

Mu'jam Arriyadh: A Comprehensive Lexicon for Contemporary Arabic Language

Afrah Altamimi, Abdulrahman Alosaimy, Halah Alharbi, Hawra Aljasim, Muneera Alhoshan, Amal Almazrua, Hanan Alharbi, Abdulrahman AIshehri, Bayan Almuqhim, Maryam Algarny, Yahya Asiri, Abdullah I. Alharbi, Saleh Albalawi, Fawziah Asiri, Sara Alhifithi, Abdulah Alfaifi

King Salman Global Academy for Arabic Language
Riyadh, Saudi Arabia

{a.altamimi, aalosaimy, hmuneef, haljasim, malhoshan, aalmazrua, halharbi, aalshehri, balmuqhim, malgarny, yasiri, aialharbi, salbalawi, fasiri, salhifithi, aalfaifi}@ksaa.gov.sa

Abstract

This paper provides an overview of Contemporary Arabic Lexicon (Mu'jam Arriyadh). It is a contemporary and inclusive Arabic dictionary that has been specifically developed to cater to the needs of both native and non-native Arabic speakers. The corpus utilized in this study is derived from the Arabic Contemporary Corpus for Analysis (ACCA), which encompasses a vast collection of 450 million words of Modern Standard Arabic spanning the previous century. Significantly, the lexicon in question prioritizes lemma-based entries over root forms, hence enhancing its user-friendliness and adaptability across different contexts. The resource offers comprehensive linguistic data pertaining to a wide array of Arabic vocabulary, encompassing morphological, morph-syntactic, and semantic aspects. The Lexicon has been developed in accordance with the ISO 24613 standard, which improves its ability to be processed by machines and facilitates the utilization of natural language processing systems. The database encompasses a range of linguistic aspects, such as synonyms, antonyms, and root forms, offering a comprehensive compilation. Mu'jam Arriyadh is a contemporary Arabic lexicon that is designed to be accessible to users, compatible with machine processing, and highly beneficial for anyone studying the language, conducting research, and utilizing natural language processing technologies.

Keywords: Mu'jam Arriyadh, ACCA, LMF

1. Introduction

Dictionaries have long been pivotal tools in preserving and advancing languages, providing a comprehensive record of a language's lexicon. For the Arabic language, the history of lexicography is rich and diverse, with many notable dictionaries that have contributed to the understanding and development of the Arabic language. Some prominent Arabic lexicons include "*Mu'jam al-'Ayn*" by *Al-Khalīl ibn Aḥmad*, "*Lisān al-'Arab*" by *Ibn Manẓūr*, and "*al-qāmūs al-muḥīṭ*" by *Al-Fayrūzābādī*, among others. These lexicons have played crucial roles in different eras of Arabic linguistic scholarship.

Arabic is a language of immense historical and contemporary significance, and the need for authoritative, up-to-date lexicons is particularly acute. Contemporary lexicography, e.g. "*mu'jam al-lughah al-'arabiyah al-mu'āṣirah*" represents a new chapter in the ongoing evolution of Arabic dictionaries.

With the advent of modern computational and linguistic technologies, electronic contemporary dictionaries are empowered to serve a broader spectrum of users. These dictionaries provide quick and efficient access to linguistic knowledge and serve as valuable resources for writers, researchers, language learners, and technology developers. Contemporary Arabic Lexicon (Mu'jam Arriyadh), an innovative linguistic

resource, is a significant addition to the tradition of Arabic lexicography.

Mu'jam Arriyadh is an ambitious linguistic project that addresses the linguistic needs of all Arabic speakers, aiming to include almost every word in contemporary usage. Its content is anchored in the realm of Modern Standard Arabic. The lexicon is restricted to standardized written usage; whereby dialectal forms are excluded unless they are attested in formal written contexts and demonstrably rooted in the classical Arabic lexical tradition. The lexicon embodies modern lexicographic construction principles, meticulously selecting, organizing, explaining, and presenting linguistic material in an easily comprehensible manner. Its construction adheres to the most modern lexicographic principles and the ISO 24613 standard for the Lexical Markup Framework (LMF). It distinguishes itself by covering a vast range of contemporary Arabic words and phrases, accommodating a diverse spectrum of users.

This paper aims to provide a comprehensive overview of Mu'jam Arriyadh, focusing on its structure, the underlying methodology, and the linguistic features it encompasses. Additionally, it will present key statistical data about the lexicon's scope, highlighting the extensive nature of this linguistic resource.

2. Related Work

Over the last decade or so, a handful of new Arabic dictionaries have emerged, each reflecting a different vision of what a modern lexicon should be. One of the earlier efforts in this wave was the Contemporary Arabic Language Lexicon (Omar, 2008). It includes around 32,300 entries, each enriched with morphological details, semantic notes, and real example sentences. In addition, it features thoughtful organization by semantic field, making it especially useful for both learners and linguists.

Then there's the Arabic Language Historical Lexicon, better known as "Alshariqah" (Gharbi, 2021). Rather than focusing on today's language, Alshariqah casts its net wide—covering every Arabic word ever documented, even those long fallen out of use. By tracing how meanings shift across centuries through textual evidence, it serves more as a historical archive than a practical reference for modern speakers.

A third notable project is the Doha Lexicon, launched in 2013 (Hamze et al., 2020). What makes Doha stand out is its diachronic lens: it follows words from their earliest appearances in ancient inscriptions all the way through classical and modern texts, mapping semantic change along a clear timeline. It's less a dictionary for everyday use and more a tool for understanding how Arabic vocabulary has evolved.

Mu'jam Arriyadh sits in a different space altogether. Like Omar's work, it's firmly rooted in present-day usage, but it goes further by being built as a fully digital, machine-readable resource from day one. In fact, to the best of our knowledge, Mu'jam Arriyadh is the first Arabic dictionary designed natively around the ISO 24613 Lexical Markup Framework (LMF). That might sound technical, but it matters: it means Mu'jam Arriyadh isn't just for people looking up words—it's structured so NLP systems can actually use it. It should be noted that Mu'jam Arriyadh differs from previously published Arabic lexicons in several methodological and structural aspects. It is fully digital, lemma-based, and structured according to the ISO 24613 Lexical Markup Framework (LMF). The lexicon integrates corpus-based frequency information, morpho-syntactic categorization, and semantic annotation in a single resource. Due to these design choices, direct quantitative or structural comparison with existing Arabic dictionaries is challenging; instead, the lexicon's methodology and data representation highlight its distinct approach to contemporary Arabic lexicography.

3. Project Description

Mu'jam Arriyadh is the result of months of close collaboration between linguists, computational

specialists, and lexicographers—each team bringing something essential to the table. From early design decisions to the final data validations, it's been a genuinely collective effort.

At its core lies the Arabic Contemporary Corpus for Analysis (ACCA)—a massive 450-million-word collection of Modern Standard Arabic drawn from over a century of written sources. What makes ACCA powerful isn't just its size, but its diversity: newspapers, academic papers, blogs, official reports, literary texts—you name it. This breadth lets us capture how Arabic has actually evolved in real-world usage, not just in theory (see Tables 1 and 2 for a breakdown of sources and regional coverage).

From the start, we built Mu'jam Arriyadh around the ISO 24613 Lexical Markup Framework (LMF)—not as a box to tick, but as a foundation for interoperability. That said, Arabic doesn't always fit neatly into international standards, so we spent considerable time adapting the framework to respect the language's morphological richness and structural nuances.

Our main goal is to create a living, usable resource—not just a static list of words. Mu'jam Arriyadh covers everything from everyday single-word terms to complex multiword expressions you'd actually encounter in modern Arabic writing. Whether you're a native speaker double-checking a rare verb form or a language learner puzzling over a collocation, Mu'jam Arriyadh is designed to meet you where you are.

And because we knew this lexicon would be used not just by people but by machines, we made sure every entry is fully machine-readable. That means rich annotations across phonology, morphology, syntax, and semantics—structured so AI models can actually understand and use them. In short, Mu'jam Arriyadh isn't just a dictionary for humans—it's built to power the next generation of Arabic language technologies.

Arab countries	Percentage
Gulf countries and Yemen	%30
Levant countries and Iraq	%25
Egypt, Libya, Sudan, Somalia, Djibouti and Comoros.	%25
Tunisia, Algeria, Morocco and Mauritania.	%20
Total	100%

Table 1: The geographical sampling frame of ACCA

What really sets Mu'jam Arriyadh apart from other Arabic electronic dictionaries—and from much of the lexicographic tradition more broadly—is how it weaves together linguistic insight and technical design. This isn't just a digitized word list; it's a

resource built from the ground up to reflect how Arabic is actually used today, while also speaking the language of modern NLP systems. In the sections that follow, we'll walk through what makes Mu'jam Arriyadh unique—not just as a reference tool, but as a bridge between human language and machine understanding.

Media	Percentage
Scientific books and verified manuscripts	%15
Journals and cultural magazines	%15
Dissertations	%7.5
Textbooks	%2.5
Official Publications (Policies, Legislation, and Reports from Local and International Organizations)	%5
Personal blog posts, websites and social media applications	%20
Newspapers and news agencies	%30
Radio and podcasts	%5
Total	100%

Table 2: The media sampling frame of ACCA

4. Methodology

4.1 Lemma-based Approach

Unlike traditional Arabic dictionaries that organize entries around trilateral roots, Mu'jam Arriyadh is built around lemmas—the canonical, dictionary forms users actually search for (e.g., *kataba* “he wrote” rather than the root *k-t-b*).

To extract lemmas at scale, we processed the entire 450-million-word ACCA corpus using Camel Tools (Obeid et al., 2020). This gave us, for every word token, both its lemma and part of speech.

Of course, Arabic's rich morphology means a single word can map to multiple possible lemmas (e.g., due to ambiguity or dialectal variation). To resolve this, we manually reviewed ambiguous cases and selected the most contextually appropriate lemma based on usage frequency and semantic coherence. As a practical threshold, we excluded any word form that appeared fewer than 100 times in the corpus—ensuring that Mu'jam Arriyadh reflects real, recurrent usage rather than rare or idiosyncratic forms.

4.2 Linguistic Foundations

A robust lexicon must go beyond surface forms. To capture the full richness of contemporary Arabic, we annotated each entry along three

interlocking dimensions: (i) Morphology: inflectional patterns, derivational forms, and root associations, (ii) Morpho-syntax: grammatical behavior—such as agreement, case marking (where relevant), and subcategorization frames, and (iii) Semantics: core meaning, sense distinctions, and contextual usage notes.

This layered approach ensures that Mu'jam Arriyadh isn't just a word list, but a structured linguistic resource—one that supports both human learners and machine-driven applications.

4.2.1 Morphological Information

We have included morphological information relating to roots and word forms. This information primarily pertains to single words, rather than compounds, as compound words often exhibit a consistent structure and do not undergo significant changes. Unlike most traditional Arabic lexicons, our approach is based on lemmatization, not on the extraction of roots. This decision has both linguistic and technical considerations.

From a linguistic standpoint, the shift towards lemma-based lexicography reflects a focus on how words are used in everyday language. The majority of individuals employ words in their forms or opt for synonyms, rather than relying exclusively on their roots.

Working with lemmas—rather than just roots—gives users something far more practical: clear definitions, accurate translations, and a real sense of how words actually function in today's Arabic. It's not just about listing forms; it's about reflecting how people really read, write, and speak. And from a technical standpoint, this approach makes searching far more intuitive. Whether you're typing a verb in the past tense or a plural noun, the system can still lead you to the right entry—no linguistic gymnastics required.

That said, we haven't abandoned roots. Even for Arabized words like “سندس” (a Persian loanword for a type of fine silk), we include its root (س ن د س)—not because it follows classical patterns, but because it helps users see connections. Roots, even unconventional ones, can be powerful tools for understanding word families, coining new terms, or grasping subtle semantic links. By preserving both lemmas and roots, we aim to honor Arabic's rich morphological tradition while building a resource that works for the way language is used now.

The lexicon provides detailed information about the morphology of nouns, adjectives and verbs. Within this paradigm, nouns and adjectives can be classified into two fundamental groups: strong *sa'hi:h* and weak *mu'atafil*. Strong nouns and adjectives encompass those that lack any vowel letters, while weak nouns and adjectives may exhibit variations influenced by specific ending

vowels like (-ى, -ي), (-ا, -اء), or (-ي). These can be respectively designated as Maqṣūr Mamdūd and Manqūṣ. Acquiring a grasp of these distinctions equips users with the knowledge needed to correctly employ nouns and adjectives in a diverse range of contexts.

Regarding verbs, Mu'jam Arriyadh classifies them into two main categories: strong verbs, known as *sa'hiḥa*, which are devoid of any vowel letters, and weak verbs. The weak verbs, known as *muḥtaṣi*, are further subdivided into two groups: those with only one vowel letter, such as "كسا" (ka'sa:), or those with two adjacent vowel letters, designated as "doubly weak verbs." The latter group, referred to as "لفيف مقرون" *læfi:f mæqrun*, includes verbs with vowel letters that are adjacent, as in "لوى" (læwa:). In contrast, the "لفيف مفروق" *læfi:f mæfru:q* category comprises verbs with vowel letters that are separated, such as "وقى" (wæqa:).

This comprehensive inclusion of morphological information extends our lexicon's utility empowering users with the knowledge needed for the accurate use of these elements in diverse linguistic contexts. It enables users to differentiate between words that maintain a consistent structure and those that may undergo variations influenced by specific vowel endings. This distinction is crucial for employing words appropriately in different linguistic contexts.

4.2.2 Morph-Syntactic Information

The lexical entries in Mu'jam Arriyadh are categorized as either single words or compounds. To enhance users' understanding of the morpho-syntactic structure of single-word entries in the lexicon, we have employed Tammam Hassan's theory (1994). Hassan's theory divides single words into seven primary linguistic categories: noun, adjective, verb, adverb, pronoun, interjections, and particles. This categorization is illustrated in Table 3, which displays the main parts of speech and their respective subs that we have relied on.

This table reflects the grammatical structure of single-word entries in the lexicon and how they are divided according to Tammam Hassan's theory. This categorization allows researchers and language learners to effectively comprehend the structure of words and vocabulary in the Arabic language.

Main POS	Number of sub-POS
Noun	7
Adjective	6
Pronoun	3
Adverb	2
Interjection	4

Verb	2
Particle	0
Total	24

Table 3: The main PoS and subcategories in the lexicon

We have chosen to employ Tammam Hassan's classification system due to its suitability for our technical criteria and the significance of its linguistic insights, which facilitate linguistic analysis at the lexical level.

For each lexical entry, the lexicon provides a conjugation table when the entry represents a single lemma. For nouns and adjectives, these tables account for gender (masculine and feminine), number (singular, dual, plural), and diminutive (diminutive, non-diminutive) forms. In the case of verbs, the tables cover gender (masculine and feminine), number (singular, dual, plural), voice (active and passive), and person (first person, second person, third person). The default form for nouns and adjectives is singular, masculine, and non-diminutive, while for verbs, it is the past tense assigned to the singular third-person masculine.

Since Mu'jam Arriyadh is lemma-based, it is necessary to generate all possible inflexions of a given lemma. This includes, for nouns, different inflexions for variations in number, gender, and diminutives. For verbs, it encompasses all possible inflexions regarding gender, number, verb aspect and voice. Although some tools can generate inflexions, such as Camel Tools, we faced some issues like not generating all possible inflexions, wrong inflexions especially for newly-coined words and they do not generate diminutives for nouns. For this purpose, we manually reviewed the inflexions of the ElixirFM (Smrz, 2007) and added the missing ones. The total inflections for nouns and adjectives is 12 and for verbs is 57, per lexical entry. Not all these inflexions are shown in the current version of Lexicon's applications, however, it used internally in our search engine.

This approach to incorporating morph-syntactic information in our lexicon ensures a comprehensive understanding of word forms and their various linguistic attributes. It is key to understanding the nuanced meanings of words within the lexicon. It provides essential clues about how words function in sentences, their syntactic roles, and their semantic associations. For example, knowing the grammatical gender of a noun or the voice of a verb can greatly impact the interpretation of a sentence. It helps establish the context in which words are used. It allows users to differentiate between homographs, words that share the same spelling but have

different meanings or grammatical roles. By understanding the morphological characteristics of a word, users can discern which specific sense or meaning is applicable in a given context.

4.2.3 Syntactic Information

Mu'jam Arriyadh places significant emphasis on the syntactic features of words and compounds due to their pivotal role in shaping the lexicon's content. Compounds represent an integral part of the lexicon, encompassing expressions that extend beyond single words. In the categorization of compounds, we adhere to Tammam Hassan's theory, which classifies them based on the initial element of the compound. This categorization offers a deeper understanding of the structural attributes of compounds and their various linguistic functions.

Compounds in Mu'jam Arriyadh are further divided into distinct categories, mirroring Hassan's classification system. Each compound category aligns with its initial element, whether it begins with a particle, an adjective, a pronoun, and so on. We compile these compound lists through a dedicated team of linguists and verify their contemporary relevance by confirming their usage in modern contexts.

Including compounds in Mu'jam Arriyadh serves several purposes. It enhances understanding, enabling readers to grasp the meanings and usage of specific terms within a language. Furthermore, it supports language learners in comprehending the linguistic structure and use of these terms.

4.2.4 Semantics Information

Mu'jam Arriyadh places significant emphasis on semantics, providing users with valuable insights into word meanings and disambiguation. The foundation of our semantic information lies in the Arabic Contemporary Corpus for Analysis (ACCA). We utilized the corpus at various stages of Lexicon development. Initially, we employed it to extract the most frequently used words as potential headwords for the Lexicon. Subsequently, for each potential headword, we extracted 100 random contexts from the corpus to serve as examples. Then, we tasked our annotators with constructing a sense list and annotating each example with a corresponding sense.

Notably, a potential headword does not always correspond to a single lexical entry. In Arabic, especially as it is often written without diacritics, the same word can possess various diacritized spellings. Additionally, a single headword may lead to multiple lexical entries based on different diacritized written forms, part-of-speech tags, root and pattern variations, or distinct sets of inflexions. For instances of the same headword

generating multiple lexical entries, please refer to Table 4. This annotation process resulted in an ordered list of senses for each headword, constituting a valuable resource for NLP applications. It can also be harnessed for word sense disambiguation, a feature we plan to employ in one-click applications for in-browser meaning queries, i.e. without the need to navigate away from the website you're currently on.

LE	POS	Root	Pattern	Inflexion	Diacritized form
بيت_1	NC	بيت	فَعْل	Singular	بَيْت
بيت_1	NC	بيت	فَعْل	Plural	بُيُوت
بيت_2	NC	بيت	فَعْل	Singular	بَيْت
بيت_2	NC	بيت	فَعْل	Plural	أَبْيَات
جامعة_1	NC	جامع ة	فاعِلَة	Singular	جَامِعَة
جامع_1	NC	جامع ة	فاعِل	Singular	جَامِع
جامع_2	AS	جامع ة	فاعِل	Singular	جَامِع

Table 4: Lexical Entry Variations in Mu'jam Arriyadh

In essence, the semantics information within Mu'jam Arriyadh offers a profound insight into the depth and breadth of word meanings and their multifaceted usages. It provides the tools necessary for effective NLP applications, enhancing users' understanding of the Arabic language's rich semantic tapestry.

4.3 Management and Structuring

The Lexical Markup Framework (LMF) serves as a model for representing machine-readable dictionaries (MRDs) and other lexical resources, such as thesauri, classification schemes, terminological databases, and lexicons. Its primary purpose is to facilitate the sharing and reuse of such resources, as well as the integration of a wide variety of lexical information (ISO 24613:2019).

We integrated LMF into our Lexicon and made slight adaptations to address specific requirements, notably for corpus-related statistics. In contrast to other initiatives that involved the conversion of existing dictionaries into LMF, Mu'jam Arriyadh adopted this framework from its inception. This strategic choice has enabled us to create a Lexicon that is more adaptable and better suited for applications in natural language processing (NLP). Consequently, the fields, like 'Definition,' are exclusively dedicated to describing a single word sense, avoiding any amalgamation with

morphological or syntactic information. For each sense in the Lexicon, we list the best examples from the corpus that illustrate its usage in that sense.

Furthermore, we incorporated additional fields for corpus-related statistics, encompassing elements such as word frequency, and collocations. Our Lexicon transcends the boundaries of a standard Lexicon, serving as a thesaurus as well. We achieved this by adding fields for synonyms, antonyms, and hypernyms, drawing upon the concept of synsets and relations from LMF. A unique feature is the inclusion of a field for the word's root, which is absent in the LMF standard. This feature carries particular significance in the context of the Arabic language, as the root forms the foundational unit of the language and plays a pivotal role in applications like stemming, morphological analysis, and information retrieval. In addition, two shared tasks were organized to advance research in this area: the recent shared task on Arabic Reverse Dictionary (Al-Matham, 2023), and the follow-up Shared Task on the Contemporary Arabic Dictionary for Reverse Dictionary and Word Sense Disambiguation (Alshammari, 2024). These initiatives provide standardized benchmarks and evaluation protocols for assessing system performance on sense retrieval and disambiguation tasks in Modern Standard Arabic, contributing to the development of high-quality lexical resources and semantic models for Arabic.

Lastly, we leveraged the 'form representation' concept from LMF, which empowers us to encompass different forms of the same word, including country-specific spellings, common misspellings, abbreviations, and acronyms.

5. Conclusion

Mu'jam Arriyadh represents a groundbreaking achievement in modern Arabic lexicography. It serves both native and non-native Arabic speakers with its comprehensive linguistic analysis, advanced technology, and adherence to modern best practices. This resource currently includes over 120,500 words—and continues to grow. Built in accordance with ISO standards, it reflects a genuine commitment to preserving the richness of Arabic while supporting its evolution for modern use.

We're currently mapping each lexical sense in our lexicon to its counterpart in Princeton WordNet. Where gaps appear—either because a sense isn't covered in WordNet or reflects a usage specific to Arabic—we're carefully documenting them. This alignment doesn't just connect our resource to a well-established semantic framework; it also makes the lexicon more interoperable with existing NLP tools and datasets.

An added benefit is that, through WordNet's multilingual extensions (like Open Multilingual WordNet), many of these senses can be linked to equivalent concepts in other languages. This opens the door to cross-lingual applications—whether for machine translation, semantic search, or comparative lexical studies.

Looking ahead, we plan to expand the lexicon's relational structure by adding richer collocational patterns and deeper semantic networks. We're also exploring how to incorporate dialectal forms that meet linguistic and standardization criteria without compromising coherence. On the technical side, we're developing reverse lookup and sense discovery features using modern NLP methods—so users can find entries not just by word form, but by meaning or context.

The resource is freely available online through a public web interface hosted by King Salman Global Academy for Arabic Language. Developers and researchers can also access the lexicon via an open API for searches, synonyms, antonyms, example sentences, and other core lexical functions. The lexicon can be accessed at <https://dictionary.ksaa.gov.sa>.

6. Bibliographical References

- Alshammari, W., Almazrua, A., Al Wazrah, A., Almatham, R., Alhoshan, M., & AIOsaimy, A. (2024, August). *KSAA-CAD Shared Task: Contemporary Arabic Dictionary for Reverse Dictionary and Word Sense Disambiguation*. In N. Habash, H. Bouamor, R. Eskander, N. Tomeh, I. Abu Farha, A. Abdelali, S. Touileb, I. Hamed, Y. Onaizan, B. Alhafni, W. Antoun, S. Khalifa, H. Haddad, I. Zitouni, B. AlKhamissi, & R. Almatham (Eds.), *Proceedings of the Second Arabic Natural Language Processing Conference* (pp. 677-685). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.arabicnlp-1.74>
- Al-Matham, R., Alshammari, W., AIOsaimy, A., Alhumoud, S., Wazrah, A., Altamimi, A., Alharbi, H., & Alaifi, A. (2023, December). *KSAA-RD Shared Task: Arabic Reverse Dictionary*. In H. Sawaf, S. El-Beltagy, W. Zaghoulani, W. Magdy, A. Abdelali, N. Tomeh, I. Abu Farha, N. Habash, S. Khalifa, A. Keleg, H. Haddad, I. Zitouni, K. Mrini, & R. Almatham (Eds.), *Proceedings of ArabicNLP 2023* (pp. 450-460). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.arabicnlp-1.39>
- Gharbi, Abdulrahman. (2021). Efforts of the Supreme Council of the Arabic Language in launching the Historical Lexicon of the Arabic

- Language Project. *Djoussour El-maaréfa*. Algeria.
- Hamze, Hassan. and Belhbib, Rachid. (2020). Introduction to *Doha Historical Lexicon of the Arabic Language*.
- Hassan, Tammam. (1994). *The meaning and structure of the Arabic Language*. Casablanca, Morocco.
- ISO 24613. (2019). Language resource management: Lexical markup framework (LMF). ISO.
- Omar, Ahamed. (2008). *Contemporary Arabic Language Lexicon*. Cairo, Egypt.
- Evaluation, Conference Proceedings). European Language Resources Association (ELRA). <https://www.aclweb.org/anthology/2020.lrec-1.868v2.pdf>
- Smrz, O. (2007, June). ElixirFM—implementation of functional arabic morphology. In *Proceedings of the 2007 workshop on computational approaches to Semitic languages: common issues and resources* (pp. 1-8).
- Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2020). CAMEL tools: An open source python toolkit for arabic natural language processing. In N. Calzolari, F. Bechet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings (pp. 7022-7032). (LREC 2020 - 12th International Conference on Language Resources and