

DISTINCTIVE FEATURES OF ARABIC DICTIONARIES: STRUCTURAL, HISTORICAL, AND LEXICOGRAPHIC PERSPECTIVES

Asranov Mirzaolim Ortiqovich

Senior Lecturer Department of Language and Literature Education
University of Business and Science Higher Education Institution

Abstract: Arabic lexicography represents one of the most ancient and systematically developed traditions in the history of world linguistics. The Arabic dictionary-making tradition, spanning over thirteen centuries, is distinguished by its unique structural organization, rooted in the trilateral root system (jidhr thulathi) that fundamentally differentiates it from Indo-European lexicographic practice. This paper investigates the principal distinctive features of Arabic dictionaries, encompassing their macrostructural arrangement (rhyme-based, root-based, and phonetic ordering systems), microstructural conventions (vocalization patterns, grammatical information encoding, and illustrative citation practices), and the evolution of specialized lexicographic genres including monolingual, bilingual, terminological, and electronic dictionaries. Special attention is devoted to classical Arabic dictionaries such as al-Khalil ibn Ahmad al-Farahidi's *Kitab al-Ayn* (8th century), Ibn Manzur's *Lisan al-Arab* (13th century), and al-Fayruzabadi's *al-Qamus al-Muhit*, as paradigmatic works that shaped the tradition. The paper further addresses the challenges Arabic poses for computational lexicography, particularly the handling of non-concatenative morphology and the diglossia between Classical Arabic (fusha) and colloquial varieties (ammiyya). Findings suggest that the Arabic lexicographic tradition offers unique methodological insights relevant to modern computational linguistics and corpus-based dictionary development.

Keywords: Arabic dictionaries, Arabic lexicography, root-based ordering, trilateral root, *Lisan al-Arab*, classical lexicography, computational Arabic lexicography, diglossia, morphological structure.

Аннотация: Арабская лексикография представляет собой одну из наиболее древних и систематически развитых традиций в истории мировой лингвистики. Традиция создания арабских словарей, охватывающая более тринадцати веков, отличается уникальной структурной организацией, основанной на системе трёхбуквенных корней (аль-джидр ас-сулясий), что принципиально отличает её от индоевропейской лексикографической практики. В настоящей статье исследуются основные отличительные особенности арабских словарей: макроструктурная организация (рифмическая, корневая и фонетическая системы расположения), микроструктурные конвенции (схемы огласовки, кодирование грамматической информации, практика цитирования), а также эволюция специализированных лексикографических жанров, включая одноязычные, двуязычные, терминологические и электронные словари. Особое внимание уделяется классическим арабским словарям — «Китаб аль-Айн» аль-Халили ибн Ахмада аль-Фарахида (VIII в.), «Лисан аль-Араб» Ибн Манзура (XIII в.) и «аль-Камус аль-Мухит» аль-Файрузабади как парадигматическим трудам, определившим развитие традиции. В статье также рассматриваются вызовы, которые арабский язык представляет для компьютерной лексикографии, в частности обработка неконкатенативной морфологии и диглоссия между классическим арабским (фусха) и разговорными диалектами (аммийя).

Ключевые слова: арабские словари, арабская лексикография, корневая лексикография, компьютерная арабская лексикография, диглоссия, морфологическая структура.



1. INTRODUCTION

The history of Arabic lexicography constitutes one of the most remarkable intellectual achievements in the annals of linguistic scholarship. Beginning in the second half of the eighth century CE, Arab grammarians and philologists embarked on a systematic endeavor to document, preserve, and analyze the vocabulary of the Arabic language — an enterprise motivated primarily by religious imperatives (ensuring correct reading of the Quran) and by concern for the preservation of classical Bedouin Arabic in the face of rapid linguistic change accompanying the spread of Islam. Over more than thirteen centuries, this tradition produced thousands of lexicographic works, ranging from comprehensive encyclopedic dictionaries to highly specialized technical glossaries.

The Arabic dictionary tradition is distinctive in several fundamental respects. Unlike the alphabetically ordered dictionaries predominant in European languages, Arabic lexicography developed multiple competing organizational principles, each reflecting different theoretical approaches to the Arabic lexicon. Moreover, the non-concatenative morphology of Arabic — where roots typically consist of three consonants from which words are derived through the intercalation of vowel patterns — necessitated organizational frameworks radically different from those suitable for languages with primarily linear word formation.

This paper examines the principal distinctive features of Arabic dictionaries from historical, structural, and methodological perspectives. It traces the evolution of the tradition from the earliest known Arabic dictionary, al-Khalil ibn Ahmad al-Farahidi's *Kitab al-Ayn*, through the medieval classical period, to modern bilingual and computational dictionaries. The analysis addresses both macrostructural features (overall organization of the dictionary) and microstructural features (internal organization of individual entries), as well as the specialized challenges Arabic presents for contemporary computational and corpus-based lexicography.

2. HISTORICAL DEVELOPMENT OF ARABIC LEXICOGRAPHY

2.1. The Earliest Phase: al-Khalil and the Phonetic Principle

The earliest systematic Arabic dictionary, *Kitab al-Ayn* ("The Book of the Letter Ayn"), was composed by al-Khalil ibn Ahmad al-Farahidi of Basra (died c. 786 CE), the same scholar credited with developing the system of Arabic prosody. The organizational principle of *Kitab al-Ayn* is phonetic-articulatory: entries are arranged according to the point of articulation of the initial root consonant, beginning with the pharyngeal /ʕ/ (ayn) — considered the deepest consonant — and proceeding forward to the labial consonants. This arrangement reflected al-Khalil's conviction that dictionary organization should mirror the natural physiology of speech production.

Within each phonetic group, al-Khalil employed the permutation principle (*taqalib al-jidhr*): all possible combinations of the root consonants were listed together, thus grouping all derivatives of a given root regardless of their phonetic surface forms. This principle of root-based grouping, first systematized by al-Khalil, proved so linguistically apt for Arabic that it remained foundational to the tradition for many centuries. The work also established the practice of providing extensive illustrative quotations from pre-Islamic poetry and Hadith literature — a convention that served to authenticate entries by grounding them in authoritative textual evidence.



2.2. The Classical Period: Rhyme-Based and Root-Based Dictionaries

The subsequent centuries witnessed the emergence of alternative organizational principles. Ibn Duraid's *Jamharat al-Lughah* (10th century) adopted a modified root-based approach, while the influential *al-Sihah* by al-Jawhari (died c. 1003 CE) introduced the rhyme-based (*qafiyah*) arrangement that would become dominant in the classical tradition. In the rhyme-based system, entries are organized primarily according to the final consonant(s) of the root, and secondarily by the initial consonant — an arrangement that greatly facilitated the composition of poetry by allowing poets and their amanuenses to locate words sharing a common rhyming consonant.

The apogee of classical Arabic lexicography was reached in the 13th century with Ibn Manzur's monumental *Lisan al-Arab* ("The Tongue of the Arabs"), a compilation of unprecedented scope containing approximately eighty thousand entries drawn from five earlier dictionaries. *Lisan al-Arab* remained organized on the rhyme-based principle but synthesized the lexicographic, grammatical, and literary-philological information accumulated over four preceding centuries. It remained the most comprehensive Arabic dictionary for nearly seven hundred years and continues to serve as an authoritative reference today. Al-Fayruzabadi's *al-Qamus al-Muhit* ("The Encompassing Ocean," 14th century) offered a more concise but highly influential alternative that gave rise to the very word *qamus* (dictionary) in Arabic and several other languages.

3. MACROSTRUCTURAL ORGANIZATION: ORDERING PRINCIPLES

The macrostructure of a dictionary refers to the overall arrangement of headwords and the principles governing that arrangement. Arabic lexicography developed three principal ordering systems, each with distinct advantages and limitations:

3.1. Root-Based (Radical) Ordering. The most linguistically principled system for Arabic, root-based ordering groups all words derived from a common triconsonantal root under a single headword representing that root. This arrangement reflects the morphological reality of Arabic, in which the root carries the core semantic content while superimposed vowel patterns and affixes express grammatical and derivational distinctions. The root-based system has the advantage of presenting the semantic family of a root comprehensively, allowing the learner or researcher to grasp the full range of meanings generated by a single root. Its disadvantage is practical: a user must first extract the root from an inflected word form before locating it in the dictionary — a task requiring non-trivial morphological knowledge.

3.2. Rhyme-Based (Qafiyah) Ordering. Arranging entries by the final consonant(s) of the root reflected the practical concerns of poets and scribes who needed to locate rhyming words. While useful for literary composition, the rhyme-based system is counterintuitive for modern readers accustomed to initial-alphabetic ordering and has largely been abandoned in contemporary lexicography.

3.3. Alphabetic (Abjad) Ordering. The adoption of strict alphabetic ordering by initial consonant, then second consonant, then third, emerged gradually and became dominant in modern Arabic dictionaries. The primary advantage of this system is accessibility: users need only know the spelling of a word to locate it, without prior morphological analysis. Modern dictionaries such as Wehr's *Dictionary of Modern Written Arabic* (1961) and *al-Mawrid* by Munir Ba'albaki adopted this approach, substantially increasing usability for non-specialist users and language learners.



4. MICROSTRUCTURAL FEATURES

The microstructure concerns the internal organization of individual dictionary entries. Arabic lexicographic entries are characterized by several distinctive conventions:

Vocalization (Tashkil): Classical Arabic script is essentially consonantal, with vowels represented by diacritical marks (harakat) that are typically omitted in normal writing. Dictionaries, however, supply full vocalization as an indispensable guide to correct pronunciation and, consequently, to correct morphological identification. The provision of tashkil in dictionary entries is a fundamental distinguishing feature of Arabic lexicography absent from most European counterparts.

Grammatical Information: Arabic dictionaries traditionally provide extensive grammatical information including the broken plural (jama al-taksir) of nouns — which must be memorized because it cannot be predicted by rule — along with the verb's verbal noun (masdar) and the pattern (wazn) according to which the root realizes its form. This grammatical density reflects the highly synthetic morphology of Arabic and the consequent importance of morphological information for correct usage.

Illustrative Quotations: A distinctive feature inherited from the earliest period is the extensive use of quotations from canonical literary and religious sources — pre-Islamic poetry (shi'r al-jahiliyya), the Quran, Hadith, and classical prose — to illustrate and authenticate lexical items. This citational practice serves simultaneously as philological documentation, literary exemplification, and normative authority, encoding in the dictionary itself the canon of correct Arabic usage.

Encyclopedic Information: Classical Arabic dictionaries, particularly the larger works, often include encyclopedic information extending well beyond strict lexical definition — geographical, botanical, zoological, and historical information is frequently embedded within entries, reflecting the encyclopedic character of medieval Islamic scholarship and the close integration of linguistic and general knowledge.

5. SPECIALIZED LEXICOGRAPHIC GENRES IN ARABIC

Beyond the general-purpose dictionary, the Arabic tradition developed a rich array of specialized lexicographic genres:

5.1. Monolingual Dictionaries (Qawamis): These constitute the mainstream of the tradition. They range from the comprehensive classical works described above to modern monolingual dictionaries such as al-Mu'jam al-Wasit produced by the Arabic Language Academy in Cairo, which sought to modernize and standardize the Arabic lexicon for contemporary use.

5.2. Bilingual and Multilingual Dictionaries: Intensive bilingual lexicography in Arabic began during the Abbasid period with works linking Arabic to Persian, Syriac, Greek, and other languages encountered through the Islamic conquests and translation movement. The modern era has produced a vast array of Arabic-English, Arabic-French, Arabic-German, and other bilingual dictionaries, the most influential of which in the Western scholarly tradition is Hans Wehr's Arabic-English Dictionary, organized on alphabetic principles and optimized for academic learners.

5.3. Terminological and Technical Dictionaries: The modern period has necessitated the development of specialized terminological dictionaries in fields such as medicine, law, science, and technology. Arabic Language Academies in Cairo, Damascus, Baghdad, and Rabat have



been instrumental in coining and standardizing technical terminology, reflecting the ongoing tension between classical Arabic purism and the pragmatic needs of modern communication.

5.4. Thematic and Onomasiological Dictionaries: Thematic dictionaries organizing vocabulary by subject or concept — analogous to Western thesauri — also have a long history in Arabic, exemplified by Ibn Sidah's *al-Mukhassas* (11th century), which organized the lexicon into semantic domains such as the human body, animals, plants, and celestial bodies.

6. ARABIC LEXICOGRAPHY IN THE DIGITAL AGE

The digitization of Arabic lexicography has introduced both new opportunities and distinctive challenges arising from the specific features of the Arabic language. Computational Arabic lexicography must address several problems without close parallel in most European language contexts:

Non-Concatenative Morphology: Standard natural language processing tools designed for concatenative morphology perform poorly on Arabic because the internal vowel patterns that distinguish grammatical categories and derivations cannot be handled by simple affixation rules. This necessitates the development of Arabic-specific morphological analyzers based on finite-state transducer technology, such as the Buckwalter Arabic Morphological Analyzer, which operate on root-and-pattern rather than stem-and-affix principles.

Orthographic Variability: Arabic orthography presents multiple sources of ambiguity: the absence of short vowels in standard text, the variable representation of certain consonants (such as hamza), and the inconsistent use of the long vowel alif maqsura create significant challenges for automatic lexical lookup. Electronic dictionaries must incorporate sophisticated preprocessing routines to normalize orthographic variation before attempting lemma identification.

Diglossia: The Arabic-speaking world is characterized by pronounced diglossia: a significant structural gap between Classical Arabic / Modern Standard Arabic (*fusha*) — the language of formal writing, education, and media — and the spoken colloquial varieties (*ammiyya*) of different regions, which differ substantially from each other and from the standard. Most traditional dictionaries document *fusha* exclusively, leaving the vast colloquial lexicon largely undocumented. Contemporary computational lexicography increasingly recognizes the necessity of dialect dictionaries and dialect-aware NLP tools, yet the field remains significantly underdeveloped relative to the linguistic reality of Arabic usage.

Digital Arabic dictionaries have expanded significantly since the early 2000s. Projects such as the Arabic WordNet, the Arabic FrameNet, and various open-source morphological lexicons have begun to address the resource gap. The introduction of machine-readable versions of classical dictionaries, including digitized versions of *Lisan al-Arab* and *al-Qamus al-Muhit*, has enabled corpus-linguistic and computational investigations that were previously impractical, opening new avenues for research in historical Arabic lexicology.

7. DISCUSSION AND CONCLUSION

The Arabic dictionary tradition presents a uniquely complex and richly developed lexicographic system that merits sustained attention from comparative and theoretical lexicographers. Several conclusions may be drawn from the foregoing analysis.

First, the structural distinctiveness of Arabic lexicography is not merely a matter of historical or cultural interest but reflects deep linguistic realities: the triconsonantal root system,



non-concatenative morphology, consonantal script, and diglossia collectively necessitate lexicographic solutions without close parallel in European traditions. The root-based macrostructure, the obligatory inclusion of vocalization, the dense grammatical information at the entry level, and the citational authority system are all direct responses to the specific properties of Arabic as a linguistic system.

Second, the historical depth of the Arabic dictionary tradition provides an extraordinary laboratory for the study of lexicographic theory and practice. The coexistence over many centuries of competing organizational principles — phonetic, root-based, rhyme-based, and alphabetic — and the sustained theoretical reflection on lexicographic method that this competition provoked make Arabic lexicography an invaluable case study for the history of linguistic thought.

Third, the challenges Arabic poses for modern computational lexicography are significant but increasingly tractable. The development of sophisticated morphological analyzers, the digitization of classical resources, and the growing recognition of diglossia as a central sociolinguistic fact are gradually producing the infrastructure necessary for high-quality computational Arabic lexicography. Collaboration between traditional philological expertise and computational methods will be essential for the continued development of this field.

In sum, Arabic dictionaries are distinguished by a convergence of structural, historical, cultural, and computational features that make the study of this tradition both intrinsically fascinating and practically important for the contemporary development of Arabic language resources and natural language processing.

REFERENCES

1. Al-Khalil ibn Ahmad al-Farahidi (8th c. CE). *Kitab al-Ayn* [Book of the Letter Ayn]. Ed. Mahdi al-Makhzumi & Ibrahim al-Samarrai. Dar al-Rashid, Baghdad, 1980–1985. 8 vols.
2. Ibn Manzur, Muhammad ibn Mukarram (1290). *Lisan al-Arab* [The Tongue of the Arabs]. Dar Sadir, Beirut. [Reprinted 1997, 15 vols.]
3. Al-Fayruzabadi, Muhammad ibn Ya'qub (14th c.). *Al-Qamus al-Muhit* [The Encompassing Ocean]. Revised edn., Mu'assasat al-Risala, Beirut, 2005.
4. Wehr, Hans (1961). *A Dictionary of Modern Written Arabic*. Ed. J. Milton Cowan. Harrassowitz, Wiesbaden. [4th edn. 1979]
5. Haywood, John A. (1960). *Arabic Lexicography: Its History and Its Place in the General History of Lexicography*. Leiden: Brill.
6. Baalbaki, Ramzi (2014). *The Arabic Lexicographical Tradition from the 2nd/8th to the 12th/18th Century*. Leiden: Brill.
7. Buckwalter, Tim & Dimah Parkinson (2011). *A Frequency Dictionary of Arabic: Core Vocabulary for Learners*. Routledge, London.
8. Dichev, Joseph (1997). 'On the Plurifunctionality and the Architecture of Entries in Arabic Dictionaries.' *Lexicology* 3(2), pp. 131–196.
9. Ryding, Karin C. (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge University Press.
10. Elgibali, Alaa (ed.) (1996). *Understanding Arabic: Essays in Contemporary Arabic Linguistics*. American University in Cairo Press.



11. Attia, Mohammed (2012). 'Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation.' PhD Thesis, University of Manchester.
12. Boulares, Mohsen & Walid Mahdi (2018). 'An Overview of Arabic Lexicographic Resources for NLP.' Proceedings of the 11th Language Resources and Evaluation Conference (LREC), Miyazaki, pp. 1105–1112.
13. Badawi, Elsaid, M.G. Carter & Adrian Gully (2004). Modern Written Arabic: A Comprehensive Grammar. Routledge, London.
14. Maalej, Zouheir (2005). 'Figurative Language in Anger Expressions in Tunisian Arabic: An Extended View of Embodiment.' *Metaphor and Symbol* 20(3), pp. 223–240.
15. Mubarak, Hamdy & Kareem Darwish (2014). 'Using Twitter to Collect a Multi-Dialectal Corpus of Arabic.' Proceedings of the EMNLP Workshop on Arabic NLP, Doha, pp. 1–7.

