Pegem Journal of Education and Instruction, Vol. 15, No. 4, 2025 (pp. 4826-4839)

RESEARCHARTICLE WWW.PEGEGOG.NET

Terminology Banks in the Arab World and Their Role in Standardizing Terminology in the Era of Digital Dominance

Hafidha Bourouba

Department of Arabic Language and Literature, Faculty of Letters and Languages,
Badji Mokhtar Annaba University, Algeria hafidha.bourouba@univ-annaba.dz

Received:12.01.2025 Accepted:17.04.2025 Published: 25.12.2025

Abstract

This study examines the role of terminology banks as digital infrastructures that manage and standardize specialized knowledge in the Arab world. Anchored in terminology science and digital terminography, it analyzes how digitization reshapes terminological management and linguistic policy. Using a descriptive–analytical and comparative approach, the research evaluates selected Arabic terminology banks in light of international models such as IATE, UNTERM, and WHO portals. Findings indicate that, despite notable progress, Arabic systems remain fragmented, technically outdated, and insufficiently AI-integrated. The paper proposes a modernization framework emphasizing interoperability, metadata harmonization, and regional cooperation to reinforce linguistic unity and digital sovereignty in the era of technological dominance.

Keywords: Terminology banks; Digital transformation; Terminography; Arab world.

1. Introduction

Terminology lies at the core of scientific and technological advancement. Specialized language not only facilitates communication but also structures how knowledge is conceptualized, categorized, and disseminated. In an era of digital globalization, the standardization of terminology ensures conceptual clarity, data interoperability, and linguistic equity across languages (Cabré, 1999; Temmerman, 2000).

The Arab world, with its rich linguistic heritage and growing scientific output, faces a dual challenge: preserving cultural specificity while adapting to global digital transformation. Traditional lexicographic tools are insufficient to manage the exponential growth of specialized vocabularies. By contrast, terminology banks - structured databases for storing and standardizing terms - serve as living infrastructures for organizing and sharing knowledge (Felber, 1984; Sager, 1990).

While international terminology banks such as IATE, UNTERM, and Term Coord exemplify technological maturity and interinstitutional coordination, most Arabic systems remain fragmented and lack cross-platform interoperability. The need for modernization is urgent.

In contemporary knowledge economies, terminology is increasingly treated as data - searchable, linkable, and reusable across platforms -rather than as static entries in reference works. This shift introduces new expectations for terminological resources: machine-readability, metadata-rich entries, version control, and semantic alignment with international standards to enable

integration with translation technologies, knowledge graphs, and scientific publishing ecosystems (Bowker & Pearson, 2002; Bowker, 2015). Without such digital affordances, terminological resources risk becoming isolated repositories that cannot effectively support modern workflows in education, research, and specialized communication.

Moreover, the rise of artificial intelligence and natural language processing has made terminological consistency a prerequisite for reliable language technologies. Terminology banks, when designed with robust conceptual modeling, can serve as high-quality training and validation resources for domain-specific machine translation, information retrieval, and automated writing assistance, reducing ambiguity and improving semantic precision in Arabic scientific discourse (Jurafsky & Martin, 2023; Bender & Friedman, 2018). In this sense, modernization is not only an infrastructural upgrade but also an enabling condition for Arabic's meaningful participation in emerging AI-driven knowledge production.

Finally, terminology banks must be understood as instruments of linguistic sovereignty. When terms are standardized through transparent, regionally coordinated mechanisms, Arabic gains the capacity not merely to receive scientific concepts through translation, but to stabilize and circulate them according to its own conceptual and morphological logic. This study therefore treats terminology banks as strategic infrastructures positioned at the intersection of language policy, digital humanities, and knowledge governance - where technical design decisions directly shape the visibility, usability, and future viability of Arabic specialized language in a global digital order.

2. Theoretical Framework

2.1. Terminology Science and Conceptual Systems

Terminology science, rooted in the work of Eugen Wüster (1979), views terminology as a structured system linking linguistic forms to conceptual relations. The General Theory of Terminology (GTT) emphasizes univocity and conceptual clarity, forming the foundation for all standardization efforts. Later models, notably Cabré's Communicative Theory of Terminology (CTT) and Temmerman's sociocognitive approach, expanded this foundation by integrating cognitive and social dimensions. Terminology is now seen as dynamic and context-dependent rather than static and prescriptive (Cabré, 1999; Temmerman, 2000). This evolution is critical in the digital era, where terminological data are continuously updated and shared through semantic web technologies. Terminology banks operationalize these theories by structuring concepts, metadata, and multilingual equivalences for digital access (Pearson, 1998; Bowker, 2015).

In practical terms, this theoretical shift reframes standardization from being a purely normative activity to an evidence-based process grounded in authentic discourse and domain usage. As specialized knowledge expands and disciplines converge, terminological systems must accommodate conceptual change, synonymy, and variation while maintaining traceable definitions and stable relationships between concepts. This balance is central to the credibility of terminology banks, which must function simultaneously as repositories of standardized forms and as adaptive systems responsive to new scientific developments.

Digital terminography further strengthens this trajectory by embedding terminological resources within computational infrastructures. When terminology banks adopt structured models—such as concept-oriented entries, persistent identifiers, and rich metadata—they become

interoperable with translation memories, corpus tools, and knowledge representation systems, enabling reusability and cross-platform integration (Bowker & Pearson, 2002; Sowa, 2000). This is particularly significant for multilingual environments, where equivalence is rarely a simple lexical pairing but a negotiated conceptual alignment requiring definitional and contextual anchoring.

Finally, terminology banks increasingly operate as strategic resources in AI-enabled workflows. High-quality terminological datasets can support automatic term extraction, domain adaptation for machine translation, and semantic consistency checks in specialized writing, thereby reducing ambiguity and improving the reliability of computational outputs. In this sense, the contemporary value of terminology science is not limited to standardization policy; it also provides a methodological backbone for building linguistically grounded digital systems capable of managing specialized knowledge at scale.

2.2. From Lexicography to Digital Terminography

Terminography differs from lexicography in its focus on specialized, domain-specific language. It aims at conceptual consistency rather than descriptive breadth (Sager, 1990). With digital transformation, terminography has evolved into digital terminography, relying on corpus-based tools, automated extraction, & metadata annotation (Bowker, 2015). In this sense, terminology banks embody the applied dimension of terminology science - functioning as dynamic repositories for terminological harmonization & multilingual translation quality assurance.

This evolution reflects a broader shift from viewing terminological work as an editorial practice to treating it as an infrastructural activity embedded in digital knowledge ecosystems. Digital terminography requires not only the collection of terms, but also the modeling of concepts through definitional structures, domain classifications, usage contexts, and equivalence relations across languages. These layers transform terminology into a form of structured knowledge that can be queried, linked, and reused, thereby enabling interoperability across institutional platforms, scientific repositories, and translation workflows (Bowker & Pearson, 2002).

Moreover, the increasing reliance on computational processing has expanded the methodological toolkit of terminographers. Corpus-driven approaches support evidence-based term selection, reveal collocational behavior and domain-specific patterns, and facilitate continuous updating as new knowledge enters the discourse. Automated term extraction and candidate validation improve scalability, but they also raise questions of quality control - requiring human expertise to resolve ambiguity, manage synonymy, and stabilize definitions in ways that preserve conceptual integrity rather than merely capturing frequency (Sager, 1990; Bowker, 2015).

Finally, digital terminography is closely linked to contemporary AI and NLP applications. Terminology banks can serve as authoritative resources for training and evaluating domain-adapted machine translation systems, improving retrieval precision in specialized search, and supporting consistent generation in technical writing and documentation. In multilingual contexts, this role becomes especially strategic: terminology banks help ensure that equivalence is conceptually grounded, reducing semantic drift and enhancing cross-language consistency, which are essential requirements for reliable communication in science, medicine, law, and emerging technologies.

2.3. Digital Transformation and Artificial Intelligence

Digitalization introduces both opportunities and challenges. Tools based on Natural Language Processing (NLP) and Artificial Intelligence (AI) allow automatic term extraction, clustering, and contextual validation. For instance, Arabic morphological analyzers such as MADAMIRA or Farasa can detect term variants, while AI-based term alignment tools enhance cross-language equivalence (L'Homme, 2020; Pasha et al., 2014; Abdelali et al., 2016). However, digital dominance also brings risks: algorithmic bias, data inconsistency, and potential loss of cultural nuance. Thus, terminological standardization must balance technological innovation with linguistic authenticity.

A major opportunity lies in moving terminology banks from "static lists" to computationally actionable resources. When terminological entries are enriched with metadata (domain labels, definitions, usage contexts, provenance, and versioning), they become compatible with automated pipelines for technical writing, information retrieval, and domain-adapted machine translation. In Arabic, this is particularly valuable because morphological variation and orthographic ambiguity can fragment term recognition; morphology-aware processing can consolidate variants and improve both extraction and indexing quality (Habash, 2010; Obeid et al., 2020).

At the same time, AI-driven term alignment and multilingual equivalence modeling require concept-level rigor rather than surface lexical matching. Cross-language alignment often fails when it treats equivalence as a one-to-one word mapping, ignoring definitional scope, domain-specific constraints, and culturally embedded conceptual boundaries. Terminology banks can mitigate this by encoding conceptual relations explicitly (preferred term, admitted variants, broader/narrower relations, and definitional notes), thereby supporting more faithful semantic alignment and reducing "translation drift" in specialized domains (Sowa, 2000; Bowker & Pearson, 2002).

Nevertheless, the deployment of AI in terminological modernization raises governance and epistemic concerns. If training data are unbalanced or poorly documented, automated extraction and clustering can amplify bias, privilege dominant registers, and marginalize regionally legitimate variants—particularly in Arabic, where dialectal and institutional diversity is substantial. For this reason, modernization efforts should adopt documentation and accountability practices (e.g., resource statements, quality audits, and transparent update policies) to ensure that computational gains do not come at the cost of representational fairness or conceptual integrity (Bender & Friedman, 2018; Mitchell et al., 2019; Bender et al., 2021).

3. Methodology

This research follows a descriptive–analytical and comparative qualitative design, appropriate for evaluating both conceptual and structural aspects of terminology banks. It also incorporates targeted empirical illustration through benchmarking-style comparisons to support analytical claims with measurable indicators of coverage, consistency, and interoperability.

3.1. Research Design

The multi-phase design integrates description, comparison, and synthesis (see **Table 1**).

Table 1: Research Design and Analytical Parameters

Stage	Objective	Method	Output

Descriptive	Analyze selected Arabic terminology banks (ALECSO, Saudi Portal, Jordanian Academy)	Document and content analysis	Identification of structures and management models
Comparative	Benchmark Arabic systems against international databases (IATE, UNTERM, WHO)	Cross-system analysis; ISO alignment	Assessment of structural and technological gaps
Integrative	Develop modernization framework	Thematic synthesis; triangulation	Strategic roadmap for digital transformation

Source: Adapted from ISO 704 (2009) and ISO 1087 (2019).

3.2. Data Sources

Data were derived from:

- Academic publications (Cabré, 1999; Felber, 1984; Bowker, 2015).
- Institutional documentation of Arabic terminology banks (e.g., ALECSO's Terminology Portal, Saudi Terminology Bank).
- ISO and W3C standards for terminological interoperability (ISO 704, ISO 12620, TBX, RDF).

3.3. Analytical Procedures

A triangulated analysis compared Arabic and international databases along four parameters:

- Conceptual and organizational structure.
- Technological infrastructure and interoperability.
- Standardization procedures and linguistic consistency.
- Accessibility and usability.

3.4. Research Questions

- How do Arabic terminology banks structure & manage terminological data?
- What are the principal differences between Arabic & global terminology systems?
- What strategies can ensure modernization & terminological harmonization?

4. Results and Discussion

4.1. The Role of Terminology Banks in Terminological Harmonization

Terminology banks represent a paradigm shift from static lexicographic resources to dynamic knowledge systems. They not only document terminology but also structure conceptual relations across languages and disciplines. As Wüster (1979) and Felber (1984) emphasized, terminological

harmonization ensures one-to-one correspondence between concepts and terms. Digital banks extend this principle into the semantic web era, where knowledge is encoded, searchable, and interconnected through metadata standards.

What distinguishes a terminology bank from a conventional reference tool is its ability to operationalize concept systems—that is, networks of relations (hierarchical, associative, and equivalence-based) that make specialized knowledge computationally navigable. When these relations are explicitly modeled through definitional entries, domain classifications, and controlled variants, terminology banks become infrastructures for coherence: they reduce ambiguity, stabilize usage across institutions, and enable consistent communication in high-stakes fields such as medicine, law, engineering, and education (Sager, 1990; Bowker & Pearson, 2002). In multilingual settings, harmonization also functions as a quality-control mechanism, ensuring that translation practices do not merely replicate words but preserve conceptual scope and disciplinary intent.

Furthermore, terminological harmonization has become inseparable from interoperability. In digital ecosystems, terms must circulate across platforms—translation memories, authoring tools, digital libraries, and AI-driven retrieval systems—and this circulation depends on standardized exchange formats and metadata conventions. When terminology banks align with standards such as TBX and adopt persistent identifiers, provenance fields, and versioning, they enable traceable reuse and prevent the fragmentation of specialized vocabulary across competing institutional silos (Bowker, 2015). From this standpoint, harmonization is not only linguistic alignment; it is also a form of information governance that determines whether Arabic terminological knowledge can be integrated into global scientific infrastructures or remain locally bounded.

Finally, the strategic value of terminology banks today lies in their capacity to function as "semantic anchors" in AI-mediated communication. As NLP and generative AI expand into technical domains, terminological consistency becomes a prerequisite for reliable automated writing, domain translation, and knowledge extraction. Well-curated terminology banks can constrain semantic drift, support disambiguation, and improve domain adaptation by offering validated concept–term mappings and contextual usage evidence. Thus, harmonization should be understood not as a conservative standardization impulse, but as a forward-looking mechanism for ensuring that digital and AI systems reproduce specialized knowledge with precision, accountability, and linguistic legitimacy.

Table 2 summarizes the multidimensional functions of terminology banks and their contributions to harmonization.

Table 2: Core Functions of Terminology Banks in Knowledge Management

Dimension	**Primary Functions**	**Contribution to Harmonization**
Linguistic	Concept definition and standardization	Ensures semantic clarity and cross-domain coherence
Technological	Data structuring, metadata management, interoperability	Enables machine-readable consistency and cross-platform exchange

Institutional	Coordination, expert validation, and linguistic policy alignment	Fosters collaborative and sustainable terminology governance

Source: Adapted from Cabré (1999), Bowker (2015), and ISO 704 (2009).

Terminology banks thus function as epistemic infrastructures integrating linguistic expertise, digital technology, and institutional policy. Their importance lies in connecting the conceptual, technical, and sociolinguistic dimensions of specialized language. In practice, this integration determines whether harmonization remains an internal scholarly ideal or becomes an operational capability embedded in translation workflows, digital education, and domain publishing.

Accordingly, the maturity of a terminology bank can be assessed by the extent to which its conceptual architecture is consistently aligned with interoperable standards and sustained governance mechanisms that ensure continuous validation and reuse.

4.2. Empirical Illustration: Arabic Terminology Banks

While the Arab world hosts several terminology initiatives, two examples highlight both progress and challenges. The ALECSO Terminology Portal, developed under the Arab League, centralizes specialized terms in education, science, and culture. Its main strength lies in regional recognition and institutional support. However, the portal remains manually updated and not yet fully compliant with TBX (TermBase eXchange) standards, limiting interoperability. In contrast, the Saudi Terminology Bank, launched by the King Abdulaziz City for Science and Technology, integrates automatic data extraction and NLP modules for Arabic morphology. Despite this innovation, its data remain inaccessible to the public, reducing potential for academic and translation use. These examples demonstrate that Arabic terminology institutions have the intellectual foundation but still lack digital openness and structural alignment with international frameworks.

From an empirical perspective, the contrast between these two cases can be read as a tension between institutional legitimacy and computational readiness. ALECSO's value is primarily epistemic and policy-oriented: it symbolizes regional consensus & provides a reference point for terminological unification across educational and cultural sectors. Yet, in a digital environment where terminological infrastructures must be continuously queried, linked, and embedded into workflows, manual updating & limited export options constrain operational impact. Without standardized exchange formats (e.g., TBX), persistent identifiers, and structured metadata (domain labels, definitions, usage contexts, provenance), a terminology bank risks remaining a "reference shelf" rather than becoming a living component of the knowledge economy (Sager, 1990; Bowker & Pearson, 2002; Bowker, 2015).

Conversely, the Saudi initiative reflects a shift toward automation, scalability, and language technology integration, indicating awareness that terminological modernization increasingly depends on NLP capabilities such as morphological normalization, variant detection, and automated candidate term extraction (Habash, 2010). However, restricted access introduces a structural limitation: closed data ecosystems prevent independent validation, reduce uptake by translators and researchers, and hinder the formation of shared regional standards. In practical terms, interoperability is not merely a technical property—it is also a governance decision about openness, reuse rights, and institutional coordination. Thus, the empirical lesson is not that one model is preferable, but that

sustainable modernization requires combining both: the legitimacy of coordinated standardization with the openness and technical compliance that make terminology computationally usable across platforms and communities.

4.3. Challenges in Digital Transformation

As summarized in **Table 3**, the main barriers to modernization involve technology, coordination, and human expertise.

Table 3: Challenges in Digital Terminology Management in the Arab World

Challenge Type	Description	Impact	Suggested Response
Technological	Absence of standardized exchange formats (TBX, RDF)	Fragmented data, lack of interoperability	Adoption of open data models
Organizational	Weak coordination among linguistic bodies	Redundancy, inefficiency	Establish regional governance body
Human Resources	Limited training in digital terminography	Low consistency, outdated methods	Introduce specialized education programs
Cultural/Conceptual	Variations in term creation and validation	Conceptual inconsistency	Context-based adaptation of global standards

Source: Synthesized from Temmerman (2000), Bowker & Pearson (2002), and ISO 12620 (2019).

Addressing these challenges requires both technological infrastructure and linguistic planning, supported by policy harmonization and long-term investment.

Equally important is establishing shared quality metrics and routine auditing mechanisms to ensure consistency, traceability, and sustainable updating across Arabic terminology platforms.

4.4. AI and NLP in Terminology Modernization

Modern terminology banks increasingly depend on AI-driven language technologies. For Arabic, tools such as MADAMIRA, Farasa, and AraBERT support automatic term extraction, morphological disambiguation, and semantic clustering. These technologies enable:

- Automated identification of candidate terms from large corpora.
- Cross-language mapping through vector-based embeddings.
- Context validation via machine learning models.

Integrating such systems into Arabic terminology banks would transform them into self-updating ecosystems, capable of learning from user input and real-time linguistic data. Yet this integration requires both technical literacy and ethical oversight to maintain linguistic authenticity.

At a methodological level, AI-assisted terminological modernization changes the locus of terminological work. Instead of relying primarily on manual compilation, experts can increasingly supervise pipelines that detect, rank, and validate terminological candidates using corpus statistics, contextual embeddings, and domain-sensitive similarity measures. In Arabic, morphology-aware preprocessing is not optional: robust segmentation and disambiguation are prerequisites for reliable extraction because surface variation can mask conceptual identity (Habash, 2010; Pasha et al., 2014; Abdelali et al., 2016). When these tools are integrated upstream, terminology banks gain the capacity to recognize variants systematically, reduce duplication, and support cleaner conceptual indexing.

However, the technical promise of embeddings and semantic clustering must be governed by conceptual rigor. Vector-based similarity can overestimate equivalence by conflating co-occurrence with definitional identity, especially in specialized domains where near-synonyms may carry distinct conceptual scope. For this reason, AI outputs should be treated as hypotheses that require terminological validation through definitions, domain constraints, and usage attestations, rather than as automatic truth. A mature modernization strategy therefore combines neural methods with concept-oriented modeling—linking extracted terms to concept entries, recording provenance and contexts, and enforcing controlled variant policies to prevent semantic drift (Sowa, 2000; Bowker & Pearson, 2002).

Finally, the move toward "self-updating ecosystems" raises governance questions that are particularly salient for Arabic. If user feedback loops or web-scale data streams are incorporated without curation, terminology banks may inadvertently privilege dominant registers, amplify institutional bias, or dilute culturally anchored conceptual distinctions. Ethical oversight is therefore not a supplementary concern but a design requirement: modernization should include documentation practices, transparency about update policies, and auditing mechanisms to ensure that computational convenience does not erode linguistic authenticity or terminological legitimacy (Bender & Friedman, 2018; Mitchell et al., 2019). In this sense, the most effective Arabic terminology bank of the future will be not merely automated, but accountable: technologically advanced while remaining conceptually disciplined and culturally faithful.

4.5. Comparative Insights from Global Best Practices

Comparing global and Arabic systems reveals sharp contrasts in governance, accessibility, and technological sophistication.

Table 4: Comparative Features of Global vs. Arabic Terminology Banks

Feature	**IATE (EU)**	**UNTERM (UN)**	**WHO Portal**	**Arabic Banks (Typical)**

Governance	Multi-institutional collaboration	Centralized UN structure	Domain-specific coordination	National or academic management
Data Format	TBX, RDF, XML	XML, multilingual ontology	RDF-based, domain-linked	Proprietary, nonstandard
Access	Fully open	Partially open	Controlled institutional access	Limited or offline
Update Frequency	Continuous	Regular	Periodic	Irregular
AI Integration	High	Moderate	Emerging	Minimal
Regional Cooperation	Strong (EU-wide)	Global	Sectoral	Weak

Source: Data adapted from TermCoord (2022), ISO (2019), and WHO Terminology Portal (2023).

The analysis shows that **openness, automation, and collaboration** are the hallmarks of successful systems. Adopting these features can guide Arabic institutions toward sustainable modernization.

4.6. Framework for Modernization and Regional Integration

Drawing from the comparative analysis, a six-pillar framework is proposed (see **Figure 1** and **Table 5**).

Figure 1: Modernization Framework for Arabic Terminology Banks (Grayscale flowchart to be visually represented in Word version)

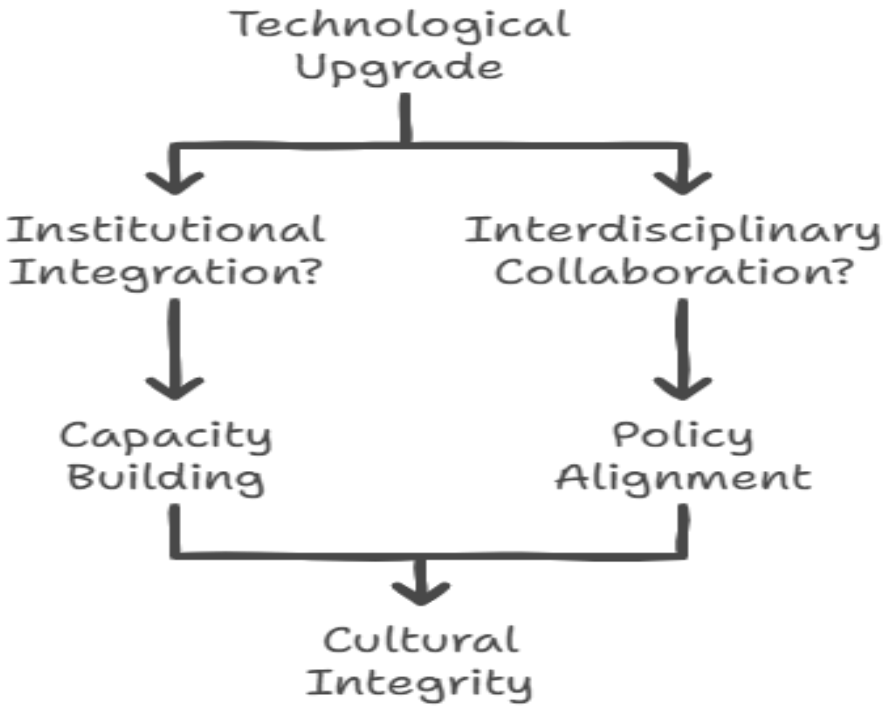

Table 5: Strategic Pillars for Modernizing Arabic Terminology Banks

Pillar	Objective	Key Actions
Technological Modernization	Upgrade databases using TBX/RDF; integrate AI term extraction	Establish cross-platform interoperability
Institutional Integration	Create Pan-Arab Terminology Network	Centralize governance and data sharing
Interdisciplinary Collaboration	Link linguists and computer scientists	Develop hybrid linguistic-technological solutions
Capacity Building	Train terminographers and computational linguists	Introduce digital terminography curricula
Policy Alignment	Embed terminology planning in national language policies	Secure long-term funding and regulation
Cultural Integrity	Maintain Arabic conceptual authenticity	Use sociocognitive approaches for equivalence

Source: Synthesized from Bowker (2015), Cabré (1999), and Temmerman (2000).

This model envisions terminology banks as collaborative, interoperable, and intelligent systems that combine human expertise with digital automation.

A practical indicator of success is whether these pillars translate into measurable outcomes open exchangeable datasets, regular update cycles, and demonstrable gains in term coverage, consistency, and cross-institution reuse.

5. Conclusion

The study demonstrates that terminology banks are not mere linguistic repositories but strategic infrastructures for knowledge organization and linguistic empowerment. In the digital age, they function as the semantic backbone of specialized communication: they stabilize concepts, discipline variation, and enable languages to participate in global knowledge circulation on equitable terms. When terminology is treated as structured data rather than dispersed vocabulary, it becomes reusable across translation, publishing, education, and AI pipelines - turning linguistic resources into operational capacity rather than symbolic heritage.

Despite commendable progress, Arabic terminology systems remain fragmented and technologically constrained. The gap is not primarily intellectual; it is infrastructural—manifesting in weak interoperability, inconsistent metadata practices, limited update governance, and uneven openness for academic reuse. To bridge this gap with global standards, the proposed six-pillar framework offers a roadmap for sustainable modernization: it links conceptual modeling to digital standards, and it connects linguistic authority to computational usability. In my view, modernization succeeds only when it respects two principles simultaneously: terminological legitimacy (conceptual integrity and institutional trust) and digital viability (exchangeability, traceability, and platform integration).

Beyond technical reform, this transformation has social and educational implications. It supports translation accuracy by providing authoritative concept–term mappings that reduce drift and inconsistency across governmental and academic discourse, particularly in rapidly evolving fields. It enhances educational accessibility by enabling digital curricula, searchable glossaries, and standardized scientific language-conditions that make learning scalable and inclusive rather than dependent on scattered resources. It also safeguards cultural identity by ensuring that Arabic develops as a language of modern knowledge on its own conceptual terms, not as a passive receiver of imported categories that may not align with Arabic's semantic architecture.

Ultimately, the future of Arabic terminology banks should not be imagined as digitized dictionaries, but as intelligent linguistic ecosystems: dynamic platforms where terminological authority, corpus evidence, and AI-enabled updating converge under transparent governance. By adopting interoperable standards, building open and auditable infrastructures, and integrating NLP responsibly, Arabic terminology banks can become engines of linguistic sovereignty—capable of sustaining modernization without sacrificing authenticity, and of enabling Arabic not only to translate knowledge, but to structure and circulate it in the digital age.

6. References

Abdelali, A., Darwish, K., Durrani, N. and Mubarak, H. (2016) 'Farasa: A fast and furious segmenter for Arabic', Proceedings of NAACL-HLT 2016: Demonstrations, pp. 11–16. Available at: https://aclanthology.org/N16-3003/ aclanthology.org

Bender, E.M. and Friedman, B. (2018) ‘Data statements for natural language processing: Toward mitigating system bias and enabling better science’, Transactions of the Association for Computational Linguistics, 6, pp. 587–604. Available at: https://doi.org/10.1162/tacl_a_00041 direct.mit.edu

Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021) ‘On the dangers of stochastic parrots: Can language models be too big?’, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21), pp. 610–623. Available at: https://doi.org/10.1145/3442188.3445922 dl.acm.org

Bowker, L. (2015) Terminology: Tradition, Transition, and Innovation. Amsterdam/Philadelphia: John Benjamins.

Bowker, L. and Pearson, J. (2002) Working with Specialized Language: A Practical Guide to Using Corpora. London: Routledge.

Cabré, M.T. (1999) Terminology: Theory, Methods and Applications. Amsterdam/Philadelphia: John Benjamins.

Felber, H. (1984) Terminology Manual. Paris: UNESCO/Infoterm.

Habash, N. (2010) Introduction to Arabic Natural Language Processing. Morgan & Claypool.

ISO 704 (2009) Terminology Work—Principles and Methods. Geneva: International Organization for Standardization.

ISO 1087 (2019) Terminology Work—Vocabulary—Part 1: Theory and Application. Geneva: International Organization for Standardization.

ISO 12620 (2019) Management of Terminological Data—Data Category Specifications. Geneva: International Organization for Standardization.

Jurafsky, D. and Martin, J.H. (2023) Speech and Language Processing (3rd ed., draft). Stanford University.

L’Homme, M.-C. (2020) Lexical Semantics for Terminology: An Introduction. Amsterdam/Philadelphia: John Benjamins.

Mitchell, M., Wu, S., Zaldivar, A., et al. (2019) ‘Model cards for model reporting’, Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT ’19). Available at: https://doi.org/10.1145/3287560.3287596 dl.acm.org

Obeid, O., Zalmout, N., Khalifa, S., et al. (2020) ‘CAMeL Tools: An open-source Python toolkit for Arabic natural language processing’, Proceedings of LREC 2020, pp. 7022–7032. Available at: https://aclanthology.org/2020.lrec-1.868/ aclanthology.org

Pasha, A., Al-Badrashiny, M., Diab, M., et al. (2014) ‘MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic’, Proceedings of LREC 2014. Available at: https://aclanthology.org/L14-1479/ aclanthology.org

Pearson, J. (1998) Terms in Context. Amsterdam/Philadelphia: John Benjamins.

Sager, J.C. (1990) A Practical Course in Terminology Processing. Amsterdam/Philadelphia: John Benjamins.

Temmerman, R. (2000) Towards New Ways of Terminology Description: The Sociocognitive Approach. Amsterdam/Philadelphia: John Benjamins.

TermCoord (2022) IATE—InterActive Terminology for Europe: Annual Report 2022. European Parliament.

Wüster, E. (1979) Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie. Bonn: Romanistischer Verlag.