

## Idioms in the Arabic-English Dictionary

Najib Ismail Jarad and Abdul-Fattah Abu-Ssaydeh  
*University of Sharjah, UAE*

**Abstract:** *Despite the considerable research carried out in the field of bilingual lexicography, several aspects of the Arabic-English dictionary targeting translators remain a largely uncharted area. One such area is “idioms” and their various manifestations in this type of dictionary. The aim of this paper was to examine three interrelated issues. First, it sought to determine if the bilingual dictionary (BD) covers this particular lexical phenomenon exhaustively and systematically in its source language component. Secondly, it examined canonical variation and its representation in the dictionary. Finally, the paper looked into the entry point of full idioms, truncated forms and canonical variants. The analysis demonstrated that a significantly large number of Arabic idioms are not listed, that the coverage of idioms lacks consistency, and that lemmatization is typically characterized by chaos and subjectivity.*

**Keywords:** full idioms, truncated idioms, canonical variation, bilingual lexicography, lemmatization

### 1. Introduction

Over the past three decades, linguistic studies at large and corpus linguistics in particular have shown significant interest in units beyond the word, namely multi-word units and collocations. Corpus-based research and non-corpus studies of collocations and phraseology both demonstrated the pervasiveness of these two lexical phenomena and confirmed their centrality in the lexis (Carter and McCarthy 2006; Sag, Bond, Copestake and Flickinger 2002; Jackendoff 1997; Granger and Meunier 2008; Atkins and Rundell 2008 and Schmitt 2005, just to mention a few). Within a few years, this interest was translated into new foci in lexicography, with the compilation of the first corpus-driven dictionary, *CollinsCobuild Advanced Learner’s English Dictionary* in 1987 ushering in a new era of monolingual dictionaries where collocations and multi-word units emerged as essential features.

However, the gap between the English monolingual dictionaries, especially the Big Four, and the bilingual dictionaries has remained so vast that, to quote Rundell (1999), “[t]he extraordinary range of lexical and grammatical information they ... include is rarely even approached by the best bilingual dictionaries available” (quoted in Granger and Lefer 2012: 682). This claim is particularly true of the Arabic-English dictionary (henceforth *AED*) vis-à-vis the English monolingual dictionary. The situation in the *AED* is further aggravated by the lack of a sophisticated lexical database comparable, for example, to Mark Davies’ Wikipedia Corpus (which comprises 1.9 billion words), an impediment which continues to severely hamper any meaningful lexicographic effort that

may be undertaken in the field of the *AED*. The absence of such a reliable, searchable Arabic database and any worthwhile parallel corpora is only one of several complex issues that need to be tackled in relation to the *AED*. This paper, however, sets for itself a much more modest objective; namely to examine the manifestations of the Arabic idioms in the *AED*.

## 2. Statement of the problem

Idioms are a colorful and fascinating aspect of any language and are so pervasive that ‘... it is difficult to speak spontaneously without lapsing into idiomatic usage (Johnson-Laird 1993: ix). They are commonly used in all types of contexts, informal and formal, spoken and written. Since idioms are so frequently encountered in both spoken and written discourse, they require special attention in bilingual dictionaries and should not be neglected or relegated to a secondary position.

Generally speaking, the *AED* users, whether these be language students, translators, teachers, researchers or lexicographers, are aware that the dictionary is outmoded, inconsistent in more ways than one, and that numerous aspects need to be developed. One of these aspects concerns idioms, in particular strategies employed in the creation of interlingual equivalence, the density of English idioms in the *AED*, the coverage of Arabic idioms, inclusion and/ or exclusion of classical and borrowed idioms, the treatment of canonical variations and, finally, lemmatization, or the point of entry in the dictionary. Most of these problems are, to some extent, unique to the *AED*. Lemmatization, however, seems to have a more universal dimension as our review of dictionaries of several types and in different languages will demonstrate, and as acknowledged by Hartmann and James (1998: 83) who state that lemmatization is "a problem awaiting a comprehensive solution (attempted by computational approaches) in connection with the wider tasks such as how to choose a suitable headword from the constituents of a fixed expression."

## 3. Statement of purpose

Idioms are multi-word units that are ubiquitous, institutionalized and semantically non-compositional. They are “syntactically heterogeneous” and “display different layers of lexical fixity” (Mulhall 2014: 907). (For other definitions of idioms, see Newmeyer 1974: 329; Larson 1984: 142; Carter 1987: 66; Fernando 1996; Glucksberg 2001: 68; Baker 2011: 67; Abu-Ssaydeh 2004). Certain features such as the permutations some idioms might permit and the “sensitivity that native speakers seem to have for judging when and how an idiom can be manipulated” (Baker 2011: 68) may be beyond the scope of the *AED*. Still, the authors maintain that a comprehensive, systematic and corpus-based approach to the idiomatic component of the *AED* can significantly enhance the quality of the dictionary and ease the burden of the Arab translator. To this end, this paper will examine three specific issues relating to the treatment of idioms in the *AED*: the **coverage** of the Arabic idioms including

borrowings, **canonical variation** and the extent to which it has been documented and **lemmatization** (or sublemmatization) of idioms.

#### 4. Methodology

This quantitative and qualitative analysis of Arabic idioms in the *AED* is based on lexicographic material found in current monolingual and bilingual dictionaries as well as on a large Arabic corpus which the authors use for the verification of data. The three major Arabic-English lexicographic works selected for analysis are Hans Wehr's *A Dictionary of Modern Written Arabic (DMWA)* (1974), Roḥi Baʿalbaki's *Al-Mawrid: A Modern Arabic-English Dictionary* (2012) and A. F. Abu-Ssaydeh's *Al-Murshid: A General Arabic-English Dictionary* (2013). For the purpose of the study, the authors randomly picked three hundred idioms from *معجم اللغة العربية المعاصرة* (or *DMSA* for short). Other relevant lexicographic sources are referred to where appropriate. Frequency and variations were calculated on the basis of data obtained from *المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية (KACST)*, a 732-million-word database that represents a fairly wide range of text types from different regions distributed over 100-year periods. The search is, unfortunately, restricted to single words and any attempt to search for collocations or idioms is cumbersome and occasionally quite frustrating. In a few cases, Google was resorted to as well to verify certain claims or support some statements. The variety of Arabic used is Modern Standard Arabic (MSA) as manifested in the press medium of *KACST* "because of its contemporaneousness, its coverage of many different topics, and the extemporaneous nature of daily reporting and editing. As a primary source of information about and from the Arab world, newspaper and magazine language reflects Arab editorial and public opinion and topics of current interest. Various subject matter and texts were covered, ranging from interviews, book reviews, feature stories, religion and culture, and sports reports, to straight news reports and editorials. In addition to newspapers, other sources used for data collection included contemporary novels and nonfiction" (Ryding 2005: xix).

#### 5. Coverage of the idioms and their meanings

According to Shvitiel (2007: 629-30) and Abdou (2011), MSA derives its idioms from three major sources:

- a. Classical Arabic: This category comprises a vast number of idioms that have been handed down to MSA from earlier periods and sources, including the Holy Quran (ألقى السمع، شد عضد، على قاب قوسين، سقط في يده)، the Tradition or Hadiths (جناح بعوضة، عض بالنواجذ على، ذو الوجهين، غنيمته) (باردة)، poetry (لا يقع له بالشنان، فتح بابا) (على نفسه، بنت العين) and sayings (على نفسه، بنت العين). These idioms are recorded in classical dictionaries, books of proverbs and some studies on rhetoric, metaphor and metonymy;
- b. Folklore and Colloquial Arabic: A number of idioms have, due to extensive use in the mass media and some literary publications, gained a

more formal status that guaranteed them a place in MSA. Among these we can identify *بييع الماء في حارة السقايبين، وتد جحا، ابن حلال، العشرة ببيضة*

- c. Borrowed idioms (Shivtiel's Calques): These came into Arabic primarily from the English language through media translation and include: *راهن على الحصان، الوجه الآخر للعملة (the other side of the coin)*، *حرب أعصاب (war of nerves)* and *العباءة والخنجر (cloak and dagger)*.

In this section, we shall examine the idioms that have been covered as well as those missing, and we shall try to determine if any correlation can be established between an idiom's inclusion/ exclusion and the source it comes from. But first, the paper will examine the cases where the compilers of the three dictionaries under review have either missed the whole meaning of the idiom or, where it has more than a single sense, one or more of the senses in which it is used in MSA.

### 5.1 Missing meanings

One of the major hurdles that face even experienced translators in dealing with idioms is recognizing that a given stretch of words constitutes an idiom and that it should be translated as such. Therefore, the first step in determining which strings to include in the dictionary is to ensure that such a string is in fact an idiom and, subsequently, to list it in the dictionary with an appropriate interlingual equivalent.

On the whole, lexical combinations that Arabic recognizes as idioms are dealt with as such except in a few cases. For example, *Al-Mawrid* fails to recognize three phrases as idioms; these are *برق خلب* and *قميص عثمان بنات الليل*. Strangely enough, *قميص عثمان* is translated as *a bloody shirt* which does not have any meaning beyond its literal sense. *KACSTAC* cites it in its idiomatic sense fifty-six times, especially in the press in contexts such as *جعل/ اتخذ من القضية بنات* to use the cause as a pretext/ an excuse (for). The second idiom, *بنات الليل*, can be used as an alternative form in the idiom *بنات الصدر/ الليل/ النفس* but it is also used on its own to mean *prostitutes, call girls, whores, sex workers*, a sense *Al-Mawrid* fails to account for. The third idiom *Al-Mawrid* overlooks is *برق خلب* which is cited in its literal sense: *lightning without rain*. Similarly, both *Al-Mawrid* and *DMWA* give the meaning *What has the ground to do with the Pleiades?* as an equivalent to the Arabic idiom *ابن الثرى من الثريا*. Neither monolingual English dictionaries nor Google supports this as a feasible equivalent.

We note, furthermore, that in cases where the idiom has more than a single meaning, one or the other of the dictionaries under consideration has, on occasion, disregarded this fact. Take, for example, the idiom *تقصير الباع* it is used in three separate senses: *a. stingy b. helpless* and *c. inexperienced*. *DMWA* and *Al-Mawrid* list the first two meanings. By comparison, *Al-Murshid* covers the three senses:

- i. *DMWA*: *powerless, helpless, impotent, weak, incapable; niggardly, stingy.*

ii. *Al-Mawrid*: powerless, impotent, helpless, impuissant, weak, incapable, unable, ineffective; niggardly, stingy, parsimonious.

iii. *Al-Murshid*: لا حَوْلَ لَهُ: **a.** helpless as a babe/ a child/ a ship without a rudder **b.** to be/ lying on one's back **c.** impotent **d.** powerless **e.** incapable **f.** of limited powers/ capabilities; مُقْتَرٍ: **a.** costive **b.** miserly **c.** niggardly **d.** stingy (inf) **e.** tight-fisted **f.** chintzy (inf) **g.** a cheapskate (inf); and قَلِيلِ الْخَبْرَةِ أَوْ الْإِطْلَاعِ: inexperienced.

The second example is the idiom رَغْمَ أَنفِهِ which has two meanings: *in spite of him* and *unwillingly or reluctantly*. Again, *DMWA* settles for the first of these senses, *Al-Murshid* cites the second sense while *Al-Mawrid* in fact provides both senses.

*DMWA*: in defiance of him, to spite him

*Al-Mawrid*: in spite of him, in defiance of him, against his will; unwillingly, reluctantly, forcedly

*Al-Murshid*: a. against one's will b. unwillingly

The third example is *ابن ساعته* to which Arabic dictionaries provide three meanings: ارتجالي، تلفاني، زائل. *Al-Mawrid*, again, cites the three senses. By comparison, *DMWA* gives the third sense while *Al-Murshid* cites only the first. A fourth idiom, the expression استنوق الجمل، can be used in one of the following two senses: *to be or become humble, low or lowly* and *to err, blunder, make a mistake*. Both *DMWA* and *Al-Murshid* provide only the second sense, whereas *Al-Mawrid* overlooks the idiom altogether. *Al-Murshid* also lists only the first of the three senses which the *DMSA* lists against طويل اليد: **a.** generous **b.** thieving, given to stealing **c.** quick to react violently.

The dictionaries do not only vary in the number of senses they give to a certain idiom, but they also miss some of the more intricate variations in the meaning. For example, the lexical items اضطرب and حبل combine to produce two idioms: اضطرب حبله (be thrown/ fall into a state of chaos and disorder) and اضطرب حبل القوم (to disagree, to hold divergent or conflicting views). In fact, the three dictionaries list only the first idiom and disregard the second. There are other examples, but the preceding discussion makes the point abundantly clear.

There is, however, one other idiom whose equivalents in the three dictionaries deserve to be scrutinized. The idiom ابن السبيل occurs in the Holy Quran in three Suras: Al-Baqara, Al-Nisa: and Al-Tawba, and the sense in which it is used is possibly best expressed, though perhaps a trifle archaically, by Edward Lane in his famous Lexicon: a traveler who wants to return to his country and finds not what will suffice him. If taken literally, this idiom means "the son of the road". But it was used in Classical Arabic in several meanings which we quote, again, from Lane's Lexicon: "he whom the road has brought (forth); wayfarer, traveler; he who travels much or often; the traveler who is far from his place of abode; the person to whom the way has become cut short (so that he is unable to continue his journey) (or) one who desires to return to his country or town and finds not what will suffice him; the traveler who is cut from his property, the person who desires to go to a country or town other than his own for a necessary affair; the guest who has been disabled from proceeding his

*journey, his means having failed him*". Consider these senses against the ones given in the three dictionaries and note how inadequate they are by comparison:

*DMWA: vagabond, tramp; wayfarer, traveler*

*Al-Mawrid: wayfarer, traveler; passer-by, walker; vagabond, tramp, vagrant, hobo, wanderer*

*Al-Murshid: traveler, wayfarer*

Furthermore, none of the equivalents proposed by Lane carries the fairly negative connotations of the English equivalents *vagabond, tramp, hobo* and *vagrant* suggested by both *DMWA* and *Al-Mawrid*.

## 5.2 Missing citations

More alarmingly, the dictionaries investigated in this paper demonstrate two major weaknesses. The first weakness is their apparent failure to list a large percentage of the idioms selected for the study, and the second is their failure to deal with them systematically. In this section, we shall focus on the idioms that are not listed in the dictionaries.

As stated earlier, the study has randomly picked three hundred idioms from the *DMSA*. Of the three hundred, it was discovered that seventy-nine idioms (or 26.3%) have not made it into the pages of any of the dictionaries under consideration. The unavoidable, and certainly worrying, question that must be addressed here is: why did three of the most comprehensive Arabic-English dictionaries overlook more than quarter of the idioms recognized by *DMSA* as an integral part of MSA?

Part of the answer undoubtedly lies in the practice of modern Arab lexicographers where it is not uncommon to rely heavily on classical dictionaries and list as part of MSA words, idioms and meanings that have survived up to the present (e.g. *لا حاجة في نفس يعقوب, أدنى من حبل الوريد, استنوق الجمل, لا* (شقى له غبار) alongside those that have virtually disappeared or become largely archaic. Indeed a casual perusal of almost any modern monolingual Arabic dictionary would immediately show that the lexicographer has, by intention or sheer gross negligence, blurred the demarcation line between Classical Arabic and MSA. The contents of *DMSA* bear witness to that. Take the idiom *كالقابض* for example; one version was cited by Ibn al-Sammak in the presence of The Caliph Harun al-Rashid, in Classical Arabic poetry and in both *حتى يبيض الديك* and its variant *حتى/ عندما يبيض الفأر*. *الصحاح في اللغة* and *لسان العرب* are classical idioms as well; we find reference to the first in Ibn al-Athir's *Ja:mi'u l-uṣool* (جامع الأصول). *ليس له جلد النمر* is cited in Al-Nuwairi's *Nihayat al-Arab fi Funun al-Adab* as *على طرف الثمام*. *غمر الرداء* is sublemmatized in *Lisa:nu l-Arab, al-ṣaḥḥa:h, al-Qa:mu:su l-Muḥi:t* and *Maqa:yi:su lugha*. None of these idioms nor *شديد جفن العين, ابو الأضياف, عفر جبينه, غمر الرداء, دنس الثياب, صحف ابراهيم, خمدت جمرتهم, ابن بطنه, لسان أفعى* are found in the press language of MSA according to *KACSTAC*. In fact, of the seventy-nine idioms that do not show up in the AED, forty-seven idioms (or almost 60%) have recorded zero matches in the database we tested. Hence the bilingual

lexicographer's reluctance to include them as part of MSA may find sympathetic audience, at least in some quarters.

Yet, an investigation of the corpus shows that other idioms that are rooted in Classical Arabic tend to make an occasional appearance in MSA as the statistics pertaining to the following idioms demonstrate: رفعت الأقدام وجفت الصحف (six citations)، حاطب ليل (six citations)، قلب الأمر ظهرا لبطن (four citations) and لا يفتق له بالشنان (three citations). Others are admittedly very rare, but they do appear in the corpus at least once: كناقل التمر إلى هجر and على طرف الثمام.

By comparison, there is another group of Classical Arabic idioms that are encountered in the corpus in statistically significant numbers (our definition of "statistically significant" being with enough frequency to warrant their inclusion in the AED). These idioms include ملء السمع والبصر (seventy-three citations)، أقرّب/أدنى من حيل الوريد (thirty-four citations)، ترك له الجميل بما حمل (twelve citations)، ما هكذا يا سعد تورّد الإيل (thirty-three citations) and امهات المؤمنين (thirty-four citations).

The other group of idioms absent from the three dictionaries derives its members from Colloquial Arabic (mostly Egyptian), but due to their regular use in the media (and to some extent in literature), they have acquired a more formal status. This fact is reflected in their frequency also in the Corpus. Examples include ابن بلد (over one hundred citations)، ابن نكتة (six citations)، سمار جحا (fifty citations)، مكتوب على الجبين (fifty citations) and ابن كيف. The last idiom, though, is a fairly odd addition to the DMSA. It does not appear in KACSDAC or *al-Mu<sup>c</sup>jamu l-Waṣīṭ* المعجم الوسيط، and the few cases where it is cited in Google are not sufficient grounds for its elevation to MSA; the only examples found for it are in Iraqi Arabic and in the text of Tharwat Abaza's novel *A Touch of Fear*.

Apart from this, we will find that the three dictionaries vary in the number and nature of idioms each lists. Of the three hundred idioms, DMWA fails to list one hundred and sixty-five idioms (55%). *Al-Mawrid*, on the other hand, overlooks a total of one hundred and ninety-two idioms (64%). The most comprehensive coverage is found in *Al-Murshid*, with ninety-nine missing idioms (33%). DMWA, though, displays one unique feature, namely, the inclusion of a disproportionately large number of outdated idioms including قرض رباطه (5)، وقر في خلد (1)، يحنو القذة بالقذة (5)، لزم الوساد (0)، في أنفه ورم (0)، ضرب وجه الأمر وعينه، كمش الإزار (0)، تقطعت به الحبال (0)، عينه فراره (0)، على أحر من جمر الغضى (0)، لين الأعطاف (0)، انتفتحت مساحره (1)، خالي الوطاب (1)، Overall, the total number of idioms missing from the three dictionaries is a cause for concern.

### 5.3 Borrowed idioms (Loan translations)

Although Classical Arabic and Colloquial Arabic form the major sources of idioms in MSA, borrowing continues to enrich Arabic with idioms from English. Hundreds of such idioms have crept into the language over the past few decades primarily through the media. But the status of these idioms varies as the case is in borrowing in general. Still, one can discern several trends at work in this regard. The first is that borrowing is a continuous process; monitoring the data would reveal that idioms borrowed from English are regularly added to

Arabic. Examples of some of the latest additions include *صيد رؤوس* *head hunter* and *خارج الصندوق (التفكير)* *thinking outside the box*. (The first does not appear in *KACSTAC* while the earliest recorded occurrence of the second is in an article published in *Al-Ahram* newspaper in 2011). Secondly, the introduction of a translated idiom into Arabic does not necessarily ensure a place for it in the more permanent part of the lexis. The idiom *أمسك الثور من قرنيه* appears four times in the corpus but it would hardly be recognized as an Arabic idiom by the majority of the native speakers of Arabic. Thirdly, the translation may produce what can be properly termed as “false friends”; *straw in the wind* is *anything that may indicate, slightly though, developments likely to occur in the future*. The Arabic idiom which sounds like a literal translation *قشة في مهب الريح* means *insignificant, unimportant*. Fourthly, trusting intuitions when the question concerns the likelihood of the occurrence of a given translation can be risky; neither author, for example, suspected that the English idiom *there is no use crying over spilt milk* would have a translation (*البكاء على الحليب/ اللبن المسكوب*) that would appear in *KACSTAC* more than sixty-five times. Nor did their intuition inform them that the translated idiom *فوق صفيح ساخن* *on a hot tin roof* would record more than one hundred matches in the said corpus. In fact, none of our postgraduates recognized the last two idioms as likely occurrences in MSA. Fifthly, there is nothing that is intrinsically different in the idioms that gain currency in MSA that sets them apart from those that remain on the margin; both categories are simply literal translations of English idioms, an approach that is basically the result of time pressure, ignorance, laziness or failure to employ a more appropriate strategy. Finally, some borrowed idioms have become so well-established in MSA that native speakers can hardly recognize them as borrowings. The idiom *القشة التي قصمت ظهر البعير* may “sound” so very Arabic, but in fact it comes from an idiom that has existed in the English language in varying forms since the seventeenth century. Webster Online Dictionary cites the year 1848 as the earliest recorded mention of a variant of this idiom in American English (*the last straw* translated into Arabic as *القشة الأخيرة*). Another version of the idiom (*the last straw breaks the laden camel's back*) is encountered also in Charles Dickens' novel *Dombey and Son* published in 1848.

Despite this fuzzy picture, most of the borrowed idioms have acquired a degree of permanence that warrants their inclusion in the AED. To determine how the AED has dealt with them, we have selected one hundred of these idioms and checked them in the three dictionaries. In each case, we confirmed that the borrowed idiom is in currency in MSA by ensuring that it occurs in *KACSTAC* at least five times. Upon examination, it was found that *DMWA* listed seven idioms (7%), *Al-Mawrid* seventeen (17%) and *Al-Murshid* ninety-four (94%). The scarcity of borrowed idioms in the first two dictionaries is attributed to the fact that *DMWA* was first published in the early sixties and *Al-Mawrid* in 1988, long before many of the borrowed idioms have established a foothold for themselves in Arabic. Neither dictionary, to the best of our knowledge, has ever undergone any updating. Some may argue that the comparison, therefore, is not fair, but one must remember that dictionaries such

as the Oxford series and Merriam Webster have stood the test of time only because of the ceaseless efforts of dedicated teams that regularly keep the dictionaries up-to-date.

### 6. Canonical variation

While binomials (*fish and chips, hammer and tongs*), proverbs (*a stitch in time saves nine*), frozen similes (*as bold as brass, as cool as a cucumber*), formulaic expressions (*once upon a time*) and some idioms (*بالبيع والنزاع، لا ناقة له ولا جمل*) tend to be syntactically and lexically fixed, it is safe to claim that a fairly large number of idioms have a propensity to display lexical, syntactic and even morphological variation (Dronov 2011: 147). This variation was noted by Sköldbberg (2004), Abu-Ssaydeh (2005), Mulhall (2010), Heid (2008), Liu (2012), Fellbaum (1993) and Abdou (2011). The occurrence of lexical variation “constitutes alternative institutionalized variants” and can be explained with reference to text type, level of formality (Abdou, 2011: 104), regionalism (Moon 1996:254) or speaker’s intention and expressive needs (Liu 2012: 112). Variation in the lexical constituency of the idiom, according to Moon (1996: 254), is a reflection of “the wooliness, indeterminacy and instability of idioms” which, to quote Moon again, “demonstrate(s) conclusively that idioms often do not have fixed forms and are formally unstable. This is a very simple, observable fact, and very important. There are immense repercussions” (p. 246). These “repercussions” continue to reverberate in at least two areas: teaching foreign languages and lexicography. As far as the latter is concerned, variations attested in the lexical and grammatical form of a given idiom need to be accurately and fully recorded in the dictionary. This, understandably, will complicate decisions relating to lemmatization and may even call into question the notion of “canonical form” altogether, at least for a number of idioms. Let us now look at variation in the idioms under investigation.

In order to evaluate the data under discussion, we should initially distinguish between canonical variations in the idiom, which represent recurrent, regular and consistent “alternative renderings” of the same idiom and the likely modifications it may undergo such as passivization, nominalization, topicalization or change of tense (Phillip 2008: 95). The latter are viewed as “embellishments, additions or reductions whose meaning and function are ultimately dependent on, and reducible to, the canonical form from which they are derived” (Phillip 2008: 96). For example, in the idiom *حصّة/ نصيب الأسد*, the words *حصّة* and *نصيب* are canonical variations since they alternate in the initial position and enjoy permanence as part of the institutionalized form of the idiom. On the other hand *انشرح صدره* should be viewed as a grammatical modification that the canonical form *شرح الصدر* may tolerate. Thus in the first instance, both *حصّة* and *نصيب* need to be recorded while the passive form of *شرح الصدر* would be left out. Moreover, we need to statistically establish on the basis of data in the corpus the validity of all the institutionalized variants (including truncated versions) and determine where the line is to be drawn between collocations on the one hand and the canonical form(s) of the idiom concerned on the other. To

illustrate, the idioms ذرف دموع التماسيح appears in *KACSTAC* forty-one times compared to its truncated form دموع التماسيح which records thirty matches. These frequencies would support a decision to list both as canonical forms. The verbs تباكى and سكب seem to be rare and stylistic variations that need not be recorded in the dictionary. Thirdly, in the case of competing canonical forms, the lexicographer needs to decide if criteria can be determined to establish which of the two should be cited in the dictionary as the canonical form.

Variation in the Arabic idiom may involve the verb ( فرس له البساط )، the noun ( حصاة/ نصيب الأسد، شرح صدره/ خاطره )، synonyms ( يعلم/ يعرف من أين )، adjectives ( ضيق ( ذات ) اليد )، addition ( وضع اللمسات الأخيرة/ النهائية )، ( يؤكل/ تؤكل الكتف )، ( يئس/ يئس )، ( من رأسه حتى (أخمص) قدميه )، non-synonymous words ( رقة )، (عند عودته/ ظهره/ ساعده، ورقة )، different word classes ( بعد علو الهمة، and عالي الهمة بعيد الهمة، and بعد الهمة )، (التين/ التوت )، (يعيد النظر and النظر )، variation at more than one point in the idiom ( دفن/ وضع رأسه في )، (التراب/ الرمل/ الرمال )، syntactic structure ( رغم أنفه، رغمًا عن أنفه، بالرغم منه، على الرغم )، borrowed idioms ( وراء ابواب مغلقة، أصاب/ اصطاد/ ضرب عصفورين بحجر )، (منه )، (واحد، أرجع/ أعاد/ رجع إلى المربع الأول، أمام/ على/ عند مفترق طرق، نق جرس/ ناقوس الخطر )، etc. Variations may at times be dauntingly numerous; an excellent example is the idiom أتى/ قضى على الأخضر واليابس. Data obtained from *KACSTAC* show that the initial position can be filled by the near-synonyms قضى، التهم، طحن، عصف ب، محرق، أكل، حصد، أخذ/ اجتاح في طريقه، اقتلع، فتك ب، حرق، محاطال. Further data derived from Google show that these candidates appear with the following frequencies:

أكل الأخضر واليابس: 34,600	أتى على الأخضر واليابس: 20,700	قضى على الأخضر واليابس: 18,500
طال الأخضر واليابس: 12,600	حرق الأخضر واليابس: 12,000	دمر الأخضر واليابس: 9,100
أحرق الأخضر واليابس: 6,090	التهم الأخضر واليابس: 6,060	حصد الأخضر واليابس: 2930
أقتلع الأخضر واليابس: 1950	عصف بالأخضر واليابس: 149	طحن الأخضر واليابس: 69
أشعل الأخضر واليابس: 8		

The two questions we would like to address here are: do the recorded idioms reflect the most statistically significant versions of the idiom? Secondly, to what extent are these variations recorded in the three dictionaries?

To answer the first question, let us consider the statistical distribution of the various versions of the idiom أطلق/ ألقى/ ترك الحبل على الغارب/ غاربه. When we googled the idiom, we got the following results:

أطلق له الحبل على الغارب: 1720	ترك له الحبل على الغارب: 2900	ألقى له الحبل على الغارب: 6980
ألقى له الحبل على غاربه: 0	ترك له الحبل على غاربه: 134	أطلق له الحبل على غاربه: 971

These statistics demonstrate that the most frequent form is ألقى له الحبل على الغارب which is indeed the form cited by the three dictionaries. But both *DMWA*

and *Al-Murshid* fail to record the second most frequent variant *ترك له الحبل على* (2900 citations). Yet, *DMWA* gives the verb *أطلق* as a variant in the initial position, but it ignores the second frequent verb *ترك*. *Al-Mawrid*, on the other hand, fails to recognize *ترك* and *أطلق* as potential competitors for the initial position. To complicate the picture further, none of three dictionaries cites the variant *أطلق/ترك له الحبل على غاربه*.

Take another example: *رجع/عاد بخفي حنين/خاوي الوفاض/خالي الوفاض*. The frequency with which the variants of this idiom appear are shown below:

رجع بخفي حنين: 22800	عاد بخفي حنين: 13500	عاد خاوي الوفاض: 8750
رجع خالي الوفاض: 4930	عاد خالي الوفاض: 4720	رجع خاوي الوفاض: 246

The analysis of our data shows that the *DMWA* is satisfied with citing *رجع بخفي حنين* as well as the verbless truncated form *خالي/خاوي الوفاض*. *Al-Mawrid* cites the most frequent form (*رجع بخفي حنين*) but it does not make any reference to any other possible variations, while *Al-Murshid* records *رَجَعَ بِخَفِي حُنَيْنٍ/ خَالِي أَوْ خَاوِي* without indicating the possibility of filling the initial position with the verb *عاد*.

Other examples of inconsistency abound. The variant forms of the idiom *أكل الأخضر واليابس* (assuming, for the sake of convenience, that this is the canonical form) show that this form is indeed the most frequent (34,600 matches in Google), followed by *أتى على الأخضر واليابس* (20,700 matches). Yet, neither *DMWA*'s nor *Al-Mawrid* notes this fact; both cite the second most frequent form. Google cites *دق ناقوس الخطر* 2640 times compared to *دق جرس الخطر* which appears 236000 times. Again, both *DMWA* and *Al-Mawrid* opt for the first of the two versions which has significantly lower frequency. In the case of the idiom *قضى نحبه/أجله*, the three dictionaries are systematic; they all cite the two possible forms *قضى نحبه* and *قضى أجله* despite the huge difference in their statistics: the first appears in Google 368,000 compared to the second which scored 2530 hits only. This is not the case in the idiom *يعرف/يعلم من أين يؤكل الكتف* while *Al-Murshid* cites the most frequent form of the idiom *عرف من أين تؤكل الكتف* which appears 36,100 times in Google, both *DMWA* and *Al-Mawrid* list the less frequent *يعلم من أين تؤكل الكتف* which is cited in Google 3,210 times. The last example involves the idiom *شمر عن ساعده/ساقه/ساعد الجد/ساعديه/للأمر*. Here, *DMWA* cites only *شَمَّرَ عَنِ سَاعِدِ الْجَدِّ* (which has a frequency of 3840) while *Al-Mawrid* cites *شَمَّرَ عَنِ سَاعِدِ الْجَدِّ* and *Al-Murshid* lists *شَمَّرَ عَنِ سَاعِدِ الْجَدِّ*. Significantly, none of the dictionaries cites the form with the highest frequency *شَمَّرَ عَنِ سَاعِدِ الْجَدِّ* which has returned 39,200 results by comparison with *شَمَّرَ* (12,700), *شمر عن ساعده* (10,600), *شمر عن ساعده* (3840) and *شمر للأمر* (489).

Let us now return to the second question: how much of the variation in the idioms is actually documented in the three dictionaries? The total number of tested idioms that display lexical variation is forty-three. Of these, the three dictionaries cite fully only three idioms: *طويل الأناة/البال/الروح* and *أخذ بيده/بناصره*, *قضى نحبه/أجله*. The remaining idioms are cited in varying degrees of

completeness. In several instances, each dictionary cites its own incomplete version of the idiom. For example, the idiom *أسقط/سقط في يده* is cited fully in the *DMWA* but the first version is cited by *Al-Murshid* (*أسقط في يده*) while the second version (*سقط في يده*) is given in *Al-Mawrid*. In the majority of cases, the three dictionaries fail to record the full form of the idiom in question, settling for a shortened version. All in all, the discrepancies, inconsistencies and incompleteness of the idioms cited in the three dictionaries are glaringly apparent and need to be addressed.

### 7. Lemmatization/ Sublemmatization of idioms

The process of lemmatizing idioms in both the monolingual and the bilingual dictionaries is not an easy task and is open to debate. The difficulty arises from the fact that the idiom is a multi-lexical entity. It may contain different grammatical categories, function words and content words. It may also display canonical lexical and syntactic variations, a feature which involves up to forty per cent of the idioms according to Moon (1998). Semantically, idioms vary in the degree of their non-compositionality and, unlike collocations, idioms are perceived as single lexical units that “carry meaning in the same way as do single words. And just as in the single words, their lexical meaning can be of different types” (Zgusta 1971: 157). This word-like function encouraged some researchers including Al-Kasimi (1977), Gouws (1991) and Botha (1992) to advocate that idioms should not be sublemmatized under any words but be lemmatized as independent entries. Other researchers in the field of theoretical lexicography have proposed alternative approaches and advanced arguments in their support. For example, Petermann (1983 quoted in Mulhall 2014) suggests listing the idiom in an “entry of one notional component with supplementary cross-references at other possible entry points” (1906). Tomaszczyk (1986) proposed that idioms (and collocations) be entered “under all major constituents. They can then be glossed under the first component and cross-referenced at the others” (291). Burger (2007) maintains that idioms should only be listed once, but does not make any reference to a specific point of entry. Botha (1992) and Lorentzen (1996) suggest a first main component entry and a noun entry strategy respectively (Mulhall 2010: 1356). Kovecses (2001) proposes an arrangement system that “should follow the presumed conceptual organization of idioms: it should indicate the target domain, the source domain and the scope of the source domain for the idioms that are based on a particular metaphor source” (113). Mulhall (2010: 1355) challenged the long-established and time-honored tradition of placing the idiom under one of its major constituents for a variety of reasons.

Instead, he proposes an extremely complex model which he bases primarily on “the intrinsic semantic and lexical features of idioms” with the entry point being chosen “on the basis of the most salient semantic or lexical feature of the (idiom group in question)” (2010: 1367). Harras and Proost’s entry model (2005) partly overlaps with Mulhall’s; they “advance an entry model for idioms, configuring their point of entry in accordance with their

semantic features; resulting in semantically opaque idioms being lemmatized and semantically interpretable idioms sub-lemmatized” (quoted in Mulhall 2014: 906).

For some inexplicable reasons, this ongoing controversy in theoretical lexicography does not seem to have significantly altered the practice entrenched in the process of compiling dictionaries. This divorce between theory and practice may suggest to the observer that lexicographers may have found an orderly and systematic approach to lemmatization/ sublemmatization of idioms. The truth of the matter is that dictionaries of idioms, general English monolingual dictionaries and general bilingual dictionaries continue to deal with the issue of idiom lemmatization in a way that can best be described as chaotic. For example, Mulhall’s review of English-Italian and Italian-English dictionaries shows that the process is largely subjective and inconsistent (2010). According to Lorentzen (1996) lexicographers tend to sublemmatize the idiom under the ‘semantically heaviest word’ (415) though no-one is certain of the sense in which “heavy” can be quantified or qualified or how the lexicographer can determine if one word is “heavier” than another. The Danish Dictionary Project sublemmatizes the idiom under the first semantically heavy noun because nouns “contribute heavily to the meaning of the phrases” (Lorentzen 1996: 417). If no noun exists then priority is given, and in the same order, to the first verb, the first adjective and the first adverb (pp. 417-418). Manser and Betsis, the authors of *Dictionary of Idioms* (2006) arrange idioms in the dictionary “in alphabetical order of keywords. A keyword in an idiom is usually its first noun, or if it has no noun, its first verb; if it has no verb then its first adjective ... but in all these cases, cross-references are included at appropriate places to help locate the entry easily” (p. 7). Finally, a study conducted by the present authors on lemmatization of idioms in both English dictionaries of idioms and in the Big Four also reveals that the approach adopted varied greatly and lacked consistency even within the same dictionary. A detailed account of the findings, however, was ruled out due to space restrictions in the present paper.

In addition to the question of lemmatization, dictionary compilers have to determine what has come to be known as the “canonical form” of the idiom. This is the citation form in the dictionary and acts as the major and legitimate source from which other versions are assumed to be derived. The problem, however, is that the notion can apply only in part to idioms. It is true that a certain percentage of the idioms in language have fossilized forms, which makes the identification and the listing of the idiom much easier. Examples would include *جزء سنمار* and *حاطب ليل، لا في العبير ولا في النفير، أحشفا وسوء كيلة*. But it is equally true that a vast number of idioms displays variations that can best be described (and treated as) canonical variations, i.e. variations that are institutionalized and have as much legitimacy to appear in the dictionary as the presumed canonized versions (Philip, 2008 quoted in *The Routledge Linguistics Encyclopedia* (idioms, page 268); Moon, 1996: 246). The idioms *أقرب من حبل الوريد، أولا وأخيرا/ وأخرا، ثاقب البصر/ النظر/ الفكر* are examples

representative of this category. Such variations need to be represented in the dictionary. This is on the one hand. On the other, even in the fairly small and limited data being investigated in this paper, it is evident that within certain idioms, a truncated version seems to enjoy enough independence and mobility to merit separate listing. *KACSTAC* clearly shows that the idioms أعطى الضوء الأخضر، أحرز/ نال قصب السبق، أخبط من حاطب ليل، جازاه جزاء سمنار، عاد الى المربع الأول، الضوء الأخضر، أخصر، قصب السبق، حاطب ليل، جازاه جزاء سمنار، عاد الى المربع الأول، ثالثة الأثافي regularly produce the versions أعطى الضوء الأخضر، أخصر، قصب السبق، حاطب ليل، جازاه جزاء سمنار، عاد الى المربع الأول، ثالثة الأثافي. Again, this fact represents a further challenge to the notion of “canonical form” and has to be expressed in the dictionary.

Let us return to the three dictionaries under discussion. Here, there are three issues that need to be discussed; the variation in the word class of the initial word, truncated versions of Arabic idioms and, finally, the point of entry selected for the idiom.

Although the data selected for analysis (400 idioms in all) may be limited by comparison to the total number of idioms in Arabic, it is sufficient to reflect certain important trends in the three dictionaries. (In some instances, however, this section will cite examples not included in the data for the sake of exemplification). One such trend is disagreement on the canonical form, especially in regards with the word class of the initial word which, more often than not, is central in the process of lemmatization. For example, the idiom قاطع طريق appears in two other forms: the noun phrase قطع الطريق and the verb phrase قطع الطريق، all of which are cited in the dictionaries under consideration. The three dictionaries also cite the verb phrase شمش أنفه أو بأنفه as well as the version beginning with the adjective شامخ الأنف as canonical variants. In instances where the canonical forms alternate between the two patterns Adjective + Noun and Noun + Noun, we note that the three dictionaries tend to cite both forms (e.g. طول الباع and أصالة الرأي، طول الأناة alongside أطول الرأي، طويل الأناة، طويل الباع). This approach is not a problem in itself for the dictionary user; the problem lies in the fact that it is not carried out consistently; both *Al-Mawrid* and *DMWA*, for example, fail to list طول اللسان and the latter does not cite طول الباع. *Al-Murshid* and *DMWA* cite the two variants نفاذ البصيرة and نافذ البصيرة whereas *Al-Mawrid* cites only the first form. *DMWA* lists لتين العريكة but not its N + N counterpart لين العريكة. In the case of the idiom أثلج/ شرح الصدر *Al-Murshid* and *Al-Mawrid* lemmatize the active form of the verb compared to *DMWA* which also canonizes the passive version of the idiom. This disagreement extends to other idioms and other word classes; both *Al-Mawrid* and *Al-Murshid*, for example, cite أخذ وأعطى while *DMWA* has decided to sublemmatize the noun phrase أخذ وعطاء.

As we stated earlier, the corpus shows that a number of the idioms in the data do appear in truncated forms. This is a significant fact for sublemmatization since such short versions of the idioms need to be sublemmatized separately in order to emphasize their autonomy. The results of the analysis show that this fact is overlooked in the majority of cases:

Table 1. Lemmatization of truncated forms of idioms

Dictionary/ Idiom	<i>DMWA</i>	<i>Al-Mawrid</i>	<i>Al-Murshid</i>
الضوء الأخضر	No	Yes	No
قصب السبق	No	No	No
حاطب ليل	Yes	No	No
جزاء سنمار	No	No	Yes
المربع الأول	No	No	No
ثلاثة الأثافي	Yes	No	Yes

Let us return now to the main issue of lemmatization/ sublemmatization and ask ourselves: is any of the theorization discussed above applied to the process of lemmatization/ sublemmatization in the A-E dictionary? Historically, both *DMWA* and *Al-Mawrid* pre-date most of the research reviewed in the preceding paragraphs. Therefore, it would be unfair to assess them on those grounds. *Al-Murshid*, which appeared recently, does not make any reference to such research findings but follows, fairly systematically, in the footsteps of other monolingual and bilingual dictionaries available on the market. Beyond this, it should be emphasized that cognitive, semantic and word-class considerations advocated by some theorists in the preceding paragraphs were of no relevance whatsoever to the three dictionaries. It follows that no idioms have been recognized as separate entities or lemmatized on their own in any of the dictionaries. Still, there are other several issues we need to look into: the point of entry in the dictionary, repetition, cross-referencing and overall consistency in the approach used in each dictionary.

*Al-Mawrid* gives no explanation to the question of lemmatization/ sublemmatization or the nature of the entry under which the idiom is listed, but it uses cross-referencing as part of its methodology “to save space and avoid repetition” (Ba<sup>°</sup>albaki: 9). As a general rule, though, *Al-Mawrid* cites the idiom and provides a cross-reference for it under one or more of its lexical constituents. For example, *ابن ساعته* is listed under *ابن* and cross-referenced under *ساعة*. The principle is applied to other examples, including *ابن السبيل، أثقل الكاهل، أخذ بيده/ بناصره، سقط في يده، اشند ساعده*, etc. However, the compiler fails to systematically adhere to this approach. The idiom *البيت الحرام* is cited fully under its two lexical constituents as is the idiom *بنات الصدر*. The idiom *ابا عن جد*, on the other hand, is sublemmatized under *أب* but is neither repeated nor cross-referenced under *جد*. The dictionary sublemmatizes *أتى على الأخضر واليابس* under the verb *أتى*, provides a cross-reference under *أخضر* but fails to cite or cross-reference the idiom under the adjective *يابس*. Similarly, the dictionary cites *أكل عليه الدهر وشرب* under *أكل*, cross-references it under *دهر* but forgets about it completely at *شرب*. In the idiom *ألقي له الحبل على الغارب*, the point of entry is, in fact, the lexical constituent in the middle, *حبل* with cross-references at the initial verb and the under last noun. It is evident, thus far, that the point of entry is neither impacted by the initial lexical constituent nor by its semantic “weight”. It

is equally obvious that the syntactic pattern of the idiom is not of any relevance either. To illustrate, let us take the Verb Phrase. The idiom *أثقل الكاهل* is cited under the verb and cross-referenced under the noun. But other idioms of the same pattern, e.g. *أثقل الصدر، اقتفى أثره*, make only one appearance under the initial verb. The same principle applies to *أطلق ساقيه للريح*. In another idiom with the same grammatical pattern, *اختلط الحابل بالنابل*, *Al-Mawrid*, oddly enough, cites the idiom under the second word *حابل* and cross-references it under both the verb and the last noun. A third example of the same grammatical pattern, *أخذ مجراه*, shows that *Al-Mawrid* cites the idiom in full under both lexical constituents. This lack of consistency is further observable in another grammatical pattern: Adjective + Noun. The idioms *ثاقب الفكر/ النظر، ثقيل الفهم* and *ثقل الدم* are sublemmatized under the adjectives without any further cross-referencing to them either under the nouns. Conversely, *ثقل الفكر، شارد الفكر* and *ثفيف الدم* are cited under the adjective and cross-referenced under the noun.

The introduction of *DMWA* does not shed any light on the question of lemmatization, so the researcher has to figure out what system, if any, has been employed in the dictionary. Like *Al-Mawrid*, *DMWA* cites *أبا عن جد* under *أبا* but does not repeat or cross-reference it under *جد*. The Noun + Noun idioms such as *أبن ساعته، ابن السبيل، بنات الدهر، بنات الصدر* are sublemmatized in full under both lexical constituents. No cross-referencing is employed in these instances, however. The citation of *أتى على الأخضر واليابس* corresponds with that found in *Al-Mawrid* as the case is in the idioms *اختلط الحابل بالنابل* and *أخذ مجراه* – the first is cited under the first noun and the second is repeated in full under both constituents. But the idioms *أثقل الكاهل، أثقل الصدر* and *أثقل الكاهل، أثقل الصدر* are all listed under the initial verb, with no-cross-referencing. The sublemmatization of the noun phrase (Adjective + Noun) is totally chaotic. The idioms *ثاقب الفكر/ النظر، بعيد النظر، ثاقب الفكر/ النظر، بعيد النظر، ثاقب الفكر/ النظر، بعيد النظر، ثاقب الفكر/ النظر، بعيد النظر* all appear under the adjective. By comparison, idioms of the same grammatical pattern *قصير الباع، بعيد الشأو، بعيد الغور، قصير الباع، بعيد الشأو، بعيد الغور، قصير الباع، بعيد الشأو، بعيد الغور* and *مشوش الفكر، سلس القيادة، شارد الفكر* are listed fully under both the noun and the adjective. The idioms *عالي الهمة، بعيد الهمة، صعب القيادة، مرهوب* and *عالي الهمة، بعيد الهمة، صعب القيادة، مرهوب* display a third approach: they all are listed under the noun and not the adjective.

In *Al-Murshid*, the arrangement employed depends on the number of words in the idiom. In principle, the idiom is not recognized as a separate entity but is sublemmatized under one of its lexical components. If the idiom consists of two words, it is sublemmatized under the entries representing both content words. Idioms consisting of more than two words are sublemmatized under the first content word. Additionally, no attempt is made to separate collocations from idioms as both categories were listed together in an alphabetical order. The motivation for this approach was purely practical as cross-referencing and/ or repetition of each idiom under every entry it contains would have increased the size of the dictionary to a point where it had to be published in two volumes. This was a prospect the compiler, one of the authors of this paper, would not entertain.

### 8. Concluding remarks

Newmark (1998: 29) underscored the importance of the bilingual dictionary for both trainee translators and practicing professionals by saying that “The bilingual dictionary is the translator's single, first and most important aid, and a translator who does not consult one when in doubt is arrogant or ignorant or both.” But to achieve this vital role in translation, the bilingual dictionary has to be exhaustive, consistent and up-to-date in its coverage of language and organization of its lexical matter, including idioms.

Unfortunately, an examination of the contents and arrangement of idioms in the major Arabic-English dictionaries demonstrates that the coverage of idioms is neither consistent nor exhaustive. It also shows that a significantly large number of Arabic idioms is not listed. Colloquial-based idioms vary from one country to another and there seems to be no agreement on which of these idioms has become formal enough to warrant inclusion in the dictionary. Apart from *Al-Murshid*, established borrowed idioms are generally overlooked and variations in the idiom's syntax and lexical constituency are neither uniformly handled nor are they universally recognized. Finally, the arrangement of idioms does not seem to follow any obvious scheme or principle, whether it be semantic, grammatical or otherwise. Briefly put, the treatment of idioms is characterized by chaos and subjectivity. To be fair to our lexicographers, though, this lack of consistency seems to be prevalent not only in A-E dictionaries but also in monolingual English dictionaries and specialized dictionaries that take the idiom as their primary domain. To quote Atkins and Rundell (2008: 168), idioms are “the most difficult multiword expressions (MWE) to handle in lexicography. In the absence of hard and fast criteria, it is well-nigh impossible to be wholly consistent.”

Despite this complex and varied nature of idioms, however, the status quo in the Arabic-English dictionary is not a fait accompli. At a time when lexicographers have at their disposal huge corpora with powerful search engines, determining the precise and pertinent features of an idiom is no longer an insurmountable obstacle. With some ingenuity, a database like *KACSTAC*, for example, can provide the researcher with tools capable, objectively and quantitatively, of identifying the exact wording of the idiom, possible incidental and canonical variations in its syntactic features and lexical constituency, frequency of its occurrence, authentic information on the time it made its debut in Arabic and its fate in the language. This type of information will enable the lexicographer to sift through idioms, separating those belonging to Classical Arabic from others that can be legitimately placed in Modern Standard Arabic. This stance, it should be emphasized, is not a call to overlook idioms that belong to Classical Arabic and may crop up in a modern texts; it is merely a reminder that such idioms are in fact archaic, old-fashioned or simply obsolete and are more appropriately documented in historical dictionaries. Borrowed idioms can be pinpointed and their frequencies identified with a degree of accuracy sufficient for an informed decision on their inclusion.

The question of variation can also be settled objectively with the help of a reasonable corpus; variations that have acquired a canonical status must be recognized and recorded in the dictionary while incidental ones can be left out. On the other hand, space limitations in paper dictionaries do not represent an issue in the age of electronic and online dictionaries; in an electronic version, each idiom can be recorded in full under every single content word regardless of the idiom's length or complexity. All this, though, cannot guarantee that the dictionary would accurately reflect the real situation in language; lexicographers need to constantly monitor the changing idiomatic scene in order to regularly record changes and keep their dictionaries up-to-date.

Najib Ismail Jarad  
Department of English Language and Literature  
University of Sharjah  
UAE  
Email: [njarad@sharjah.ac.ae](mailto:njarad@sharjah.ac.ae)

Abdul Fattah Abu Ssaydeh  
Department of English Language and Literature  
University of Sharjah  
UAE  
Email: [abussaydeh@sharjah.ac.ae](mailto:abussaydeh@sharjah.ac.ae)

## References

- Abdou, Ashraf.** (2011). *Arabic Idioms: A Corpus-based Study*. London: Routledge.
- Abu-Ssaydeh, Abdul-Fattah.** (2005). 'Variation in multi-word units: The absent dimension'. *Studia Anglica Posnaniensia*, 41: 125-147.
- Abu-Ssaydeh, Abdul-Fattah.** (2004). 'Translation of English idioms into Arabic'. *Babel*, 50 (2): 114-131.
- Al-Kasimi, Ali.** (1977). *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill
- Atkins, Beryl T. Sue and Michael Rundell.** (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Baker, Mona.** (2011). *In Other Words: A Coursebook on Translation*. London: Routledge.
- Botha, Willem.** (1992). 'The Lemmatization of expressions in descriptive dictionaries'. In Hannu Tommola, Krista Varantola, Tarja Salmi-Tolonen & Jurgen Schopp (eds.). EURALEX '92 Proceedings I-II. Euralex International Congress, Tampere, August 4-9, 1992. Department of Translation Studies, University of Tampere. pp. 493-502
- Burger, Harald; Dmitrij Dobrovolskij, Peter Kuhn, and Neal R. Norrick,** (eds.), (2007). *Phraseologie / Phraseology: An International Handbook of Contemporary Research*, Volume 2. Berlin: Walter de Gruyter.
- Carter, Ronald and Michael McCarthy.** (2006). *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge: Cambridge University Press.
- Carter, Ronald.** (1987). *Vocabulary: Applied Linguistic Perspectives*. London: Allen and Unwin.
- Dronov, Pavel.** (2011). 'Idiom modifications in bilingual dictionaries of idioms'. *Jezikoslovlje*, 12 (2): 147-163.
- Fellbaum, Christine.** (1993). 'The determiner in English idioms'. In Cristina Cacciari and Patrizia Tabossi (eds.), *Idioms: Processing, Structure, and Interpretation*. New Jersey: Lawrence Erlbaum Associates Inc.
- Fernando, Chitra.** (1996). *Idioms and Idiomaticity*. Oxford: Oxford University Press.
- Glucksberg, Sam.** (2001) *Understanding Figurative Language: From Metaphors to Idioms*. Oxford: Oxford University Press.
- Gouws, Rufus.** (1991). 'Toward a lexicon-based lexicography dictionaries'. *Journal of the Dictionary Society of North America*, 13: 75-90.
- Granger, Sylviane and Marie-Aude Lefler.** (2012). 'Towards more and better phrasal entries in bilingual dictionaries'. In *Proceedings of the 15th EURALEX International Congress*, R. Vatvedt Fjeld and J.M. Torjusen (eds), 682–692. Oslo: UiO.

- Granger, Sylviane and Fanny Meunier.** (2008). 'Phraseology in language learning and teaching: where to from here?' In Fanny Meunier and Sylviane Granger. (eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: Benjamins, pp. 247-252.
- Harras, Gisela and Kristel Proost.** (2005). 'The Lemmatisation of idioms'. In Henrik Gottlieb, Jens Erik Mogensen and Arne Zettersten (eds.), *Symposium on Lexicography XI, Proceedings of the Eleventh International Symposium on Lexicography*. Copenhagen, May 2-4 2002 Tübingen: Max Niemeyer, pp. 277-291.
- Hartmann, Reinhard R, K. and Gregory James.** (1998). *Dictionary of Lexicography*, London, England: Routledge.
- Heid, Ulrich.** (2008). 'Computational phraseology: An overview'. In Sylviane Granger and Fanny Meunier (eds.), *Phraseology: an interdisciplinary perspective*. Amsterdam/Philadelphia: Benjamins, pp. 337-360.
- Jackendoff, Ray.** (1997). 'Twistin' the night away'. *Language*, 73 (3): 534-59.
- Johnson-Laird, P.N.** (1993). 'Foreword'. In Cristina Cacciari and Patrizia Tabossi. (eds.), *Idioms: Processing, Structure, and Interpretation*. New Jersey: Lawrence Erlbaum Associates
- Kövecses, Zoltán.** (2001). *An English-Hungarian Idiom Dictionary for Learners of English*. Budapest: MPL Könyv Kft.- Librotrade.
- Larson, Mildred.** (1984). *Meaning-based Translation*. New York: University Press of America.
- Liu, Zhengyuan.** (2012). 'Analysis of idiom variation in the framework of linguistic subjectivity'. *English Language Teaching* 5 (6): 105-113. <http://www.ccsenet.org/journal/index.php/elt/article/viewFile/17471/11643>
- Lorentzen, Henrik.** (1996). 'Lemmatization of multi-word lexical units: In which entry?'. In M. Gellerstam et al. *Euralex 1996 Proceedings*. 415-421.
- Martin, Manser and Andrew Betsis.** (2013). *English Idioms Dictionary*. GLOBAL ELT.
- Moon, Rosamund.** (1998). 'Frequencies and forms of phrasal lexemes in English'. In Cowie, Anthony Paul. (ed.), *Phraseology. Theory, Analysis, and Applications*, Oxford: Oxford University Press, pp. 79-100.
- Moon, Rosamund.** (1996). 'Data, description, and idioms in corpus lexicography'. In Gellerstam, Martin et al. (eds.) *Euralex '96 Proceedings*: 245-256.
- Mulhall, Chris.** (2010). 'A semantic and lexical-based approach to the lemmatisation of idioms in bilingual Italian-English dictionaries'. *EURALEX 2010 Proceedings*. [http://www.academia.edu/2284537/A\\_Semantic\\_and\\_Lexical-based\\_Approach\\_to\\_the\\_Lemmatisation\\_of\\_Idioms\\_in\\_Bilingual\\_Italian-English\\_Dictionaries](http://www.academia.edu/2284537/A_Semantic_and_Lexical-based_Approach_to_the_Lemmatisation_of_Idioms_in_Bilingual_Italian-English_Dictionaries)
- Mulhall, Chris.** (2014). 'The treatment of tripartite Italian idioms in monolingual Italian and bilingual Italian-English dictionaries'. *Proceedings of the*

- XVI EURALEX International Congress: The User in Focus*, pp. 905-913.
- Newmark, Peter.** (1988). *A Textbook of Translation*. London: Prentice Hall.
- Newmeyer, Fredrick.** (1974). 'The regularity of idiom behavior'. *Lingua*, 34: 327-342.
- Phillip, Gill.** (2008). 'Reassessing the canon: Fixed phrases in general reference corpora'. In Sylviane Granger and Fanny Meunier (eds.), *Phraseology: an interdisciplinary perspective*. Amsterdam/Philadelphia: Benjamins, pp. 95-108.
- Ryding, Karin.** (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger** (2002). 'Multi-word expressions: A pain in the neck for NLP'. <http://lingo.stanford.edu/pubs/WP-2001-03.pdf>
- Shivtiel, Avihai.** (2007). 'Phraseology'. In Versteegh, K. et al. (eds.) *Encyclopedia of Arabic Language and Linguistics*. Vol. III. Leiden: Brill, pp. 628-634
- Schmitt, Norbert.** (2005). 'Formulaic language: Fixed and varied'. *ELIA* 6 (6): 13-39
- Sköldberg, Emma.** (2004). *Cards on the Table: Variations in Content and Expression in Swedish Idioms*. PhD dissertation. Gothenburg: Meijerbergs institut för svensk etymologisk forskning, University of Gothenburg, Sweden. <https://gupea.ub.gu.se/handle/2077/16279>
- Tomaszczyk, Jerzy.** (1986). 'The bilingual dictionary under review'. In Mary Snell-Hornby (ed.), *ZüriLEX 1986 Proceedings Euralex International Congress*. Zurich: University of Zurich. Tübingen: Gunter Narr Verlag, pp. 289-297.
- Zgusta, Ladislav.** (1971). *Manual of Lexicography*. Praha: Academia.

### Dictionaries

- Abu-Ssaydeh, Abdul-Fattah.** (2013). *Al-Murshid: A General Arabic-English Dictionary*. Amman/Jordan: Da:r al-sharq wa l-gharb for Publishing and Distribution.
- Baalbaki, Rohi.** (2012). *Al-Mawrid: A Modern Arabic-English Dictionary*. Da:r el-ilm LilMalayi:n.
- Omar, Ahmed Mukhtar.** (2008). *Mu'jam al-lughah al-'Arabiyah al-mu'a:širah* [Dictionary of Contemporary Arabic]. Cairo: 'a:lam al-kutub.
- Rundell, Michael.** (2007). *MacMillan English Dictionary for Advanced Learners*. London: Macmillan.
- Siefring, Judith.** (1999, 2004). *The Oxford Dictionary of Idioms*. Oxford: Oxford University Press.
- Summers, Della.** (2005). *Longman Dictionary of Contemporary English*. 5<sup>th</sup> edition. Harlow: Pearson/Longman.

**Wehr, Hans.** (1974). *A Dictionary of Modern Written Arabic*. Librairie Du Liban and Macdonald and Evans.

*Cambridge Advanced Learner's Dictionary*. (2008). 3<sup>rd</sup> edition. Cambridge: Cambridge University Press.

*Longman Idioms Dictionary*. (1998). Harlow: Longman.