

Chapter 2

Lexical Semantics and Morphological Productivity in Arabic: Incorporating Patterns in Pedagogical Lexicography

David Stringer¹

1. Introduction

The project of compiling the *Doha Historical Dictionary of the Arabic Language*, currently underway, provides an opportunity to create not only a comprehensive, etymological dictionary for speakers of Arabic, but also learner dictionaries for the teaching and learning of Arabic, whose primary readers are nonnative speakers. Current pedagogical dictionaries are invariably either outdated, lacking in coverage, or somewhat superficial. For example, the most commonly used learner dictionary of Arabic in the United States is the 4th edition of the *Hans Wehr*, published in 1994, but not updated since 1979. The *Oxford Arabic Dictionary* (2014) has more contemporary language but has a narrow focus on Modern Standard Arabic (MSA) and is less comprehensive. Other recent pedagogical dictionaries, such as the *Oxford Essential Arabic Dictionary* (2010) or Awde and Smith's (2022) *Arabic Practical Dictionary*, are very useful word lists, but contain few examples or the kind of grammatical information considered typical in learner dictionaries for other languages. In order to take full advantage of this unique moment in dictionary research, it is worth going back to basics to reconsider what might be the best foundational principles for a new Arabic pedagogical lexicography.

In this paper, I briefly review the development of some of the main pedagogical dictionaries for English and Japanese, the former being the largest and most diverse range of dictionaries for nonnative speakers of any language, and the latter for insights regarding non-phonemic writing systems and examples of grammatical information beyond English. I then consider how research on the mental lexicon – the actual word bank in our brain – sheds light on the relations between words, and therefore on possible relations to be captured in dictionaries. Finally, I turn to the question of what might be considered important aspects of lexical grammar that might be included in a modern,

¹**D. Stringer:** Corresponding Author(✉)
Associate Professor, Second Language Studies
Director of Undergraduate Studies, Second Language Studies
Indiana University Bloomington, USA
Email: ds6@iu.edu

pedagogical dictionary of Arabic, with a focus on recent research in Arabic lexical semantics and its implications for our understanding of verbal and deverbal morphology.

2. What We Can Learn From Learner Dictionaries

Before considering lexicography for Arabic pedagogy, I examine the development of Monolingual Learner Dictionaries (MLDs) for English and bilingual Japanese-English dictionaries, as these traditions provide us with failures and successes that can inform current research. English MLDs have been developed over the last eighty years or so in order to facilitate language learning and usage for those who speak English as a second language (L2), now considered to number more than 1 billion people (Eberhard et al., 2023). The first English dictionaries for nonnative speakers were developed in the 1930s and 40s to support language teaching in Japan. *The New Method English Dictionary* (1935) was closely followed by the more comprehensive *Idiomatic and Syntactic English Dictionary* (1942), republished in 1948 and in twenty subsequent editions under the name of the *Oxford Advanced Learner's Dictionary* (OALD). This dictionary had no serious competitor until the publication of the *Longman Dictionary of Contemporary English* (LDOCE) in 1978, which considerably shook things up and paved the way to a new kind of international market for learner dictionaries. Innovations of the LDOCE included modern vocabulary, idioms, and colloquial terms; a greater inclusion of American English; examples of authentic speech and writing (partly sourced from Quirk's Survey of English Usage); a simplified grammar; and a defining vocabulary of only 2000 words. Over the next decade, a further revolution was on the way, and with the publication of the *Collins COBUILD English Dictionary* (CCED) in 1987, corpus-based lexicography became the norm, informing the subsequent publication of the *Cambridge International Dictionary of English* (CIDE) (1995), the *Longman Dictionary of American English* (LDAE) (1997, 2nd ed.), the *Macmillan English Dictionary for Advanced Learners* (MEDAL) (2002), and *Merriam-Webster's Advanced Learner's English Dictionary* (MWALED) (2008). These publications arguably constitute the foundation for the large variety of English learner dictionaries available today. For historical context, see Landau (2001), and for a critical examination of the lexicographical issues involved in the creation of these competing texts and their revised editions, see Chi (2022).

High-quality MLDs can be distinguished from other kinds of dictionaries in various ways. They usually provide more information on grammar, verb constructions, and idiomatic language than native-speaker dictionaries (in line with larger, bilingual foreign-language dictionaries). For example, entries in the *Longman Dictionary of English Language and Culture* (LDELC, 2005) indicate if an English noun is typically count or mass (e.g., [C] countable, [U] uncountable, [C, U] both uses), or show typical transitivity patterns of verbs (e.g., [I] intransitive, [T] transitive, [I, T] both uses). Similarly, entries may note if adjectives are typically attributive, or whether they lack comparative or superlative forms. Annotations provide a range of information on the complementation of verbs, nouns, and adjectives: for example, [+to-v] and [+v-ing] for infinitives or gerunds; [+ (that)] for obligatory or optional complementizers; [+wh-] for interrogative complementizers; [+adv/prep] for adverbial or prepositional complements; and [+obj (i)+obj (d)] for double objects, among other distinctions. Needless to say, the dictionary contains a glossary of codes with examples². More recently, MLDs have started to move toward more the more explicit spelling out of grammar codes (Yamada, 2022).

Also generally incorporated into learner dictionaries are sections separate from the main word list, either at the beginning, middle or end of the volume, with supplementary information on grammar, style, and sometimes word lists in particular lexical fields. Separate grammar sections may include topics such as irregular verbs, tense and aspect morphology, phrasal verbs, adjectives and adverbs, articles and quantifiers, active and passive forms, and contractions, with considerable differences in presentation across dictionaries. Of relevance to the discussion below concerning productive Arabic morphology is the inclusion in many English learner dictionaries of separate sections on derivational morphology. For example, there is often an annotated

² Examples from the LDELC (2005) include:

[+to-v] shows that a word can be followed by an infinitive verb with **to**: I want to leave early today. / an **attempt** to reach an agreement. / We're **ready** to go.

[+v-ing] shows that a verb can be followed by another verb in the **-ing** form: I **like** playing football. / We **watched** them playing football.

[+(that)] shows that a word can be followed by a clause beginning with **that**, but the word **that** can be left out: He **knew** he would be late for work. / I'm **sorry** you failed your exam.

list of Greek and Latin roots, as these act as bound bases in significant numbers of words, e.g., *bio* ‘life’ (*autobiography*, *biology*, *biotechnology*, etc.), or *hydr* ‘water’ (e.g., *dehydrate*, *hydrant*, *hydroelectric*, etc.). Similarly, there is often an annotated list of affixes – note that prefixes can usually be found together in the headwords of an alphabetic dictionary (with the exception of mutating prefixes, e.g., *il-*, *im-*, *in-*, *ir-*), but suffixes are scrambled. Style sections often present rules of punctuation, or give guidelines to help with writing emails, formal and informal letters, CVs and resumés. Lexical field supplements can include, for example, body parts, cooking terms, plants and trees (LDAE, 2014); money, numbers, weights and measures (MWALED, 2008); or cars, computing, homes and houses (OALD, 2020).

Other distinguishing features of such dictionaries are a controlled vocabulary of 2-3000 words for definitions, and generous sprinklings of examples to show use in context. This latter feature is also tied to the concept of learner dictionaries as supporting the *encoding* as well as the *decoding* function of dictionaries, that is, they are intended to guide the production of utterances and sentences, not just their comprehension. This is a feature of all pedagogical dictionaries, including dictionaries for children. Another feature shared between MLDs and children’s dictionaries is the existence of versions for different levels of user, whether in terms of schooling (elementary, high-school, college), or L2 proficiency (beginner, intermediate, advanced).

One notable development in MLDs has been the gradual move toward strict alphabetical ordering. Previously, entries were presented according to a nesting principle, by which various related words and phrases would fall under the same headword, thus making semantic and morphological relationships more transparent. The OALD maintained this tradition until the 6th edition (2000), but since that time almost all MLDs have prioritized alphabetic lists. Thus, for example, in the MWALED (2008), TEA and TEACUP are separated by 15 entries, while TEACUP and TEASPOON are separated by 35 entries, as semantically unconnected as TEAM and TEARDROP. However, to a greater or lesser degree, MLDs are in fact mixed systems that involve at least some redirection to nested entries: thus BLOW SOMEONE A KISS, located under BLOW, steers the reader to KISS, and BLOW SOMEONE’S MIND points to MIND. As Chi (2022) observes, the embrace of the strict alphabetic approach may smooth the process of modern methods of computing production, but it is not clear that it is pedagogically sound. It elevates the decoding function, allowing quicker access to the

search term, but it arguably interferes with encoding function, as word families are split apart, hindering vocabulary acquisition.

Unlike English MLDs, all Japanese learner dictionaries are bilingual, either unidirectional with a list of words in Japanese and definitions in English, or bidirectional (Japanese-English, English-Japanese). One of the biggest differences for learners of Japanese is that the orthography is not a phonemic alphabetic system (with both consonants and vowels) like English, but rather a mixed system involving *kanji*, which are logographic characters mostly of Chinese origin, and two syllabaries, consisting mainly of consonant-vowel combinations: *hiragana*, used primarily for grammatical morphemes and some native or long-established Japanese words, and *katakana*, commonly used for foreign loan words (of which there are many), and onomatopoeia. The kana systems are phonetically transparent and can be mastered in a matter of months: each has 46 basic characters, which are augmented by a small system of functional marks and diacritics. Of relevance to the current discussion is the fact that most content words in Japanese are represented by kanji, which have at least two basic pronunciations and are initially phonologically opaque to language learners. Consider the following sentence:

- (1) 花子はフェリーで海を渡った。
Hanako wa ferī de umi o watatta.
Hanako TOP ferry by sea-ACC cross-PAST
“Hanako crossed the sea by ferry.”

Here, the words for the female name *Hanako*, *sea* and *cross* are in kanji, the loan word *ferī* is in katakana, and the function words for the topic marker *wa*, the postposition *de*, the accusative case marker *o*, and the past tense maker *ta* are in hiragana. So for a language learner who has mastered kana, but not yet kanji, it is not clear from the orthography how to pronounce three of the four content words in this sentence.

In order to find a word such as 海(*umi*) ‘sea’ in a traditional Japanese dictionary, one needs to recognize that it is a complex character, and the ‘radical’, or core component, is 氵, which means water. Words in kanji dictionaries are grouped by radicals, so the organization is fundamentally semantic rather than phonological. There are 214 traditional radicals, with some variants; Japanese school children use these as the basis to memorize, at a minimum, just over 2000 Kanji to achieve basic literacy;

university students are typically familiar with upwards of 3000; and it is possible to keep expanding one's knowledge of characters: the classic *Dai Kan-Wa Jiten* (1955-1960) kanji dictionary has more than 50,000 individual kanji and about 530,000 compounds.

More general Japanese dictionaries, which contain not only kanji but also words written in hiragana or katakana, follow the conventional order of the syllabaries, so if you know how a word is pronounced, you can easily find it by turning to its place in the syllabary, much like looking up a word in an alphabetic dictionary. However, without prior knowledge of how to pronounce a character like 海 (*umi*) 'sea' a learner cannot find it in a phonetically organized dictionary. Hence dictionaries for young Japanese children and some dictionaries for foreigners make use of a system of *furigana*, that is, writing hiragana or katakana above the characters to render them phonetically transparent. Adding furigana to example (1) results in example (2).

(2) ^{はなこ}花子は^{うみ}フェリーで^{わた}海を渡った。

If a reader were not to know the meaning of 海, the furigana うみ (*u-mi*) indicates the place in the dictionary to find this definition. Japanese learner dictionaries that make use of furigana include *Kodansha's Furigana Japanese Dictionary* (1999) and the more comprehensive *Kodansha's Communicative English-Japanese Dictionary Bilingual Edition* (2006). Other learner dictionaries use the Latin alphabet for transliteration, but this arguably slows down the acquisition of literacy; just like Japanese children, foreign learners can make use of a tried-and-tested pedagogical tool which will ultimately be discarded as a superscript at the advanced level. It is worth noting that the kana syllabaries are not a recent invention or in any way inferior to the logographic writing system; the invention of kana is sometimes attributed to the Shingon Buddhist priest Kūkai (9th century), and the classic *Genji Monogatari* 'The Tale of Genji' (11th century), arguably the world's first novel, was written entirely in kana.

The problem of how to facilitate dictionary consultation for foreigners who have not yet attained an advanced level also applies to Arabic, despite the fact that the abjad (consonant alphabet) is considerably easier to master than Japanese orthography. The lack of marking for vowels or consonant length in standard Arabic orthography hinders lexical acquisition for learners, so pedagogical dictionaries and grammar books enhance the text. As with Japanese, there is a home-grown solution, this time in the form of

tashkīl (diacritics). Also parallel to Japanese, this system is well-established in texts for children, and it has a long tradition in the culture, being used especially for texts that require careful pronunciation, such as the Koran (7th century), hadiths, and poetry. Learner dictionaries that are vowelled include the *Al-Mawrid Modern Arabic-English Dictionary* (1995) and the *Oxford Arabic Dictionary* (2014), but several dictionaries rely exclusively on transliteration in the Latin alphabet, including the *Arabic Practical Dictionary* (2022) and the *Hans Wehr* (1994). As with Japanese, it can be argued that this short-term solution that in fact hinders progress toward the goal of being entirely comfortable with reading in Arabic.

Locating words alphabetically is not a problem in Arabic, but the question remains of whether to organize entries primarily by root or by strict alphabetic order. This mirrors the earlier discussion of English learner dictionaries, concerning whether the nesting principle or absolute ordering by alphabet is best (recall that the former favors decoding while the latter facilitates encoding). As with English, there are good arguments for maintaining the root /nesting approach (as found in, e.g. *Hans Wehr*, 1994), but, as noted by Wightwick and Gaafar (2018: 8-9), this may not be able to withstand the advance of the alphabet in the age of computerization.

This discussion has deliberately focused mostly on basic principles of pedagogical lexicography that apply to both traditional dictionaries and electronic or online dictionaries, but it would be remiss not to acknowledge the revolution in lexicography as dictionaries have gone digital. Even before the arrival of widespread access to the internet, electronic dictionaries had become the norm for advanced learners of Japanese. First introduced in the 1970s, by the 1990s thin, portable electronic dictionaries by companies such as Sharp, Casio, and Seiko were available preloaded with multiple dictionary systems, so that, for example, a kanji dictionary, a bidirectional phonetic dictionary, and a thesaurus were accessible through the same interface. However, at that time most were oriented primarily to Japanese native speakers, did not have furigana, and were less useful to beginners and intermediate learners. Nevertheless, they provided the blueprint for the online age. Modern internet-based Japanese dictionaries such as jisho.org have the ability to allow searches using radicals, touch screen writing, kana or Latin-alphabet look-up, or voice input, and contain proper words, regional dialects, and links to Wikipedia articles; a smartphone can of course contain multiple apps for different dictionary functions.

The development of mobile apps and online reference websites offers exciting possibilities for teachers and learners of Arabic, although current resources fall short in several ways. For example, several of the most common offline apps (e.g., *Arabic Almanac*, *Al-Mawrid*, *Lane's Lexicon*) and mobile apps for Android and iOS do little more than reproduce outdated paper dictionaries. However, just as with Japanese internet-based dictionaries, existing technology allows for multiple forms of look-up, using either roots or alphabetic ordering. Traditional dictionaries become more expensive with additional paper and ink, and more inconvenient with additional weight. The lack of such constraints allows for extensive incorporation of Arabic dialects, extra grammatical information, more examples of usage, a thesaurus, and the option of coloring *tashkīl* (note that MS Word already offers a simple way to color diacritics).

Despite the promise of technology, the basic observations above regarding best practices for learner dictionaries remain essentially unchanged, and the fundamental work of lexicography comes prior to any online instantiation of lists of headwords and their definitions. In the following section, we turn to how theoretical investigations of the mental lexicon can inform our knowledge of the relations between words, with a view to identifying the kinds of semantic components and relations that could be usefully incorporated into a pedagogical dictionary.

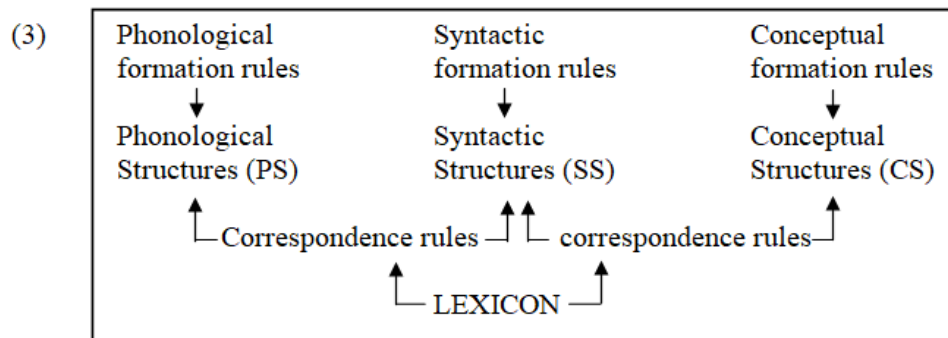
3. The Mental Lexicon Transcribed: From Network to List

In a fundamental sense, a dictionary of contemporary language is an attempt to reflect the collective mental lexicons of everyone in a speech community. Yet a dictionary must, by its nature, drastically simplify the complexity of organization in every individual person's mental lexicon. While traditional dictionaries are organized in lists, it makes sense to think of the mental lexicon as an interactive network with multiple types of connections between elements. In order to develop a theory of how this network might be organized, it is important to understand the various ways in which words can semantically relate to one another. One source of evidence for connections between lexical entries comes from semantic priming experiments. A word is "primed" when activation of a related word results in a secondary activation of the word in question, which enables fast retrieval and integration into syntax. For example, in a word recognition task, a response to a target (e.g., *boy*) is faster when it is preceded by a semantically related prime (e.g., *girl*) compared to an unrelated prime (e.g., *telephone*). Evidence from semantic priming, as well as word association tests, slips of the tongue,

and lexical recall in cases of aphasia (language deficits following brain damage) makes clear that there are various types of relations between lexical items, and that the organization of the mental lexicon is significantly more complicated than a simple word list. Connections between lexical entries involve “lexical fields” (groupings by conceptual association, according to areas of knowledge or regular co-occurrence in real-world situations), as well as lexical relations such as synonymy, antonymy, polysemy, and hyponymy (for more detailed discussion of lexical relations, see Stringer, 2019; and for an accessible introduction to the use of semantic priming in L2 research, see McDonough and Trofimovich, 2009).

A comprehensive representation of such connections is difficult in a traditional dictionary, due to constraints of space, although learner dictionaries have attempted to integrate at least some of this information, in ways that are highly selective and somewhat unsystematic. For example, entries in the MWALED (2008) occasionally include one or two synonyms with helpful explanations of differences in use and examples of usage (e.g., HIGH points the reader to TALL and LOFTY, explaining that *high* is used of things, *tall* is used of both people and things, and *lofty* indicates a great or impressive height). However, the advent of electronic and online dictionaries makes it much easier to cross reference words in terms of lexical relations that to a significant extent mirror the organization of words in the mental lexicon. Indeed, manual consultation of a dictionary-thesaurus can sometimes result in faster recall of related words, as every writer knows.

Not only is the mental lexicon a complex network, but each individual word can be analyzed as a relation, rather than a thing. Much research in lexical semantics assumes some version of Jackendoff’s (1997, 2010) theory of parallel architecture; on this approach, at a minimum, a word consists of a phonological representation, linked to an independent syntactic representation, linked to an independent conceptual representation. Moreover, the generation of phonological, syntactic, and conceptual forms occurs in modular fashion, so that, for example, only phonological primitives and combinatorial rules are used to create phonological representations. A lexical item is created through correspondences between these distinct representations, as in (3).



(adapted from Jackendoff, 1997, pp. 39; 100)

This conception of representational modularity at the lexical interface with other cognitive systems is compatible with evidence from phonological and syntactic priming, which show co-activation of lexical items and constructions even without any semantic connection. In an early experiment on phonological priming conducted by Tanenhaus et al. (1979), when participants heard the sentence *They all rose*, both the verb (‘stood up’) and the noun (‘type of flower’) were primed. Interestingly, closed-class and open-class lexical items may belong to separate networks. When Shillock and Bard (1993) compared priming effects for open- and closed-class lexical items with identical phonology, they found that closed-class modals such as *would* had no priming effect related to open-class items such as *wood*.

Given that many Arabic words consist of both a root and a pattern, it is important to consider whether discontinuous, derivational affixes are co-activated in a similar way to roots. Boudelaa and Marslen-Wilson (2001) found that words with a shared root will, in fact, prime each other whether the semantic relationship is transparent (e.g., *iħtaraqa-ħaraqa* ‘get burned’-‘burn’) or opaque (e.g., *taqaadama – taqaddama* ‘become older’-‘progress’). Interestingly, Boudelaa and Marslen-Wilson (2011) found that priming was determined exclusively by the productivity of the root, not the word pattern, with highly productive patterns failing to prime in the context of unproductive roots. Although further investigation is necessary, this appears to mirror the priming differences in English between lexical and functional morphology.

Another major aspect of grammar with direct relevance to lexical organization is how the linking together of words in semantic classes allows for the prediction of whether individual words can participate in particular syntactic processes. These semantic groupings are on the basis of semantic components that play a role in grammar: what Pinker (2013 [1989]) calls the “grammatically relevant semantic subsystem”. Early work in both cognitive linguistics (e.g., Talmy, 1980) and conceptual

semantics (e.g., Jackendoff, 1990) established that many concepts recur as syntactic features in the world's languages, while others appear never appear to be grammaticalized. Drawing on resource materials from the MIT Lexicon Project in the 1980s, Levin (1993) provides the most copious and convincing evidence for a lexical semantic system underlying English grammar, using meaning components that are found cross linguistically to sort 3000 English verbs into approximately 85 semantic classes. The genius of Levin's approach is that she does not rely on intuition or typological guesswork, but rather uses syntactic evidence to identify semantic components. For example, she shows that while the verbs (a) *cut*, (b) *crack*, (c) *stroke*, and (d) *whack* may seem conceptually similar at first glance, detailed analysis reveals that *crack* and *stroke* may not be used in the 'conative' construction (e.g., *Harry cut at the pastry*), *crack* may not be used in the 'body-part ascension' construction (e.g., *Sally cut Harry on the arm*), and *stroke* and *whack* may not be used in the 'middle' construction (e.g., *This surface cuts easily*), while *cut* is grammatical in all three environments (Levin, 1993: 6-7). The semantic elements that appear relevant to this distribution appear to be conflated as follows.

- (4) a. cut: [CAUSE, CHANGE OF STATE, CONTACT, MOTION]
- b. crack: [CAUSE, CHANGE OF STATE]
- c. stroke: [CONTACT]
- d. whack: [CONTACT, MOTION]

If this analysis is correct, then predictions can be made as to the syntax of verbs that share the same semantic features. Such predictions are borne out with the syntactic distribution of (a) *cut*-type verbs (*scratch, hack, slash, etc.*); (b) *crack*-type verbs (*rip, break, snap, etc.*); (c) *stroke*-type verbs (*tickle, pat, touch, etc.*); and (d) *whack*-type verbs (*kick, hit, tap, etc.*), leading to the conclusion that lexical semantic features do play a determining role in possibilities of syntax. In the following section, we will consider how recent work in Arabic morphology and lexical semantics might suggest the addition of useful layers to pedagogical dictionaries based on semantic subclasses.

4. Meaning and Morphological Productivity: Simplifying Semantics for Inclusion in Pedagogical Dictionaries

One question of continuing debate in Arabic linguistics is how consonantal roots and word patterns interact to produce Arabic words, often following rules which are productive to a greater or lesser degree. In traditional Arabic grammar, even though certain writers such as Ibn Jinnii and Ibn Farris (10th C) did assume a basic meaning for the root, this was not seen as the foundation for word formation (Owens, 1988). However, many modern grammarians do see roots and patterns as discontinuous morphemes, stored in the lexicon as bounded bases, which merge to form words, as shown in the famous analysis of McCarthy and Price (1990:5).

(5)	Vowel Melody	u	i	perfect passive	<i>kuttib</i>
					‘to be made to write’
	CV skeleton			causative (Form II)	
			∨		
	Root			‘write’	
		k	t	b	

The evidence for bases in root-and-pattern morphology seems at first contradictory. Sometimes processes apply to stems, rather than roots. For example, Heath (2003) observes how separation of roots from vowels in nouns such as *xubz* ‘bread’, *kalb* ‘dog,’ and *ḍahab* ‘gold’ does not leave vowels with any recognizable pattern, so these are best analyzed as atomic stems that can serve as bases for derivation. Similarly, Ratcliffe (1997) demonstrates that broken plurals, as in *kilaab* ‘dogs’, *dafaatir* ‘notebooks’, and *salaatiin* ‘sultans’, involve the insertion of the long vowel /aa/ into the second syllable of the singular form, triggering vowel changes, rather than there being some kind of consistent word-level vowel pattern. In another much-cited example, Benmamoun (1999) illustrates how certain nouns of place preserve the vowel patterns of the imperfective verb stems they are derived from, seen in *yu-ḥallim* ‘he teaches’ → *mu-ḥallim* ‘teacher’, and *yu-saaʿid* ‘he assists’ → *mu-saaʿid* ‘assistant’.

However, other evidence points in the opposite direction; there are clear arguments that native speakers do access roots in a range of morphological phenomena, revealing the psychological reality of roots independently of words. As noted earlier, Boudelaa and Marslen-Wilson (2001, 2011, *inter alia*) obtained experimental results that demonstrate the role of roots in lexical priming. Along the same lines, Prunet (2006) reviews an impressive array of evidence from psycholinguistics, neurolinguistics and lexical studies that arrives at the same conclusion. In their paper on hypocoristics

(nicknames) in Jordanian Arabic, Davis and Zawaydeh (2001) argue convincingly that roots are extracted from proper names, leaving affixes and epenthetic vowels behind, resulting in, e.g., *Basma* → *Bassum*; *Muḥammed* → *Ḥammuud*; *ʔamdʒad* → *Maḏʒuud*.

A synthesis of these positions is offered by Ratcliffe (1998, 2006), who argues that Arabic contains both rule-based morphology, applying a change to a source word with resulting changes in output; and templatic morphology, involving root extraction and regular word patterns. An example of the former is the insertion of /aa/ to form broken plurals, as discussed above; an example of the latter is the regular pattern evident in the formation of elative adjectives, such as *sahl* → *ʔashal* ‘easier/easiest’, *kabiir* → *ʔakbar* ‘bigger/biggest’, and *ḏaahil* → *ʔaḏʒhal* ‘more/most ignorant’ (Ratcliffe, 2006: 76).

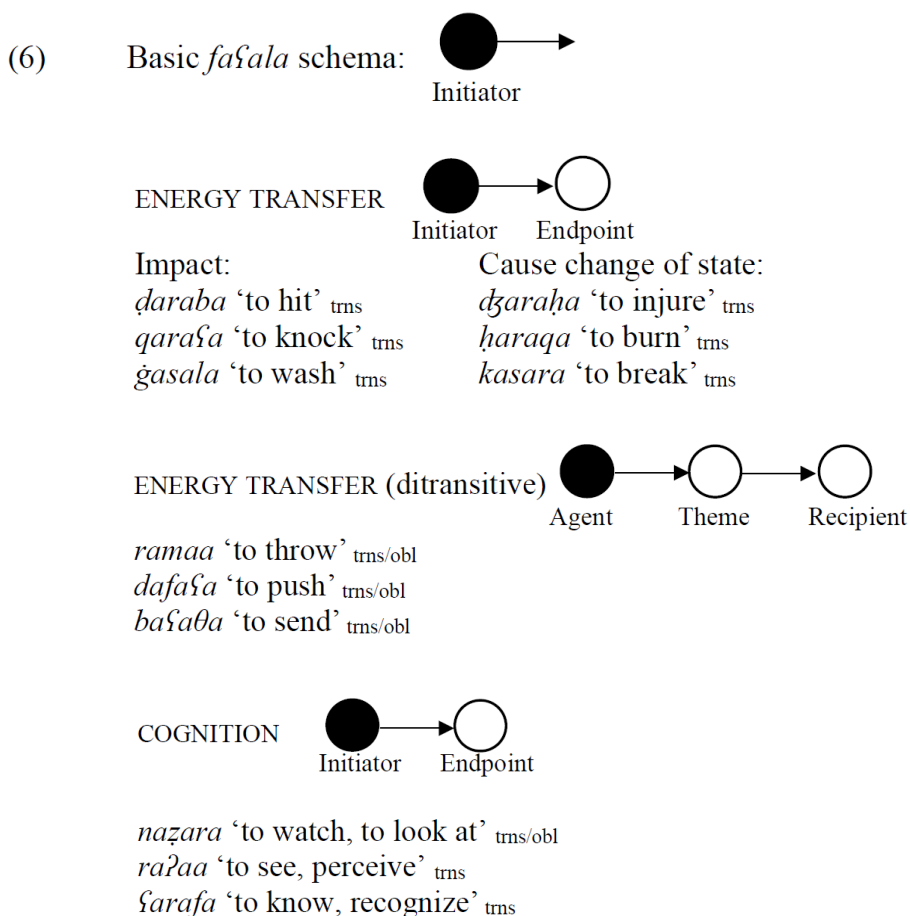
While the status of roots has been the subject of ample debate over the last few decades, the patterns that slot into roots to create the wonderful exuberance of Arabic vocabulary have received considerably less attention³. Yet one could argue that Arabic pedagogical dictionaries might benefit from grammatical scaffolding that sheds light on regularities in word patterns and encourages learning by analogy. To be sure, dictionaries have long made reference to the classic verb forms given Roman numerals in the Western Arabist tradition, e.g., Form I, *faʕala*; Form II, *faʕʕala*; Form III, *faaʕala*, etc⁴. However, they receive no explanation, no semantic analysis, and no overt motivation for inclusion, whether in earlier, classic, or more modern publications – that is, from Salmoné’s (1972 [1889]) *Advanced Learner’s Arabic-English Dictionary*, through the *Hans Wehr* (1961-1994 [1979]) to the *Oxford Arabic Dictionary* (2014). They are simply listed as derived forms with the assumption that the learner will be exposed to this grammar elsewhere.

A recent, in-depth investigation of the lexical semantics of Arabic verbs by Glanville (2018) gives a sense of the equal importance of roots and patterns in word

³ Notable exceptions are Fassi Fehri’s (2003) generative analysis of the two causative templates, Forms II and IV, and Gafos’s (2018) morphophonological stem-based approach to derivation. The semantics of derivation is not discussed in any depth in otherwise excellent overviews of Arabic grammar, such as Aoun et al., 2010, and Benmamoun and Bassiouney, 2018, and is absent from research on the L2 acquisition of Arabic (e.g., Alhawary, 2018). The acquisition of derivational morphology is arguably understudied across languages in the field of Second Language Acquisition (SLA).

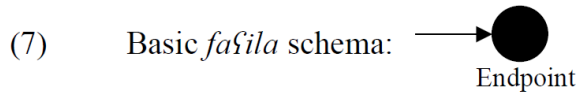
⁴ Only Forms I-X are productive in modern Arabic. The verb *faʕala* ‘to do’ is used in Arabic as a model to show derived forms, in which case it is completely bleached of meaning.

creation. Not only does this research shine a light on our theoretical understanding of regularities in the mental lexicon, but it also provides generalizations with potential applications in lexicography or pedagogy. For example, Granville argues at some length that the three distinct patterns in the perfective of Form I verbs – *faʕala*, *faʕila*, and *faʕula* – can each be analyzed in terms of a unified semantics. While *faʕala* verbs are often described as agentive, a more accurate and more extensive categorization is that they have subjects that are ‘initiators’ that do not absorb a force. By contrast, *faʕila* verbs, which have previously been discussed in terms of experiencer subjects, are here defined as verbs for whom the most important thematic role is that of endpoint. Finally, *faʕula* verbs can be characterized as predicating a state. These observations may at first seem too theoretical for inclusion in a pedagogical dictionary, but the use of visual schemas (which Granville elaborates based on Croft, 1990) and selected examples could make these patterns fairly transparent.





xaradʒa ‘to exit’_{obl}
daxala ‘to enter’_{trns/obl}
balagʾa ‘to reach’_{trns}



EXPERIENCED STATES

ħazina ‘to become sad’_{int}
saʿida ‘to become happy’_{int}
ħabila ‘to become pregnant’_{int}

INFLUENCE

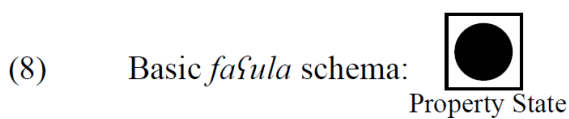
ħahika ‘to laugh’_{int/obl}
faʿila ‘to fail’_{int}
ġariqa ‘to drown; to sink’_{int}

OBJECT IN/ON SUBJECT

ħafiẓa ‘to memorize’_{trns}
ʃariba ‘to drink’_{trns}
labisa ‘to dress; to wear’_{int/trns}

SEQUENCE

tabiʿa ‘to follow, be subordinate to’_{trns}
laħiqa ‘to come after, catch up with’_{trns}



STATIVE VERBS

ṣaġura ‘to be small’_{int}
kabura ‘to be big’_{int}
qaruba ‘to be near’_{obl}

(examples adapted from Glanville, 2018, pp. 30-45)

Glanville’s (2018) work is concerned primarily with verb patterns and includes investigations of reflexivity, symmetry, causation, and repetition. However, derivation extends beyond verb classes to pedagogically important patterns in the formation of nouns and adjectives. It seems clear that this kind of lexical semantic analysis requires a labelling system that is not only more fine-grained than verbal Forms I-XV, but also broader, in order to capture derivational patterns across word classes. A natural solution would be to place a greater emphasis on the patterns introduced earlier in terms of *faʕala*, *faʕʕala*, *faaʕala*, etc. These patterns, known as *al-awzaan* (also termed forms, measures, or binyanim) use the abstract root √ʕʕl to host the discontinuous morpheme that is added to the root. It is already familiar to native Arabic speakers (who often have no knowledge of the Roman numeral system for this purpose) and is ideal for incorporation into the grammar sections of learner dictionaries as it extends beyond verb classes as in the examples below.

- (9) *faaʕil* pattern: ACTIVE PARTICIPLES
 √ktb ‘write’ *kaatib* ‘writer’
 √tħn ‘grind’ *taahin* ‘grinder’
- (10) *maʕʕal* pattern: NOUNS OF PLACE
 √ʕbd ‘worship’ *maʕbad* ‘temple’
 √dxl ‘enter’ *madxal* ‘entrance’
- (11) *faʕiil* pattern: STATIVE ADJECTIVES (cf. example 8)
 √sgr ‘small’ *ʕagiir* ‘small’
 √kbr ‘big’ *kabiir* ‘big’

Interestingly, while no current pedagogical dictionary includes such grammatical scaffolding, the oldest Arabic-English dictionary reviewed for this project – Salmoné’s (1972 [1889]) *Advanced Learner’s Arabic-English Dictionary* – does include a table of numbered derivations based on the root √ʕʕl and is a precursor to the suggestion made in this paper. However, this innovative system is idiosyncratic, somewhat desultory, and has no correspondence with the modern Forms I, II, III, etc.

An alternative possibility is to describe word-types by using higher-frequency words as labels for the type, e.g., *ʕariba*-type verbs, *madxal*-type nouns, and *ʕagiir*-type adjectives. This is the strategy adopted by Levin (1993) as she analyzes the syntax and semantics of dozens of English verb classes in terms of, e.g., *put*-verbs,

send-verbs, *build*-verbs, etc. However, it is arguably better for learners to become familiar with the √fʕl system as it is already in place in the minds of native speakers.

5. Conclusion

The current context of collaborative scholarship emanating from the *Doha Historical Dictionary* project has generated optimal conditions for the development of updated, systematic, materials to facilitate the teaching and learning of Arabic. Given the lessons learned over decades of pedagogical lexicography in other languages such as English and Japanese, advances in lexical semantic theory, and recent insightful analyses of Arabic root-and-pattern morphology, a number of ideas seem particularly worthy of consideration.

The incorporation of systematic grammatical scaffolding in the more successful English MLDs suggests an expansion of what is currently available in Arabic learner dictionaries. The purpose of folding grammar into the dictionary is not to reduplicate everything in grammar books, but rather to provide useful, succinct advice about how to use words in context, and to allow the learner to make generalizations based on existing vocabulary. Additional layers of grammatical information may be provided in a supplementary grammar section as well as incorporated into entries as appropriate. The decision whether to use native, universally understood pronunciation aids with historical pedigree that are also found in children's dictionaries (furigana in Japanese, *tashkīl* in Arabic) or transliteration in the Latin alphabet is an important one; it is arguably best practice to keep transliteration for short-term-use tourist and business wordlists and not for pedagogical dictionaries whose purpose is the advance the learner's knowledge of the language. Similarly, the organization of headwords by root rather than strict alphabetical order carries the advantage of familiarization with native-speaker practice; moreover, it supports the encoding function of pedagogical dictionaries. One of the more exciting aspects of the development of mobile apps and online reference websites is that they remove the necessity of several of these either-or decisions. Current software makes possible multiple types of look-up, by root or by alphabet, in Arabic or in transliteration. The addition of *tashkīl* can also be made optional in digital dictionaries, with the possibility of color coding at no extra production cost. Another recommendation for traditional dictionaries that may be implemented more efficiently in electronic versions is the choice to go beyond direct

consultation to include links to synonyms, antonyms, or lexical fields, in a way that mirrors more closely the proximity of words in the mental lexicon.

Finally, it seems evident that pedagogical dictionaries would be immeasurably improved by the incorporation of lexical semantic generalizations regarding derivational morphology. Roots and patterns are the warp and weft of lexical fabric in Arabic. This weaving together of two basic morphological forms is fundamental not only to verb patterns but can be found throughout derivational processes in the language. This phenomenon is therefore an excellent candidate for inclusion in a supplementary grammar component, as well as in the codification of at least some major classes of headwords. Making language learners aware of lexical semantic patterns in discontinuous morphology provides them with an important noticing tool not only in the classroom but also for acquisition in contexts of immersion.

References

- Alhawary, M.T. (Ed.). (2018). *The Routledge Handbook of Arabic Second Language Acquisition*. New York: Routledge.
- Aoun, J., Benmamoun, E., & Choueiri, L. (2010). *The Syntax of Arabic*. Cambridge: Cambridge University Press.
- Al-Mawrid Modern Arabic-English Dictionary* (1995). R. Baalbaki (Ed.). Beirut: Dar El-Ilm Lilmalayin.
- An Advanced Learner's Arabic-English Dictionary* (1972 [1889]). H.A. Salmoné (Ed.). Beirut: Librairie du Liban.
- Arabic Practical Dictionary* (2022, 2nd ed.). N. Awde & K. Smith (Eds.). New York: Hippocrene Books.
- Benmamoun, E. (1999). Arabic morphology: The central role of the imperfective. *Lingua* 108: 175-201.
- Benmamoun, E. & Bassiouney, R. (Eds.) (2018). *The Routledge Handbook of Arabic Linguistics*. New York: Routledge.
- Boudelaa, S., & Marslen-Wilson, W. D. (2001). The time-course of morphological, phonological and semantic processes in reading Modern Standard Arabic. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-Third Annual Meeting of the Cognitive Science Society* (pp. 138-43). Edinburgh, Scotland.

- Boudelaa, S., & Marslen-Wilson, W.D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes*, 26(4-6): 624-652.
- Cambridge International Dictionary of English* (1995). P. Procter (Ed.). Cambridge: Cambridge University Press.
- Collins COBUILD English Dictionary* (1987). J. Sinclair (Ed.). London: Harper Collins.
- Chi, A. (2022). Researching pedagogical lexicography. In H. Jackson (Ed.), *The Bloomsbury Handbook of Lexicography* (2nd ed.) (pp. 145-164). New York: Bloomsbury.
- Croft, W. (1990). Possible Verbs and the Structure of Events. In S.L. Tsohatzidis (Ed.), *Meaning and Prototypes* (pp. 48-73). London: Routledge.
- Dai Kan-Wa Jiten* (The Great Chinese–Japanese Dictionary) (1955-1960). T. Morohashi (Ed.). Revised and enlarged ed. 1984–1986. Tokyo: Taishukan.
- Davis, S. & Zawaydeh, B.A. (2001). Arabic hypocoristics and the status of the consonantal root. *Linguistic Inquiry* 32(3): 512-520.
- Eberhard, D. M., Simons, G.F., & Fennig, C.D. (Eds.). (2023, 26th ed.). *Ethnologue: Languages of the world*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Fassi Fehri, A., (2003). Transitivity, causativity and verbal plurality. In J. Lecarme, J. Lowenstamm, and U. Schonsky (Eds.), *Research in Afroasiatic Grammar II : Selected Papers from the Fifth Conference on Afroasiatic Languages* (pp. 151-185). Amsterdam: John Benjamins. Republished in Fassi Fehri, A.(2012). *Key Features and Parameters in Arabic Grammar*, Ch.2. Amsterdam: John Benjamins.
- Gafos, A. (2018). Stems in Arabic morphology and phonology. In E. Benmamoun & R. Bassiouney (Eds.), *The Routledge Handbook of Arabic Linguistics* (pp. 62-86). New York: Routledge.
- Glanville, P.J. (2018). *The Lexical Semantics of the Arabic Verb*. Oxford: Oxford University Press.
- Hans Wehr Dictionary of Modern Written Arabic* (1994 [1979], 4th ed.). M. Cowan (Ed.). Wiesbaden: Harrassowitz.

- Heath, J. (2003). Arabic derivational ablaut, processing strategies, and consonantal “roots”. In J. Shimron (Ed.), *Language Processing and Acquisition in Languages of Semitic, Root-Based, Morphology*. (pp. 115-129). Amsterdam: John Benjamins.
- Idiomatic and Syntactic English Dictionary* (1942). A.S. Hornby, E.V. Gatenby, & H. Wakefield (Eds.). Tokyo: Kaitakusha.
- Jackendoff, R. (1990). *Semantic Structures*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- Jackendoff, R. (2010). *Meaning and the Lexicon: The Parallel Architecture 1975–2010*. Oxford: Oxford University Press.
- Kodansha's Communicative English-Japanese Dictionary Bilingual Edition* (2006). P. Sharpe (Ed.). Tokyo: Kodansha International.
- Kodansha's Furigana Japanese Dictionary* (1999). M. Yoshida & Y. Nakamura (Eds.). Tokyo: Kodansha International.
- Landau, S.I. (2001, 2nd ed.). *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL: University of Chicago Press.
- Longman Dictionary of American English* (1997, 2nd ed.). W. Nichols (Ed.). Harlow: Longman.
- Longman Dictionary of American English* (2014, 5th ed). Harlow: Pearson Education.
- Longman Dictionary of Contemporary English* (1978). P. Procter (Ed.). Harlow: Longman.
- Longman Dictionary of English Language and Culture* (2005, 3rd ed.). Harlow: Pearson Education.
- Macmillan English Dictionary for Advanced Learners* (2002). Oxford: Macmillan Education.
- McCarthy, J.J. & Prince, A. (1990). Prosodic morphology and templatic morphology. In M. Eid & J.J. McCarthy (Eds.), *Perspectives on Arabic Linguistics II: Papers from the Second Annual Symposium on Arabic Linguistics* (pp. 1-54). Amsterdam: John Benjamins.
- McDonough, K. & Trofimovich, P. (2009). *Using priming methods in second language research*. New York: Routledge.

- Merriam-Webster's Advanced Learner's English Dictionary* (2008). Springfield, MA: Merriam Webster.
- The New Method English Dictionary*. (1935). M.P. West & J.G. Endocott (Eds.). London: Longmans, Green.
- Owens, J. (1988). *The Foundations of Grammar: An Introduction to Medieval Arabic Grammatical Theory*. Amsterdam and Philadelphia: John Benjamins.
- Oxford Advanced Learner's Dictionary*. (2020, 10th ed.). D. Lea & J. Bradbery (Eds.). Oxford: Oxford University Press.
- Oxford Arabic Dictionary* (2014). T. Arts (Ed.). Oxford: Oxford University Press.
- Oxford Essential Arabic Dictionary* (2010). R. Al-Jabari (Ed.). Oxford: Oxford University Press.
- Pinker, S. (2013 [1989]). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.
- Prunet, J-F. (2006). External evidence and the Semitic root. *Morphology* 16: 41-67.
- Ratcliffe, R. (1997). Prosodic templates in a word-based morphological analysis of Arabic. In M. Eid & R. Ratcliffe (Eds.), *Perspectives on Arabic Linguistics X* (pp. 147-171). Amsterdam: John Benjamins.
- Ratcliffe, R. (1998). *The 'Broken' Plural Problem in Arabic and Comparative Semitic: Allomorphy and Analogy in Non-Concatenative Morphology*. Amsterdam: John Benjamins.
- Ratcliffe, R. (2006). Analogy. In K. Versteegh (Ed.), *Encyclopedia of Arabic Language and Linguistics Vol. I* (pp.153-162). Leiden and Boston: Brill.
- Shillock, R., & Bard, E.G. (1993), Modularity and the processing of closed class words. In G. Altman and R. Shillock (eds.), *Cognitive Models of Speech Processing: The Second Sperlonga Meeting* (pp. 163-185). Hove: Lawrence Erlbaum.
- Stringer, D. (2019). Lexical semantics: Relativity and transfer. In N. Erdogan & M. Wei (Eds.), *Applied Linguistics for Teachers of Culturally and Linguistically Diverse Learners* (pp. 180-203). Hershey, PA: IGI Global.
- Talmy, L. (1980). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language Typology and Syntactic Description, Vol. 3*. (pp. 36-149). New York: Cambridge University Press.

- Tanenhaus, M.K., Leiman, J.M., & Seidenberg, M.S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Reasoning and Verbal Behavior*, 18, 427-440.
- Wightwick, J. & Gaafar, M. (2018, 3rd ed.). *Arabic Verbs and Essentials of Grammar*. New York: McGraw-Hill.
- Yamada, S. (2022). Monolingual learners' dictionaries – past and future. In H. Jackson (Ed.), *The Bloomsbury Handbook of Lexicography* (2nd ed.) (pp. 165-191). New York: Bloomsbury.

Cite this Chapter

Stringer, D. (2023). Lexical semantics and morphological productivity in Arabic: Incorporating patterns in pedagogical lexicography. In M. Larouz, H. Laabidi, M. Mliless & F. Akki (Eds.), *Proceedings of the Transformations of Meaning and Investments in the Field of Applied Linguistics: The Doha Historical Dictionary of Arabic Language as a Case* (pp: 17- 38). Meknes, Morocco: University Moulay Ismail Press.