Journal of Modern Languages
Vol. 25, (2015)

Evaluation of Google and Bing online translation of verb-noun collocations from English into Arabic

Hussein Soori
University of Malaya, Malaysia

Su'ad Awab
University of Malaya, Malaysia

Abstract

This article aims to investigate and evaluate the translation of verb-noun collocation in English into Arabic Google and Bing online translation engines. A number of sentences were used as a testing dataset to evaluate both engines. Human translations by three bilingual speakers were used as a gold standard. A simple evaluation metric was proposed to calculate the translation accuracy of verb-noun collocations. The results showed that Bing scored a verb-noun collocation value of 0.72 with a trend estimation ranging between 0.65 and 0.67. Google scored a verb-noun collocation value of 0.75 (3% higher than Bing) with a trend estimation ranging between 0.63 and 0.85. The results also showed that, in most cases, the Arabic translation output of both engines produced a one verb synonym which did not collocate with the different nouns in the testing data sentences. These results indicate that Google and Bing, so far, have not been able to resolve the verb-noun collocability problem in their Arabic output. This study and its results may shed some light on the problem and to develop new methods to improve Arabic verb noun collocability in the output translation of current machine translation engines.

Keywords: English-Arabic MT, verb-noun collocations, machine translation evaluation, evaluation metric, Arabic verb synonymy

1. Introduction

The rationale behind this study lies in the fact that improving the output of any translation engines requires one to have adequate knowledge of the degree of accuracy that can be achieved by a translation engine. Commercial information technology companies tend for competitive reasons not to publish detailed literature and research results about the methods, algorithms and sources used in the making of their translation engines. One way to evaluate the effectiveness of these engines is to compare the output of these engines, and testing it, against a set of gold standard translations for evaluation purposes.

This article investigates Arabic verb-noun collocation in Google (2015) and Bing (2015) online translation engines. It also evaluates the translation accuracy of verb-noun collocations in the Arabic output of these two widely used translation engines. In this article, a number of sentences are used as a testing dataset. For evaluation purposes, human translations by three bilingual speakers are used as gold standard. In addition to that, a simple evaluation metric is proposed to calculate the percentage of verb-noun translation value. Evaluation metrics is applied equally to both Google and Bing translations of the testing datasets. The results of this study are discussed and illustrated with figures and line charts.

2. Statement of the Problem

This article evaluates verb-noun collocation in English-Arabic Machine Translation (MT) engines. The problem this article aims to highlight is the semantically inaccurate output that is generated when a wrong synonymous verb is collocated with a noun by translation engines in the Arabic Target Language (TL).

The problem can be explained in the following examples 1-4 below (Hornby & Cowie, 1980: *stand*, entry 6: [to be placed in an upright position]. A human translation is provided in Arabic for illustration with the English back translation below the Arabic translation.

Table 1. Human translation of examples 1-4

1. The carpenter's wife stood the ladder against the wall.	اسندت زوجة النجار السلم على الجدار [The carpenter's wife stood the ladder against the wall]
2. The woman stood the little girl on the chair.	اوقفت المرأة الفتاة الصغيرة على الكرسي [The woman stood the little girl on the chair]
3. They stood the tent near a tree.	نصبوا الخيمة بقرب شجرة [They stood the tent near a tree]
4. She stood the toy on the bookshelf.	اوقفت الدمية على رف الكتب [She stood the toy on the bookshelf]

The problem here is related to the synonymous choice of the TL verb where the equivalent of the SL verb, *stood* in 1-4 in Table 1 is the three different synonyms in Arabic: اسند, *asnada*, اوقف, *awqafa*, and نصب, *nasaba*. This is found frequently in many verbs in dictionaries, for which derivations of the same sense for the SL verb may be deployed in many synonymous equivalent TL verbs.

The researchers have tested examples 1-4 given above by inserting them in Google and Bing online translation engines. The translations are shown in Table 2 below.

Table 2. Google and Bing translations of examples 1-4

Google online	**Bing online**
زوجة النجار وقفت سلم ضد الجدار [The carpenter's wife stood a ladder against the wall]	وقفت زوجة النجار السلم ضد الجدار. [The carpenter's wife stood the ladder against the wall]
وقفت امرأة فتاة صغيرة على الكرسي. [The woman stood a little girl on the chair]	امرأة وقفت فتاة صغيرة على كرسي. [The woman stood a little girl on a chair]
وقفوا خيمة بالقرب من شجرة. [They stood the tent near a tree]	وقفوا الخيمة قرب شجرة. [They stood the tent near a tree]
وقفت لعبة على رف الكتب. [She stood the toy on the bookshelf]	وقفت اللعبة على الرف. [She stood the toy on the shelf]

Obviously, what we see in the above Google and Bing translations is a classic case of total ignorance of noun-verb collocation in the TL. In Arabic the verb synonym نصب, *nasaba* is used to stand a tent, pillar or a pole, etc., the verb

synonym اسند, *asnada* is used to stand a ladder and the verb synonym اوقف, *awqafa* is used to stand an object that has two legs or to stand a human. Instead, the verb stand is replaced in both translation engines in the TL by the verb اوقف, *awqafa* in all translations. Thus, Google and Bing assign only one TL verb synonym in translating the sentences 1-4 above; hence, in this article verb-noun collocability in Arabic serves as the criterion of evaluating Google and Bing.

3. Literature Review: Evaluation of Machine Translation Engines

The accuracy of an online translator is usually evaluated by comparing the results against human judgments. Most evaluation methods depend on a number of scales: adequacy, fluency, intelligibility, fidelity and informativeness. However, different evaluation methods use different scales depending on their purpose.

In MT evaluation, the scales of evaluation are called metrics. A metric is a measurement that evaluates the machine translation output and its quality. Banerjee *et al*. (2005) describe attributes of a good metric. According to them, a good metric is one that correlates highly with human judgment, and should be consistent in terms of the equality of scales put on the engines at hand and the way results are obtained from these engines (*i.e.*, consistency of the gold standards and the testing data used for evaluation), and work with different text domains.

One of the first metrics used to evaluate MT engines is Word Error Rate (WER). WER is derived from the Levenshtein Distance as introduced by Vladimir Levenshtein in 1965 (Levenshtein, 1966). WER metric calculates the distance between the words of the system output and the words of the reference translation divided by the length of the reference translation. This metric tries to determine the optimal alignment between the MT output and the reference translation, with each word in the MT output aligning to either 1 or 0 words in the reference translation, and vice versa. In cases in which there is no counterpart found to a reference word, it is labeled as 'deletion'. The alignment of a word from the MT output to nothing is labeled as an 'insertion'. If a reference word matches the MT output word, then it is labeled a match or a substitution. Thus, the WER is the

total of the number of substitutions, insertions and deletions, divided by the number of words in the reference translation. The equation is written as follows.

wer = substitutions + insertions + deletions / (all divided by reference-length)

However, the WER metric does not allow calculation of synonymous words.

Another one of the most common metrics to evaluate machine translations is BLEU. BLEU is based on the concept of closeness of the output of a MT system to a reference or a human translation of the same text. This closeness is determined by a modified n-gram precision as proposed by Papineni *et al.* (2002). The metric is based on counting the number of common words in the candidate translation and the reference translation, then, dividing the number of common words to the total number of words in the candidate translation. However, any higher or lower number of n-grams (words or word variants) is penalized which affects the overall evaluation of the output of a given MT engine. According to BLEU metric, a good engine is good only if it produces the same number of n grams in the reference sentence. In other words, the translation of a short sentence should be a short TL sentence, and the translation of a long sentence should be a long TL sentence.

The BLEU metric does not work well for calculation of synonymous words, as one word in the SL may not allow more than a one synonym in the TL and vice versa.

In 2002, the US National Institute of Standards and Technology (NIST) presented a modified version of the metric used in BLEU. The new version is known as NIST. NIST metric tries to compute how informative a particular n-gram is, where a low frequency of a particular n-gram means yielding a higher weight, while a high frequency of a particular n-gram means yielding a lower weight (Goutte, 2006). In other words, rarer words or phrases, such as *impressive observations*, are given more weight than some other words or phrases that are less rare such as, *to the*. NIST metric also differs from BLEU in terms of not penalizing slightly longer counterpart translations. Although both BLEU and NIST are widely used, these two evaluation metrics are language independent in the sense that they both ignore linguistic features.

Qin *et al.* (2009) conducted an experiment to test n-gram co-occurrence in BLEU and NIST. Their experiment involved replacing the words in the reference translation with synonyms. Their study shows that both metrics have low evaluation rates. Their experiment also revealed that the enhancement of both BLEU and NIST is correlated to human evaluation.

To overcome the weaknesses of the above mentioned metrics, a new metric was proposed by Lavie *et al*. (2004), known as METEOR. The concept of METEOR is rather different from that of the above metrics in that all the other metrics relied on unigram precision (the two SL and TL identical word strings) only, while METEOR gives more weight to unigram precision and unigram recall. This gives the application more room, in the TL output translation, to deal with morphological variants (words with an identical stem) and synonymous words.

This metric would have suited the purpose of evaluating the verb-noun collocation translations studied in this article. Unfortunately, the METEOR developing group stated on their website that their stem and synonym matches do not cover Arabic yet, as shown in Table 3.

Table 3. Fully supported languages in METEOR (from Denkowski & Lavie, 2011)

Language	**Exact Match**	**Stem Match**	**Synonym Match**	**Paraphrase Match**	**Tuned Parameters**
English	Yes	Yes	Yes	Yes	Yes
Arabic	Yes	No	No	Yes	Yes
Czech	Yes	No	No	Yes	Yes
French	Yes	Yes	No	Yes	Yes
German	Yes	Yes	No	Yes	Yes
Spanish	Yes	Yes	No	Yes	Yes

No reason is given on their website, but a possible explanation for not including the stem and synonym matches for Arabic is language specific. Morphologically, Arabic is one of the highly inflectional languages, and semantically, Arabic is rich in vocabulary and synonymous words. In order to include a strong stemming tool in translation evaluation software, one needs to

come up with a strong morphological analyzer first. However, the scale of challenge is very high when it comes to computational stemming of Semitic languages such as Arabic and Hebrew. Arabic has an unusual system of inflection, as various stems can be formed from the root.

The verb system in Modern Standard Arabic has ten standard forms and four exceptional forms: hamzated, weak, quadriliteral, and doubled verbs. By multiplying that by 90 conjugation rules for each verb form, one can imagine the nature of complexity researchers face when dealing with stemming in Arabic.

Computational segmentation in Arabic involves many different levels. Many attempts have been made to overcome this problem. Buckwalter's stemmer (2002) contains three categories: a prefix, a stem, and a suffix aiming to minimize the size of the prefix, suffix and stem dictionaries. The task here then becomes to list all the possible prefixes and suffixes that come with every stem. On the other hand, tree banks decision of splitting or segmentation is based on parts of a prefix, parts of a suffix, and a stem which can still have parts from both, the prefix and suffix.

In English, a verb lexeme appears in three or four different forms plus the present participle. In German, the verbs have seven to ten verb forms for one lexeme, and additional four for the present particle (Breidt, 1993). In Arabic, a verb lexeme appears in a dictionary in the form of third person singular masculine of the past tense in the fourteen forms mentioned above (ten standard forms plus four exceptional forms). Roughly speaking, every Arabic verb has around 122 variants.

Another difference that poses a challenge in Arabic text processing is that the lexeme is agglutinated with the prefixes, infixes, suffixes, prepositions, articles and pronouns. The affixes form tense, number, mood, aspect, and so on in the case of verbs, and number, gender, definitiveness, case and the possessive clitic, in the case of nouns.

These inflectional features have been studied and subcategorized. However, so far, no strong and complete morphological tool has been made available to resolve the morphological complexity in Arabic.

In conclusion, WER, BLEU, and NIST all lack the ability to evaluate synonymous words as appropriate matches. METEOR too has not been able to include stem and synonymous matches for Arabic yet. This calls for a new automatic metric that evaluates the TL (Arabic translations), and deals with morphological variants (words with an identical stem) of a word, and with synonymous matches in the Arabic language.

In section 6, a simple metric is proposed to evaluate the translations of the testing data in the present study.

4. Testing Data

The testing dataset contains sentences from various domains. The verbs in the testing dataset in this study are selected from among those verbs that pose a problem for machine translation (MT) when translated from English into Arabic. The testing dataset list is made of groups of two or more sentences that contain verbs with the same lexical meaning in the Source Language (SL), but require different synonymous verbs in the Target Language (TL) due to the collocational relations between the noun and the verb in the Arabic translations.

The testing data list contains sentences with verb-noun collocates and it is divided into two parts. The first part contains sentences where the noun is an object and the second part where the noun is a subject. The lexical meanings of the SL verbs are taken from the *Oxford Advanced Learners' Dictionary* (2013). A list of sentences chosen as testing data sets is shown in Table 4. A human translation of this list is discussed in the following section.

Table 4. Testing dataset

Verbs that collocate with object nouns
1. The verb, *close*: End (verb) 5: to end or make sth end.
1a. The director *closed* the deal.
1b. The investigator *closed* the case
2. The verb, *put on:* Put sth on (verb) 1: to dress yourself in sth.
2a. The man *put* the shoes *on*.
2b. The man *put* the shirt *on*.
2c. The boy *put* a cap *on*.
3- The verb, *pass:* Law/ proposal (verb) 14: to accept a proposal/ law by voting.
3a. The parliament *has passed* the law.

3b. The parliament *has passed* the proposal.

4. The verb, *win*: (verb T) 3: to achieve or get sth that you want, especially by your own efforts.
4a. The Counselor *won* the support for his proposal.
4b. The activist *won* the admiration of many people in his battle against cancer.

Verbs that collocate with subject nouns

5- The verb, *drop:* Become weaker/ less (verb I,T) 4: to become or make sth weaker, lower or less.
5a- The price *dropped* dramatically
5b- The temperature *dropped* tonight.
5c- The singer's voice *dropped* suddenly.
5d- The Dutch team *dropped* to the fifth place.

6- The verb, *fall:* Happen/ occur (verb I) 9: to come quickly and suddenly.
6a- A sudden silence *fell* upon the room.
6b- Darkness *fell* quickly tonight.

7- The verb, *fall:* (verb I) 11: to move in a particular direction or come in a particular position.
7a- My eye *fell* on a strange object.
7b- The regime *fell* after 1989.

5. Human Translation of Testing Data

When evaluating an output of a given MT engine, one always needs a reference or, as it is called in MT, a gold standard. The gold standard comes in the form of proposed or suggested translations by human translators to validate the TL translations of the testing data. This gold standard also serves as an evaluator to the output of the MT engine at hand.

For the purpose of this study, three fluent bilingual speakers and experienced translators of Arabic and English were asked to validate the verb-noun collocates in the list of dataset sentences twice. The three experienced translators are teachers of translation with over 10 years experienced. Two of them are specialized in literary translation and one in legal translation. The three of them are authorized licensed translators. The first time, each of them was given the dataset in Table 4 and the Arabic translation of the testing data sentences without the TL verbs. Each was asked to suggest an accurate TL translation of the verbs that collocates with the nouns based on the SL sentences in Table 4. The verb translations they suggested are listed in Table 5.

Table 5. The first set of suggestions by the three bilingual speakers

SENTENCE (V+N)	NOUNS	VERB TRANSLATIONS BY BILINGUAL SPEAKER 1	VERB TRANSLATIONS BY BILINGUAL SPEAKER 2	VERB TRANSLATIONS BY BILINGUAL SPEAKER 3
1a *closed+deal*	الإتفاق	عقد	اغلق	ابرم
1b *closed+case*	القضية	اغلق	اغلق	اغلق
2a *put on+shoes*	حذاء	انتعل	انتعل	انتعل
2b *put on+shirt*	قميص	ارتدى	لبس	ارتدى
2c *put on+cap*	كوفية	وضع (على راسه)	ارتدى	ارتدى
3a *passed+law*	قانون	اقر	سن	مرر
3b *passed+ proposal*	مقترح	سن	مرر	مرر
4a *won+support*	دعم	كسب	نال	كسب
4b *won+admiration*	إعجاب	كسب	نال	نال
5a *dropped+price*	سعر	هبط	نزل	انخفض
5b *dropped+ temperature*	درجة الحرارة	انخفض	انخفض	هبط
5c *dropped+voice*	صوت	خَفَتَ	خَفَتَ	انخفض
5d *dropped+team*	منتخب	تراجع	نزل	تراجع
6a *fell+silence*	صمت	ساد	خيم	ساد
6b *fell+darkness*	ظلام	خيم	خيم	حل
7a *fell+eye*	عين	وقع	وقع	وقع
7b *fell+regime*	نظام	سقط	سقط	وقع

One of the drawbacks of setting a gold standard in MT is only one option is considered as valid. In the case of the suggested verb translations in Table 5, this is not true because in many instances, the three of them gave different verb synonyms as a collocate for the same noun in the same sentence as, for example, in sentences 1a, 3a, 5a. The different verb synonyms suggested by the three speakers for the same sentences are all valid as good collocates in Arabic.

Next, the three bilingual speakers were given the dataset for the second time, but this time with all the different verbs translations earlier suggested by all of them in Table 5. The bilingual speakers were then asked to choose one or two synonymous verb/s that best suit/s the noun in the TL sentence. The lists of the verb synonyms chosen by them are presented in Table 6.

Table 6. The second set of synonyms suggested by the three bilingual speakers

SENTENCE (V+N)	NOUNS	VERB TRANSLATIONS BY BILINGUAL SPEAKER 1	VERB TRANSLATIONS BY BILINGUAL SPEAKER 2	VERB TRANSLATIONS BY BILINGUAL SPEAKER 3
1a *closed+deal*	الإتفاق	عقد	اغلق/ عقد	ابرم
1b *closed+case*	القضية	اغلق	اغلق	اغلق
2a *put on+shoes*	حذاء	انتعل	انتعل	انتعل
2b *put on+shirt*	قميص	ارتدى	لبس	ارتدى
2c *put on+cap*	كوفية	وضع (على راسه)	ارتدى	ارتدى
3a *passed+law*	قانون	سن/ اقر	سن	مرر
3b *passed+ proposal*	مقترح	سن	مرر	مرر
4a *won+support*	دعم	نال/ كسب	نال	كسب/ نال
4b *won+ admiration*	إعجاب	كسب	نال	نال
5a *dropped+price*	سعر	انخفض/ هبط	نزل	انخفض
5b *dropped+ temperature*	درجة الحرارة	انخفض/ هبط	انخفض	انخفض/ هبط
5c *dropped+voice*	صوت	خَفَتَ	خَفَتَ	انخفض
5d *dropped+team*	منتخب	تراجع	نزل/ تراجع	تراجع
6a *fell+silence*	صمت	ساد	خيم	ساد
6b *fell+darkness*	ظلام	خيم	حل	حل
7a *fell+eye*	عين	وقعت	وقعت	وقعت
7b *fell+regime*	نظام	سقط	وقع/ سقط	سقط

In comparison with Table 5, we may see that in Table 6, a few verb synonyms were chosen unanimously by all three bilingual speakers, as in 1b, 2a and 7a. Some verb synonyms were chosen by two speakers as in the case of the verb synonym (ارتدى) in 2b which was chosen only by speakers 1 and 3. Some other near synonyms were also suggested by the three speakers in Table 6. Accordingly, the highest count of every verb synonym for every sentence was chosen and added as a gold standard as shown in Table 7 below. The other lower count verbs will also be used and added to the metric as suggested verbs.

This gold standard will be used to evaluate the results obtained from the translations by Google and Bing online translation engines. The other near synonymous verb translations suggested by the bilingual speakers in the second time in Table 6 will also be used in the evaluation metric in the following sections. Both the gold standard verb synonyms and the other suggested synonyms are sorted and listed in Table 7.

Table 7. Gold standard verbs and other suggested verb synonyms

SENTENCE (V+N)	NOUNS	GOLD STANDARD VERB SYNONYM	OTHER SUGGESTED VERB SYNONYMS
1a *closed+deal*	الإتفاق	عقد	ابرم/ اغلق
1b *closed+case*	القضية	اغلق	-
2a *put on+shoes*	حذاء	انتعل	-
2b *put on+shirt*	قميص	ارتدى	لبس
2c *put on+cap*	كوفية	ارتدى	وضع (على راسه)
3a *passed+law*	قانون	سن	مرر/ اقر
3b *passed+proposal*	مقترح	مرر	سن
4a *won+support*	دعم	نال	كسب
4b *won+admiration*	إعجاب	نال	كسب
5a *dropped+price*	سعر	انخفض	نزل/ هبط
5b *dropped+temperature*	درجة الحرارة	انخفض	هبط
5c *dropped+voice*	صوت	خَفَتَ	انخفض
5d *dropped+team*	منتخب	تراجع	نزل
6a *fell+silence*	صمت	ساد	خيم
6b *fell+darkness*	ظلام	حل	خيم
7a *fell+eye*	عين	وقعت	-
7b *fell+regime*	نظام	سقط	وقع

6. A Verb-Noun Collocation Evaluation Metric

This metric evaluates verb-noun collocation translations and calculates the verb-noun collocation value in a sentence. The maximum collocation value is (1.0). The metric is composed of two variables, *N*, which stands for the noun, and *V*, which stands for the verb. Since both components comprise the concept of collocation, and neither of them can stand alone, both *N* and *V* are given equal value (0.5). *N* value is either 0.0 or 0.5. *V* is made of four variables: GSV, which stands for Gold Standard Verb, SV, which stands for Suggested Verb, USV, which stands for Unsuggested Verb, and 0V, which stands for Zero Verb (i.e. verb is not present in a sentence). The values assigned to GSV, SV, USV, and 0V are 0.5, 0.3, 0.1, and 0.0, respectively.

The formula is written as follows:

$$NVCOLV = N + V$$

$$0 \geq NVCOLV \leq 1$$

$$N = \begin{pmatrix} 0.5 \text{ when the noun is present in a sentence} \\ 0.0 \text{ otherwise} \end{pmatrix}$$

$$V = \begin{cases} \text{if GSV } 0.5 \\ \text{if SV } 0.3 \\ \text{if USV } 0.1 \\ \text{if 0V } 0.0 \end{cases}$$

The verb-noun-collocation-value (NVCOLV) is equal or bigger than 0 and equal or smaller than 1. The *N* has only two possibilities, either present in the sentence with a value of 0.5, or not present with a value of 0.0. The values for GSV and SV will be given to those gold standard and suggest verb synonyms in Table 7 above. Any other verb synonym not in Table 7 will be considered unsuggested verb, USV, and will be given the value (0.1). If the verb is not present, 0V, then its value is 0.0.

This suggested evaluation metric will be used to evaluate Google and Bing translations in the following section.

7. Testing Data Against Google and Bing online Translators

In this section, the testing datasets are translated by Google and Bing online translators. This is followed by an evaluation of the verb-noun collocation translations using the metric described in section 6 above for both translation engines.

7.1 Google Translation

Google is one of the most widely used free online machine translators in the world. More than 200 million people use Google Translator every day. According to Google Translator, it covers 80 languages and uses a pattern matching method.

Translations of the testing datasets by Google are shown in Table 8. The numbering of sentences in this sub-section follows the numbering in Table 4.

Table 8. Google translation of testing dataset

SL Verb+ Noun Collocates in Sentences	**Google Translations**
1- The verb ***close***	أغلق المدير الصفقة.
a) The director *closed* the deal.	[The director *closed* the deal]
b) The investigator *closed* the case.	إغلاق المحقق في القضية.
	[The investigator *closure in* the case*]
2- The verb ***put on***	وضع الرجل الأحذية على.
a) The man *put* the shoes *on.*	[The man *put* the shoes *on (on* as preposition)]
b) The man *put* the shirt *on*.	وضع الرجل على قميص.
c) The boy *put* a cap *on*.	[He *put* the man on a shirt *(on* as preposition)]
	وضع قبعة على صبي.
	[He *put* a cap on a boy *(on* as preposition)]
3- The verb ***pass***	وقد أقر البرلمان القانون.
a) The parliament *has passed* the law.	[And the parliament *has passed* the law]
b) The parliament *has passed* the proposal.	وقد أقر البرلمان هذا الاقتراح.
	[And the parliament *has passed* this proposal]
4 - The verb ***win***	فاز المستشار الدعم لاقتراحه.
a) The Councillor *won* the support for his proposal.	[The Councillor *won* the support for his proposal]
b) The activist *won* the admiration of many people in his battle against cancer.	فاز الناشط إعجاب كثير من الناس في معركته ضد السرطان.
	[The activist *won* the admiration of many people in his battle against cancer]
5- The verb ***drop***	انخفض سعر كبير.
a) The price *dropped* dramatically	[The price *dropped* dramatically]
b) The temperature *dropped* tonight.	انخفضت درجة الحرارة هذه الليلة.
c) The singer's voice *dropped* suddenly.	[The temperature *dropped* tonight]
d) The Dutch team *dropped* to the fifth place.	انخفض صوت المغني فجأة.
	[The singer's voice *dropped* suddenly]
	تراجع المنتخب الهولندي إلى المركز الخامس.
	[The Dutch team *dropped* to the fifth place]
6- The verb ***fall***	ساد صمت مفاجئ على الغرفة.
a) A sudden silence *fell* upon the room.	[A sudden silence *fell* upon the room]
b) Darkness *fell* quickly tonight.	حلول الظلام بسرعة هذه الليلة.
	[Falling of darkness quickly tonight*]
7- The verb ***fall***	سقطت عيني على كائن غريب.
a) My eye *fell* on a strange object.	[My eye *fell* on a strange object]
b) The regime fell after 1989.	سقط النظام بعد عام 1989.
	[The regime *fell* after 1989]

The Arabic translations of almost all the sentences in the dataset in Table 8 contain one verb synonym, and, in most of these sentences, the nouns did not collocate well with the verb synonym, with the exceptions of (انخفض and تراجع)

for the verb *drop* in 5c and 5d, and (ساد and حلول) for the verb *fall* in 6a and 6b. Although, in the case of (حلول), one cannot say that it is the right equivalence because (حلول), which is 'falling', is a noun. In other words, the verb in the SL is translated into a noun in the TL which produced, grammatically, an ill-formed sentence in Arabic. The back translation of Google's translation of 6b is, *Falling of darkness quickly tonight**.

Translating a verb as a noun can also be seen in 1b were *closed* is translated as إغلاق, closure/ closing) instead of اغلق (closed). This also produced a grammatically ill-formed sentence in Arabic. The back translation of Google's translation of 1b is, *The investigator closure in the case**.

Some additions are also found in the TL as underlined in the following sentences.

1b. The investigator closed the case. إغلاق المحقق في القضية.

3a. The parliament has passed the law. وقد أقر البرلمان القانون.

3b. The parliament has passed the proposal. وقد أقر البرلمان هذا الاقتراح.

In 1b above, the preposition في, 'in' is added, so the output in Arabic came out as *The investigator closure in the case*.

In 3b, هذا, 'this' is added, so the TL output came out as, *The parliament has passed this proposal*.

In 3a and 3b, the preposition و, 'and' is added in the beginning of both sentences. Consequently, the output of both Arabic sentences came out as, *And the parliament has passed the* law and *And the parliament has passed the proposal*, respectively.

The preposition و, 'and' in Arabic remains a problem in Arabic natural processing until today. According to Habash (2010), this problem presents a challenge in computational linguistics in many areas of text processing, such as text matching, data retrieval and data mining. In most cases, typists ignore putting a space after this letter. As a result, it is linked to the following word, and forms a totally different word. Such typing mistakes are found frequently in Arabic. Eventually, the texts are included in corpora and produce inaccurate results as

shown in the following example where adding a space can change the meaning all together.

(و جد) = *And he worked hard.*

(وجد) = *He found.*

Speakers of Arabic can tell the difference by intuition or if the text is annotated, which is very rare in Modern Standard Arabic. However, electronic parsers, morphological analyzers, and corpus extraction software are not able to do this and as a result, many morphological, parsing and translations problems occur.

In the researchers' opinion, this may possibly be the reason behind the addition of the letter (و) in sentences 3a and 3b above in Google translation.

As mentioned above, the Arabic translations of almost all the sentences in the datasets produced a one synonym TL verb. One example can be seen in sentences 2a, 2b and 2c. Google gave one Arabic synonym وضع for the phrasal verb *put on* in the three sentences below.

2a. The man *put* the shoes *on*. وضع الرجل الأحذية على.

2b. The man *put* the shirt *on*. وضع الرجل على قميص.

2c. The boy *put* a cap *on*. وضع قبعة على صبي.

In addition to that, Google does not seem to recognize phrasal verbs as in 2a, 2b and 2c above, in which the second part of the phrasal verb *on* is translated as the preposition of place على, 'on' instead of being part of the verb. To prove this, three extra sentences were inserted to be translated by Google. Google gave the same synonym (وضع) in the following sentences with the non-phrasal verb *put*.

The man *put* the shoes on the table. وضع الرجل الأحذية على الطاولة.

The man *put* the shirt on the table. وضع الرجل القميص على الطاولة.

The man *put* the cap on the table. وضع الرجل غطاء على الطاولة.

By implementing the metric mentioned above on Google translations, the evaluation results came out as shown in Table 9 below.

Table 9. Google translation collocation value

Collocation Value (Google Translator)						
Sentence (V+N)	***Noun***	***GSV***	***SV***	***UV***	***0V***	***Col. Value***
1a *closed+deal*	1	0	1	0	0	0.8
1b *closed+case*	1	0	0	0	1	0.5
2a *put on+shoes*	1	0	0	1	0	0.6
2b *put on+shirt*	1	0	0	1	0	0.6
2c *put on+cap*	1	0	1	0	0	0.8
3a *passed+law*	1	0	1	0	0	0.8
3b *passed+proposal*	1	0	0	1	0	0.6
4a *won+support*	1	0	0	1	0	0.6
4b *won+admiration*	1	0	0	1	0	0.6
5a *dropped+price*	1	1	0	0	0	1
5b *dropped+temperature*	1	1	0	0	0	1
5c *dropped+voice*	1	0	1	0	0	0.8
5d *dropped+team*	1	1	0	0	0	1
6a *fell+silence*	1	1	0	0	0	1
6b *fell+darkness*	1	0	0	0	1	0.5
7a *fell+eye*	1	0	0	1	0	0.6
7b *fell+regime*	1	1	0	0	0	1
Total Value						**0.752941**

According to the metric, the collocation value for Google translation is 0.75. Figure 1 below shows the result of Google's translation value.

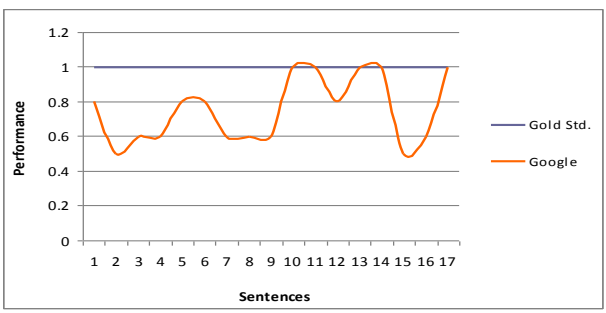

Figure 1. Google translation value

7.2 Bing Translation

Bing Translator is an online machine translator that supports 50 languages and is powered by Microsoft Translator. According to Microsoft Translator (2015), Bing Translator is a state-of-art statistical machine translation system. It allows users to translate only 2 million characters per month.

Bing's translations of the testing datasets are shown in Table 10. The numbering of sentences in this section follows the numbering of the testing dataset sentences in Table 4.

Table 10. Bing translation of testing dataset

SL Verb+ Noun Collocates in Sentences	**Bing Translations**
1- The verb ***close*** a) The director *closed* the deal. b) The investigator *closed* the case.	واختتم المدير الصفقة. [And the director *closed* the deal] المحقق إغلاق القضية. [The investigator closure of the case*]
2- The verb ***put on*** a) The man *put* the shoes *on*. b) The man *put* the shirt *on*. c) The boy *put* a cap *on*.	الرجل وضع الأحذية على. [The man put the shoes on* (*on* as preposition)] الرجل وضع القميص على. [The man put the shirt on* (*on* as preposition)] الصبي وضع قبعة على. [The boy put a cap on* (*on* as preposition)]
3- The verb ***pass*** a) The parliament *has passed* the law. b) The parliament *has passed* the proposal.	وقد أقر البرلمان القانون. [And the parliament *has passed* the law] وقد أقر البرلمان الاقتراح. [And the parliament *has passed* the proposal]
4 - the verb ***win*** a) The Councillor *won* the support for his proposal. b) The activist *won* the admiration of many people in his battle against cancer.	وفاز المستشار دعما لاقتراحه. [And the Councillor *won* support for his proposal] الناشط إعجاب الكثير من الناس في معركته ضد مرض السرطان. [The activist the admiration of many people in his battle against cancer*]
5- The verb ***drop*** a) The price *dropped* dramatically. b) The temperature *dropped* tonight. c) The singer's voice *dropped* suddenly. d) The Dutch team *dropped* to the fifth place.	الأسعار انخفضت بشكل كبير [The price *dropped* dramatically] انخفضت درجة الحرارة هذه الليلة. [The temperature *dropped* tonight] صوت المطرب انخفض فجأة. [The singer's voice *dropped* suddenly] أسقطت المنتخب الهولندي للمركز الخامس. [She *dropped* the Dutch team to the fifth place]
6- The verb ***fall*** a) A sudden silence *fell* upon the room. b) Darkness *fell* quickly tonight.	سقط صمت مفاجئ عند الغرفة. [A sudden silence *fell* upon the room] هبط الظلام بسرعة الليلة. [Darkness *fell* quickly tonight]
7- The verb ***fall*** a) My eye *fell* on a strange object. b) The regime fell after 1989.	سقطت عيني على كائن غريب. [My eye *fell* on a strange object] سقط النظام بعد عام 1989. [The regime *fell* after 1989]

The Arabic verb translations of the three sets of sentences 2, 3 and 7 in Table 10 produced a one-word TL verb synonym. In most these sets, the nouns did not collocate well with the verb synonym.

In Bing translation, we can see a case where a verb is translated as a noun in 1b. The verb *closed* is translated into إغلاق, 'closure or closing'. Interestingly, the same problem occurred in Google translation for the same sentence (refer to section 7.1 above).

Another observation in Bing translation is translating an intransitive verb into a transitive verb, as in:

5d. The Dutch team *dropped* to the fifth place.

أسقطت المنتخب الهولندي للمركز الخامس.

The back translation of the Arabic sentence is, *She dropped the Dutch team to the fifth place.*

Some additions of the preposition و, 'and' are also found in the TL Bing translations as underlined in 3a and 3b below (see the back translation in the Table 10).

3a. The parliament has passed the law.

وقد أقر البرلمان القانون.

3b. The parliament has passed the proposal.

وقد أقر البرلمان الاقتراح.

Based on these findings, the researchers are led to believe that both Google and Bing are using some shared corpora. For a discussion of the possible reason for the addition of "و", see section 7.1 above.

Like Google, Bing does not recognize phrasal verbs as can be seen in 2a, 2b and 2c in the following sentences (see the back translation in the Table 10).

2a. The man *put* the shoes *on*. الرجل وضع الأحذية على.

2b. The man *put* the shirt *on*. الرجل وضع القميص على.

2c. The boy *put* a cap *on*. الصبي وضع قبعة على.

On is translated as the preposition of place على instead of being part of the phrasal verb, *put on*.

In the Bing translation, only one case occurs in which the verb is dropped from the TL sentence (see 4b).

> The activist won the admiration of many people in his battle against cancer.
>
> الناشط إعجاب الكثير من الناس في معركته ضد مرض السرطان.

The verb *won* is deleted in the Arabic translation and the output is a verb-less sentence. Thus, the Arabic translation came out as, *The activist the admiration of many people in his battle against cancer**.

After implementing the metric mentioned above on Bing translations, the evaluation results came out as shown in Table 11.

Table 11. Bing translation collocation value

Bing Translation Collocation Value						
Sentence (V+N)	**Noun**	**GSV**	**SV**	**USV**	**0V**	**Col. Value**
1a *closed+deal*	1	0	0	1	0	0.6
1b *closed+case*	1	1	0	0	1	1
2a *put on+shoes*	1	0	0	1	0	0.6
2b *put on+shirt*	1	0	0	1	0	0.6
2c *put on+cap*	1	0	1	0	0	0.8
3a *passed+law*	1	0	1	0	0	0.8
3b *passed+proposal*	1	0	0	1	0	0.6
4a *won+support*	1	0	0	1	0	0.6
4b *won+ admiration*	1	0	0	0	1	0.5
5a *dropped+price*	1	1	0	0	0	1
5b *dropped+temperature*	1	1	0	0	0	1
5c *dropped+voice*	1	0	1	0	0	0.8
5d *dropped+team*	1	0	0	1	0	0.6
6a *fell+silence*	1	0	0	1	0	0.6
6b *fell+darkness*	1	0	0	1	0	0.6
7a *fell+eye*	1	0	0	1	0	0.6
7b *fell+regime*	1	1	0	0	0	1
Total Value						**0.723529**

According to the metric, the collocation value for the Bing translations is 0.73. The result of the Bing translation is shown in the line chart in Figure 2.

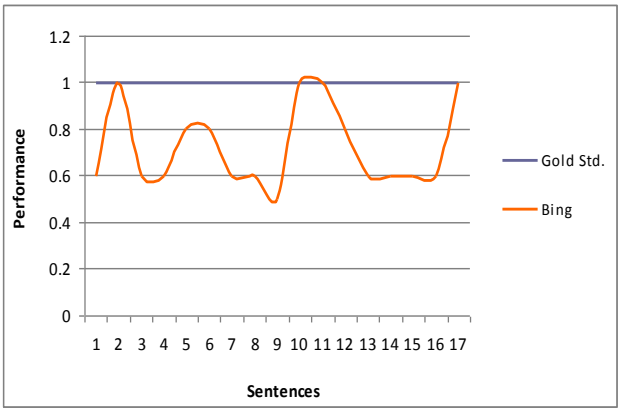

Figure 2. Bing translation value

The discussion on the comparison of the collocation values of Google and Bing will be taken up in the following section.

8. Results

In total, the number of the sentences in the testing data is seventeen. The same evaluation metric was used for the seventeen sentences to investigate the collocation value of the two MT engines: Google and Bing. Figure 3 below shows the total collocation evaluation value for the two translation engines.

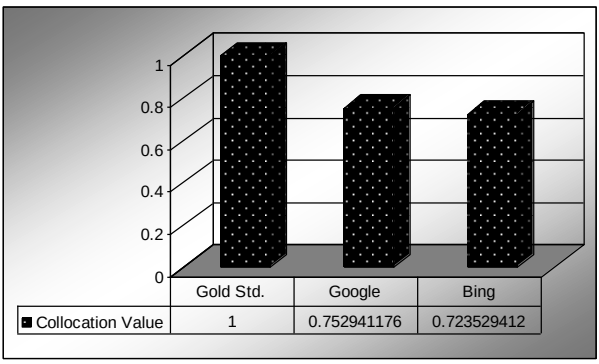

Figure 3. Collocation evaluation value

According to the metric, the collocation value for Google translation is 0.75. This result is shown in Figure 4. In the plot area, we may see that the trend-line is approximately between 0.63 and 0.85, which shows the average performance rate for Google translator.

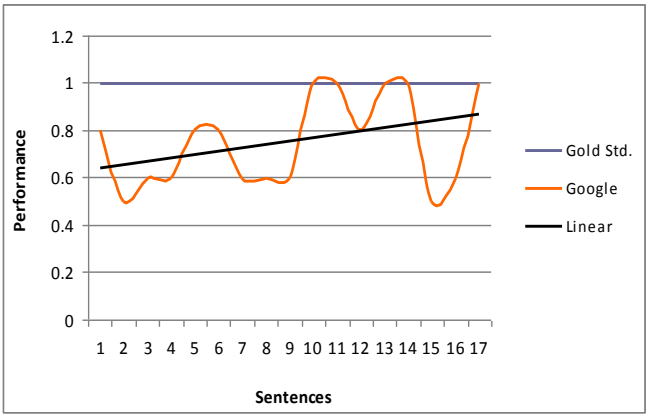

Figure 4. Trendline of Google translation

The translation value for Bing is 0.72. As shown in the plot area of Figure 5, the trend-line is approximately between 0.65 and 0.67, which shows the average performance rate for Bing. Thus, the translation value rate is less than Google's translation value by 0.03.

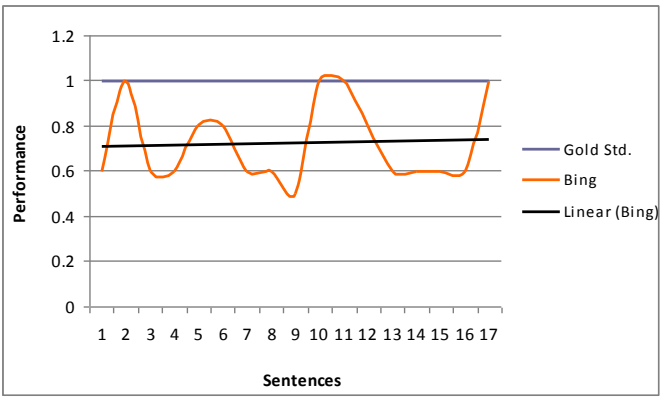

Figure 5. Trendline of Bing translation

9. Conclusion

The simple evaluation metric proposed in this article to evaluate verb-noun collocations in the translation outputs of the two translation engines (Google and Bing) proved to be effective. It correlated with human judgment in the form of the suggested translations provided by the three bilingual speakers. The metric showed consistency in terms of the scales that were used and the way results were similarly obtained for Google and Bing.

Google showed a performance rate of 0.75. In the Arabic translation provided by Google, the translations of five sets from the testing data produced a one-word TL verb synonym, which, according to the metric, did not collocate with the different nouns in most sentences. Google translated the verb as a noun which produced ill-formed sentences in Arabic in two sentences. According to the evaluation metric, this placed the verb evaluation into the no verb category. Google's translations of the phrasal verbs in English assigned the lexical meanings (denotative semantic meaning) of the SL non-phrasal form of the verbs to the TL translation. For example, in translating the verb *put on*, Google translator assigned the TL lexical meaning (denotative semantic meaning) for the non-phrasal verb *put*. Five verbs in Google translations were put under the gold standard verb category, four verbs under the suggested verb category, six verbs under the non-suggested verb category and two verbs under the no-verb category according to the suggested metric.

Bing showed a performance rate of 0.72, lower than Google by 3%. In the Arabic translation provided by Bing, the translations of three sets from the testing data produced a one-word synonym TL verb. Bing translated the verb as a noun in one sentence which produced an awkward structure in Arabic. According to the evaluation metrics, this places the verb evaluation into the no verb category. Bing produced no counterpart of the verb in the Arabic translation in one sentence. Like Google, Bing's translation of the phrasal verbs in English, assigned the lexical meanings (denotative semantic meaning) of the SL non-phrasal verbs to the TL translation. Four verbs in Bing's translation were under the gold standard verb category, two verbs under the suggested verb category, nine verbs under the

non-suggested verb category, and two verbs under the no-verb category according to the metric.

The implications of these numbers show that while Google performed slightly better than Bing by 3%, both engines still do not recognize verb synonymy and, consequently, verb-noun collocations in their TL output in the English into Arabic translations. Google and Bing also do not recognize phrasal verbs. The performance trend estimation for Google and Bing translation – for all the sentences in the dataset – is presented in one chart as shown in Figure 6.

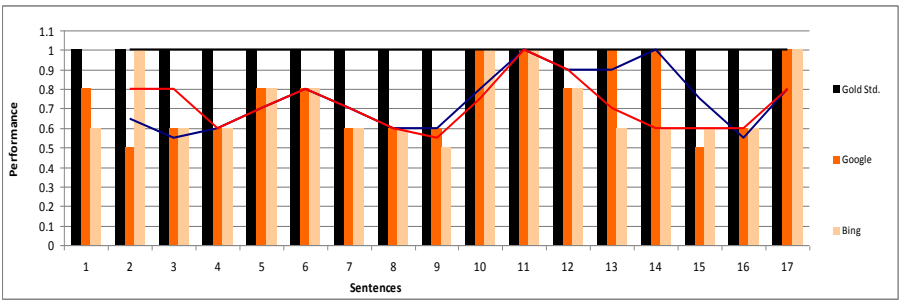

Figure 6. Translation performance chart

The results of this study indicate that Google and Bing MT engines, so far, have not been able to resolve the verb-noun collocability problem in their Arabic output. It is hoped that the findings of this article may shed some light on the problem and to develop new methods to improve Arabic verb noun collocability in the final output of current MT engines.

References

Banerjee, S., & Lavie, A. (2005, June). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* (pp. 65-72).

Bing Translator (2015). Available from http://www.bing.com/translator

Breidt, E. (1993, June). Extraction of VN-collocations from text corpora: A feasibility study for German. In *Proceedings of the Workshop on Very Large Corpora: Academic and Industrial Perspectives, Columbus, USA*.

Buckwalter, T. (2002). Buckwalter Arabic Morphological Analyzer Version 1.0, Linguistic Data Consortium (LDC) catalog number LDC2002L49 and ISBN 1-58563-257-0.

Denkowski, M., & Lavie, A. (2011). METEOR 1.4 automatic machine translation evaluation system. CMU Language Technologies Institute. Retrieved from https://www.cs.cmu.edu/~alavie/METEOR/README.html#about

Google Translator (2015). Available from https://translate.google.com

Goutte, C. (2006). Automatic evaluation of machine translation quality. Presentation at the European Community, Xerox Research Centre Europe, on January, 27, 2006.

Habash, N. Y. (2010). Introduction to Arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, *3*(1), 1-187.

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.

Lavie, A., Sagae, K., & Jayaraman, S. (2004). The significance of recall in automatic metrics for MT evaluation. In *Machine Translation: From Real Users to Research* (pp. 134-143). Springer Berlin Heidelberg.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707-710.

Microsoft Translator (2015). Available from http://www.microsoft.com/en-us/translator

Oxford Advanced Learner's Dictionary (2013). Available from: http://oald8.oxfordlearnersdictionaries.com/dictionary/draw_1

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 311-318). Association for Computational Linguistics.

Qin, Y., Wen, Q., & Wang, J. (2009, September). Automatic evaluation of translation quality using expanded N-gram co-occurrence. In *Natural Language Processing and Knowledge Engineering, 2009. NLP-KE 2009. International Conference on* (pp. 1-5). IEEE.

About the Authors

Hussein Soori holds a Ph.D. degree from the Faculty of Languages and Linguistics, University of Malaya, Malaysia. His research areas include information retrieval, natural language processing, machine translation, morphological segmentation, and word sense disambiguation.
Email: soori2003@hotmail.com

Su'ad Awab (Associate Professor) is a lecturer at the Faculty of Languages and Linguistics, University of Malaya, Malaysia. Her specialized areas are corpus linguistics, phraseology, and language technology. She is currently researching on authenticating translations of hadith on social media.
Email: suad@um.edu.my