

ONTOLOGY OF MACHINE TRANSLATION IN THE CONTEXT OF COMPUTATIONAL LINGUISTICS

 Aida Haouchi¹¹ University of Bejaia-Algeria
aida.haouchi@univ-bejaia.dz

Abstract: This study examines the ontological status of machine translation (MT) within computational linguistics, focusing on how Arabic-oriented MT systems formalize morphological, syntactic, semantic, and pragmatic knowledge, and asks to what extent such systems approximate human linguistic cognition. The study adopts a qualitative, comparative document- and system-analysis design. Four MT systems representing the principal architectural paradigms used for Arabic — rule-based (Golden El Wafi), online multilingual (ArabTran), statistical-to-neural (Google Translate), and generative AI-based (ChatGPT Translator) — were purposively selected and analyzed against a shared framework of linguistic-processing criteria (morphological analysis, syntactic parsing, semantic representation, and pragmatic/cultural adequacy) derived from the computational linguistics and Arabic NLP literature. The four systems differ systematically in how they operationalize linguistic knowledge. Rule-based systems encode explicit morphosyntactic rules and perform reliably on structural regularity but struggle with idiomatic and pragmatic meaning; neural and generative systems capture contextual and idiomatic nuance more fluently but remain dependent on training-data coverage and are comparatively opaque in how they represent linguistic structure. Across all four systems, Arabic's root-and-pattern morphology, the routine absence of diacritics in everyday text, and its syntactic flexibility remain the most persistent sources of translation ambiguity. Machine translation is best understood ontologically as a computational formalization of linguistic knowledge rather than a replication of human cognition. No single architecture currently resolves the full range of challenges Arabic poses, and human oversight remains indispensable, particularly for pragmatic, cultural, and literary content. The study offers a structured, criterion-based comparative framework for evaluating Arabic MT systems and reaffirms the necessity of sustained collaboration between linguists and computer scientists in Arabic natural language processing research, tool development, and pedagogy.

Keywords: Ontology; machine translation; computational linguistics; Arabic natural language processing; comparative system analysis; interdisciplinary collaboration

How to cite this article:

Haouchi, A. (2026). Ontology of machine translation in the context of computational linguistics. *Journal of Studies in Language, Culture, and Society (JSLCS)*, 9(2), 146-159.

¹ Corresponding author : Aida Haouchi Authors' ORCID ID: <https://orcid.org/0009-0004-2038-4280>

1. Introduction

Advances in computing have reshaped research methods across the humanities and sciences, and language study is no exception. Early investigations into how information, linguistic and non-linguistic, could be transmitted through mechanical and electronic systems trace to the cyberneticists of the 1940s, whose mathematical theory of communication (Shannon, 1948; Shannon & Weaver, 1964) supplied the formal apparatus — channels, signals, redundancy — later adapted by computational approaches to language. That mathematical foundation is distinct from, though foundational to, the linguistic project of formally representing meaning, which draws instead on the theories and methods of modern linguistics (Shtaytiyah, 2008; Chomsky, 1965).

Computational linguistics emerged as the discipline that applies artificial intelligence and formal methods to language processing (Shtaytiyah, 2008), bridging theoretical linguistics and computer science to build systems capable of analyzing and generating natural language. Its progress depends on close collaboration between linguists and engineers, since neither discipline alone can capture the full complexity of human language (El-Haj Saleh, 2007).

Arabic poses distinctive challenges for this enterprise. Its root-and-pattern morphology, flexible word order, and reliance on diacritics for full disambiguation complicate both parsing and machine translation (MT) (Ali, 1988; El-Haj Saleh, 2000). Early Arabic computational-linguistics work, exemplified by rule-based tools such as Golden El Wafi, embedded explicit morphosyntactic rules into translation software (El-Haj Saleh, 2000; Fattoum, 2024). More recent neural and generative systems, including Google Translate's transformer-based architecture and ChatGPT Translator, instead learn statistical regularities from large multilingual corpora, aiming to capture contextual and idiomatic nuance that rule-based systems typically miss (Bahdanau, Cho, & Bengio, 2015; Devlin, Chang, Lee, & Toutanova, 2019).

This study asks what these two families of systems reveal, jointly and separately, about the ontological status of machine translation: what kind of linguistic knowledge they formalize, how that formalization is achieved, and where its limits lie. Rather than treating MT as a settled technical artifact, the study approaches it as an evolving epistemic object shaped by the continuing negotiation between linguistic theory and computational method (El-Haj Saleh, 2007; Fattoum, 2024).

2. Methodology

2.1. Statement of the problem

Despite the growing everyday use of Arabic MT tools, the literature offers few structured, criterion-based comparisons of how different MT architectures handle the language's specific structural properties, and the ontological status of MT — what these systems formalize, and to what limits — remains under-theorized for Arabic. This study addresses that gap through a comparative analysis of four representative systems.

2.2. Research questions

RQ1. What is the ontological status of machine translation as a formalization of linguistic knowledge within computational linguistics?

RQ2. How do rule-based, hybrid, statistical/neural, and generative-AI architectures differ in their treatment of Arabic morphology, syntax, semantics, and pragmatics?

RQ3. Which structural features of Arabic — root-and-pattern morphology, diacritic omission, syntactic flexibility — pose the most persistent challenges across these architectures?

RQ4. In what domains, and to what extent, can current MT systems approximate versus fall short of human linguistic cognition?

RQ5. What forms of interdisciplinary collaboration between linguists and computer scientists are necessary to advance Arabic MT?

2.3 Objectives of the study

Accordingly, the study aims to: (1) characterize the ontological status of MT within computational linguistics; (2) compare how four representative Arabic-capable MT systems operationalize morphological, syntactic, semantic, and pragmatic knowledge; (3) identify the Arabic-specific structural features that most consistently challenge these systems; and (4) draw implications for interdisciplinary training and collaboration in Arabic NLP.

3. Methodology

3.1 Research Design and Rationale

The study adopts a qualitative, comparative document- and system-analysis design, a method consistent with the established tradition of comparative MT scholarship (Hutchins & Somers, 1992; Koehn, 2009). This design suits the study's aim, which is conceptual and comparative — mapping how different architectures formalize linguistic knowledge — rather than benchmarking translation performance on a held-out test set.

3.2 Corpus of Analysis: Selection of Systems

Four systems were purposively selected to represent the principal architectural paradigms currently used for Arabic: a rule-based desktop system (Golden El Wafi), an online multilingual dictionary/translation portal (ArabTran), a statistical-to-neural system with large-scale deployment (Google Translate), and a generative pre-trained transformer system (ChatGPT Translator). These systems span the typology from direct machine translation to increasingly interactive, refinable output.

3.3 Analytical Framework

Each system was analyzed against four criteria drawn from the computational linguistics and Arabic NLP literature (Habash, 2010; Jurafsky & Martin, 2021; Koehn, 2009): (a) morphological processing — how the system handles Arabic's root-and-pattern structure; (b) syntactic processing — how it parses word order and agreement; (c) semantic and pragmatic processing — how it resolves ambiguity, idiom, and register; and (d) architecture and accessibility — rule-based versus data-driven design, and offline versus online deployment. Evidence for each criterion was drawn from the developers' documented design principles and from published evaluations in the Arabic NLP literature cited throughout this article.

3.4 Limitations of the Design

This is a descriptive, criterion-based comparison grounded in documented system architecture and published literature; it is not an empirical evaluation using a controlled bilingual test corpus, automatic metrics (e.g., BLEU, COMET), or human adequacy and fluency ratings. This limitation is revisited in Section 10, together with a proposed design for a follow-up corpus-based study.

4. Linguistics and Computational Processing: Theoretical Foundations

4.1 Language Computing and Linguistic Representation

Language computing examines how natural-language expressions can be constructed, analyzed, and represented within computational frameworks (Kornilova, 2015). Its central challenge is linguistic representation: formalizing the meaning of words, phrases, and sentences so that a machine can process them reliably, despite natural language being contextual, polysemous, and culturally embedded (Manning & Schütze, 1999; Bird, Klein, & Loper, 2009). Arabic's root-and-pattern morphology sharpens this challenge, since related word forms can differ only in vocalic pattern — a distinction that is frequently unmarked in ordinary written text (Ali, 1988; Fattoum, 2024).

Progress in this area depends on integrating two kinds of expertise: comprehensive linguistic knowledge of phonology, morphology, syntax, semantics, and pragmatics (Shtaytiyah, 2008; El-Haj Saleh, 2000), and computational expertise in algorithm design and NLP techniques (Kornilova, 2015; El-Haj Saleh, 2007; Jurafsky & Martin, 2021). Neither is sufficient alone: without linguistic grounding, computational procedures remain formally correct but semantically empty; without computational implementation, linguistic theory cannot be operationalized at scale (Gruber, 1993; Sowa, 2000).

4.2 The Foundations of Computational Linguistics

Computational linguistics is inherently interdisciplinary, drawing on theoretical linguistics, computer science, mathematics, and logic (Darbellay, n.d.). Arab scholars have contributed substantially to this convergence for Arabic specifically: Ali (1988) argued that automating Arabic morphological analysis is essential for text generation, dictionary development, and information retrieval, while El-Haj Saleh (2007) emphasized that effective computational systems require linguists to supply structural and semantic expertise that engineers then implement algorithmically. This body of work has been supported by the development of large annotated Arabic resources, including the Penn Arabic Treebank (Maamouri, Bies, Buckwalter, & Mekki, 2004; Bies, Maamouri, Buckwalter, & Mekki, 2012) and comparable corpora for dialectal Arabic (Elfardy & Diab, 2012; Zaghoulani, 2014), which supply the annotated training and evaluation data that both rule-based and data-driven Arabic MT systems ultimately rely on (Dorr & Zajic, 2001).

Institutional divisions between linguistics and computer science departments continue to complicate this collaboration in practice (Darbellay, n.d.; Shtaytiyah, 2008), but the convergence of computational rigor and linguistic insight remains a precondition for models that move beyond symbolic manipulation toward genuine engagement with meaning (Lakoff & Johnson, 1980; Rumelhart & McClelland, 1986).

5. The Ontology of Machine Translation

5.1 Defining Machine Translation Ontologically

Machine translation can be defined ontologically as a computational system that formalizes linguistic knowledge into structured representations, enabling a machine to analyze and generate natural-language text (Fattoum, 2024). Ontology, in the technical sense used here, refers to the organization of linguistic concepts — roots, patterns, syntactic roles, semantic relations — into hierarchical and relational structures that a system can query and manipulate (Gruber, 1993; Noy & McGuinness, 2001; Guarino, Oberle, & Staab, 2009). Within this framing, language is treated not as an abstract symbolic system but as a domain that can be modeled through algorithms, logical rules, and, increasingly, data-driven statistical mechanisms (Shannon, 1948; Shannon & Weaver, 1964).

5.2 Typology of Machine Translation Systems

MT systems are conventionally classified by the degree of human intervention involved: (1) direct machine translation, in which the system produces output with no human involvement in the translation process itself; (2) machine-aided translation (MAT), in which software supports a human translator (e.g., through terminology databases or translation memory); and (3) machine-aided human translation / interactive machine translation (MAHT/IMT), in which the system and the human translator iteratively refine output together (Kornilova, 2015). The four systems compared in this study span this typology, from largely direct translation (Golden El Wafi, ArabTran) to increasingly interactive, refinable output (Google Translate, ChatGPT Translator).

6. Comparative Analysis of Arabic Machine Translation Systems

6.1 ArabTran

ArabTran is a free, online multilingual translation portal supporting more than 130 languages, including Arabic (Shtaytiyah, 2008). Its translation logic relies primarily on dictionary lookup and shallow rule-based mapping rather than deep morphosyntactic modeling. This makes it broadly accessible but limited in handling Arabic's syntactic flexibility and semantic ambiguity beyond the lexical level. While it offers immediate access and broad language coverage, its structural depth remains minimal compared to systems incorporating explicit linguistic modeling.

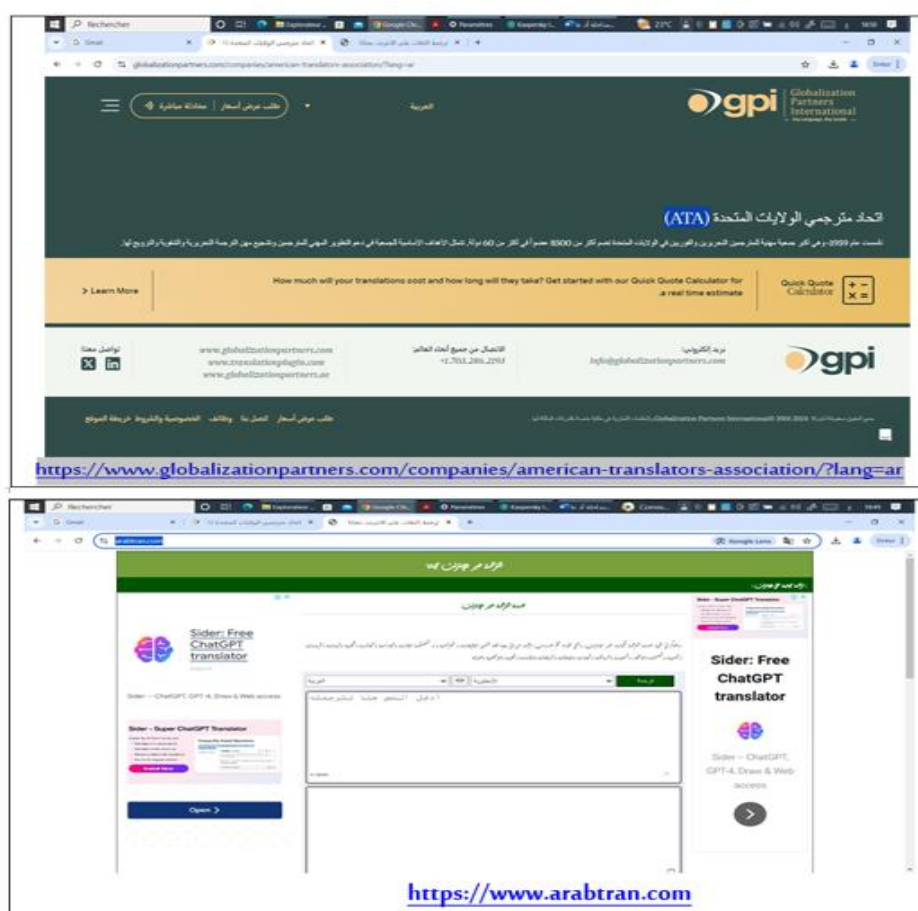


Figure 1

ArabTran as a model

6.2 Google Translate

Launched in 2006, Google Translate has evolved from rule-based and statistical methods toward neural machine translation built on transformer architectures (Hutchins & Somers, 1992; Koehn, 2009; Bahdanau et al., 2015; Vaswani et al., 2017). It processes whole sentences as coherent units rather than word by word, improving fluency and contextual accuracy for language pairs with large parallel corpora (Jurafsky & Martin, 2021). Its performance depends on the scale and quality of the multilingual corpora it is trained on, which can introduce domain inconsistency and, for morphologically rich languages such as Arabic, residual pragmatic ambiguity (Habash, 2010; Fattoum, 2024). The translation process involves multiple technical stages, including text preprocessing, tokenization, semantic representation, alignment, decoding, and post-processing, each requiring both computational expertise (algorithm design, neural network optimization) and linguistic expertise (syntax, semantics, pragmatics, discourse analysis). Despite its sophistication, it functions as a support tool rather than a substitute for professional human translation, particularly for domain-specific or culturally sensitive text (Baker, 1992; Lopez, 2008).



Figure 2

Google Translate as a model

6.3 Golden El Wafi

Golden El Wafi is an offline, rule-based desktop application for Arabic–other-language translation. Its structured algorithms perform explicit morphological analysis — for example, distinguishing related root-and-pattern forms such as ‘alima (عَلِمَ), ‘ulima (عَلَّمَ), and ‘ilm (عِلْم), which can otherwise appear ambiguous to systems without linguistic modeling (Habash, 2010; Shtaytiyah, 2008) — alongside syntactic parsing of subject–verb agreement and noun–adjective concord (Jurafsky & Martin, 2021). At the semantic level, the software evaluates meaning by relying on contextual cues and lexicographic resources to select appropriate equivalents in the target language.

Unlike statistical or neural MT systems, Golden El Wafi primarily adopts a rule-based translation approach, embedding syntactic, morphological, and semantic rules directly into the system. This approach has proven particularly effective for highly inflected languages such as Arabic, where purely data-driven models often struggle with structural complexity. Its offline operation offers data privacy and pedagogical transparency, since users can trace how a given output was derived from explicit rules. It also serves as an educational tool, helping students, translators, and researchers understand how Arabic can be represented computationally at both theoretical and applied levels (Ali, 1988; Habash, 2010).

Its principal limitation is reduced ability to handle idiomatic expression, pragmatic nuance, and culturally embedded meaning, and its language coverage is narrower than large cloud-based systems (Baker, 1992; Lopez, 2008; Fattoum, 2024). Consequently, human intervention remains necessary, particularly in literary or professional translation contexts.

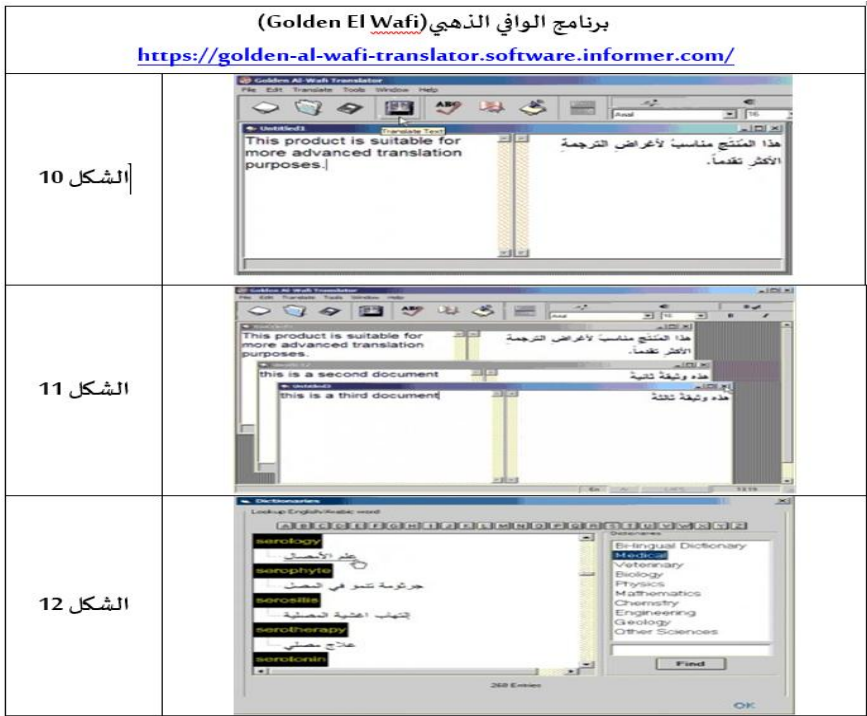


Figure 3
Golden El Wafi as a model

6.4 ChatGPT Translator

ChatGPT Translator, built on the Generative Pre-trained Transformer (GPT) architecture, represents the generative-AI paradigm in MT (Vaswani et al., 2017; Devlin et al., 2019; Brown et al., 2020). Rather than seeking word-for-word correspondence, it aims to produce contextually appropriate output that reflects the communicative intent of the source text, drawing on tokenization, contextual representation, and probabilistic text generation (Radford et al., 2019; Jurafsky & Martin, 2021).

The system demonstrates advanced understanding of syntax, semantics, pragmatics, and discourse context, leading to more nuanced translations. It supports interactive refinement, allowing users to request clarification or rephrasing, and shows comparatively strong handling of idiom and register. Its multilingual coverage spans dozens of languages, and the model can be adapted or fine-tuned for specific domains, enabling rapid responsiveness to specialized translation needs.

Its limitations mirror those of large language models generally: performance depends on training-data coverage and quality, and outputs can be inconsistent for rare languages, specialized terminology, or culturally dense content. Occasional inaccuracies or stylistic inconsistencies may also occur, particularly in culturally dense or idiomatic contexts, which is why human oversight remains necessary for professional, legal, or literary translation (Baker, 1992; Fattoum, 2024).

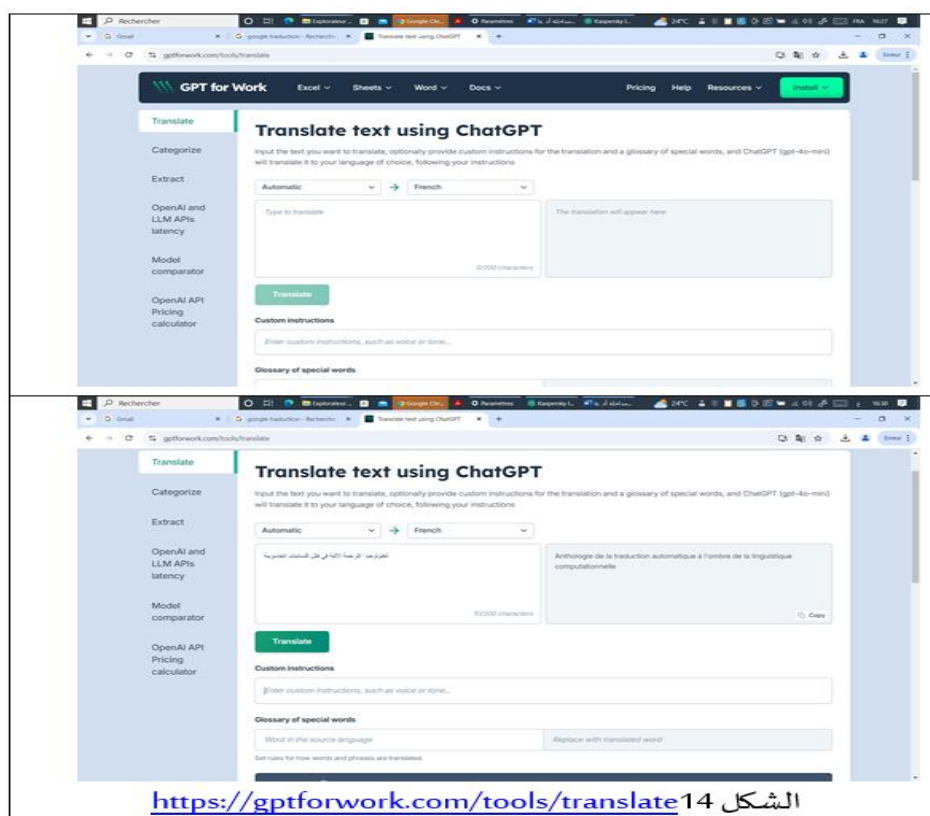


Figure 4

ChatGPT Translator as a model

6.5 Comparative Synthesis

Table 1 summarizes the four systems along the criteria set out in Section 3.3.

Table 1.

Comparative overview of four Arabic machine translation systems

System	Architecture	Arabic-specific handling	Key strength	Key limitation
ArabTran	Rule/dictionary-based, multilingual web portal (130+ languages)	Basic morphological lookup	Broad language coverage; free, immediate access	Limited handling of syntactic and semantic complexity
Google Translate	Statistical → neural (transformer-based) MT	Root-and-pattern recognition supported by large parallel corpora	Fluency, contextual accuracy, scalability	Domain inconsistency; possible corpus bias in low-resource domains
Golden El Wafi	Rule-based, offline desktop software	Explicit root-and-pattern and syntactic rule encoding	Structural precision; data privacy; pedagogical transparency	Weak on idiomatic/pragmatic meaning; narrow language coverage
ChatGPT Translator	Generative pre-trained transformer (GPT)	Context-aware semantic mapping; handles idiom via learned patterns	Fluency; interactive refinement; idiomatic sensitivity	Training-data dependency; occasional inconsistency; needs professional oversight

Read together, the comparison suggests a broad trade-off: rule-based systems (Golden El Wafi) offer structural transparency and precision at the level of morphology and syntax but limited pragmatic flexibility; data-driven systems (Google Translate, ChatGPT Translator) offer greater fluency and idiomatic sensitivity but at the cost of interpretability and consistency across domains. ArabTran, as a lightweight dictionary-based portal, sits below both in structural depth but remains valuable for its accessibility and coverage.

7. Linguistic and Computational Challenges in Processing Arabic

7.1 Morphological and Syntactic Processing

Morphology is central to digital processing of Arabic, since word forms are generated through root-and-pattern derivation rather than simple affixation (Shtaytiyah, 2008). The computer requires detailed knowledge of linguistic structures for each language to determine correctness. Syntactic modules must additionally detect and, where possible, correct grammatical and diacritical errors, sometimes drawing on external linguistic resources for validation (Shtaytiyah, 2008).

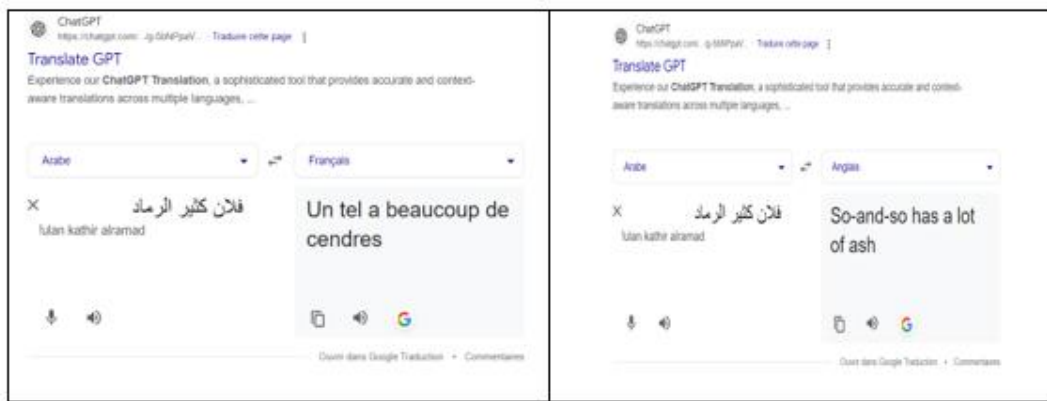


Figure 5

ChatGPT Translator Morphological and Syntactic Processing as a model

7.2 Semantic and Pragmatic Processing

Beyond structure, systems require detailed knowledge of how meaning is constructed in context. Lexical-semantic processing involves mapping word senses, consulting relevant lexicographic resources, and accounting for semantic change over time (Kornilova, 2015; Fattoum, 2024). This is where rule-based and data-driven systems diverge most sharply: the former encode fixed semantic rules, while the latter infer probable meaning from usage patterns in training data — an approach that captures nuance more flexibly but less predictably.

7.3 The Limits of Replacing Human Cognition

Across all four systems, idiomatic Arabic expressions remain a recurring source of mistranslation, underscoring the continued necessity of human oversight (Fattoum, 2024). Reliability issues generally trace to insufficient linguistic programming rather than to computational power alone (El-Haj Saleh, 2007), reinforcing the case for sustained collaboration between linguists and engineers (Darbellay, n.d.).

8. Discussion: Answering the Research Questions

RQ1 (ontological status). Machine translation is best understood as a computational formalization of linguistic knowledge into structured, queryable representations (Section 5.1), rather than as a neutral technical process or a replica of human understanding. The MT process begins with receiving text, converting it into a mathematical representation of the language, which reflects word patterns, syntactic roles, and morphological structures (Fattoum, 2024). The machine then converts the mathematical code into the target language, establishing a translation formula (Kornilova, 2015). Within computational linguistics, language is no longer treated as a purely symbolic or abstract phenomenon, but as a system that can be modeled through algorithms, logical rules, and data-driven mechanisms (Shannon, 1948; Shannon & Weaver, 1964).

RQ2 (architectural differences). The four systems formalize morphological, syntactic, semantic, and pragmatic knowledge differently: rule-based systems (Golden El Wafi) encode this knowledge explicitly and transparently; hybrid/dictionary-based systems (ArabTran) rely on shallower lexical mapping; and data-driven systems (Google Translate, ChatGPT Translator) infer it statistically from large corpora, trading transparency for fluency and idiomatic range (Section 6).

RQ3 (Arabic-specific challenges). Root-and-pattern morphology, the routine omission of diacritics in everyday text, and Arabic's relatively flexible word order recur as challenges across all four systems, though each architecture manages them differently (Section 7.1–7.2). The absence of diacritics in most written Arabic further increases ambiguity, making it difficult for machines to distinguish between multiple meanings derived from the same root (Fattoum, 2024).

RQ4 (limits relative to human cognition). Current systems approximate human performance reasonably well for structurally regular, non-idiomatic text, but fall short wherever cultural, pragmatic, or literary meaning is dense — domains where human translators retain a clear advantage (Section 7.3). Human translators possess cultural awareness, pragmatic competence, and contextual sensitivity that machines cannot yet replicate (El-Haj Saleh, 2007).

RQ5 (interdisciplinary collaboration). Advancing Arabic MT requires structured collaboration between linguists, who supply and validate linguistic rules and semantic judgments, and computer scientists, who implement and scale these insights algorithmically (Section 9). Without this interdisciplinary foundation, machine translation systems would remain limited to superficial word substitution rather than meaningful translation.

9. Implications for Interdisciplinary Collaboration and Training

The analysis reinforces the case for interdisciplinary education that combines linguistics and computational sciences. Algerian and wider Arab-world researchers have called for graduate programs that establish a shared technical language between linguists and engineers (El-Haj Saleh, 2000), producing professionals capable of bridging theoretical linguistics and applied computational methods. Such educational initiatives can create professionals capable of integrating linguistic rules into computational algorithms with precision.

Despite the capabilities of tools such as Golden El Wafi and ChatGPT Translator, human expertise remains irreplaceable wherever cultural, pragmatic, or literary interpretation is required (Fattoum, 2024; Ali, 1988). Computational linguistics in the Arab world is a pioneering field, expanding recently with more researchers combining computer science and linguistics. Applications include MT, automatic error correction, language teaching, document processing, and artificial speech synthesis (Shtaytiyah, 2008; Fattoum, 2024).

10. Limitations and Directions for Future Research

This study is descriptive and comparative rather than empirical: its conclusions rest on documented system architecture and the published Arabic NLP literature, not on a controlled bilingual test corpus or automatic/human evaluation metrics. Generative AI systems such as ChatGPT Translator also evolve rapidly, so architectural comparisons of this kind are time-bound and should be periodically revisited.

A natural next step is a corpus-based evaluation study that submits a shared Arabic–English/English–Arabic test set — stratified by genre (news, literary, technical, dialectal) — to all four systems, scored using automatic metrics (e.g., BLEU, COMET) alongside native-speaker human ratings of adequacy, fluency, and pragmatic appropriateness, with a structured error-typology analysis of morphological, syntactic, and pragmatic failures.

11. Conclusion

This study has examined the ontological status of machine translation within computational linguistics through a comparative analysis of four Arabic-capable MT systems spanning the rule-based, dictionary-based, statistical-to-neural, and generative-AI paradigms.

The comparison shows that these architectures formalize linguistic knowledge in systematically different ways, with a persistent trade-off between structural transparency and pragmatic fluency, and that Arabic's morphological richness and diacritic ambiguity remain unresolved challenges across all four.

Machine translation should accordingly be understood as a valuable but partial formalization of linguistic knowledge — a supportive and complementary tool rather than a substitute for human translators, particularly in culturally and pragmatically dense contexts. Sustained collaboration between linguists and computer scientists, supported by dedicated interdisciplinary training, remains the most promising path toward Arabic MT systems that are both computationally robust and linguistically faithful.

Declaration of AI Use: This manuscript discusses AI-based translation systems. However, no artificial intelligence tools were used to conduct the comparative analyses or to derive the conclusions presented in Sections 6 through 8; these sections are based solely on the author's interpretation of the cited literature. Furthermore, no AI tools were utilized at any stage of the conceptualization, analysis, writing, or final preparation of this work.

Ethical Declaration: This study was conducted in full compliance with the fundamental ethical principles and methodological standards of academic research. It adopts a rigorous scientific approach and incorporates the established perspectives of researchers operating at the intersection of linguistics and computational science.

Funding: This research received no specific grant from any funding agency, commercial entity, or non-profit organization.

Author Contributions: The author is the sole contributor to all phases of this study, including conceptualization, study design, literature review, data collection, data analysis and interpretation, manuscript drafting, and subsequent revisions.

References

- Ali, N. (1988). *Al-lugha al-‘arabiyya wa al-ḥāsūb [Arabic language and the computer]*. Ta‘rīb Publishing House
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). *Neural machine translation by jointly learning to align and translate*. In Proceedings of the 3rd International Conference on Learning Representations (ICLR). <https://doi.org/10.48550/arXiv.1409.0473>
- Baker, M. (1992). *In other words: A coursebook on translation*. Routledge. <https://doi.org/10.4324/9780203133590>
- Bies, A., Maamouri, M., Buckwalter, T., & Mekki, W. (2012). *Arabic treebank design and use*. In Proceedings of the ACL Workshop on Arabic Treebanks.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., ... Amodei, D. (2020). *Language models are few-shot learners*. In Advances in Neural Information Processing Systems (Vol. 33, pp. 1877–1901). Curran Associates.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

- Darbellay, F. (2011). Vers une théorie de l'interdisciplinarité? Entre unité et diversité. *Nouvelles perspectives en sciences sociales*. 7(1)-65-87. <https://doi.org/10.7202/1007082ar>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- Dorr, B. J., & Zajic, D. (2001). Natural language processing for Arabic. In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL).
- Elfardy, H., & Diab, M. (2012). *Synchronous context-free grammar for dialectal Arabic*. In Proceedings of COLING 2012: Technical Papers.
- El-Haj Saleh, A. R. (2000). *Buḥūth wa dirāsāt fī al-lisāniyyāt al-‘arabiyya [Research in Arabic linguistics]*. Movem.
- El-Haj Saleh, A. R. (2007). *Al-naẓariyya al-Khalīliyya al-ḥadītha [Modern Khalilian theory]*. Center for Scientific and Technical Research for Arabic Language.(CRSTDLA), Center Papers, No. 4.
- Fattoum, F. Q. (2024). Difficulties in machine translation in Arabic: The pragmatic level as a model. *Egyptian Journal of Language Engineering*, 11(1). https://ejle.journals.ekb.eg/article_352396_33598009b688926511ef12d45c9f7cce.pdf#page=1.00&gsr=0
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. <https://doi.org/10.1006/knac.1993.1008>
- Guarino, N., Oberle, D., & Staab, S. (2009). *What is an ontology?* In S. Staab & R. Studer (Eds.), *Handbook on ontologies* (2nd ed., pp. 1–17). Springer. https://doi.org/10.1007/978-3-540-92673-3_0
- Habash, N. Y. (2010). *Introduction to Arabic natural language processing*. Morgan & Claypool Publishers. <https://doi.org/10.2200/S00277ED1V01Y201008HLT010>
- Hutchins, W. J., & Somers, H. (1992). *An introduction to machine translation*. Academic Press.
- Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing* (3rd ed.). Pearson. <https://web.stanford.edu/~jurafsky/slp3/>
- Koehn, P. (2009). *Statistical machine translation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815829>
- Kornilova, A. (2015, October 6). *A brief introduction to computational semantics*. Stanford University. <http://www.akornilo.com/computational-semantics/>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226470993.001.0001>
- Lopez, A. (2008). Statistical machine translation. *ACM Computing Surveys*, 40(3), Article 8, 1–49. <https://doi.org/10.1145/1380584.1380586>

- Maamouri, M., Bies, A., Buckwalter, T., & Mekki, W. (2004). *The Penn Arabic Treebank: Building a large-scale annotated Arabic corpus*. In Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology (Technical Report KSL-01-05)*. Stanford Knowledge Systems Laboratory.
http://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. OpenAI Technical Report.
https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press.
<https://doi.org/10.7551/mitpress/5236.001.0001>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shannon, C. E., & Weaver, W. (1964). *The mathematical theory of communication*. University of Illinois Press.
- Shtaytiyah, S. S. (2008). *Al-lisāniyyāt: Al-majāl, al-wazīfa, wa al-manhaj [Linguistics: Field, function, methodology]*. Al-Kitab Al-Hadith.
- Sowa, J. F. (2000). *Knowledge representation: Logical, philosophical, and computational foundations*. Brooks/Cole.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998–6008). Curran Associates.
- Zaghouani, W. (2014). The design of a comparable corpus for Arabic dialects. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*.