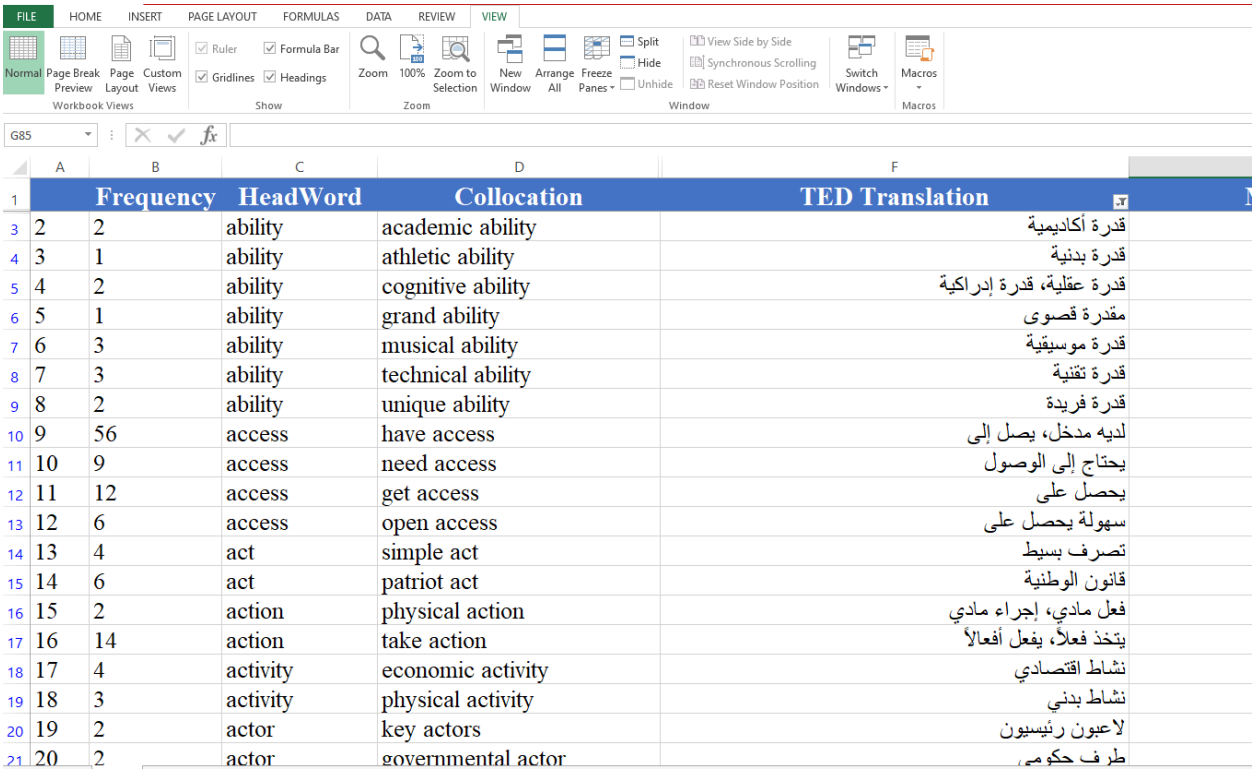


image 5: using different settings to locate collocations

6- The resulted collocations were extracted into an Excel sheet, and cleaned from any duplicated values.



	A	B	C	D	F
1		Frequency	HeadWord	Collocation	TED Translation
3	2	2	ability	academic ability	قدرة أكاديمية
4	3	1	ability	athletic ability	قدرة بدنية
5	4	2	ability	cognitive ability	قدرة عقلية، قدرة إدراكية
6	5	1	ability	grand ability	مقدرة قصوى
7	6	3	ability	musical ability	قدرة موسيقية
8	7	3	ability	technical ability	قدرة تقنية
9	8	2	ability	unique ability	قدرة فريدة
10	9	56	access	have access	لديه مدخل، يصل إلى
11	10	9	access	need access	يحتاج إلى الوصول
12	11	12	access	get access	يحصل على
13	12	6	access	open access	سهولة يحصل على
14	13	4	act	simple act	تصرف بسيط
15	14	6	act	patriot act	قانون الوطنية
16	15	2	action	physical action	فعل مادي، إجراء مادي
17	16	14	action	take action	يتخذ فعلاً، يفعل أفعالاً
18	17	4	activity	economic activity	نشاط اقتصادي
19	18	3	activity	physical activity	نشاط بدني
20	19	2	actor	key actors	لاعبون رئيسيون
21	20	2	actor	governmental actor	طرف حكومي

Figure 5: Collocations extracted on Excel sheet

7- A quality check was conducted to ensure the validity of the collocations and that each collocation is related to the right headword. The validation process involved consulting *Oxford English Collocations Dictionary*.
8- The resulted collocations were arranged alphabetically according to the headword to form the glossary. The glossary shows each collocation along with its frequency and its Arabic translation produced by Ted talks translators.
9- In case of mistranslations or left out collocations, the researchers suggested a translation after consulting specialized dictionaries and expert linguists.
10- The total number of collocations resulted from the 1,000 headwords is 3733.
11- The full glossary is available at *https://figshare.com/s/dec71ce53741ca8b1341* . Table4 below shows a sample of the glossary:

Table 4: Sample of glossary

ID	Frequency	Headword	Collocation	Arabic Translation
1	37	ability	academic ability	قدرة أكاديمية
2	2	ability	athletic ability	قدرة رياضية
3	1	ability	cognitive ability	قدرة إدراكية
4	2	ability	grand ability	مقدرة قصوى
5	1	ability	have ability(ies)	لديه قدرة (ات)
6	3	ability	musical ability	قدرة موسيقية
7	3	ability	technical ability	قدرة تقنية
8	2	ability	unique ability	قدرة فريدة
9	56	access	get access	يحصل على
10	9	access	have access	يصل إلى
11	12	access	need access	يحتاج الوصول
12	6	access	open access	مُتاح للجميع
13	4	act	patriot act	فعل وطني
14	6	act	simple act	تصرف بسيط
15	2	action	physical action	إجراء مادي
16	14	action	take action	يتخذ اجراءً، يتصرف
17	4	activity	economic activity	نشاط اقتصادي
18	3	activity	physical activity	نشاط بدني
19	2	actor	governmental actor	جهة حكومية
20	2	actor	great actor	ممثل بارع

4. Results and Discussion

4.1. Categories of collocations

After applying the software tool in question to the parallel corpus and concluding all the steps to process the data as highlighted in the previous section, four major categories of the Arabic collocations were detected in the data: collocations successfully rendered into natural Arabic; collocations whose translations could not be found in the Arabic TED corpora; mistranslated collocations; collocations considered collocations only in English.

4.2. Collocations successfully rendered into natural Arabic

The first category of collocations is collocations that TED translators managed to render into natural Arabic equivalents. This category constitutes most of the glossary. The translations in this category seem to follow the linguistic system of the Arabic language and, therefore, they read smoothly to the Arabic audience. Moreover, the researchers consulted Arabic linguists to review and proofread the translations. The following are some examples on successful translations carried out by TED translators.

Table 5: Examples on successful translations carried out by TED translators

Frequency	English Collocation	Arabic Translation
6	digital camera	كاميرا رقمية
2	university campus	حرم جامعة
2	conference table	طاولة الاجتماعات
3	armed conflict	صراع مُسلح
12	developing country	دولة نامية
25	credit card(s)	بطاقة ائتمان
8	world cup	كأس العالم
2	economic damage	ضرر اقتصادي
9	opening day	يوم الافتتاح
2	development index	مؤشر التنمية
3	physical disability	إعاقة جسدية
6	realize dream(s)	تحقيق الحلم
26	solar energy	طاقة شمسية
5	have faith	يتحلى بالإيمان
4	look familiar	يبدو مألوفاً
4	cease fire	وقف إطلاق النار

4.3. Left out collocations

The second category of collocations is those whose translations were not found in the Arabic TED corpora. The total number of collocations in this category is 47 only. In this case, the researchers provided Arabic translations for these colocations depending on the context in which they were used in the corpus. See table 6 below for more examples about this category.

Table 6: Examples on left out collocations

Frequency	Collocation	Translation
1	broken arm	ذراع مكسورة
1	survive attack	ينجو من الهجوم
1	make attempt	يجري محاولة
1	blood results	نتائج تحليل الدم
1	glass boat	قارب زجاجي
1	boundary fence	سياج حدودي
1	national channel	قناة محلية
1	diagnostic characteristics	خصائص تشخيصية
1	charge fees	يتقاضى رسوماً

4.4. Mistranslated collocations

The third category includes collocations that were mistranslated by TED translators. These collocations do not seem natural within the cultural and the linguistic systems of the Arabic language. Arabic linguists reviewed the collocations in this category and helped the researchers in providing alternative Arabic translations based on the context in which they were used. Table 7 below includes some examples which belong to this category:

Table 7: Mistranslated collocations

Frequency	English Collocation	TED translation	Alternative translation
4	bad apple	تفاح فاسد	شخص سيئ
2	big yes	نعم كبيرة	يوافق بشدّة
2	blowing wind	تهب الرياح	رياح عاتية
3	take train	يأخذ القطار	يستقل القطار
6	paper towel	منشفة ورقية	مناديل ورقية
5	high technology	تكنولوجيا عالية	تكنولوجيا متطورة
1	wildly spread	ينتشر بعنف	ينتشر إلى حد كبير

4.5. Collocations rendered into a single word

The fourth category involves those entries that are considered collocations only in English. When rendered into Arabic, these collocations become single words and, therefore, they are not considered collocations in Arabic. Examples of these entries are as demonstrated in table 8 below:

Table 8: Collocations rendered into single words

Frequency	Collocation	Translation
3	front desk	الاستقبال
3	full stop	نقطة
3	capital city	عاصمة
40	young woman	شابة
4	take a walk	يتمشى
4	take turns	يتناوب
6	get upset	ينزعج
10	get tired	يتعب

4.6. Strategies of translating collocations

TED translators seem to have used three main strategies to translate English collocations used in the corpus. The first strategy is maintaining the meaning and the form of the English collocation while achieving a natural translation in Arabic. For example, the collocation *free fall* was rendered into سقوط حر.

The second strategy is literal translation where translators transferred the form of the collocations producing awkward Arabic renditions. For example, the collocation *take train* was rendered as يأخذ قطاراً which sounds unnatural in Arabic. The last strategy of translating collocations detected in corpus is omission. Some collocations, as explained earlier, were left out and in this case the researchers provided suitable renditions based on their context.

5. Conclusion

Having an excellent knowledge in collocations is an indication of fluency and proficiency; therefore, it is of great importance for translators to have reliable and various resources to consult when translating collocations such as collocation dictionaries, online collocations databases, and collocations glossaries.

There are few English language dictionaries on collocations such as the *Oxford English Collocation*, yet these resources are still considered scarce and not as comprehensive as other traditional dictionaries. As for the Arabic language, there are very few attempts to compile specialized dictionaries or glossaries for Arabic collocations. The most recent publications in this area are only three dictionaries: *Al-Hafiz Dictionary of Arabic Collocations* (2004), *Dar Al-Alam Dictionary of Collocations* (2007), and *Talal Abu-Ghazaleh Collocations Dictionary* (2012).

The translation industry has been evolving rapidly due to the technological advancements and machine learning;

consequently, the translation process nowadays depends heavily on CAT tools and Translation Memories (TMs), and most translation projects are large in volume and the time needed to handle such translations are relatively short, so there is a clear need for bilingual dictionaries specialized in collocations available to the public domain. Such projects will benefit a huge number of translation students, language learners and professional translators.

The researchers studied the first 1,000 unique headwords only out of 39,288. Therefore, it is important to emphasize that the current project is an ongoing work; more headwords will be included along with their translations following the same process detailed in this paper. As explained before, the corpus used to compile this glossary is available to the public which means the project can be contributed to by anyone. This project is only the starting point; once completed, it will be a helpful resource for language learners and translators.

Using technology in translation and language studies is a relatively new field of study where many new ideas and projects can be implemented.

References

Al-Maany. (n.d.). https://www.almaany.com/ Accessed 19 October 2020.

Bani-Younes, M. A. (2015). "Cultural and Sociolinguistic Issues in English-Arabic Translation of Collocations." *Studies in Literature and Language*, 10(6), 53-58.

Conklin, K. and Schmitt, N. (2012). "The Processing of Formulaic Language." *Annual Review of Applied Linguistics*, 32, 45–61.

Crystal, D. (1991). *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell Publishers.

Dash, N. S. and Ramamoorthy, L. (2019). *Corpus and Dictionary Making*. In: Dash, N. S. and Ramamoorthy, L. (ed.) *Utility and Application of Language Corpora*, 121-138. Springer.

Dweik, B. and Abu Shakra, M. (2011). "Problems in Translating Collocations in Religious Texts from Arabic into English." *The Linguistics Journal*, 5(1). 5-41.

Firth, John R. (1957). *Papers in Linguistics*. Oxford: Oxford University Press.

Ghazala, H. (1995). *Translation as Problems and Solutions*. ELGA publication.

Halliday, M.A.K. (1966). *Lexis as a Linguistic Level*. In: Bazell, C.E., Catford, J.C., Halliday, M.A.K. and Robins, R.H. (eds). *In Memory of J. R. Firth*. 148–163. London: Longman.

Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English*. London: Longman. J.

Lambert, B. (2004). *Statistical Identification of Collocations in Large Corpora for Information Retrieval.* University of Illinois at Urbana-Champaign.

Mahmoud, A. (2005). "Collocation Errors Made by Arab Learners of English." *The Asian EFL Journal*, 5, 1-9.

Nofal, K. (2012). "Collocations in English and Arabic: A Comparative Study." *English Language and Literature Studies*, 2(3), 75-93.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Wray, A. (2002). *Formulaic Language and the Lexicon*. New York: Cambridge University Press.