

## Digitization of the Arabic linguistic heritage: Problems, challenges, and prospects of linguistic research

Larouane Nora <sup>1</sup>

<sup>1</sup> Hassiba Benbouali University of Chlef

Email: [n.larouane@univ-chlef.dz](mailto:n.larouane@univ-chlef.dz)

**Abstract---**The study aims to shed light on the importance of digitizing the Arabic linguistic heritage and its role in preserving the linguistic and cultural identity of the nation, and highlight its potential in the development of contemporary linguistic research, and the research poses a problem that: to what extent can the digitization of the Arabic linguistic heritage contribute to its preservation, revival, and activate its presence in modern linguistic studies, and what challenges hinder its transition to a digital environment. The study also reviews the technical, methodological and cultural challenges facing digitization projects, such as the diversity of Arabic fonts, the lack of human expertise, and the absence of unified standards, highlighting Arab and international experiences and lessons learned from them, in order to provide a clear scientific vision for the development of digitization projects of Arab heritage to serve scientific research and education.

**Keywords---**digitization, heritage, Arabic language.

### Introduction

The Arabic linguistic heritage is a deep-rooted knowledge and cultural asset, which has contributed over many centuries to the construction of linguistic, grammatical and semantic lessons, and formed the basis on which successive generations relied in understanding the structure of Arabic and its development. With the accelerated digital transformations, traditional paper preservation methods are no longer sufficient to preserve this heritage or ensure the sustainability of its use, which has made the digitization of heritage an urgent necessity imposed by the challenges of the Times and the requirements of modern linguistic research.

---

### How to Cite:

Nora, L. (2026). Digitization of the Arabic linguistic heritage: Problems, challenges, and prospects of linguistic research. *The International Tax Journal*, 53(1), 93–99. Retrieved from <https://internationaltaxjournal.online/index.php/itj/article/view/500>

The International tax journal ISSN: 0097-7314 E-ISSN: 3066-2370 © 2026

ITJ is open access and licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Submitted: 06 April 2025 | Revised: 05 November 2025 | Accepted: 31 December 2025

The importance of this topic stems from the fact that it combines two complementary dimensions: The first is related to preserving the linguistic identity of the nation by preserving its original sources of manuscripts, dictionaries, grammatical and semantic books, and the second is related to the development of contemporary linguistic research by providing digital materials capable of automated analysis and investing in artificial intelligence and natural language processing techniques. Thus, digitization is not limited to being a technical process, but is a strategic scientific project that transcends the boundaries of documentation to establish a new knowledge space that revives heritage and puts it at the service of researchers.

This intervention falls within the fourth axis of the National Forum on "rooting linguistic heritage in contemporary linguistic thought – reality and challenges", and aims to present a scientific approach that highlights the reality of digitizing the Arabic linguistic heritage, evaluating the efforts exerted, analyzing the technical and methodological challenges facing Arab institutions in this field, while looking forward to the prospects that can contribute to its development.

#### **The problem of studying:**

To what extent does the digitization of the Arabic linguistic heritage contribute to its preservation, revival and activation of its presence in contemporary linguistic studies? What are the most prominent technical, methodological and cultural challenges that hinder the achievement of an integrated Arab digital project.

#### **Objectives of the study:**

- \* Highlighting the scientific value of linguistic heritage and the importance of digitization.
- \* Analysis of the impact of digitization in modern linguistic research and the development of its tools.
- \* Presenting the most important Arab and international efforts in digitizing the linguistic heritage.
- \* Identify the challenges facing digitization projects in the Arab world.
- \* Foreseeing the prospects for the development of this field in order to serve the contemporary linguistic lesson.

**Approved methodology:** this study is based on:

- Descriptive-analytical approach to the analysis of digitization concepts and the reality of Arab projects.
- A comparative approach to comparing Arab efforts with leading global experiences.

#### **1-the concept of Arabic linguistic heritage and its scientific dimensions :**

The Arabic linguistic heritage represents the basic pillar of linguistic, grammatical, phonetic and semantic Sciences in the Arab culture, and it is a reservoir of knowledge that has preserved the scientific and cultural identity through centuries, "the Arab heritage is not just a record of events," it is a living force that runs in the present and directs the future! "It is not a static historical material, but an influential element in the construction of modern Arabic linguistic thought and the development of its approaches.

From the very beginning, scientists have taken care of establishing a solid linguistic structure that preserves Arabic and regulates its laws, and Ibn Jinni describes the essence of the language by saying: "they are voices through which every people expresses their purposes." <sup>2</sup>linking the language with its function makes the preservation of linguistic heritage a necessity to understand the phonetic and synthetic foundations on which Arabic was based, especially in light of the development of linguistic analysis tools and artificial intelligence.

Studies point to the importance of preserving heritage because of its connection with religious and cultural knowledge, al-Suyuti pointed to this, saying: "know that the book of Allah Almighty is the origin and complexity of Science, and the Arabic language is the vessel of its understanding and the key

to its doors"<sup>3</sup>, which confirms that linguistic heritage is not an abstract linguistic science, but is a carrier of Arab and Islamic culture, and represents a basis for understanding religious texts, which makes its maintenance part of the preservation of the cultural and scientific identity of the Arab nation.

Heritage books are a foundational stage for the Arabic language lesson, because they represent the first structure that cannot be surpassed, because they laid the foundations for the language lesson. <sup>4</sup> it is an indispensable reference for any modern development or linguistic project.

Linguistic manuscripts occupy a special place in this heritage, as it is " an honest mirror of the efforts of scientists, the virtue of which is shown only by their realization and revival.their importance is <sup>5</sup> multiplied thanks to modern technologies that transform them from isolated texts locked in manuscripts to open resources that can be browsed, searched and automated. Based on these insights, three basic dimensions of the Arabic linguistic heritage can be identified:

**1. The cognitive dimension :**

Linguistic heritage represents the largest field of concepts, terms, and theories from which various Arabic linguistic trends originated, and modern linguistic phenomena cannot be explained without reference to their roots in this heritage.

**2. The cultural dimension :**

The linguistic heritage carries the nation's valuable and cognitive memory, and embodies its civilized character through the ages, which makes its preservation and preservation part of the preservation of identity itself.

**3. Applied scientific dimension :**

Heritage materials today have become the basis for the construction of corpora Arabic linguistic blogs, the development of automated processing applications, OCR visual recognition algorithms, morphological and semantic analysis, which makes its digitization a strategic step for modern linguistic research.

**2-efforts to digitize the Arabic linguistic heritage:**

Over the past two decades, the Arab world has witnessed accelerated steps towards the digitization of linguistic heritage, whether at the level of national libraries or university projects and scientific research centers, with the aim of preserving manuscripts and rare books and facilitating access to them for researchers.studies have indicated that the manuscript digitization project at the National Library of Algeria "reflects an institutional will to transform heritage into searchable and easily accessible digital data"<sup>6</sup>. this shows that digitization is not a purely technical process, but a knowledge and strategic project linking technology and institutional policies to ensure the sustainability of heritage.

The use of artificial intelligence technologies in the processing of Arabic manuscript texts provides the possibility of analyzing the derivation and semantic distribution in a way that was not available in the past, "and opens up wide horizons for the study of linguistic heritage and the extraction of grammar and digital lexicons"<sup>7</sup>, because investing in technology enhances the researcher's ability to access information quickly and accurately, while preserving the authenticity of the texts.

Several libraries and research centers have launched important digital programs, such as the Arab digital library project, which seeks to provide a digital copy of the Arab heritage, focusing on ancient manuscripts and books.despite the financial and technical challenges, this project allows researchers to benefit from digital texts for research and analysis. it is a strategic step to enhance cooperation between Arab institutions, because digitization facilitates the retrieval of manuscripts and expand the circle of access to them, which raises the level of Arab Scientific Research.

In the same context, reference can be made to the Saudi digital library, which adopted advanced scanning methods and accurate cataloging of books and manuscripts, which allowed them to be entered into automatically searchable databases, and studies show that "digital projects require the integration of

human experience and technical expertise to ensure data quality and sustainability"<sup>8</sup>, this indicates that digitization is a precise scientific process that needs coordination between technology and the expertise of researchers and calligraphers.

At the institutional level, cooperation between universities and libraries is the key to the success of digitization projects, as joint agreements with European libraries, such as the British Library, allowed the exchange of experiences and technologies, and contributed to the training of Arab specialists on the latest methods of scanning and digital processing of texts. This enhances the ability of Arab libraries to develop their digital projects and reach international standards, " he added.<sup>9</sup>

This experience shows that the exchange of experiences and technologies between countries can reduce the digital divide and ensure the sustainability of Arab heritage at the global level.

### **3-digitization and automated processing of Arabic texts**

Automated processing of Arabic texts constitutes the second stage after the digitization of manuscripts, as it allows converting digital texts into searchable and analyzable data using artificial intelligence techniques, Optical Character Recognition (OCR), semantic and morphological analysis programs, because it "provides unprecedented possibilities for analyzing the derivation and distribution of semantic concepts, which contributes to the construction of advanced digital dictionaries and huge text databases"<sup>10</sup>, thus the researcher exceeds the limitations that he faced in traditional paper copies.

The diversity of Arabic fonts and the decoration of manuscripts pose a great challenge to automated indexing, as these texts require the integration of human experience with digital technology to ensure the accuracy of the data"<sup>11</sup>, which highlights that the success of digitization projects depends on the balance between technological tools and traditional linguistic knowledge, which makes the automated processing of texts more efficient for scientific research.

Modern OCR technologies have helped to convert printed and Manuscript texts into a digital format amenable to processing, where researchers can now conduct complex searches, analyze derivations, extract semantic relationships between words, because digitization and automated processing turn the Arab heritage into a living material, which can be analyzed and used in the development of linguistic and Educational Studies"<sup>12</sup>, digitization and automated processing is a strategic step to reproduce Arabic linguistic knowledge.

At the level of practical applications, some Arab universities have launched digital platforms containing comprehensive text databases (Corpora) for ancient and modern texts, with the possibilities of statistical, grammatical and semantic analysis, and global reports indicate that "the creation of accurate text databases for Arabic texts opens the way for in-depth linguistic analyses that were not previously possible"<sup>13</sup>, reflecting that digitization and automated processing enable the Arab researcher to deal with heritage in an innovative way, which contributes to the development of educational curricula and advanced digital educational tools.

### **4-technical and methodological challenges in the projects of digitization of linguistic heritage :**

Despite the great progress in digitizing the Arabic linguistic heritage, these projects face a range of challenges ranging from technical, financial, human and methodological, which hinders the optimal use of heritage in scientific research and education.

1. Technical challenges: they are manifested in the difficulty of dealing with decorated texts or damaged manuscripts because they "require advanced OCR algorithms, as the diverse Arabic calligraphy and fine decorations are an obstacle to accurate automated indexing"<sup>14</sup>, because technology alone is not enough, but human expertise must be involved to ensure the accuracy of the data and the continuity of its reliability.
2. Financial challenges: a major obstacle to the implementation of large-scale digitization projects, the report of the Algerian Ministry of Culture indicated that the cost of scanning digital manuscripts and

building the necessary infrastructure may reach millions of dollars, which requires the adoption of long-term funding plans.

This shows that the success of digital projects is linked to financial stability and the provision of sustainable resources to ensure the continuation of digitization and automated processing processes.

3. Human challenges: it lies in the lack of qualified experts in Arabic calligraphy and modern technologies, which makes the continuous training of specialists an urgent necessity " because this limits the ability of institutions to digitize heritage systematically, and requires advanced training programs to enhance competencies, so it is necessary to invest in the human element.
4. Methodological challenges: related to the absence of unified standards for converting manuscripts into searchable and analyzable digital texts, this reduces the ability of researchers to take full advantage of digital texts and creates disparities between Arab projects. This shows that the development of clear protocols for digitization and automated processing is necessary to unify Arab efforts and ensure interchangeability and integration between different databases.

In addition, cultural and institutional factors pose a challenge to digitization, as some manuscripts require approvals and careful verification before being entered into digital databases, which prolongs the digitization process and makes it complicated compared to printed books, because the digitization of Arab heritage is not limited to copying or scanning. This confirms that the success of digitization projects depends on the cooperation between technology and scientific expertise to ensure the quality of digital content.

#### **5-international experiences and lessons learned in the digitization of linguistic heritage:**

Libraries and international research institutions have witnessed successful projects to digitize the linguistic heritage, which can be used to develop the Arabic experience, and the British digital library is a prominent example, as it has digitized millions of books and manuscripts, with accurate indexing and an advanced search interface, which allowed quick and accurate access to texts. The British Library report notes that: the comprehensive digitization of manuscripts enhances academic research and enables the analysis of texts via advanced digital tools<sup>15</sup>, and this experience demonstrates the importance of combining advanced technology and careful institutional organization to ensure the success of the digital project.

The Google Books project reflects a global experience in digitizing old and modern printed books, focusing on publicly available texts, and allows quick search within digital texts. According to the project report, the creation of accurate digital copies of printed texts expands the circle of knowledge and allows unprecedented linguistic and cognitive analysis<sup>16</sup>, and this indicates that digitization is not just memorization, but enables the development of scientific research and analysis of texts in an advanced way.

In the context of the Arab experiences, the Arab digital library stands out, which aims to provide a digital version of the Arab heritage, but it still faces the challenges of financing and automated text processing. This reveals that international experiences provide clear lessons, the most important of which are: the need for unified standards, technology investment, and human qualification to ensure the success of digitization projects in the Arab world.

International experiences have also shown that cooperation between research institutions and international libraries increases the efficiency of projects and promotes the exchange of experiences, which is missing in many Arab projects, making them less competitive and aligned with international standards. According to the UNESCO report:

International cooperation and the exchange of technical expertise contribute to the development of heritage digitization projects and ensure their sustainability<sup>17</sup>, This reduces the digital gap and enhances the quality of Arab projects.

## Conclusion

The research concluded a number of results, which we summarize in the following:

- The Arabic linguistic heritage includes grammatical, semantic and rhetorical manuscripts and books, and preserves the nation's scientific and cultural memory, making it a fundamental pillar for all contemporary linguistic studies.
- Digitization is not just a technical process, but a cultural and Knowledge project that contributes to the preservation of linguistic identity and the promotion of scientific research.
- Arab and international experiences have shown that the success of digital projects depends on the integration of technology, institutional policies, qualified human competencies, and the development of unified standards for digital transformation and automated text processing.
- Digitization provides free access to texts, enables advanced analytical tools, which enhances the quality of scientific research and develops modern educational and linguistic applications. At the same time, Arab projects face financial, human, technical and methodological challenges, including: the difficulty of dealing with illuminated manuscripts, the lack of qualified experts, the absence of unified standards, and the high cost of digital infrastructure.

Based on these data, the following recommendations can be made to promote digitization projects in the Arab world:

1. Develop unified Arabic standards for digitization to ensure the quality of digital texts and standardize procedures between different institutions.
2. Expanding training programs in the field of Arabic calligraphy, digital linguistic analysis, OCR technologies and automated processing.
3. Development of open digital platforms for researchers that allow free access to digital texts and conduct advanced linguistic and lexical analysis.
4. Strengthen cooperation between Arab and international institutions to exchange experiences and technologies, and benefit from successful experiences globally.
5. Allocate sustainable financial resources to ensure the continuity of digitization projects and the maintenance of the Arabic linguistic heritage

In conclusion, it can be said that the digitization of the Arabic linguistic heritage represents a strategic knowledge and civilization investment, linking the past and the present, and opens up wide horizons for scientific research, education and innovation in the Arabic language, and the ultimate goal remains to transform this heritage into a usable living material, benefiting researchers, teachers and students, to serve the development of Arabic linguistic thought in the digital age.

## References:

- 
- <sup>1</sup> Ahmed Amin, dawn of Islam, Dar Al-Alam for millions, Cairo, 2002.P. 5.
  - <sup>2</sup> Ibn Jinni, characteristics, investigation of Muhammad Abdus Salam, House of scientific books, Beirut, 1995. P. 33
  - Fikr, Damascus, 2008.P. 6-Mazhar in Language Sciences, Dar Al-Suyuti, Al-Al <sup>3</sup>
  - <sup>4</sup> Tamam Hassan, the Arabic language, its meaning and its building, renaissance library, Cairo, 2010. P. 12.
  - <sup>5</sup> Abdessalam Haroun, the investigation of texts, the House of Arab Thought, Cairo, 2015.P. 11.
  - <sup>6</sup> Souad tebbert, manuscripts digitization project of the National Library of Algeria, Journal of linguistic studies, 2020, P.45.
  - <sup>7</sup> Muhammad Elias boumedini, applications of artificial intelligence in Arabic manuscript text processing, 2019, P.32.
  - <sup>8</sup> Jamila Riwaq, digitizing Arabic calligraphy: challenges and prospects, Arab heritage Magazine, 2018.P. 23.

- 
- <sup>9</sup> Digital collections of the British Library, 2022
- <sup>10</sup> Muhammad Elias boumedini, applications of artificial intelligence in Arabic manuscript text processing, 2019, P. 45.
- <sup>11</sup> Jamila Riwaq, digitization of Arabic calligraphy: challenges and Prospects, 2018, p.30.
- <sup>12</sup> Abdullah Younis, digitization of Arab heritage: foundations and challenges, Journal of Informatics, 2020, P.22.
- <sup>13</sup> Google Books project, 2021
- <sup>14</sup> Mohamed Elias boumedini, applications of artificial intelligence in the text processing of Arabic manuscripts, Algeria, 2019, P. 51.
- <sup>15</sup> Digital collections of the British Library, 2022
- <sup>16</sup> Google Books project, 2021
- <sup>17</sup> UNESCO, World Digital Library report, 2020.