

Dirassat & Abhath
The Arabic Journal of Human
and Social Sciences



مجلة دراسات وأبحاث
المجلة العربية في العلوم الإنسانية
والاجتماعية

EISSN: 2253-0363
ISSN : 1112-9751

Linguistic Engineering of the Arabic Language in Light of Digital Development

الهندسة اللسانية للغة العربية في ضوء التطور الرقمي

Nour Sakina

Doctoral student, Ibn Khaldoun University Tiaret, sakina.nour@univ-tiaret.dz

نور سكيينة

طالبة دكتوراه، جامعة ابن خلدون - تيارت، كلية اللغة والأدب العربي قسم اللغة والأدب العربي، مخبر الدراسات النحوية واللغوية بين التراث والحداثة في الجزائر، sakina.nour@univ-tiaret.dz

تاريخ القبول : 2025/09/13

تاريخ الاستلام: 2025/07/07

Abstract:

This study examines the capacity of the current linguistic situation of Arabic to meet the demands of digitization in planning, education, and research, while questioning whether Arab linguists possess the tools and knowledge to analyze the language in ways that support digital content. It focuses on the linguistic engineering of Arabic within digital development, exploring opportunities for technological integration and ways to enhance teaching and applications through modern tools. The methodology combines descriptive, critical analysis with a review of the literature and practical examples that reveal gaps in digital use. Findings indicate a serious shortage of linguistic resources and technical tools, alongside challenges of dialectal variation, underscoring the need for systematic planning and collaborative efforts. The study concludes by stressing the importance of adopting a scientific, methodical approach to advance Arabic linguistic engineering in line with the requirements of the digital age.

Keywords: linguistic engineering, digital content, Arabic language, linguistic features.
المخلص:

تبحث هذه الدراسة في قدرة الوضع اللغوي العربي الراهن على تلبية متطلبات الرقمنة في مجالات التخطيط والتعليم والبحث، مع التساؤل حول ما إذا كان اللسانيون العرب يمتلكون الأدوات والمعرفة اللازمة لتحليل اللغة بما يدعم المحتوى الرقمي. وتركز على هندسة اللغة العربية في سياق التطور الرقمي، من خلال استكشاف فرص دمجها تقنياً وسبل تحسين تدريسها وتطبيقاتها عبر أدوات حديثة. يجمع المنهج بين التحليل الوصفي والنقدي ومراجعة الأدبيات والأمثلة التطبيقية التي تكشف عن فجوات في الاستخدام الرقمي. وتشير النتائج إلى نقص حاد في الموارد اللغوية والأدوات التقنية، إلى جانب تحديات التعدد اللهجي، مما يبرز الحاجة إلى تخطيط منهجي وجهود تعاونية. وتخلص الدراسة إلى التأكيد على أهمية تبني مقاربة علمية ومنهجية للنهوض بهندسة العربية بما يواكب متطلبات العصر الرقمي.

الكلمات المفتاحية: الهندسة اللغوية؛ المحتوى الرقمي؛ اللغة العربية؛ السمات اللغوية.

1. INTRODUCTION

It has become necessary today to intensify efforts and focus more on integrating information technology with the Arabic language and within Arab countries. Contemporary linguistics has witnessed a qualitative leap in its theoretical and methodological dimensions, as a result of its interaction with highly specialized scientific disciplines within the cognitive sciences framework. It relies on scientific hypotheses that question how knowledge systems are represented in their diverse intellectual trajectories and seeks rigorous methods to reformulate essential questions, aiming to uncover the algorithmic foundations of human knowledge by observing information processing in the human brain.

With the technological boom and advances in artificial intelligence, the Arabic language requires modern, comprehensive technologies to

avoid challenges and keep pace with rapid changes in the digital age. In this context, the research paper titled Linguistic Engineering of the Arabic Language in Light of Digital Development opens the horizon for a legitimate scientific inquiry into Arabic and digitization, posing a fundamental question: Does the current linguistic situation of Arabic meet the requirements of digitization in planning, learning, and research? Put differently: Do Arab linguists possess the knowledge and tools to analyze the linguistic features of Arabic both form and function in ways that support the development of Arabic digital content?

2. Linguistic Engineering

Linguistic engineering, also known as linguistic informatics or language engineering, is the integration of two fields: linguistics and computer science, through the automated

processing of living languages. This field includes various domains such as research in linguistic sciences, audio-visual communication-both digital and non-digital-and advanced language education.⁽¹⁾

AL-WA'AR also defined it as the science that studies human language as a flexible tool for processing by machines (electronic computers) The principles of this discipline are drawn from general linguistics at all its analytical levels—phonological, syntactic, and semantic—as well as from artificial intelligence, logic, and mathematics.⁽²⁾

3. Linguistic Engineering of Arabic and the Foundations for Developing Digital Content

The epistemological foundations of linguistic theories have become the subject of debate from computational-engineering perspectives, giving rise to platform linguistics. This new generation seeks formalized descriptions of linguistic resources by relying on automated processing techniques. Natural language processing (NLP) varies depending on the language being processed, despite similarities between some languages. Arabic, for example, is a fusional and derivational language, with a complex and rich morphological structure. Its affixes attach to the root, and most written texts omit diacritics. These and other challenges make Arabic particularly demanding for researchers. To develop advanced applications for processing Arabic-similar to what exists for English-a high-performance technical platform with a suite of essential tools is required, including:

1. A text segmentation tool that divides text into sentences, each sentence into words, and each word into its root and affixes.
2. A morphological analyzer that performs detailed morphological analysis and clarifies the word's structure.

3. A part-of-speech tagging tool that identifies the grammatical category of words in context.
4. A syntactic parser that performs syntactic analysis of sentences.
5. A named-entity recognition and classification tool that identifies and classifies proper names in Arabic text.

Despite research in this field and the emergence of some of these tools in academic work, researchers face significant challenges and obstacles focused on two main issues:

- The lack of financial support for this type of research.
- The scarcity of appropriate programming and linguistic resources needed to build these tools.

4. Formal Study of Linguistic Units

Formal syntactic studies face numerous challenges as they seek to bridge linguistics and logic. This is a scientific endeavor focused on describing the relationships among linguistic symbols and abstracting their possible interpretations. It aims to describe linguistic expressions and their rules while also striving to build models capable of simulating sentences from linguistic data within a set of computable symbols.

Formal representation is a procedural process that provides symbolic rules capable of offering functional descriptions of grammatical rules based on an attribute-value pairing. This approach reduces variables by generalizing partial descriptions and determining their functional meanings. Such formalism also allows the creation of logical programs that simulate linguistic systems while providing economic benefits that support the automated processing of natural languages.

This kind of processing is a subfield of artificial intelligence, branching from computer

science, and it overlaps significantly with linguistic sciences, which provide the necessary linguistic descriptions for computers. As previously noted, this field enables us to develop software that can analyze, simulate, and understand natural languages.

Computational linguistics primarily aims to enable humans to interact with machines using their natural languages. Researchers in this field work on developing computational tools that facilitate this interaction, which are generally divided into:

- **Analytical tools**, which enable the extraction of information.
- **Generative tools**, which enable the creation of texts in a given language based on knowledge represented as elements linked by various relations.

This processing can be divided into four stages:

- **Morphological processing**: focuses on identifying word types, including pronouns and other morphological information.
- **Syntactic (structural) processing**: deals with the relationships among words, sentence structure, and related syntactic information.
- **Semantic processing**: focuses on understanding the meaning of sentences through logical connections with information.⁽³⁾

This aspect relies on establishing a set of formal rules. However, the scientific requirements for constructing a rigorous model fundamentally demand defining units through mathematical and logical forms. Within this framework, Chomsky defines language as ⁽⁴⁾a (finite or infinite) set of sentences, each of finite length, constructed out of a finite set of elements. This definition is formal and

general, encompassing both natural and artificial languages.

5. Analysis of Linguistic Features: A Case Study of the Phonological and Morphological Levels

5.1 Automated Processing of the Phonological Level in Arabic – The Case of Phonetic Alignment

Phonetic alignment is a linguistic phenomenon that involves the harmony and compatibility of sounds within words and sentences, contributing to overall phonetic coherence. In the automated processing of Arabic, phonetic alignment can be employed in various ways:

Phonetic Imitation: This refers to using sounds that mimic and convey meaning—for example, employing harsh or explosive sounds to express anger or enthusiasm. Phonetic imitation enhances the stylistic and literary qualities of texts, framing their semantic and emotional values.

Phonetic Variation: Using a variety of sounds reflects the diversity of emotions and feelings, such as employing both harsh and soft sounds to convey a range of emotions from sadness to joy.

Semantic Phonetic Symbolism: Each sound in the Arabic language has its own distinctive identity, which influences the listener psychologically during articulation. Sound coloration is an aesthetic stylistic feature designed to evoke and engage.

Advanced Audio Technologies: For example, Nokia's OZO Audio technology promises to radically transform audio experiences on mobile devices through an advanced array of microphones and software. This technology enables three-dimensional sound that allows the listener to hear everything as if seated with the other person during a call.

Through these and other methods, phonetic alignment can be effectively used in the automated processing of Arabic to convey and highlight meanings within texts.⁽⁵⁾

Modern technologies can be used to improve phonological harmony (phonotactic or phonetic compatibility) in Arabic through several innovative approaches:

1. **Machine Learning and Artificial Intelligence:** Machine learning algorithms help develop phonetic models capable of accurately simulating phonological harmony. These models can be trained on large datasets of Arabic speech to improve pronunciation accuracy and recognize phonetic patterns.
2. **Multimedia:** The use of audiovisual media, such as educational videos and animations, can enhance students' understanding of phonological harmony. These resources allow learners to hear correct pronunciation and observe how sounds are articulated.
3. **Mobile Applications:** Developing educational apps focused on improving pronunciation and phonological harmony can be highly effective. Such applications can include interactive exercises that help users practice Arabic sounds in an engaging and enjoyable manner.
4. **Voice Interaction:** Integrating speech recognition technologies into education can provide learners with an interactive experience, allowing them to speak to the system and receive immediate feedback on their pronunciation.
5. **Speech Training:** Developing specialized training programs that focus on improving pronunciation and phonological harmony using advanced technologies such as virtual reality can create immersive learning environments.

By leveraging these technologies, phonological harmony in Arabic can be significantly enhanced, contributing to improved educational quality and more effective communication.

The importance of phonological harmony lies in two key aspects:

- **Achieving Cohesion:** Phonological harmony helps create cohesion between words, enhancing the aesthetic quality of literary and poetic texts.
- **Guiding Meaning:** It serves as an effective tool for guiding meaning, as the interaction of sounds can influence the interpretation of a text.⁽⁶⁾

5.2 Automatic Processing of the Morphological Level in Arabic - The Morphological Structure of the Word as a Model

Arabic morphological analyzers are generally constructed on different theoretical foundations. Some rely on the root-and-pattern (root-template) system, while others adopt the concept of the stem or lemma. In addition, certain approaches combine both within a hybrid framework that has proven effective in the field of Arabic natural language processing.

However, the development of morphological analyzers cannot be separated from the syntactic component. A morphological analyzer is a computational linguistic tool designed to analyze the internal structure of a word and present its morphological and certain syntactic features.⁽⁷⁾ It identifies the word's part of speech (whether it is a noun, verb, or particle), its derivational status (whether it is inflected or uninflected / derived), its morphological pattern (meter), its root, as well as any prefixes and suffixes attached to it.

The tagging process ("étiquetage") requires integrating the morphological and syntactic levels through morpho-syntactic

features that annotate lexical entries within a sequential framework. This framework takes into account the essential stages of automatic processing of Arabic texts.⁽⁸⁾

Word	Noun, Verb, Particle
Type of Word	Non-derived (basic), Derived
Noun Type	Past, Present, Imperative
Verb Type	Preposition, Conjunction, Interrogative
Particle Type	ROOT
Word Root	STEM
Word Stem	Conjunctions, Prepositions, Interrogative Particles
Existing Prefixes and Their Meanings	Attached Pronouns, Case/Inflectional Markers
Existing Suffixes and Their Meanings	Noun, Verb, Verbal Noun
Morphological Pattern	

Despite the efforts made to develop morphological analyzers for the Arabic language, attention to the syntactic level remains very limited, making it difficult to speak of any comprehensive scientific projects for fully integrated Arabic syntactic parsers. Consequently, some research initiatives have sought to lay the foundations for their development. One example is the Al-Irfan Lexicon project⁽⁹⁾, which aims to advance Arabic natural language processing toward a broader horizon where morphological, syntactic, and semantic analyzers can interact. This vision is grounded in a theoretical framework based on empirical foundations and computational approaches.⁽¹⁰⁾

The output of a morphological analyzer when analyzing the word (wahm)

Analysis	Fully Vowelled Form	Dictionary Entry	Morphological Structure	Functional Structure	English Translation
Analysis 1	وَهُمْ	وَهُم	وَهُم	Noun	Imagination / delusion
Analysis 2	وَهُمُ	و + هُم	Prefix + Noun	Noun with conjunction "wa"	Concern / interest
Analysis 3	وَهُمُ	و + هُم	Prefix + Pronoun	Pronoun (they, masc. plural)	They (masculine plural)
Analysis 4	يُؤْهِمُ	وَهُم + ي	Verb stem + prefix	Imperfect verb, 3rd person masculine singular	Be-deluded
Analysis 5	يُؤْهِمُ	و + هُم	Prefix + Verb stem	Imperfect verb, 3rd person masculine singular with conjunction "wa"	Intend / be ready
Analysis 6	أَوْهَمَ	وَهُم + أ	Verb stem + prefix	Perfect verb, 3rd person masculine singular	Imagine / suppose

Types of Spelling Errors Handled by Arabic Spell Checkers in Texts:

- Confusion between hamzat al-qat' (glottal stop) and hamzat al-wasl forms. (e.g., "استقبال" instead of "استقبال")
- Confusion between the alif maqṣūra (ى) and the final yā' (ي). (e.g., "مستشفى" instead of "مستشفى")
- Errors in writing medial and final hamza. (e.g., "شيء" instead of "شيء")
- Confusion between the tā' marbūṭa (ة) and the hā' (ه). (e.g., "سيارة" instead of "سيارة")
- Omitting spaces between two words. (e.g., "كتاب الله" instead of "كتاب الله")
- Missing letters in the word. (e.g., "يحيى" instead of "يحيى")
- Repetition or addition of extra letters. (e.g., "المصطفين" instead of "المصطفين")
- Transposition of letters within a word. (e.g., "ينتشر" instead of "ينتشر")
- Letters pronounced dialectally in a way that differs from their standard spelling. (e.g., "الضبي" instead of "الطبي")
- Confusion between adjacent letters on the keyboard. (e.g., "فارورة" instead of "قارورة")
- Confusion between letters entered with a composed keystroke on the keyboard. (e.g., "غيمان" instead of "إيمان")

6. RESULTS AND DISCUSSION

- Arabic lacks adequate linguistic resources and specialized technological tools, which significantly hinders its ability to adapt to digitization and advanced natural language processing.
- Some progress has been achieved in computational processing of Arabic at certain levels (orthographic, morphological, syntactic, semantic, and lexical), but other critical areas remain underdeveloped, especially pragmatics, corpus development, and machine translation quality.
- Traditional perspectives continue to restrict Arabic from evolving in line with modern computational linguistics and artificial intelligence innovations.
- Active engagement in computational data resources, digital platforms, and software development is essential; without it, Arabic will struggle to advance in linguistic engineering.
- Languages that do not secure their presence in the field of artificial intelligence risk marginalization and decline, making it urgent to adopt scientific, systematic approaches to strengthen Arabic in the digital era.

7. CONCLUSION

The study affirms that Arabic digital content is deeply affected by the transformations of the contemporary scientific revolution led by artificial intelligence. Despite notable progress in computational treatment at several linguistic levels—orthographic, morphological, syntactic, structural, semantic, and lexical—significant gaps remain at the pragmatic level, in text analysis, corpus development, and in enhancing machine translation quality. Addressing these challenges requires sustained

research and innovation. The findings emphasize that no language can advance in linguistic engineering without active engagement in computational resources and digital platforms; yet Arabic continues to be limited by traditional perspectives that slow its integration into natural language processing. In the context of the third millennium, dominated by artificial intelligence and information technologies, the survival and progress of Arabic depend on its effective presence in this domain, without which it risks marginalization and gradual decline.

4. Bibliography List :

1. *The Phonetic Miraculousness in the Qur'an (A Look at the Works of Classical and Contemporary Arab Scholars)*, Sayyid Ali

⁽⁸⁾- Ibid., p. 232.

⁽⁹⁾- Al-Irfan Lexicon is an Arabic linguistic resource that includes all Arabic lexical entries described according to the requirements of NOOJ at both the morphological and syntactic levels. It is currently used to develop numerous engineering applications and can be accessed on multiple computational platforms that support the Arabic language.

⁽¹⁰⁾- Haidar Ghadban, *Arabic Linguistics: Visions and Horizons*, 1st ed., 2019, Alam Al-Kutub Al-Hadith, Jordan, p. 232.

Mir Louhi, Majid Al-Najjar, *Ahl al-Bayt Journal*, Issue 9.

2. Haidar Ghadban, *Arabic Linguistics: Visions and Horizons*, 1st ed., 2019, Alam Al-Kutub Al-Hadith, Jordan.
3. Haidar Ghadban, *Arabic Linguistics: Visions and Horizons*, 1st ed., 2019, Alam Al-Kutub Al-Hadith, Jordan.
4. Saleh bin Hamad Al-Suhaibani, *A Preliminary Proposed Framework for a Project on Teaching Arabic to Non-Native Speakers for Specific Purposes (Diplomatic Mission Members in Riyadh)*, *Journal of Teaching Arabic to Speakers of Other Languages*, Issue 11, Institute of Arabic Language, Africa International University, Sudan, January 2011.
5. Mazen Al-Waer, *Fundamental Issues in Modern Linguistics*, Damascus, Dar Al-Atlas, 1988.
6. Michel Zakaria, *Linguistics (Modern Language Science), Principles and Figures*.

⁽¹⁾-Haidar Ghadban, *Arabic Linguistics: Perspectives and Horizons*, 1st ed., 2019, Alam Al-Kutub Al-Hadith Publishing, Jordan.

⁽²⁾-Mazen Al-Wa'ar, *Fundamental Issues in Modern Linguistics* (Damascus: Dar Al-Atlas, 1988), p. 406.

⁽³⁾- Dr. Michel Zakaria, *Linguistics (Modern Language Science): Principles and Key Figures*, p. 80.

⁽⁴⁾- Dr. Michel Zakaria, *ibid.*, p. 80.

⁽⁵⁾- See: Saleh bin Hamad Al-Suhaibani, *A Preliminary Proposed Framework for a Project to Teach Arabic to Non-Native Speakers for Specific Purposes (Members of Diplomatic Missions in Riyadh)*, *Journal of Arabic for Non-Native Speakers*, Issue 11, Institute of Arabic Language – International University of Africa, Sudan, January 2011, pp. 117–118.

⁽⁶⁾- *Phonetic Miraculousness in the Holy Qur'an: A Review of the Works of Classical and Contemporary Arab Scholars*, by Sayed Ali Mir Louhi and Majed Al-Najjar, *Ahl al-Bayt Journal*, Issue 9, p. 23.

⁽⁷⁾- Haidar Ghadban, *Arabic Linguistics: Visions and Prospects*, 1st ed., 2019, Alam Al-Kutub Al-Hadith, Jordan, p. 232.