ISSN 1799-2591 (print); ISSN 2053-0692 (online)
Theory and Practice in Language Studies, Vol. 16, No. 8, pp. 2569-2581, August 2026
DOI: https://doi.org/10.17507/tpls.1608.03

English to Arabic Translation of Medical Texts: A Review of Research (2015-2025)

Yara Alhassan Alahmed*
School of Languages, Literacies and Translation, Universiti Sains Malaysia, Penang, Malaysia

Mohamed Abdou Moindjie
School of Languages, Literacies and Translation, Universiti Sains Malaysia, Penang, Malaysia

***Abstract*—Translating medical texts from English to Arabic is challenged by terminological inconsistency, register variation, and uneven methodological practices, with direct implications for patient safety and medical education in Arabic-speaking countries. This study adopted a systematic literature review methodology guided by the PRISMA framework. Five major databases—Scopus, ERIC, SpringerLink Nature, Nature, and BMC—were searched to identify relevant studies published between 2015 and 2025. A total of 22 studies were included and analysed to map publication trends, translation procedures, the use of digital technologies, and implications for medical education. The findings show that most research has focused on instrument development, with particular attention to patient safety and the translation of clinical instruments and leaflets. The analysis further reveals that only a very limited number of studies have examined the use of technology and artificial intelligence in this field. The review suggests that AI-based tools, when used alongside human-led translation workflows, may help expand access to high-quality medical translations for Arabic-speaking populations.**

***Index Terms*—medical translation, machine translation, AI in translation, English–Arabic translation**

I. Introduction

Translating medical terminology from English into Arabic is a high-stakes task with direct clinical, educational, and ethical implications. In contemporary healthcare systems, accurate translation is essential not only for patient safety but also for effective research communication, medical education, and equitable access to care. Demand has increased as Arabic-speaking populations rely more heavily on bilingual medical records, clinical trial documentation, and patient-facing materials. Despite its importance, this practice is constrained by linguistic, cultural, and technical factors that may compromise clarity and accuracy. These challenges are particularly pronounced in Arabic due to the coexistence of Modern Standard Arabic and regional variants, as well as the absence of fully unified medical terminology across Arabic-speaking healthcare systems.

A central challenge lies in terminological inconsistency and ambiguity. Many medical terms are polysemous or used variably across subfields, increasing the risk of misinterpretation. For example, serious mistranslations, such as rendering *myocardial infarction* or *dementia* as symptoms rather than diagnoses, illustrate the clinical consequences of terminological inaccuracy (Almahasees & Husienat, 2024). Competing labels for the same medical concept further hinder knowledge transfer, affecting both researchers and healthcare professionals (Martin et al., 2022). Beyond terminology, cultural mediation complicates the task. Translators must balance scientific precision with culturally appropriate communication, and failure to achieve this balance may result in significant clinical consequences (Ødemark & Engebretsen, 2022; Song, 2024; Latif, 2020; Farahat, 2025).

Arabic's morphological richness and syntactic flexibility introduce additional layers of complexity. Errors may arise when tense, voice, or syntactic scope are misinterpreted, particularly in dense medical discourse. These challenges are often amplified in machine-translation output, where systems struggle to maintain semantic and syntactic coherence in specialised contexts (ElSabagh et al., 2025). The extensive use of abbreviations and acronyms without explicit expansion presents further risks, especially when overlapping meanings exist across medical specialties (Bell & Torres, 2024; Mohammed, 2022). Hasani et al. (2025) show that avoiding abbreviations in AI-generated summaries improves doctor–patient communication.

Without systematic training in terminology management, such ambiguity may compromise patient safety.

The consequences of mistranslation are well documented. Inaccuracies in dosage instructions, diagnostic descriptions, or symptom reporting can directly endanger patient outcomes (Bell & Torres, 2024; Song, 2024). Although machine-translation tools such as Google Translate and Microsoft Copilot offer speed and accessibility, they lack sufficient domain sensitivity and contextual awareness for high-stakes medical use (Al-Jarf, 2024). These risks may be amplified in English–Arabic translation, where structural differences, morphological complexity, and terminology variation increase the likelihood of inaccurate or ambiguous outputs. Studies on AI-assisted interpretation suggest potential benefits in limited

* Corresponding Author.

contexts but highlight the need for continuous human oversight to mitigate risks related to literal translation, dialectal variation, and ethical accountability (Brandenberger et al., 2025). Hybrid approaches combining translation memories and example-based methods have shown efficiency gains; however, they remain limited when applied to specialised medical content (Amini et al., 2024; Fan & Zhang, 2025).

To address these challenges, existing scholarship highlights the importance of robust quality-assurance mechanisms and sustained professional development. Peer review, systematic validation, and continuous training are widely recognised as essential for maintaining translation reliability (Williams, 2024; Mohammed, 2022). A domesticating approach that adapts terminology to culturally familiar expressions while preserving clinical accuracy has been proposed to enhance patient comprehension (Yateem et al., 2023). In addition, recent research advocates the development of standardised Arabic medical-terminology resources to reduce inconsistency and promote safer communication across healthcare settings (Yateem et al., 2023; Almahasees & Almahasees, 2025).

Despite the urgency of the issue, research on English–Arabic medical translation remains fragmented. Many studies focus on isolated terms, specific instruments, or individual technological tools, while relatively few offer a systematic account of translation procedures across different types of medical texts. This lack of synthesis limits the development of evidence-based guidelines for translators and healthcare institutions.

The present study addresses this gap by systematically reviewing research on English–Arabic medical translation published between 2015 and 2025. It synthesises evidence on terminological, cultural, and technological challenges, examines the application of translation procedures, and evaluates the roles and limitations of human- and machine-assisted translation.

II. Literature Review

A. Theoretical Framework in Translation Studies

Medical translation is guided by established frameworks in translation studies that explain how meaning and function transfer between languages. Structural differences between English and Arabic require careful procedural and linguistic decisions. Vinay and Darbelnet (1995) identify procedures such as borrowing, transposition, and modulation to address structural differences effectively. Newmark (1988) distinguishes semantic translation, which prioritises accuracy, and communicative translation, which prioritises clarity for readers. House (2015) emphasises register equivalence across field, tenor, and mode, providing a basis for evaluating whether translations preserve communicative function and meaning in specialised medical contexts.

B. Growing Demand for Medical Translation

The demand for accurate English–Arabic translation of medical texts is increasing, particularly for patient records, clinical trials, and clinical instruments. International agencies have acknowledged both the opportunities and the risks in this domain. The World Health Organization (WHO, 2025) highlights the potential of artificial intelligence (AI) and large language models (LLMs) in clerical and administrative tasks, including electronic health record entries, drafting visit notes, and simplifying jargon for patient-friendly communication. At the same time, WHO warns that "serious errors could arise" from inaccuracies or hallucinations, stressing that such functions should not be fully automated and that human oversight remains essential.

C. Accuracy and Reliability of Machine Translation

Empirical studies show mixed results in the use of machine translation (MT) in healthcare. In emergency department materials, Google Translate preserved meaning in 82.5% of short instructions, but performance varied widely across languages, reinforcing warnings against relying on MT alone for patient-critical text (Taira et al., 2021). A comparative study of discharge instructions reported 90% or higher sentence-level accuracy for Spanish and Chinese but lower accuracy for Russian, with the authors recommending disclosure, human review, and provision of the original-language text as part of best practices in clinical education (Kong et al., 2025). Accuracy drops further for low-resource languages, such as Haitian Creole, raising concerns about equity and access (Brewster et al., 2024).

For plain-language patient education pamphlets, MT scored above 3 out of 5 in many domains, yet consistently underperformed relative to professional translation, supporting calls for curriculum modules on plain-language authoring and post-editing (Ugas et al., 2025). A clinical systematic review likewise advises that AI be used as a supplement, often as a last resort, and never as a replacement, aligning with training in ethics and professional values (Genovese et al., 2024).

D. Translation and Validation of Clinical Instruments

Patient-reported outcome measures (PROMs) are a cornerstone of standardised, patient-centred evaluation, and their translation requires rigour. Jamjoom et al. (2023) document the widespread use of translated, adapted, and validated PROMs in Arab spine research. Sawaqed and Suoub (2021) confirm the reliability of the Overactive Bladder Symptoms questionnaire in tracking treatment response across adult patients. Almohaya et al. (2024) highlight the importance of rigorous workflows, including forward–back translation, expert review, and cognitive debriefing, in Arabic PROM adaptations to ensure conceptual equivalence and cultural acceptability. These examples illustrate the level of methodological discipline now expected in the translation of clinical instruments and trial outcomes.

E. Lexical and Terminological Challenges

Terminological accuracy is among the most persistent challenges in English–Arabic translation of medical texts. Ambiguous terms and acronyms often create inconsistency, with recommendations to expand acronyms at first mention to reduce ambiguity (Bell & Torres, 2024; Almahasees & Husienat, 2024). Translators require biliterate competence, search skills, and familiarity with domain-specific technology to manage uncommon acronyms (Bell & Torres, 2024).

In safety-critical contexts, mistranslations of diagnoses, unclear dosing, or inconsistent abbreviations can lead to adverse outcomes (Mohammed, 2022; Almahasees & Husienat, 2024). In Jordanian healthcare centres, Almahasees and Almahasees (2025) found that inconsistent physiotherapy terminology, particularly in cardiopulmonary and musculoskeletal subdomains, confused patients and undermined adherence. Interviews with physiotherapists confirmed that inconsistency weakens patient education and engagement, strengthening the case for a standardised Arabic physiotherapy lexicon. These challenges are particularly significant in Arabic-speaking contexts, where variation between Modern Standard Arabic and regional usage, inconsistent terminology development, and limited standardisation across healthcare systems increase the risk of ambiguity and misunderstanding in medical communication.

F. Syntactic and Stylistic Issues

Beyond terminology, English and Arabic diverge significantly in structure and style. Translators must account for differences in clause structure, information packaging, and verbal versus nominal preference. Sharkas (2024) shows that nominalisation in Arabic translations of patient-information leaflets serves three functions: signalling formality, depersonalising agency, and increasing abstractness. These functions align with genre conventions in the target culture and highlight the need for comparative corpus studies of Arabic expert–lay and institutional–user genres.

Syntactic ambiguity in English, such as multi-parse sequences, can yield divergent Arabic realisations, and dynamic equivalence often produces clearer translations than strict correspondence (Sharkas, 2024). Medical texts also reflect cultural-linguistic conventions. For example, drug inserts address both clinicians and laypersons, requiring careful adjustment of technicality without loss of accuracy (Khayyal, 2022). In Arabic medical texts, these structural differences directly influence clarity, authority, and accessibility, making register-appropriate choices essential for ensuring effective communication between healthcare providers and Arabic-speaking patients.

G. Roles and Limits of Machine vs. Human Translation

Machine translation offers efficiency and consistency in standardised contexts, but its limitations are widely recognised. Research emphasises its role as a supportive tool rather than a replacement, particularly in high-stakes scenarios such as consent or individualised counselling (Fan & Zhang, 2025; Adiel et al., 2023). Human oversight remains indispensable, with deep post-editing identified as essential and pre-editing as a useful but secondary step (Matheson & Kollmer, 2024; Zappatore & Ruggieri, 2024).

Asynchronous document translation in well-resourced language pairs, such as English–Spanish or English–Chinese, can be effective with post-editing, but synchronous or live communication requires greater caution and restriction to lower-risk content (Zappatore & Ruggieri, 2024). Polysemy further complicates translation, as MT often misses subtle sense distinctions (Attieh, 2024; İsmayilli, 2024; Adiel et al., 2023). Automated quality assurance tools are not sufficient substitutes for human expertise. For instance, Matheson and Kollmer (2024) show how a correct translation (“27 years of age”) was wrongly flagged as a numerical error, illustrating how tools without contextual understanding can mislead.

H. Emergence of AI in the Translation of Medical Texts

Healthcare AI translation has evolved from the ad hoc use of broad machine translation to targeted evaluations of large language models and neural MT, with growing emphasis on safety, governance, and educational deployment. Early pragmatic studies on Google Translate (GT) for emergency department discharge instructions reported mean adequacy of approximately 82.5% across seven languages, but with wide language variability and significant risks to meaning fidelity, emphasising MT as an adjunct to professional translation within clinical contexts (Taira et al., 2021). Medication instruction research identifies the highest-risk surface: GT accuracy declined sharply for counselling points (for example, Arabic 54.1% and Spanish 38.2%), and approximately 29% of incorrect outputs were extremely clinically significant or potentially fatal, requiring close human surveillance for drug-related information (Cornelison et al., 2021).

Head-to-head comparisons with LLMs show incremental improvement but continued safety limitations. In paediatrics, ChatGPT-4 and GT were equivalent to human translation in Spanish and Portuguese discharge instructions but fell short of professional translators for Haitian Creole, with substantially fewer clinically meaningful errors, illustrating the equity gap for low-resource languages (Brewster et al., 2024). A more general comparison of GPT-4 to GT in patient-specific emergency department instructions reported generally high sentence-level accuracy (for example, 97% for Spanish, 95% for Chinese, and 89% for Russian), but potentially unsafe mistranslations ranged from ≤1% at the sentence level to up to 6% when assessed across full instruction sets, where a single error can be clinically significant (Kong et al., 2025). Implied safety measures include disclosure of MT or LLM usage, presentation of the original-language output, human evaluation before use, avoidance of jargon, secure handling of data, and teach-back to patients (Kong et al., 2025).

Syntheses of the field confirm the overall pattern: neural MT dominates, LLMs widen the toolkit, but proper deployment is contingent upon in-domain corpora, controlled language, post-editing, and task-based safety checks (Zappatore & Ruggieri, 2024). Broader assessments of ChatGPT and LLMs in healthcare confirm these assertions, noting

their utility for drafting, summarising, and translating, balanced by risks of hallucination, bias, and privacy issues, and thus the need for open governance and human oversight (Sallam, 2023).

I. Implications for English–Arabic Translation

The findings support a tiered approach: (i) allow supervised AI or MT for low-stakes, templated educational materials, with disclosure and post-editing; (ii) require obligatory human review for routine written instructions; and (iii) avoid AI-only use for high-stakes content, such as medication orders, consent forms, and individualised counselling. For Arabic specifically, where register differences and dialect variation complicate mapping, priority should be given to plain-language writing, terminology control through glossaries, and bilingual clinical verification before dissemination. This approach maximises access advantages while safeguarding patient safety and aligns with current governance guidance.

III. Methodology

This study adopts a systematic literature review (SLR) design to synthesise research on the translation of medical texts between English and Arabic. Systematic reviews are widely used in education and applied linguistics research to identify trends, evaluate practices, and map gaps in an emerging field. The review follows established principles of transparency, rigour, and replicability as outlined by Kitchenham (2004), Gough et al. (2012), and the PRISMA framework (Moher et al., 2009).

A. Research Questions

The review is guided by the following research questions:

RQ1. What are the emerging trends in the translation of medical texts from English into Arabic between 2015 and 2025 (year-wise, area of study, and domain of medical text)?

RQ2. How have translation procedures been applied in the translation of medical texts from English into Arabic in the past decade?

RQ3. What technologies have been used for translating medical texts from English into Arabic?

RQ4. What recommendations have been proposed for improving the translation of medical texts from English into Arabic?

B. Subjects / Corpus of the Review

As this study is a systematic literature review, the "subjects" are published research articles rather than human participants. The review corpus consists of peer-reviewed journal articles that address the translation of medical texts between English and Arabic or Arabic and English.

Following the application of inclusion and exclusion criteria, a total of 22 full-text studies published between 2015 and 2025 were identified as relevant and included in the final analysis. These studies represent diverse medical domains, translation procedures, and methodological approaches.

C. Instruments

The instruments used for this review include:

- Academic electronic databases: Scopus, Springer Nature, ERIC, Nature, and BMC
- Structured search strings developed using Boolean operators
- Clearly defined inclusion and exclusion criteria
- Title, abstract, and full-text screening procedures
- A data extraction framework for recording publication year, domain, translation procedures, technologies used, and recommendations

The search strings, including keywords and limiters, are presented in Table 1 under the "Search Terms" column.

Table 1
Results of Initial Search

Search Terms	Search Limiters	Databases	Frequency distribution of primary results
"Medical translation" OR 'medical discourse" OR "biomedical translation" AND "translation" OR "interpreting" OR "translation procedures" AND "machine translation" OR "AI translation" OR "post-editing" AND "English to Arabic" OR "Arabic to English"	1. Document type: Journal Articles 2. Language: English 3. Year Range: 2015- 2025	Scopus	45
		Springer Nature	13
		ERIC	7
		Nature	45
		BMC	86
		Total	196

D. Data Collection Procedures

Data collection followed a systematic multi-stage process. Boolean operators were used within each database to retrieve relevant studies. The initial search yielded 196 records across the five databases. After removing seven duplicate records, 189 studies were screened.

During title and abstract screening, 132 records were excluded because they were either unrelated to medical translation or did not involve the English–Arabic language pair. An additional 35 records were excluded due to closed access. Ultimately, 22 full-text articles met all eligibility criteria and were included in the review. Table 1 summarises the search strategy, databases, and frequency of retrieved records to enhance transparency and replicability, in line with established guidelines for systematic literature reviews (Kitchenham, 2004; Moher et al., 2009).

The selection process followed PRISMA guidelines, and the flow of study identification, screening, and inclusion is illustrated in Figure 1.

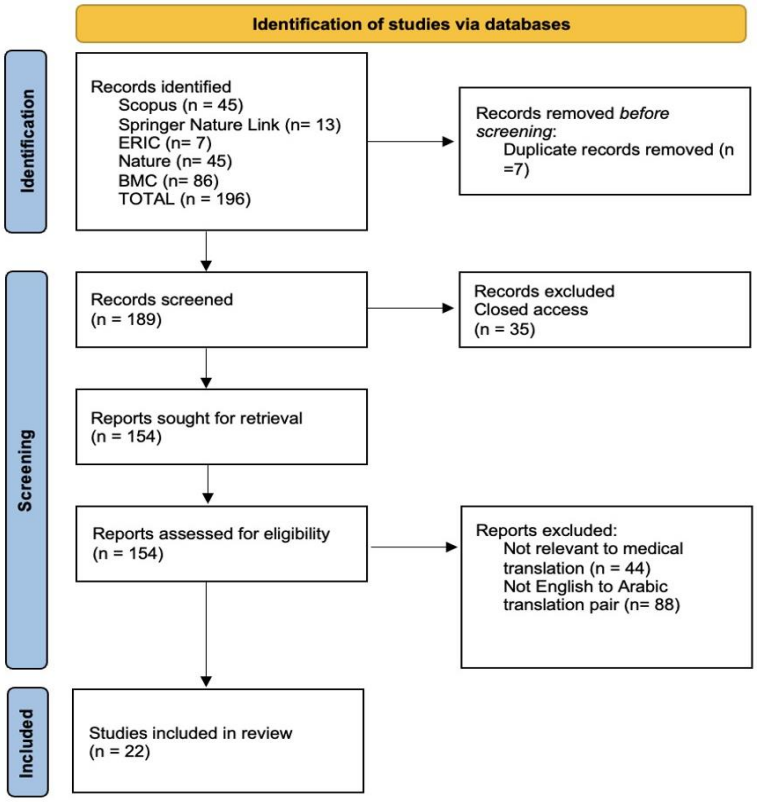

Figure 1. Prisma 2020 Flow Diagram

IV. Results and Discussion

To answer the research questions, an in-depth analysis of the included studies was conducted to identify patterns in publication trends, areas of study, domains of medical texts, translation procedures, and the use of technology in English–Arabic medical translation.

A. Year-Wise Distribution

The year-wise distribution of research studies was analysed to identify emerging trends in English–Arabic medical translation. The findings show that research activity has increased notably in recent years, with the highest number of studies published in 2025 (n = 5, 22.73%). A substantial number of studies were also published in 2021 (n = 4, 18.18%) and 2024 (n = 4, 18.18%), indicating sustained research interest during this period. The year 2023 accounted for three studies (13.64%), while 2022 and 2018 each recorded two studies (9.09%).

In contrast, very limited research output was observed in the earlier years of the review period, with only one study published in 2015 (4.55%) and one in 2016 (4.55%). This uneven distribution suggests that English–Arabic medical translation has gained stronger scholarly attention only in recent years, particularly alongside growing concerns about patient safety, instrument validation, and technology-assisted translation. Figure 2 provides a graphical representation of the year-wise distribution of the reviewed studies.

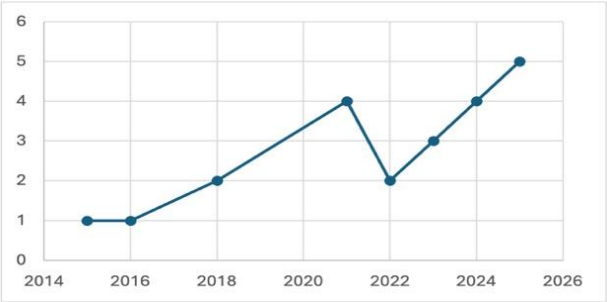

Figure 2. Year-Wise Distribution

B. Area of Study

Analysis of the reviewed articles revealed five major areas of study addressed in English–Arabic medical translation research. As shown in Table 2, infectious diseases and public health (n = 5, 22.73%) and paediatrics and nursing care (n = 5, 22.73%) constituted the largest thematic clusters. These studies focused on areas such as COVID-19 communication, falls, overactive bladder, foot health, paediatric care, and nursing practice, reflecting strong links between translation research and applied healthcare needs.

TABLE 2
AREA OF STUDY

	Area of Study	Number of Articles	Topics covered
1	Chronic & Non-Communicable Diseases	4	Diabetes, Cancer, Osteoarthritis, Tumours
2	Mental & Developmental Health	4	Mental Health, Autism, Disability
3	General Medicine & Health Terminology	4	Medicine (General), Medical Terminology, Nutrition
4	Infectious Diseases & Public Health	5	COVID-19, Falls, Overactive Bladder, Foot Health
5	Paediatrics & Nursing Care	5	Paediatrics, Nursing

Chronic and non-communicable diseases (n = 4, 18.18%) also received considerable attention, particularly in relation to diabetes, cancer, osteoarthritis, and tumour-related care. Similarly, mental and developmental health (n = 4, 18.18%) emerged as a key area, covering the translation of assessment tools for mental health, autism, and disability. General medicine and health terminology (n = 4, 18.18%) addressed broader issues such as terminology consistency, nutrition, and general medical discourse. Figure 3 illustrates the distribution of studies across these areas, highlighting the breadth of domains in which English–Arabic medical translation research is applied.

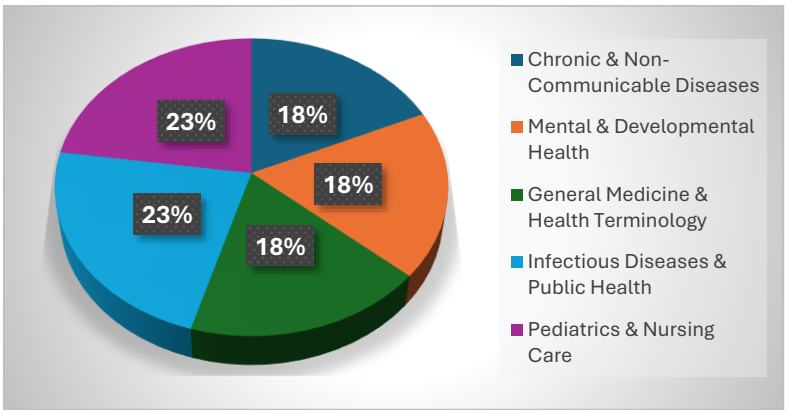

Figure 3. Area of Study

C. Domain of Medical Texts

To further understand the research focus, the domains of medical texts translated from English into Arabic were examined in detail. As presented in Table 3, instrument validation studies dominated the field, accounting for 14 of the included articles (68%). These studies primarily focused on the translation, cultural adaptation, and validation of questionnaires, scales, and clinical instruments for use with Arabic-speaking populations.

TABLE 3
DOMAIN OF MEDICAL TEXTS TRANSLATED FROM ENGLISH TO ARABIC

Domain	Number of Articles	Citation
Instrument Validation Studies	14	(Batais et al., 2021)
		(Halaweh et al., 2016)
		(Ziouziou et al., 2018)
		(Alajmi & Al-Ghamdi, 2021)
		(Sawaqed & Suoub, 2021)
		(Awad et al., 2025)
		(Obeid et al., 2024)
		(Masoud & Mosli, 2023)
		(Alshammari et al., 2023)
		(Amri et al., 2022)
		(Alshaygy et al., 2025)
		(Elboim-gabyzon et al., 2015)
		(Maryem et al., 2023)
		(Alharbi et al., 2024)
AI & Technology in Medical Translation/Interpretation	3	(Qiu et al., 2024)
		(Brandenberger et al., 2025)
		(Gilad et al., 2025)
Public Health Communication & Health Literacy	2	(Tjaden et al., 2022)
		(Halboub et al., 2021)
Standardization & Cross-Cultural Screening	4	(Halboub et al., 2021)
		(Huda et al., 2024)
		(Halaweh et al., 2016)
		(Al-Jarf, 2018)

Other domains were represented to a lesser extent. Studies on AI and technology in medical translation or interpretation accounted for three articles (14%), while public health communication and health literacy were addressed in two studies (9%). Standardisation and cross-cultural screening were also examined in two studies (9%). Figure 4 presents a graphical overview of this distribution. It is worth noting that one study by Halboub et al. (2021) addressed two domains simultaneously, focusing on both public health communication and standardisation and cross-cultural screening.

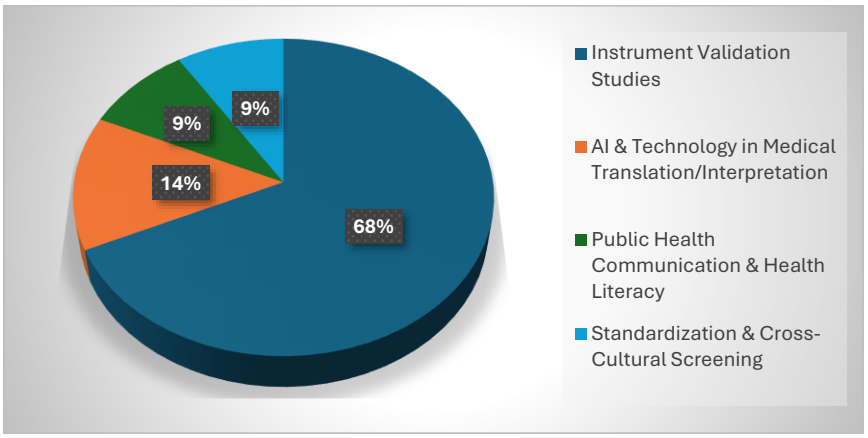

Figure 4. Domain of Medical Texts Translated From English Into Arabic

These findings indicate a strong research emphasis on producing reliable Arabic versions of clinical instruments, reflecting practical healthcare needs across Arabic-speaking contexts. This reflects the specific needs of Arabic-speaking healthcare systems, where validated Arabic instruments are essential for reliable clinical assessment and patient communication.

D. Translation Procedures

A range of translation procedures was identified across the reviewed studies. As shown in Table 4, forward–back translation (FBT) was the most frequently employed procedure (n = 13), confirming its role as the standard approach for the cross-cultural adaptation of medical instruments. Expert or committee review was explicitly reported in six studies, often as a complementary step to strengthen content validity.

TABLE 4
TRANSLATION PROCEDURES ADOPTED

Translation Procedures	Number of Articles	citation
Forward + Back translation (FBT)	13	(Batais et al., 2021) (Halaweh et al., 2016) (Ziouziou et al., 2018) (Alajmi & Al-Ghamdi, 2021) (Alzeer et al., 2025) (Sawaqed & Suoub, 2021) (Awad et al., 2025) (Obeid et al., 2024) (Alshammari et al., 2023) (Amri et al., 2022) (Alshaygy et al., 2025) (Elboim-gabyzon et al., 2015) (Maryem et al., 2023)
Expert/committee review (explicit mention)	6	(Masoud & Mosli, 2023) (Amri et al., 2022) (Huda et al., 2024) (Elboim-gabyzon et al., 2015) (Maryem et al., 2023) (Alharbi et al., 2024)
Not mentioned	3	(Tjaden et al., 2022) (Halboub et al., 2021) (Huda et al., 2024)
Back-translation only (reported standalone)	3	(Halaweh et al., 2016) (Ziouziou et al., 2018) (Alajmi & Ghamdi, 2021)
Cognitive/face validation	2	(Masoud & Mosli, 2023) (Alharbi et al., 2024)
Linguistic translation (unspecified method)	1	(Gilad et al., 2025)
Machine/AI tool named	1	(Qiu et al., 2024)
LLM/AI (general)	1	(Brandenberger et al., 2025)
Terminology strategies (not a workflow)	1	(Al-Jarf, 2018)

Three studies relied mainly on back-translation without further methodological detail, while three did not specify the translation procedure used. Less commonly reported approaches included cognitive or face validation (n = 2), general linguistic translation (n = 1), terminology-based strategies such as borrowing or transliteration (n = 1), and the use of machine translation or large language models (n = 2). Figure 5 provides a graphical overview of the distribution of translation procedures.

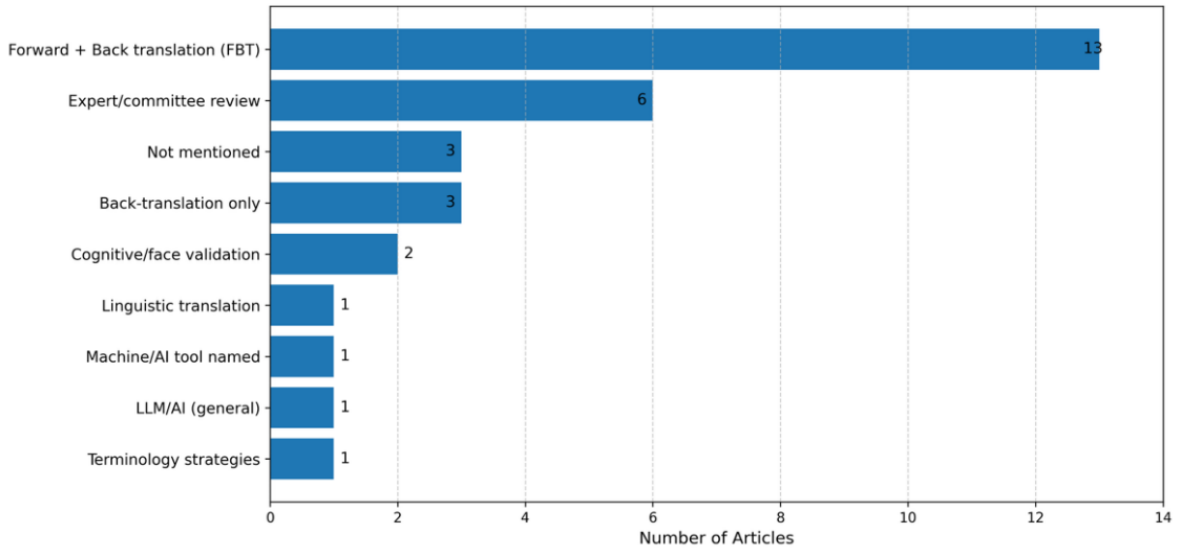

Figure 5. Translation Procedures

These findings demonstrate strong alignment with internationally recognised cross-cultural adaptation and psychometric frameworks. The inclusion of expert review and cognitive validation reflects increasing awareness of the need to ensure conceptual equivalence and semantic accuracy, rather than relying solely on literal translation. At the same time, variation in reporting practices suggests that translation workflows are often adapted flexibly to local contexts and available resources within Arabic-speaking research settings.

Taken together, many studies repeat similar purposes and procedures, largely relying on standard FBT-based adaptation workflows for instrument validation across medical domains.

E. Technology Used for Translation

Only two studies explicitly examined the use of technology in English–Arabic medical translation. One study focused on the development and evaluation of multilingual medical large language models (Qiu et al., 2024), highlighting their potential to support cross-lingual understanding of complex medical texts. The second study examined the use of Google Translate in non-urgent paediatric emergency consultations through staged, low-acuity clinical interactions, in which clinicians conducted simulated consultations and a professional interpreter acted as the caregiver (Brandenberger et al., 2025). The analysis combined performance testing with a detailed legal and policy review, identifying generally high accuracy in controlled contexts while also drawing attention to challenges related to dialect variation, pronoun use, and broader medico-legal issues, such as liability, consent, and ethical responsibility.

Overall, these two studies represent 9.1% of the reviewed literature. While limited in number, they provide important insights into both the potential and the constraints of AI-based tools in Arabic medical translation. The findings suggest that research on translation technology in this context remains exploratory, with cautious adoption driven by concerns related to linguistic complexity, patient safety, privacy, and legal accountability. Table 5 summarises the translation technologies identified in the reviewed studies and their frequency of use.

TABLE 5
TECHNOLOGY USED FOR TRANSLATION

Technology Used	Number of Articles	citation
Multilingual medical large language models (LLMs)	1	Qiu et al. (2024)
Google Translate / AI-based interpretation	1	Brandenberger et al. (2025)

F. Critical Interpretation of Trends

The observed trends reflect the strong influence of clinical priorities on English–Arabic medical translation research. The dominance of instrument validation studies (n = 14) indicates that translation is primarily driven by the need to adapt clinical tools for Arabic-speaking populations. This reflects the essential role of translated instruments in diagnosis and patient evaluation. However, this emphasis has limited attention to linguistic and procedural aspects of translation. Although forward–back translation was widely used (n = 13), few studies examined how translation procedures affect meaning or clarity in Arabic medical texts. Terminology-focused research remains limited (n = 4), despite persistent inconsistency in Arabic medical terminology. Similarly, the relatively small number of AI studies (n = 6) reflects ongoing safety and legal concerns. Overall, these patterns highlight the need for stronger theoretical grounding and greater focus on Arabic-specific translation challenges.

V. CONCLUSION

This systematic review offers the first comprehensive mapping of English–Arabic medical translation published between 2015 and 2025, bringing together evidence on translation practices, domains of application, and emerging technological developments. The findings show that the field continues to rely primarily on traditional human-led workflows, particularly forward–back translation, expert panel review, and cognitive validation. This reflects the ongoing need for accuracy, cultural appropriateness, and clinical safety in medical communication. Despite this methodological stability, the persistence of terminological inconsistency across medical subfields remains a major barrier to consistent and reliable communication. Divergent Arabic equivalents, inconsistent use of technical terms, and the absence of unified glossaries point to an urgent need for stronger standardisation efforts across healthcare institutions. This review also contributes to translation studies by providing the first systematic synthesis of English–Arabic medical translation research across multiple domains, including clinical instruments, terminology, and emerging AI technologies. By situating translation procedures within both clinical and linguistic contexts, the study clarifies the structural and terminological factors shaping translation practices and identifies priorities for future theoretical and empirical research.

Although artificial intelligence is increasingly discussed in the broader medical translation landscape, its adoption in English–Arabic medical translation remains very limited. Only two studies explicitly examined technological tools, and their findings confirm that while AI and machine translation may support low-risk or administrative tasks, they remain immature for high-stakes clinical communication. Challenges such as dialectal sensitivity, contextual misinterpretation, and medico-legal risks underline the importance of cautious and supervised implementation. These insights highlight the need to integrate technology thoughtfully within established human-led workflows rather than viewing it as a replacement.

From a practical and clinical perspective, the review underscores the need for clearer guidelines on translation procedures, expanded psychometric testing across diverse Arabic-speaking populations, and stronger collaboration between linguists, clinicians, and healthcare educators. The development of standardised medical terminology resources and the inclusion of translation awareness in medical and allied health training can further strengthen communication and patient safety. Overall, the evidence points to growing recognition of the importance of reliable Arabic medical translation,

alongside a clear need for more targeted research, richer methodological reporting, and careful integration of emerging technologies to support safe, consistent, and culturally grounded healthcare communication.

A. Implication of the Study

From a practical standpoint, the results of this review carry important implications for improving English–Arabic medical translation across clinical practice, education, and policy. The dominance of instrument-validation studies indicates that Arabic-speaking healthcare systems rely heavily on translated clinical tools to support diagnosis, screening, and patient follow-up. This pattern highlights the need for continued development and refinement of Arabic versions of established medical instruments, ensuring that healthcare providers can generate accurate and culturally appropriate patient data. Healthcare institutions are therefore encouraged to adopt validated instruments whenever available and to follow standardised workflows when adapting new tools.

Persistent challenges related to medical terminology also emerge as a key concern, particularly variation across countries, specialties, and even within the same research field. This observation underscores the importance of establishing unified medical terminology resources that can support clinicians, translators, and researchers alike. Shared glossaries and terminology guidelines can reduce confusion, improve clarity, and strengthen communication across Arabic medical texts in different regions.

Regarding translation technologies, only a small number of studies addressed their use, yet their findings point to the growing relevance of AI in this area. While AI and machine translation tools may offer useful support in low-risk situations or educational settings, human oversight remains essential in clinical contexts. This reinforces the need for clear institutional policies defining when such technologies can be used, how accuracy should be verified, and what safeguards are required to protect patient safety.

At a broader systemic level, the review points to the need for coordinated national and regional frameworks that support high-quality translation of medical materials. Establishing clear standards, reviewing translation workflows, and encouraging collaboration between translators and healthcare professionals can contribute to a more unified and accessible healthcare environment for Arabic-speaking populations.

B. Recommendations

Building on the findings of this review, several priorities emerge for advancing English–Arabic medical translation research and practice. Greater collaboration between translators, clinicians, and researchers is needed to support the development and maintenance of shared terminology resources and updated translation guidelines. Coordinated initiatives at national and regional levels would help promote consistency and improve the quality of translated medical material.

Future research should extend beyond the translation and validation of clinical instruments to include a wider range of medical genres. Greater attention to consent forms, clinical guidelines, digital health systems, and patient education materials would provide a more comprehensive understanding of translation challenges across real healthcare contexts.

Stronger methodological transparency is also needed. Studies should clearly document translation procedures, validation steps, and decision-making processes. Multi-centre and cross-country research designs would further strengthen the generalisability of findings and allow comparison across different Arabic-speaking settings.

The increasing availability of machine translation and AI tools also creates new research opportunities. Future studies should evaluate these technologies specifically in English–Arabic medical contexts, with attention to accuracy, safety, and appropriate clinical use. Clear evaluation frameworks will help determine where such tools can be used safely and effectively.

Finally, sustained investment in professional training is essential. Expanding training opportunities in medical terminology, translation methods, and technology use will help prepare translators and healthcare professionals to meet the growing demand for accurate and reliable Arabic medical translation.

REFERENCES

[1] Adiel, M., Elsadig, M. A., Altigani, A., Mohamed, Y. A., Ahmed, B. E. S., & Elhassan, S. M. O. (2023). Accuracy and problems of machine-based translation in contrast to human-based translation when rendering health awareness texts versus poetry texts. *Academic Journal of Interdisciplinary Studies.* https://doi.org/10.36941/ajis-2023-0108

[2] Al-Jarf, R. (2018). Multiple Arabic equivalents to English medical terms: Translation issues. *International Linguistics Research*, *1*(1), 102–110. https://doi.org/10.30560/ilr.v1n1p102

[3] Al-Jarf, R. (2024). Artificial intelligence and its practical applications in the digital economy. In *Artificial Intelligence and Its Practical Applications in the Digital Economy* (pp. 220–235). Springer. https://doi.org/10.1007/978-3-031-71429-0_17

[4] Alajmi, M., & Al-Ghamdi, S. (2021). Translation and validation of the Arabic version of the osteoarthritis quality of life questionnaire (OAQoL) in Saudi patients with osteoarthritis. *Health and Quality of Life Outcomes*, *19*, Article 193. https://doi.org/10.1186/s12955-021-01741-9

[5] Alharbi, K., Alamri, A. A., & Gassas, R. (2024). Translation, cultural adaptation, and validation of the Arabic version of the Student Evidence-Based Practice Questionnaire (S-EBPQ). *BMC Medical Education*, *24*, Article 312. https://doi.org/10.1186/s12909-024-06259-4

[6] Ali, A. F., Alhebshi, S. H. S., & Alaqad, M. H. (2025). Evaluation of AI medical interpretation. *LinguAmerta*, *2*(1), 1–8. https://doi.org/10.32722/jla.v2i1.7821

[7] Almahasees, R., & Almahasees, Z. (2025). Breaking language barriers in physiotherapy: Ensuring consistent English-to-Arabic translations for better patient care. *Scientific Reports*, *15*, Article 4209. https://doi.org/10.1038/s41598-025-90765-y
[8] Almahasees, Z., & Husienat, I. (2024). A comparative analysis of terminological inconsistency in scientific translation from English into Arabic across different medical fields. *Russian Journal of Linguistics*, *8*(3), 25–40. https://doi.org/10.22363/2521-442X-2024-8-3-25-40
[9] Almohaya, H. S., Bakhsh, H. R., Sheeha, B. B., Aldhahi, M. I., & Alhasani, R. (2024). Translation and cultural adaptation into Arabic of PROMIS measures. *Healthcare*, *12*, Article 211. https://doi.org/10.3390/healthcare12020211
[10] Alshammari, S., Alshwieer, M. A. M., Dammas, S. S., Alrasheed, A. M., Alasmari, M. A., Mohammed, M., & Alahmari, A. (2023). Arabic translation, cross-cultural adaptation, and validation of Foot Health Status Questionnaire among Saudi individuals with plantar fasciitis. *Journal of Orthopaedic Surgery and Research*, *18*, Article 42. https://doi.org/10.1186/s13018-023-04202-9
[11] Alshaygy, I., Alageel, M., Alaseem, A., Alogayyel, M., Griffin, A., Shaheen, M., Alsudairi, A., Murrad, K., Alqahtani, S., Alqahtani, R., Alawami, F., Algarni, N., Albishi, W., & Aljassir, F. (2025). Translation and cross-cultural adaptation of the Toronto extremity salvage score system into Arabic and its validity. *Journal of Orthopaedic Surgery and Research*, *20*, Article 715. https://doi.org/10.1186/s13018-025-05980-0
[12] Alzeer, H. M., Alkhunein, S. M., Alakeel, S. A., Alhumaidan, O., Elbarazi, I., & Aldhwayan, M. (2025). Validity and reliability of the Arabic national nutrition plans checklist. *Scientific Reports*, *15*, Article 5362. https://doi.org/10.1038/s41598-025-89928-8
[13] Amini, M., Ravindran, L., & Lee, K.-F. (2024). Implications of using AI in translation studies: Trends, challenges, and future directions. *Translation Studies Review*, *6*(1), 740–754.
[14] Amri, M. I., Alzhrani, M. M., Alanazi, A. D., Alqahtani, M. M., & Kashoo, F. Z. (2022). Cross-cultural adaptation and validation of the Arabic version of the foot function index in patients with chronic lateral ankle instability. *Journal of Foot and Ankle Research*, *15*, Article 21. https://doi.org/10.1186/s13047-022-00527-6
[15] Arraji, M., Iderdar, Y., Mourajid, Y., Guennouni, M., Boumendil, K., Korrida, A., El Khoudri, N., Ifleh, M., Khalis, M., & Chahboune, M. (2023). Translation and validation of the General Medication Adherence Scale. *BMC Nursing*, *22*, Article 302. https://doi.org/10.1186/s12902-023-01316-9
[16] Attieh, A. A. S. B. (2024). Challenges of translating American and British forensic terms and texts into Arabic. *Journal of Translation Studies*, *11*, 1–20.
[17] Awad, E., Malaeb, D., Sakr, F., Dabbous, M., Iskandar, K., Khatib, S., Obeid, S., Hallit, S., & Fekih-Romdhane, F. (2025). Psychometric properties of the Arabic version of the eco-guilt and eco-grief scales. *Scientific Reports*, *15*, Article 11735. https://doi.org/10.1038/s41598-025-90765-y
[18] Batais, M. A., Alosaimi, F. D., AlYahya, A. A., Aloofi, O. A., Almashouq, M. K., Alshehri, K. S., & Alfraiji, A. F. (2021). Translation, cultural adaptation, and evaluation of the psychometric properties of an Arabic diabetes distress scale. *Saudi Medical Journal*, *42*(5), 509–516. https://doi.org/10.15537/smj.2021.42.5.20200286
[19] Bell, A., & Torres, P. A. (2024). Translating medical reports: Challenges and quality assurance. *The Linguist*, *33*(1), 30–32. https://doi.org/10.56012/aibj6031
[20] Brandenberger, J., Stedman, I., Stancati, N., Sappleton, K., Kanathasan, S., Fayyaz, J., & Singh, D. (2025). Using artificial intelligence-based language interpretation in non-urgent paediatric emergency consultations: A clinical performance test and legal evaluation. *BMC Health Services Research*, *25*, Article 123. https://doi.org/10.1186/s12913-025-12263-1
[21] Brewster, D., Jones, K., & Patel, S. (2024). Evaluating ChatGPT and Google Translate in clinical translation. *Journal of Medical Internet Research.* https://doi.org/10.2196/50012
[22] Cornelison, M., Brown, J., & Smith, T. (2021). Machine translation accuracy in medication instructions. *BMJ Open.* https://doi.org/10.1136/bmjopen-2021-056789
[23] Elboim-Gabyzon, M., Agmon, M., Azaiza, F., & Laufer, Y. (2015). Translation and validation of the Arab version of the Late-Life Function and Disability Instrument: A cross-sectional study. *BMC Geriatrics*, *15*, Article 51. https://doi.org/10.1186/s12877-015-0046-8
[24] ElSabagh, A. A., Azab, S. S., & Hefny, H. A. (2025). Neural computing and applications in medical language processing. *Neural Computing and Applications*, *37*(10), 7015–7048. https://doi.org/10.1007/s00521-025-11020-z
[25] Fan, R., & Zhang, Y. (2025). Role of AI in translation and interpretation. In *Advances in AI Translation Technologies* (pp. 95–118). IGI Global.
[26] Farahat, T. (2025). Cultural sensitivity in Arabic medical translation. *International Journal of Translation Studies*, *5*, 1–15.
[27] Genovese, A., Borna, S., Gomez-Cabello, C. A., Haider, S. A., Prabha, S., Forte, A. J., & Veenstra, B. R. (2024). Artificial intelligence in clinical settings: a systematic review of its role in language translation and interpretation. *Annals of Translational Medicine*, *12*(6), 16. https://doi.org/10.21037/atm-24-162
[28] Gilad, T., Lithwick Algon, A., Vakil, E., Kitany, S., Gharra, R., Higaze, R., Leavitt, V. M., & Saban, W. (2025). Breaking geographic and language barriers in neuropsychology: Online administration of MoCA in diverse populations. *NPJ Digital Medicine*, *8*, Article 379. https://doi.org/10.1038/s41746-025-01787-9
[29] Gough, D., Thomas, J., & Oliver, S. (2012). Clarifying differences between review designs and methods. *Systematic Reviews*, *1*, Article 28. https://doi.org/10.1186/2046-4053-1-28
[30] Halaweh, H., Svantesson, U., Rosberg, S., & Willén, C. (2016). Cross-cultural adaptation of Arabic Falls Efficacy Scale. *Medical Principles and Practice*, *25*(1), 1–7. https://doi.org/10.1159/000441128
[31] Halboub, E., Al-Ak'hali, M. S., Al-Mekhlafi, H. M., & Alhajj, M. N. (2021). Quality and readability of web-based Arabic health information on COVID-19: An infodemiological study. *BMC Public Health*, *21*, Article 531. https://doi.org/10.1186/s12889-021-10218-9
[32] Hasani, E., Richter, S., Juratli, T. A., Buszello, C. H., Prem, M. G., Willkommen, S., Sandi-gahun, S., Yasin, I., & Polanski, W. H. (2025). Layperson-Friendly AI Translation of Medical Documents to Improve Doctor – Patient Communication: Protocols

for the AI-INFOCARE and AI-MEDTALK Randomized Controlled Trials. *JMIR Research Protocols*, *14*. https://doi.org/10.2196/77204
[33] House, J. (2015). *Translation Quality Assessment*. Routledge. https://doi.org/10.4324/9781315752839
[34] Huda, E., Hawker, P., Cibralic, S., John, J. R., Hussain, A., Mendoza Diaz, A., & Eapen, V. (2024). Screening tools for autism in culturally diverse paediatric populations. *BMC Pediatrics*, *24*, Article 610. https://doi.org/10.1186/s12887-024-05067-5
[35] Jamjoom, A. B., Gahtani, A. Y., Alzahrani, M. T., Baydhi, L. M., Albeshri, Ahmad, S., & Sharab, M. (2023). Arabic-Translated Versions of Patient-Reported Outcome Measures Utilized in Spine Research : A Review of Validated Studies. *Cureus*, *15*(10), 1–11. https://doi.org/10.7759/cureus.46303
[36] Khayyal, O. (2022). Teaching and learning scientific translation: Problems, challenges, and solutions. *British Journal of Translation, Linguistics and Literature*, *2*(3), 38-48. https://doi.org/10.54848/bjtll.v2i3.38
[37] Kitchenham, B. (2004). *Procedures for performing systematic reviews.* Keele University.
[38] Kong, M., Fernandez, A., Bains, J., Milisavljevic, A., Brooks, K. C., Shanmugam, A., Avilez, L., Li, J., Honcharov, V., Yang, A., & Khoong, E. C. (2025). Evaluation of the accuracy and safety of machine translation of patient-specific discharge instructions: a comparative analysis. *BMJ Quality and Safety*, 1–9. https://doi.org/10.1136/bmjqs-2024-018384
[39] Martin, A. K., Green, T. L., McCarthy, A. L., Sowa, P. M., & Laakso, E.-L. (2022). Healthcare terminology confusion. *Journal of Multidisciplinary Healthcare*, *15*, 765–772. https://doi.org/10.2147/JMDH.S342197
[40] Masoud, M., & Mosli, M. (2023). Arabic translation and validation of pediatric sleep questionnaire. *BMC Pediatrics*, *23*, Article 50. https://doi.org/10.1186/s12887-022-03820-2
[41] Matheson, I., & Kollmer, B. (2024). Quality assurance in medical translation. *Medical Writing*, *33*(1), 34–36. https://doi.org/10.56012/qjes2321
[42] Mohammed, S. (2022). Inconsistency of translating medical abbreviations and acronyms into the Arabic language. *Russian Journal of Linguistics*, *6*(3), 67–77. https://doi.org/10.22363/2521-442X-2022-6
[43] Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097
[44] Newmark, P. (1988). *A Textbook of Translation.* Prentice Hall.
[45] Obeid, S., Hallit, R., Haddad, C., Hallit, S., & Hallit, N. (2024). Psychometric properties of the Arabic Stress Numerical Rating Scale. *Clinical Interventions in Aging*, *19*, 209–220. https://doi.org/10.2147/CIA.S401363
[46] Ødemark, J., & Engebretsen, E. (2022). Medical knowledge translation challenges. *Humanities and Social Sciences Communications*, *9*, Article 71. https://doi.org/10.1057/s41599-022-01088-6
[47] Qiu, P., Wu, C., Zhang, X., Lin, W., Wang, H., Zhang, Y., Wang, Y., & Xie, W. (2024). Multilingual language model for medicine. *Nature Communications*, *15*, Article 6263. https://doi.org/10.1038/s41467-024-52417-z
[48] Sallam, M. (2023). The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations. *Healthcare*, *11*(887), 1–34. https://doi.org/10.3390/healthcare11060887
[49] Sawaqed, F., & Suoub, M. (2021). Arabic linguistic validation of Overactive Bladder Symptom Score. *Scientific Reports*, *11*, Article 661. https://doi.org/10.1038/s41598-020-79974-9
[50] Sharkas, H. (2024). Nominalization in Arabic translation. *Translation & Interpreting*, *16*(1), 1–16. https://doi.org/10.12807/ti.116201.2024.a01
[51] Song, C. (2024). Ethical norms in medical translation technology. *Pacific International Journal*, *7*(5), 174–181. https://doi.org/10.55014/pij.v7i5.717
[52] Taira, B. R., Kreger, V., Orue, A., & Diamond, L. C. (2021). A Pragmatic Assessment of Google Translate for Emergency Department Instructions. *Journal of General Internal Medicine*, *36*(11), 3361–3365. https://doi.org/10.1007/S11606-021-06666-Z
[53] Ugas, M., Anna, M., Tan, J., Umakanthan, B., Hill, C., Tse, K., Cashell, A., Muraj, Z., Giuliani, M., & Papadakos, J. (2025). Evaluating the feasibility and utility of machine translation for patient education materials written in plain language to increase accessibility for populations with limited English proficiency. *Patient Education and Counseling*, *131*, 108560. https://doi.org/10.1016/j.pec.2024.108560
[54] Vinay, J. P., & Darbelnet, J. (1995). *Comparative Stylistics of French and English*. John Benjamins Publishing.
[55] Williams, H. (2024). Harmonising linguistic validation with AI: Precision, efficiency, and the human touch in patient-reported outcome translation. *The Linguist Magazine*, *33*(1), 66–69. https://doi.org/10.56012/lxmw9690
[56] World Health Organization. (2024). *Ethics and governance of artificial intelligence for health.* WHO.
[57] Yateem, N. A., Hijazi, H., Saifan, A. R., Ahmad, A., Deh, R. M., Alrimawi, I., Rahman, S. A., Subu, M. A., & Ahmed, F. R. (2023). Quality and safety issue: language barriers in healthcare, a qualitative study of non- Arab healthcare practitioners caring for Arabic patients in the UAE. *BMJ Open*, *13*, e076326. https://doi.org/10.1136/bmjopen-2023-076326
[58] Ziouziou, I., Touzani, M. A., Karmouni, T., El Khader, K., Koutani, A., & Andaloussi, A. I. (2018). Arabic translation of Bladder Cancer Index. *African Journal of Urology*, *24*(2), 93–98. https://doi.org/10.1016/j.afju.2017.09.007

Yara Alhassan Alahmed is a Ph.D. candidate at Universiti Sains Malaysia (USM), Penang, Malaysia. She earned her M.A. in Translation and Interpreting Studies from USM and holds a Diploma in Translation from Sana'a University. She also obtained a bachelor's degree in English Language and Translation from the University of Science and Technology, Yemen. Her research focuses on medical translation, translation and culture, literary translation, and translation quality assessment. Alongside her academic work, she is a CELTA-qualified English instructor. She worked as a freelance translator and an EFL instructor.

Mohamed Abdou Moindjie is a polyglot and holds a Ph.D. in Literary Translation. He is currently a Senior Lecturer in Translation Studies at the School of Languages, Literacies, and Translation, Universiti Sains Malaysia. From 2016 to 2018, he served as the Chairperson of the Department of Foreign Languages at Université des Comores. His research interests include translation studies, English studies, Arabic studies, French studies, and multilingualism.