Profiling of (Ir)reversible Binomials in Translated Arabic Texts
A corpus-based study

Abdelhamid Elewa

Imam Mohamed Ibn Saud Islamic University, Saudi Arabia; Al-Azhar University, Egypt

eleewah@gmail.com

Mohamed Elaskary

Hankuk University of Foreign Studies

elaskary9@yahoo.com

Abstract: A binomial, or binomial pair, is a pair of words conventionally linked by a lexical link such as a conjunction or a preposition, for example, "day and night", "life and death", "dos and don'ts", etc. Binomials are very common in all languages. Binomials can be reversible or irreversible; i.e. they share/do not share the same word order in the Source Language or Target Language. Binomials play a very import role in learners' or speakers' competence; the more binomials you are familiar with the more competent one would be. In this paper, we will examine binomials in the Arabic language; we will use corpus-based approaches to explore how binomials are reversible in Arabic and English. In an attempt to find a practical-digital way for language learners and translators to rely on when they come across conjoined words, mainly binomials, different approaches to binomial extraction have been explored. Corpus linguistic techniques and Mutual Information[1] statistics are used to

---

[1] Mutual information is "one of many quantities that measures how much one random variable tells us about another. It is a dimensionless quantity with (generally) units of bits, and can be thought of as the reduction in uncertainty about one random variable given knowledge of another" (Latham & Roudi, 2009).

test the relations between the different parts of binomials; the most frequent binomials and conjoined phrases are analyzed in an Arabic text coupled with their translations into English. By calculating the significant binomials, mutual information helps us to identify what to examine in the concordance lines. Based on the analysis of the concordances of the data, some rules have been proposed, something which would enable language learners and translators to make generalizations and reach firmer conclusions regarding the order of parts within binomials and whether they are reversible or irreversible when translated from one language into another.

Keywords: Binomials, irreversibility of binomials, corpus-based analysis, conjoined phrases, Arabic/English translation

ملخص: الثنائيات اللغوية أو الأزواج اللغوية تتكون عادة من كلمتين يربط بينهما حرف عطف غالبا، على سبيل المثال: الليل والنهار، الموت والحياة والمسموحات والممنوعات الخ. الثنائيات اللغوية شائعة تجد في جميع اللغات، ولاسيما في النصوص الدينية يمكن أن تكون الثنائيات اللغوية قابلة للمواقفة اللغوية مع اللغة المترجم منها/إليها أو لا؛ أي أنها تشارك/لا تشارك في نفس ترتيب الكلمات في لغة المصدر أو اللغة المترجم إليها بتلعب الثنائيات اللغوية نور ألمهم في كفاءة المتعلمين أو المتحدثين، وكلما زاد عدد الثنائيات اللغوية التي يعرفها الشخص كلما تحسنت كفاءة اللغوية في هذا البحث سنبحث في الثنائيات اللغوية في اللغة العربية، سنستخدم المقاربات المستندة إلى التحليل القائم على لسانيات المتون أو علم متن اللغة لاستكشاف كيفية ترجمة الثنائيات اللغوية للغتين العربية والإنجليزية. في محاولة لإيجاد طريقة رقمية تقربعملية المتعلمي اللغة المترجمين للاعتماد عليها عندما تكون هناك ثنائيات لغوية تتم استكشاف طرق مختلفة لاستخراج الثنائيات اللغوية تم استخدام تقنيات اللغات المصدر وإحصاءات المعلومات المتبادلة لاختبار العلاقات بين مختلف أجزاء الثنائيات اللغوية سيتم تحليل أكثر عبارات الثنائيات اللغوية والعبارات المترابطة في نص عربي مع ترجمتها إلى اللغة الإنجليزية. ومن خلال حساب الثنائيات اللغوية المهمة تساعدنا المعلومات المتبادلة في تحديد ما يجب فحصه في خطوط التوافق وبناء على تحليل توافق البيئات، اقترحنا بعض القواعد وهو الأمر الذي سيمكن متعلمي اللغة والمترجمين من الحصول على صورة علمة لمعنى الثنائيات اللغوية والتوصل إلى استنتاجات أكثر تماسكا فيما يتعلق بترتيب الأجزاء داخل الثنائيات اللغوية وما إذا كانت قابلة للعكس أولا عند ترجمتهم من لغة إلى أخرى.

كلمات مفتاحية: الثنائيات اللغوية، التحليل القائم على لسانيات المتون أو علم متن اللغة، العبارات المرتبطة، الترجمة العربية الإنجليزية

1. Introduction

The general trend of translation studies to move away from a linguistics-based approach toward more sociological and culture-oriented approaches (Bassnett, 2002; Bassnett & Lefevere, 1990; Chamberlain, 1992; Snell-Hornby, 1990; Venuti, 1995). Despite this general trend, we the authors of this article believe that studies with linguistical orientation should be continued and appreciated.

Against this background, we think that 'the binomial construction' and its (ir)reversibility should be explored because they are abundant and hence this could improve the awareness, productivity and output of learners and translators. Binomials are very common in almost every language (see also: Al-Jarf, 2016; Al-Otaibi, 2021). The first thing that a learner does when s/he comes across a word s/he does not know is to look for word meaning in a dictionary. However, a great deal of word meaning may not be available in dictionaries or in grammar books. Native speakers of a given language tend to use and reuse chunks of words in which words are combined together in rigid or looser bonds. However, Stubbs (1996, p. 172) argues that the intuition of native-speakers about how words are combined is not reliable.

This paper aims to improve learners' proficiency and raise translators' awareness of binominals, which are prefabricated/prepacked chunks of language,[2] as are collocations, idioms, compounds, clichés and institutionalized expressions. Examples of binomials include phrases such as "cats and dogs", "day and night", and "boys and girls." The paper is based on the idea that a major part of language is the ability to comprehend and produce or reproduce lexical phrases as indivisible chunks of words. We did our best to provide learners and translators with a more practical way in word selection, tested with corpus linguistic techniques. These techniques could be utilized to help learners/translators improve their awareness and choice of lexical items, something which would help them render better translations.

When combined together words may become part of what is called "multiword units"; this includes collocations, binomials, idioms, compounds, clichés, proverbs and institutionalized expressions, depending on how rigid the relationship that holds between the different parts of the structure. We are going to focus on one area of word combination, particularly the collocation of conjoined words or what is commonly called "binomials". Some scholars claim that cultural norms govern the order of parts in the binomial pairs. For instance, they argue that "man" is to be fronted in phrases like "Mr. and Mrs.", "man and woman", "boys and girls", because language may be gender-biased in some communities. Others say this process is a universal common linguistic feature; there is a long list of phrases which are reversed

---

2   To put it in Carter's (1987, p. 59) terms, "language production consists of piecing together such ready-made 'pre-fabricated' units appropriate to a situation and that lexical acquisition may involve the learning of complete collocational chunks of language."

in translation on no grounds except language use such as "black and white", "cat and mouse" or "fish and chips". More interestingly, Malkiel (1959, p. 143) noted that the parts of binomials may be ordered differently across languages like "East and West" that is reversed in German to "West und Ost". Therefore, this paper tries to use corpus techniques to examine this phenomenon in Arabic and English quantitatively to find out on what grounds the different parts of binomials are ordered.

2. Classification of Lexical Items

Generally, words can have more than one meaning. Relying on a dictionary as a single source of translation does not work all the time. Learners of English as a second language or for specific purpose still give much weight to dictionaries believing that there is a one-to-one correspondence between words, but they are often misled or left without a convincing answer. A classification of lexical phrases, therefore, can pinpoint some features of language use which are not attainable in dictionaries. Linguists have classified the types of lexical phrases and called them "multi-words or poly-words". Nattinger (1980) classified such lexical phases into: (1) polywords (fixed phrases whose meanings are often not analyzable by the regular grammatical rules); (2) phrasal constraints (relatively fixed phrases that can undergo some variation)' (3) deictic locutions (phrases of low variability to monitor conversation); (4) sentence builders (phrases that are highly variable); (5) situational utterances (usually sentences uttered on certain occasions and (6) verbatim texts (fixed phrases such as aphorisms, proverbs etc.). Alexander (1989), on the other hand, mentioned five categories: (1) idioms; (2) discourse-structuring devices; (3) proverbs; (4) catchphrases and (5) quotations/allusions.

Considering the functions of such phrases, Cowie (1988) highlighted two major groups: (1) phrases with pragmatic functions (e.g., "good morning", "how are you") and (2) phrases with semantic/idiomatic functions (e.g., "kick one's heels", "pass the buck"). Kjellmer (1991) looked at the nature of such phrases and gave three types: (1) fossilized; (2) semi-fossilized and (3) variable phrases. Drawing on such classifications, Lewis (1993) gave the following types: (1) polywords: two or three words that could be semantically opaque or transparent; (2) collocations: fixed collocations which are considered a kind of polyword and (3) institutionalized expressions that are employed for social

communication purposes. For instance, we frequently hear fixed expressions which might be ungrammatical like, just a moment, or grammatical like give me a second, decorating our language for communication. Many phrases or sentences, short or long, follow this pattern, such as "by and large", "let us now turn to", "another way of putting it is", 'strictly speaking', etc. (for further details, see Quirk & Greenbaum, 1973). On the other hand, a collocation is a group of words that occur together more often than by chance. Below, we are going to discuss one category of collocations: binomials.

3. Binomials

The long-established definition of binomials dates back to Malkiel (1959, p. 113): "the sequence of two words pertaining to the same form-class, placed on an identical level of syntactic hierarchy, and ordinarily connected by some kind of lexical link". Looking at this definition, one can identify possible binomials across languages as two conjoined words that have the same word category and have a parallelistic syntactic structure. With this definition we can exclude three conjoined lexemes that is called trinomials, in addition to other types of multi-word units like idioms, collocation, etc.

The question then is on what basis can we know that a given phrase is a binomial and then find out what goes with what? Actually, binomials can be identified intuitively, semantically, lexically or quantitatively. McIntosh and Halliday (1966, p. 194) argue that our experience of the meanings that a given word has in a certain context sheds light on what words or range it co-occurs with. For example, the lexical item *fish* is likely to coincide with chips and other items, but the bond between the two items "fish and chips" is fixed. This is due to our experience with such items in a variety of contexts. Firth (1957) views such a phenomenon as a relation of mutual expectancy and as an inseparable part of the native speaker's knowledge of his own language, i.e., competence (Emery, 1988). However, one cannot simply figure out what is more frequent or typical in language use without extensive research because introspection is not easily measurable. On the other hand, by using advanced technology and corpus linguistic techniques one can assess the problem more accurately and quickly.

To assess a given binomial, we can resort to semantics. In this regard, Cruse (2000, 1986) makes a distinction between two types of semantic co-

occurrence restrictions: (1) selectional restrictions (which can be defined as "semantic co-occurrence restrictions which are logically necessary"), and (2) collocational restriction (which is defined as "co-occurrence restrictions that are irrelevant to truth conditions") (Cruse, 1986, p. 279). In the following example, the verb *die* needs to be preceded by a (+animate) grammatical subject; this is called selectional restriction:

> John died
> The tree leaves died
> *The book died

> Further, semantic requirements are needed in sentences like:

> John kicked the bucket.
> *The cow kicked the bucket.
> *The tree kicked the bucket.

The lexical item "kick the bucket" requires in addition to the (+animate) feature another restriction, which is (+human). Restrictions of this type are called collocational restrictions. In short, the semantic approach tries to define collocations by the actual meanings they have and by the usefulness of combinations of words in different contexts.

The lexical approach[3], on the other hand, concentrates on the language as a complete unit; it does not make a distinction between grammar and vocabulary. This approach differs from the semantic one in that the latter tends to account for all the relations that hold among lexical occurrences 'in a semantically motivated way' (as in Cruse, 1986; Emery, 1988). In other words, the lexical approach looks at collocation, for example, as a matter of combinatorial process without giving any explanation. It does not explain why a given lexical item collocates with another lexical item (Lehrer, 1974, p. 176).

These approaches to the identification of multi-word units make the phenomenon language-specific. Translators and language learners will have no fixed criteria for their lexical choice except to be as fluent as native speakers of that language. So, the proper use of collocation is one of the

---

3   The lexical approach not only deals with individual words, as might be understood, but also with larger units, i.e., the word combinations that we store in our minds.

features that makes the style of a native speaker different from that of a learner (Hoey, 2003). In the same vein, Hoey (2005) argues that all speakers across the world subconsciously associate multi-word units like collocation and colligation (as well as binomials) with certain situations, genres, and styles. This is because knowledge of binomials is always related to an individual's mental lexicon. Therefore, we may be aware of the phenomenon spontaneously without prior knowledge, according to Hoey's Lexical Priming Theory (2005). Sometimes, we do not know why two words are bonded together more often than by chance. But we will try to look at another aspect of this bonding relation in terms of interlanguage treatment in an attempt to argue that this phenomenon is cross-linguistic. Crystal (1987, pp. 161-162) notes that collocations in general differ greatly between languages, and provide a major difficulty in mastering foreign languages. By the same token, binomials could behave similarly across languages.

3.1 (Ir)reversibility of Binomials

Many studies on binomials addressed the irreversibility of the different parts of binomials from different perspectives: intuition, linguistics, and corpus linguistics. For instance, Abraham (1950) relied on morphology and semantics as the main principles of ordering the two words of a binomial. Morphologically, words of fewer number of syllables are fronted. From a semantic point of view:

> [t]he desirable usually precedes the undesirable, the more important the less important, the light the dark, the masculine the feminine, the positive the negative, the principal the subsidiary, the greater the smaller, the near the far, the top the bottom, the present the future. (Abraham, 1950, p. 284)

Nöth (1993) also offers more semantic explanations in an attempt to logically reveal symmetrical coordinates, i.e., irreversible binomials, giving more semantic constraints such as "more dynamic before less dynamic", "more active before less active", "useful before useless", "sympathy before antipathy", "good before bad", "beautiful before ugly", "interesting before uninteresting", "valuable before worthless", and "superior before subordinate" (see also Motschenbacher, 2013).

Phonologically, Cooper and Ross (1975, p. 79) argue that the fronted

word is that which contains a shorter vowel, fewer or less obstruent initial consonants, more front vowel, fewer (or more obstruent) final consonants. Added to the list of phonological constraints, Parker (2003) includes another principle for fronting words of more sonorous sounds while Benor and Levy (2006) front those with more stressed syllables. Such constraints are proposed by analysis of anecdotal examples, except for Benor and Levy (2006) who conduct an empirical study that consists of four stages: (1) retrieval of binomial hits from a corpus search; (2) proposing a list of constraints: phonological, semantic, and frequency-based; (3) coding and quantitative analysis; and (4) application of three models for data analysis: traditional optimality theory, stochastic optimality theory, and logistic regression. Later, Mollin (2012) retrieved for analysis more than 500 binomials from a large-scale corpus British National Corpus (BNC) with the aim of finding out the principles of reversibility or reversibility of binomials. She argues that the semantic factors explain the preferred order of binomials to the greatest degree, followed by metrical ones, word frequency and some phonological factors (Mollin, 2012, p. 96).

This phenomenon is heavily and extensively discussed in linguistics. However, few attempts are made to explore it in other languages and cultures than English through translation. For instance, Ebeling and Ebeling (2015) examine cross-linguistic data in Norwegian and English to spot how one binomial "more or less" and its Norwegian equivalent "mer eller mindre" are used in an English-Norwegian Parallel Corpus. They found that binomials are not used as equivalent to each other in the Corpus. The English binomial "more or less" is overused in the English translations, overlooking some shades of meaning in the Norwegian data.

Hussein and Lingwood (2011) examine the strategies used by Arab students in translating binomials from English into Arabic. They find that students were usually unable to properly translate binomials for lack of specialized dictionaries.

Various kinds of constraints have been proposed to account for the irreversibility of binomials monolingually and bilingually, using qualitative measures since the 1950s. With the advent of corpus linguistic tools and techniques, the irreversibility or formality of binomials can be quantified. The explanations, discussed above, for ordering the internal parts of binomials have not been replicable in translation. Therefore, this paper examines the possible criteria and constraints of translating binomials into Arabic.

3.2 Binomials Between Arabic and English

As mentioned above, the criterion for identifying binomials is hard to achieve with a casual look at either the source text (ST) or target text (TT). We may not realize that a phrase we have just said is a collocation unless we think deeply about other possible alternatives for such a phrase. The way words are conjoined may differ from one language to another, because one may find the same chunk of words carrying the same meaning but ordered in a different way. This could bring to light some common features of what and how people order their thoughts. Dickins et al. (2002, p. 71) state, "an important area of collocation is the use of conjoined phrases on the pattern (X and Y)". Thus, English tends to say "knives and forks" rather than "forks and knives", and "pots and pans" rather than 'pans and pots'.

They did not provide further explanations regarding the basis for following this type of fixed pattern in English or in any other language. Therefore we cannot expect the translator to follow this pattern unless s/he becomes as fluent as a native speaker of the target language because there are no criteria to build on except language use.

Many conjoined words are reversed in translation on no grounds except language use. For example, in Arabic, we have phrases like "almawt wa alhayah [death and life]" not "life and death", "allay wa lanahar [night and day]" not "day and night", "alkilab wa alqitat [dogs and cats]" not "cats and dogs", "ashab alnofoudh wa alaghniya [the powerful and rich]" not "the rich and powerful", "yameen wa shamal [right and left]" not "left and right", "min wa 'ila [from and to]" not "to and from".

Hence, one can notice, in the above examples, that the items of collocation are not placed in their Arabic order. In the above examples the second part of the collocation is reversed. The words *death, night, dogs, powerful, right,* and *from* are placed at the end. This is how Arab people say it; it thus appears that it is language use which determines what comes with what. So, is frequent occurrence of conjoined phrases in a given language the main reason for determining which item comes first? To find out the answer to this question, we used corpus linguistic techniques to investigate such conjoined phrases in both languages.

3.3 Data Description

The data analysed in this paper is part of a pilot project of a systematic parallel corpus aiming to collect Arabic texts and their translations into English. For convenience, two texts are used in this study: Arabian Nights, (alternatively rendered as: *The Thousand and One Nights*) and its translation into English. The work appeared in 850 A.D. during the Abbasid age. Some chroniclers claim that the work was originally written in Persian and then translated and reworked completely into Arabic to leave no Persian traces which might have contradicted Islamic thought. Because the author of the original work is anonymous and its origin is unknown, we will consider this work original for the changes and refinement which made it tantamount to non-translated texts. The English version was published in 1811 A.D. by Jonathan Scott and was posted on the internet on January 24, 2009 as an eBook on www.gutenberg.net. The total number of words of the Arabic work is 531,099 words while the English version is 548,405 words.

3.4 Tools for Analysis

We used a concordancer called Monoconc to search the frequency of each phrase. Although the program makes use of statistics for collocation retrieval, we did the calculations outside the program. The fact that Arabic is written without diacritics signs required additional laborious work. In other words, we had to search for all the possible forms of a given word separately. For example, if you search for a word with its diacritics signs using Monoconc, it will give you all the exact occurrences of that word in the corpus disregarding the other possible forms of that word or even the character variations such as *alif* with/without *hamza* and dotted/un-dotted *yaa'*. This makes it difficult to arrive at all possible forms of words under examination or work out the various conventions of Arabic writing, let alone the universal problem of polysemy and homographs. Because such problems cannot be solved automatically, manual editing is required.

3.5 Analysis and Discussion

The first step was to compile two lists of the conjoined phrases. The lists were extracted from the two corpora according to the following pattern "X and/

or Y". Before we began analysing the significant collocations of the conjoined phrases, we noticed that the Arabic list of conjoined phrases was longer. Taking a closer look at the nature of both languages, we noticed that Arabic employs coordination more extensively than the English language.

Secondly, having extracted all forms of the conjoined phrases, we had to list all instances of binomials and eliminate all irrelevant hits from the Arabic output. For example, we manually eliminated instances where the form is not a conjoined phrase as a result of morphological ambiguity in Arabic. These irrelevant instances appeared in the raw version of the list because of the absence of short vowels in Arabic which makes the process of any computational analysis of Arabic quite challenging and sometimes ambiguous. Consider for example the three letters-word ورد *wrd* which can be lexicalised as a verb وَرَدَ *warada* [came, have been mentioned in], a noun وَرْدٌ *ward* [flower], a noun وِرْدٌ *wird* [watering place], وَرَدَّ *wa radda* [and he replied]. Also, we eliminated all the instances of conjoined sentences such as "let me embrace you and give you the first marks of a father's love".

Thirdly, after discarding all the irrelevant combinations, we used Mutual Information (MI) statistics to test how significant each conjoined phrase is, thus helping us identify interesting patterns. For example, if two or more words showed up in our corpus a number of times, we could examine how far such a pattern is interesting by comparing their joint probability with chance, i.e., to count the number of occurrences of the combination with the number of the occurrences of each word independently. Words with large MI scores are likely to be more interesting (Church et al., 1991).

The formula as introduced by Church et al. for two given words reads:

$$MI(w_1, w_2) = \log_2 \frac{P(w_1, w_2)}{P(w_1)P(w_2)}$$

The MI compares probabilities of *x* and *y* together with probabilities of *(x)* and *(y)* independently. Church and Hanks (1990, p. 23) argue, If *p (x, y)* is bigger than *p(x) p(y)*, then it is evidence that there is more likely a genuine association. If *p (x, y)* equals or is less than *p(x) p(y)*, then we can predict no interesting association.

To use this formula to find significant phrases in our corpora, we first counted the frequency of the items of the phrase separately and then

altogether. In the following table (n) stands for the corpus size, f (x) for the frequency of the first word, f (y) for the frequency of the second word and f (x, y) for the frequency of the two words together.

Table 1: The top 30 conjoined phrases in the English corpus

F (x, y)	F (x, y)	F (x)	F (y)	MI
you and I	25	7428	10373	-4.07
gold and silver	19	354	80	6.93
him and his	16	5617	7535	-3.85
go and see	15	562	799	2.61
her and the	14	4964	34346	-6.05
me and the	14	4264	34346	-5.83
Deen/Din and the (like in Salah El-Din)	13	664	34346	-3.26
mother and the	13	548	34346	-2.98
he and his	12	9572	7535	-5.03
you and your	12	7428	2649	-3.16
me and that	12	4264	6791	-3.72
father and mother	12	482	548	3.05
him and you	11	5617	7428	-4.37
come and see	11	632	799	1.99
mother and sister	11	548	154	4.57
wife and children	11	458	158	4.97
him and that	10	5617	6791	-3.38
me and you	10	4264	7428	-4.11
now and then	10	493	1013	1.87
prince and princess	9	1334	983	0.32
mother and daughter	9	548	394	2.93
hundred and ninety	8	154	10	9.89
up and down	7	987	604	1.10
go and fetch	7	562	39	5.86
mother and son	6	548	666	1.58
bread and water	6	58	227	6.38
bows and arrows	6	25	30	10.51
officers and ladies	5	136	206	5.02
happiness and prosperity	5	123	31	7.90
read and write	5	81	30	8.55

(n = 548,405)

Table 2: The top 30 conjoined phrases in the Arabic corpus

	F (x, y)	F (x, y)	F (x)	F (y)	MI
سمعاً وطاعة	gladly and obeying	215	1819	266	6.29
الحسن والجمال	goodness and beauty	189	201	434	8.58
الأكل والشرب	food and drink	173	932	586	5.80
وقته وساعته	timely and on the spot	100	688	497	5.69
يميناً وشمالاً	right and left	71	154	110	9.53
الذهب والفضة	gold and silver	55	291	119	8.13
عز وجل	glory and exalted	55	70	60	11.04
ليل ونهار	day and night	42	493	321	5.55
حباً وكرامة	love and honour (lit: heartily)	41	142	52	9.94
الصيد والقنص	hunting and hawking/angling	35	76	38	11.24
الأمراء والوزراء	princes and ministers	33	126	96	8.91
الصلاة والسلام	prayers and peace	25	124	389	6.52
هو ومن	he and whom	18	1105	13837	-2.26
الزمان وسالف	time and in the past	18	904	21	7.39
الدنيا والآخرة	this life and the afterlife	15	322	62	7.05
أشجار وأنهار	trees and rivers	12	135	58	8.08
أمه وأخته	his mother and sister	9	174	110	6.37
جاء وقال	he came and he said	8	430	4776	-0.53
القيل والقال	tittle and tattle/ gossip	8	8	8	14.43
الشموع والقناديل	candles and lanterns	8	43	24	10.42
المكان والوزير	the place and the minister	7	799	1206	0.36
خمسة وعشرون	twenty and five	7	157	53	7.21
الخليفة وجعفر	khalifa and jaafar	6	708	175	3.10
الويل والثبور	woe and grief	4	63	4	11.45
ألذه وأعذبه	most delicious and sweetest	3	29	21	9.76
الدموع والدم	tears and blood	3	146	72	5.65
أهل وجيران	family and neighbors	3	636	25	5.06
التهليل والتكبير	exulting and praising	3	5	16	12.69
الأطباء والحكماء	doctors and physicians	2	13	58	8.87
الحمار والثور	the donkey and the bull	2	187	34	5.79

(n = 531,099)

Following Barnbrook (1996, p. 90), we will consider phrases that occur at least three times within the span to be relevant for the range of analysis. This is because words that occur just once or twice can give spuriously high

significance scores as shown in Table 2 above where the phrases "al-atibaa wa alhukamaa" [doctors and wise men] and "al-himar wa al-thawr" [the donkey and the bull] occurred just twice yet they show a high MI score.

In the above tables, the higher the MI, the more significant the association between the two parts of the phrase. We therefore discarded the insignificant pairs of words and tried to carry out another test to see if the two items of the significant conjoined phrase are in complementary distribution, i.e., they can be used interchangeably. To do this we counted all the possible occurrences of the phrases under examination in a reverse way. In other words, we attempted to see if any interesting results occur if we transpose the two items of the phrase from (x, y) to (y,x); i.e. "out and in instead of in and out). In Tables 3 and 4, we counted the number of times that the word (y) appears before the word (x). If we spotted any significant result for the transposed phrase, we definitely ruled out the phrase, as shown in the following two tables.

Table 3: Word order in the English conjoined phrases

F (x)	F (y)	F(x, y)	F (y, x)
gold	silver	19	2
go	see	15	0
father	mother	12	0
wife	children	11	0
come	see	11	0
mother	sister	11	0
now	then	10	0
mother	daughter	9	0
prince	princess	9	0
hundred	ninety	8	0
up	down	7	0
go	fetch	7	0
mother	son	6	0
bow	arrows	6	
bread	water	6	0
happiness	prosperity	5	0
read	write	5	0
officers	ladies	5	1

Table 4: Word order in the Arabic conjoined phrases

X		Y		F (x, y)	F (y, x)
سمعاً	gladly	طاعة	obeying	215	0
الحسن	goodness	الجمال	beauty	189	3
الأكل	food	الشرب	drink	173	0
وقته	timely	ساعته	on the spot	100	0
يمين	right	شمالاً	left	71	0
الذهب	gold	الفضة	silver	55	1
عز	glory	جل	exalted	55	0
ليل	night	نهار	Day	42	2
حباً	love	كرامة	honour	41	0
الصيد	hunting	القنص	hawking/angling	35	0
الأمراء	princes	الوزراء	ministers	33	11
الصلاة	prayers	السلام	Peace	25	0
الزمان	time	سالف	ancient past	18	0
الدنيا	this life	الآخرة	the afterlife	15	0
أشجار		أنهار		12	1
أمه	his mother	أخته	his sister	9	1
القيل	tittle	القال	tattle	8	0
الشموع	candles	القناديل	lanterns	8	4
المكان	the place	الوزير	the Minister	7	0
خمسة	five	عشرون	twenty	7	0
الخليفة	Khalifa/Caliph	جعفر	Jaafar	6	0
الويل	Woe	الثبور	grief	4	0
ألذه	most delicious	أعذبه	sweetest	3	0
الدموع	tears	الدم	blood	3	0
أهل	family	جيران	neighbours	3	0
التهليل	praising God	التكبير	exalting God	3	1
الأطباء	doctors	الحكماء	physicians	2	0
الحمار	the donkey	الثور	the bull	2	0

In an attempt to give reasons for the word order in conjoined phrases, Fenk-Oczlon (2001, 1989) argued that the key element in determining what goes before what is frequency, so items of high frequency should come first. However, the lists above show the inconsistency of such a statement. For example, the word *go* in "go and see" occurred in our corpus 562 times, whereas the word *see* occurred 799, and the word *bows* (25 times) in "bows and arrows" is less frequent than *arrows* (30 times). Similarly, Church et al. (1989) also noted that word order in phrases like" doctor and nurses", "man and woman" is fixed, "illustrating a wide variety of biases ranging from sexism to syntax". Such an argument is not conclusive simply because we have many word pairs which fall outside this pattern. For example, in phrases like "ladies and gentlemen", the feminine form occurs first. Also, syntax does not account for word order in conjoined phrases since items in such phrases have the same syntactic category N+N, V+V, and Prep+Prep.

4. Results

By utilizing a corpus linguistic methodology, we managed to determine some rules of how words and phrases are conjoined. This bonding relation is noticed in many languages across the globe. In the field of Arabic/English translation, no inclusive platform had previously been proposed to enable translators, language learners and researchers to make generalisations and reach firmer conclusions on the behavior of conjoined phrases between both languages. Based on the above analysis, we are proposing new rules for determining the order of the two parts of binomials, or what comes before what.

First of all, in English and Arabic we mainly start a phrase with the positive item as in "alkhayr wa alshar" [good and evil], "taw'an wa karkha" [willy-nilly], "alhubb wa alkurh" [love and hate], "al'ijab wa alsalb" [positive and negative]. However, we can see that that is not the case at all times and on many occasions word order in both languages do not match. For example, the "almawt wa lhayah" [death and life] versus the English "life and death", and "almawat wa almilad" [death and birth] versus "birth and death". Through extensive research we can offer an explanation why such permutation took place. For example, people tend to reuse bits and pieces of sacred books, proverbs, and clichés as in the following phrases "almawt wa alhayah" [life

and death". In the holy Qur'an Allah states, "alladhi khalaqa almawta wa lahayata" [Who has created death and life that He may test you] (Qur'an 67:2). Some interpreters of the Qur'an argued that death comes first because everything is created from nothing.

Secondly, one item of the phrase may be fronted for phonological reasons. For example, "alqitat wa alkilab" [cats and dogs], and "min damihi wa lahmih" [his own flesh and blood], "al-mujab wa as-saalib" [plus and minus], the word starting with (or containing more) voiceless sounds is to be fronted.

Thirdly, fronting can take place for morphological reasons as well. For example, the phrase "taht wa fawq" is translated into "up and down". For "alddakhil wa lakharij" [in and out], "ashaabu al-nufoudh wa al-aghniyaa" [the rich and powerful], "min wa ila" [to and from], "ath-tharwah wa ash-shuhrah" [fame and fortune], "alshuwak wa as-sakakin" [knives and forks], and "al-mar'ah wa ar-rajul" [man and woman], the shorter form (in terms of graphemes) is the one to be fronted.

Fourthly, words with higher vowels come first as shown in Figure 1.

Figure 1: Order of binomials on the English Vowel Chart
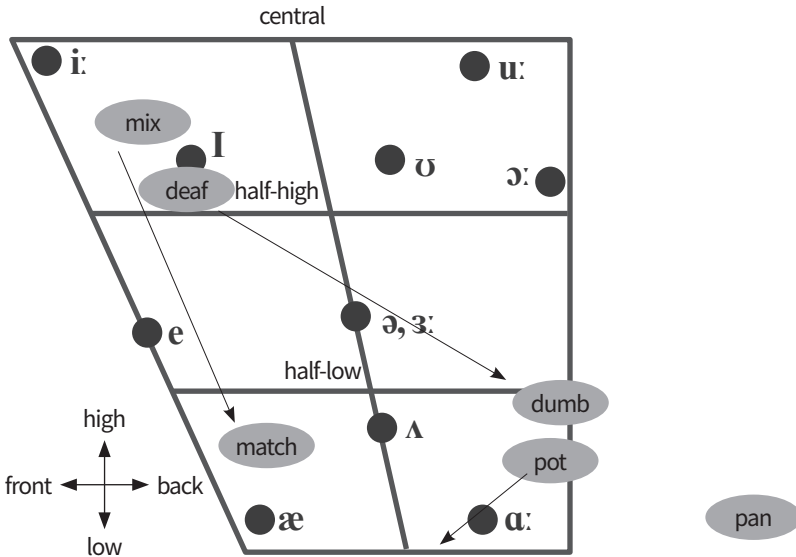
For instance, the word *pot* in "pot and pan" contains a vowel which is higher than the vowel in *pan*. The same applies to "as-summ wa al-bukm" [deaf and dumb], "imzij wa waffiq" [mix and match], "hadhaa wa dhaak" [this and that]. Also words with long vowels or diphthongs occur finally as in "al-ʿum wa al-bint" [mother and daughter], "abiyadh wa aswad" [black and white]. If the two words contain long vowels or diphthongs, the lower vowel is to be fronted such as "al-layal wa an-nahaar" [day and night].

5. Conclusion

Corpus-based approaches have been used to explore how binomials are reversible or irreversible in Arabic and English. The different criteria or constraints on the formulaic nature or the irreversibility of binomials have been explored since the 1950s to find a practical way for language learners and translators to use or translate conjoined words. Many linguists and researchers proposed qualitative measures within one language or languages belonging to the same family, based on a linguistic analysis for the extraction of binomials. Others opted for frequency-based constraints using corpus linguistic techniques.

In this paper, we used corpus linguistics and MI score statistics to test the relations between the different parts of binomials in an Arabic text and its translation into English. The most frequent binomials and conjoined phrases were then analyzed. Using these two tools enabled us to identify what to examine in the concordance lines by calculating the significant binomials. Having identified the significant binomials some rules were proposed to help language learners and translators make generalizations and understand potential constraints for the order of parts in the two languages under examination.

To do the analysis, we first used corpus linguistic techniques and MI statistics to identify the significant binomials. This test helped us identify interesting patterns, i.e., to decide what conjoined phrases should be candidates for binomials. We presupposed words with large MI scores were likely to be more interesting. We then carried out a transposition test to find out whether the phrase under examination is an irreversible binomial or not. Finally, we suggested some rules for translating such phrases which can be followed when translating from Arabic into English and vice versa. These

proposed rules can contribute to the quality of human translation and can also help those interested in machine translation to figure out an appropriate algorithm to implement these rules when building or adjusting machine translation programs.

References

Abraham, Richard D. (1950). Fixed order of coordinates: A study in comparative lexicography. *The Modern Language Journal*, *34*(4), 276-287.

Al-Jarf, Reima. (2016). Translation of English and Arabic binomials by advanced and novice student translators. In Larisa Ilynska & Marina Platonova (Eds.), *Meaning in Translation: Illusion of Precision* (pp. 281-298). Cambridge Scholars Publishing.

Al-Otaibi, Ghuzayyil. (2021). Religious binomials in Hebrew and Arabic: A review of literature. *International Journal of Linguistics, Literature and Translation*, *4*(3), 209-215.

Alexander, Richard. (1989). Fixed expressions, idioms and collocations revisited. In Paul Meara (Ed.), *Beyond Words: British Studies in Applied Linguistics* (Vol. 4, pp. 15-24).

Barnbrook, Geoff. (1996). *A Practical Introduction to the Computer Analysis of Language*. Edinburgh University Press.

Bassnett, Susan. (2002). *Translation Studies*. Routledge.

Bassnett, Susan and André Lefevere (Eds.). (1990). *Translation, History, and Culture*. Pinter.

Benor, Sarah Bunin and Roger Levy. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language*, *82*, 233-278.

Carter, Ronald. (1987). *Vocabulary: Applied Linguistic Perspectives*. Allen & Unwin.

Chamberlain, Lori. (1992). Gender and the metaphorics of translation. In Lawrence Venuti (Ed.), *Rethinking Translation: Discourse, Subjectivity, Ideology* (pp. 57-74). Routledge.

Church, Kenneth, William Gale, Patrick Hanks and Donald Hindle. (1991). Using statistics in lexical analysis. In Uri Zernik (Ed.), *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon* (pp. 115-164). Lawrence Erlbaum Associates.

Church, Kenneth and Patrick Hanks. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22-29.

Cooper, William E. and John R. Ross. (1975). World order. In Robin E. Grossman, L. James San & Timothy J. Vance (Eds.), *Papers from the Parasession on Functionalism* (pp. 63-111). Chicago Linguistic Society.

Cowie, Anthony P. (1988). Stable and creative aspects of vocabulary use. In Ronald Carter & Michael Mccarthy (Eds.), *Vocabulary and Language Teaching* (pp. 126-139). Longman.

Cruse, David Alan. (2000). *Meaning in Languages: An Introduction to Semantics and*

Pragmatics. Oxford University Press.

Cruse, David Alan. (1986). *Lexical Semantics*. Cambridge University Press.

Crystal, David. (1987). *The Cambridge Encyclopedia of Language*. Cambridge University Press.

Dickins, James, Sandor Hervey and Ian Higgins. (2002). *Thinking Arabic Translation, A Course in Translation Method: Arabic to English*. Routledge.

Ebeling, Jarl and Signe Oksefjell Ebeling. (2015). An English-Norwegian contrastive analysis of downtoners, more or less. *Nordic Journal of English Studies*, *14*(1), 62-89.

Emery, Peter. (1988). *Body-Part Collocations and Idioms in Arabic and English* [Doctoral Dissertation]. University of Manchester.

Fenk-Oczlon, Gertraud. (2001). Familiarity, information flow, and linguistic form. *Typological Studies in Language*, *45*, 431-448.

Fenk-Oczlon, Gertraud. (1989). Word frequency and word order in freezes. *Linguistics*, *27*, 517-556.

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

Hoey, Michael. (2005). *Lexical Priming: A New Theory of Words and Language*. Routledge.

Hoey, Michael. (2003). What's in a word? *MED Magazine*, (10). http://www.macmillandictionary.com/MED-Magazine/August2003/10-Feature-Whats-in-a-word.htm

Hussein, Riyad F. and Richard Lingwood. (2011). Strategies used in translating English binomials into Arabic. *Babel*, *57*(2), 168-184.

Kjellmer, Goran. (1991). A mint of phrases. In Karin Aijmer & Bengt Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 111-127). Longman.

Latham, Peter E. and Yasser Roudi. (2009). Mutual information. *Scholarpedia*, *4*(1), 1658. Retrieved November 6, 2022 from http://www.scholarpedia.org/article/Mutual_information

Lehrer, Adrienne. (1974). *Semantic Fields and Lexical Structure*. North-Holland.

Lewis, Michael. (1993). *The Lexical Approach*. Language Teaching Publications.

Malkiel, Yakov. (1959). Studies in irreversible binomials. *Lingua*, *8*, 113-160.

McIntosh, Angus and Michael Alexander Kirkwood Halliday. (1966). *Patterns of Language. Papers in General, Descriptive and Applied Linguistics*. Longmans.

Mollin, Sandra. (2012). Revisiting binomial order in English: Ordering constraints and reversibility. *English Language & Linguistics*, *16*(1), 81-103.

Motschenbacher, Heiko. (2013). Gentlemen before ladies? A corpus-based study of conjunct order in personal binomials. *Journal of English Linguistics*, *41*(3), 212-242.

Nattinger, James R. (1980). A lexical phrase grammar for ESL. *TESOL Quarterly*, *14*, 337-344.

Nöth, Winfried. (1993). Iconicity of symmetries and asymmetries in syntactic coordination. In Christoph Küper (Ed.), *Von der Sprache zur Literatur. Motiviertheit im sprachlichen und*

im poetischen Kode. [From language to literature. Motivation in the linguistic and in the poetic code] (pp. 23-36). Stauffenburg.

Parker, Steve. (2003). The psychological reality of sonority in English. *Word*, *54*(3), 359-399.

Quirk, Randolf and Sidney Greenbaum. (1973). *A University Grammar of English*. Longman.

Snell-Hornby, Mary. (1990). Linguistic transcoding or cultural transfer? A critique of translation theory in Germany. In Susan Bassnett & André Lefevere (Eds.), *Translation, History, and Culture* (pp. 79-86). Pinter.

Stubbs, Michael. (1996). *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Blackwell.

Venuti, Lawrence. (1995). *The Translator's Invisibility*. Routledge. https://doi.org/10.4324/9780203360064

Professional Profile

Abdelhamid Elewa, a PhD from Manchester University, United Kingdom (UK), is Associate Professor at the Faculty of Languages and Translation, Al-Imam University, Saudi Arabia. He translated more than 20 books from English into Arabic and vice versa. He also published a number of peer-reviewed papers on different areas of translation studies and linguistics. His research interests focus on corpus linguistics, critical discourse analysis, and semiotics

Mohamed Elaskary, a PhD from the University of Exeter, UK, is Full Professor at Hankuk University of Foreign Studies, South Korea. He formerly was Assistant Professor at Languages and Languages and Translation, Al-Azhar University, Egypt. His research interests include comparative literature, Arabic studies, Islamic studies, and Mediterranean studies.