



ISSN: 2663-9033 (Online) | ISSN: 2616-6224 (Print)

Journal of Language Studies

Contents available at: <https://jls.tu.edu.iq/index.php/JLS>

Evaluating Artificial Intelligence (AI) Translation Performance in Diglossic Iraqi Courtroom Discourse: A Comparative Study of ChatGPT and Microsoft Translator (MST)

Asst. Lect. Fatimah Awad Mkhlaif^{*1}

Fatima.muklif112@tu.edu.iq

Department of English Language, College of Arts, Tikrit University, Tikrit, Iraq
&

Asst. Lect. Osamah Mohammed Abdullah

Ossama.m.a23@tu.edu.iq

Department of Translation, College of Arts, Tikrit University, Tikrit, Iraq

Received: 01/12/2025, **Accepted:** 06/01/2026, **Online Published:** 30/06/2026

Abstract

This research examines how artificial intelligence translation systems deal with diglossic language in Iraqi courtroom discourse. Iraqi courtrooms commonly involve a mixture of Modern Standard Arabic and Iraqi Colloquial Arabic, which creates serious challenges for machine translation systems. The study compares the performance of ChatGPT and Microsoft Translator (MST) in translating 30 instances of authentic Iraqi courtroom dialogues collected from publicly available trial videos on YouTube. The data were manually transcribed and analyzed using both qualitative and quantitative methods, with special focus on diglossic shifts and pragmatic meaning. The results show that ChatGPT achieved higher accuracy in interpreting diglossic shifts, correctly identifying about seventy three percent, while MST achieved about sixty seven percent. However, both

^{*1} **Corresponding Author:** Fatimah Awad Mkhlaif, Email: Fatima.muklif112@tu.edu.iq

Affiliation: Tikrit University – Iraq.

© This is an open access article under the CC by licenses <http://creativecommons.org/licenses/by/4.0>



systems failed to provide fully reliable sentence level translations suitable for legal use. The findings confirm that current AI translation tools cannot be safely used independently in courtroom settings and must be supported by human translators to avoid serious legal misunderstandings.

Keywords: Arabic diglossia, Iraqi courtroom discourse, Machine translation, ChatGPT, Microsoft Translator (MST)

Abbreviations: Microsoft Translator (MST), Modern Standard Arabic (MSA), Iraqi Colloquial Arabic (ICA), Chat Generative Pre-trained Transformer (ChatGPT)

تقييم أداء الترجمة بالذكاء الاصطناعي في الخطاب القضائي العراقي الازدواجي دراسة مقارنة بين ChatGPT و Microsoft Translator

م. م. فاطمة عواد مخلف

قسم اللغة الإنكليزية، كلية الآداب، جامعة تكريت، العراق

م. م. اسامه محمد عبد الله

قسم الترجمة، كلية الآداب، جامعة تكريت، تكريت، العراق

المستخلص

تبحث هذه الدراسة في كيفية تعامل أنظمة الترجمة بالذكاء الاصطناعي مع الازدواج اللغوي في الخطاب القضائي العراقي، اذ تتميز قاعات المحاكم العراقية باستخدام مزيج من العربية الفصحى الحديثة والعربية العامية العراقية، مما يخلق تحديات كبيرة امام أنظمة الترجمة الآلية. تقارن الدراسة بين أداء ChatGPT و MST في ترجمة ثلاثين مثالا من حوارات قضائية عراقية حقيقية تم جمعها من مقاطع محاكم منشورة علنا على منصة يوتيوب، حيث تم تفرغ البيانات يدويا وتحليلها باستخدام مناهج نوعية وكمية مع تركيز خاص على التحولات اللغوية الازدواجية، والمعنى التداولي. اظهرت النتائج ان ChatGPT حقق دقة اعلى في تفسير التحولات الازدواجية، اذ تمكن من التعرف بشكل صحيح على ما يقارب ثلاثة وسبعين بالمئة، بينما حقق MST حوالي سبعة وستين بالمئة، ومع ذلك فشل النظامان في تقديم ترجمات موثوقة بالكامل على مستوى الجملة تكون صالحة للاستخدام القانوني. تؤكد النتائج ان ادوات الترجمة الحالية المعتمدة على الذكاء الاصطناعي لا يمكن استخدامها بشكل مستقل في سياقات المحاكم، ويجب دعمها بمرجمين بشريين لتجنب سوء فهم قانوني خطير.

الكلمات المفتاحية: الازدواج اللغوي في العربية، الخطاب القضائي العراقي، الترجمة الآلية، ChatGPT، MST

1. Introduction

1.1 Problem of the Study

Multiple documented challenges exist for machine translation systems when processing legal text because legal language includes intricate structures which need precise word choices and full comprehension of both situational context and pragmatic aspects (El-Farahaty, 2015; Abdul-Mageed et al.,2024). The problems with language usage become more severe when diglossic systems exist as seen in the Arab world where Modern Standard Arabic (MSA) functions as the official language for legal procedures yet Iraqi Colloquial Arabic (ICA) dominates the natural courtroom conversations between witnesses and lawyers and their clients (Ferguson, 1959; Jabbari, 2013).

Research conducted before this study demonstrates that AI-based translation systems primarily use standardized MSA corpora for training which results in poor performance when dealing with dialectal Arabic and code-switching and spontaneous speech (Abdul-Mageed et al.,2024; Khondaker et al.,2023). Studies of neural machine translation and large language models demonstrate that systems fail to understand negation and pragmatic elements and they struggle with interrupted or unfinished statements which occur in mixed-register discourse (Jiao et al.,2023; Zhu et al.,2023).

The legal system faces major problems because court translation mistakes result in incorrect witness statement interpretations which endangers the entire legal framework (Radam, 2020). The research investigates how MST (MST) and ChatGPT perform when translating Iraqi courtroom dialogue which contains diglossic elements to show how Arabic diglossia affects modern AI translation systems.

1.3 Aims of the Study

The research assesses MST and ChatGPT translation accuracy through their processing of authentic Iraqi courtroom dialogues which contain natural language shifts between Modern Standard Arabic and Iraqi Colloquial Arabic. The research investigates these main objectives:

- The research identifies and organizes and examines the translation mistakes which both systems produce when they encounter diglossic language changes.
- The research evaluates system performance through its content understanding capabilities and its flexibility when handling different language structures.

- The research examines how translation mistakes affect legal understanding and judicial fairness and machine translation system reliability during court proceedings.
- The research provides essential knowledge which will help developers create better translation systems that work well in multilingual settings with diglossic dialects.

1.2 Hypothesis

This study hypothesizes that:

- MST and ChatGPT will demonstrate significantly lower translation accuracy when processing diglossic Iraqi courtroom discourse than when translating Modern Standard Arabic (MSA) segments alone. The research establishes particular predictions which state that.
- Both systems will correctly translate a higher proportion of MSA-only segments than Iraqi Colloquial Arabic (ICA) segments.
- The translation accuracy of the systems decreases when users switch between Modern Standard Arabic (MSA) and Iraqi Colloquial Arabic (ICA) at high speed.
- ChatGPT will achieve better overall accuracy than MST in understanding diglossic shifts because it possesses superior abilities to process context and pragmatics.

1.4 Value of the Study

The research provides direct evidence about how present-day AI translation systems do not perform well when translating legal texts which contain diglossic language. The research delivers developers with functional implementation guidance through its evaluation of MST and ChatGPT performance on genuine Iraqi courtroom audio which demonstrates methods to boost their systems through dialect-based training and code-switching identification. The research shows legal professionals and judicial institutions that AI translation mistakes in witness statements and negations and pragmatic elements produce meaning changes which affect court operations so human oversight must be used for courtroom translation systems. The research delivers practical evidence to linguists and translation scholars about Arabic diglossia and natural spoken dialogue effects on neural machine translation systems and large language models which enhances existing understanding of diglossia and legal discourse and machine translation assessment in Arabic languages that have limited representation.

1.5 Limits of the Study

The research study contains multiple restrictions which need to be addressed. The public source recordings of courtroom sessions contain uneven audio quality which creates challenges for accurate transcription work. The research results focus on Iraqi Arabic but their applicability remains uncertain for other Arabic language varieties. The research uses qualitative and quantitative case studies. The performance of both systems will change because their software receives periodic updates which introduce new system behaviors. The study is looking into trial sessions of Iraqi Courtrooms available publically on Youtube. One of the limitations is to conduct the study on real time trial sessions; as it is very difficult to receive approvals to conduct the study – not to mention the sensitivity of cases discussed in the courtroom. Therefore, the research depends on publically available videos of trials on Youtube. It studies the switching between the MSA and ICA varieties. The MT used in this study are ChatGPT-4 and MST (MST) in 2025.

2. Literature Review:

All legal systems across the world need exact and clear language usage to function properly in their administration of justice. The Arab world experiences additional obstacles because diglossia establishes a sociolinguistic system where Modern Standard Arabic (MSA) serves as the High (H) variety and local dialects function as Low (L) varieties which never intersect in usage (Ferguson, 1959; Fathi, 2024). The Iraqi courtroom shows this pattern through its normal court procedures. The legal system uses MSA to write official documents and court decisions and official agreements but Iraqi Colloquial Arabic (ICA) serves as the main language for witness statements and legal discussions and casual conversations (Radam, 2020). The functional separation between these two languages exists beyond procedural boundaries because they have developed into distinct linguistic systems which sometimes become impossible to understand between each other (Jabbari, 2013). Machine translation tools operate in professional environments to assist users who need to communicate through different languages. MST (MST) together with its equivalent systems enable people to communicate through instant language translation services. However, the architectural foundation of these tools is predominantly based on corpora of formal, standard language. The speech patterns of diglossic communities create a major technological obstacle because their H to L code-switching does not count as an error since it forms an essential part of their communication skills. The theoretical framework diglossia has been thoroughly studied by Ferguson (1959) and Fishman (1967) yet there is a scarcity of research about this topic in Iraq (Jabbari, 2013). There is a scarcity of research that directly assesses the performance of neural machine translation systems against the spontaneous, code-switched, and high-stakes discourse of an authentic diglossic legal setting. The research investigates this knowledge deficiency through its examination of the following research

questions. How effectively do MST and ChatGPT process spontaneous diglossic shifts between MSA and Iraqi Colloquial Arabic in authentic courtroom discourse? The system produces two primary categories of linguistic and pragmatic mistakes which affect legal precision through their unique characteristics. The research results indicate how well modern machine translation systems function when used independently in legal systems that use two languages. The research examines actual courtroom dialogues to demonstrate that Arabic legal communication through diglossia presents an ongoing problem which current machine translation systems fail to address. The discovered errors create multiple problems which extend past basic mistakes because they could result in significant errors when analyzing witness statements and their purpose. The following sections examine previous research about diglossia and legal discourse before describing the research approach and presenting findings and drawing conclusions.

2.1. Concept of Diglossia

2.1.1 Definitions and Theoretical Frameworks

The social language pattern of diglossia occurs when one community uses two separate language systems which function as independent social communication systems. Ferguson (1959) established the traditional definition which shows how High (H) language serves official purposes through written communication in institutions but Low (L) language functions for casual face-to-face conversations. In Arabic-speaking societies, this distinction is most clearly reflected in the use of Modern Standard Arabic (MSA) as the H variety and regional colloquial dialects, such as Iraqi Colloquial Arabic (ICA), as L varieties.

People learn colloquial language varieties from their family environment before they acquire MSA through their school education and media contact. The dual acquisition process leads to a major language difference between these two varieties because they have distinct sound systems and word choices. Research shows that Modern Standard Arabic (MSA) contains different phonological patterns from colloquial Arabic which create difficulties for language processing and new word learning because MSA words sound very different from how native speakers speak (Saiegh & Saiegh, 2025, pp. 1533–1535).

Fishman (1967) developed Ferguson's model through his research which showed that diglossia functions by allocating social domains instead of using different languages. The system of diglossia exists because institutions have established rules which place MSA in official contexts including educational settings and media outlets and legal proceedings but colloquial dialects continue to rule in casual everyday speech. Fathi (2024) establishes functional specialization as the core element of Arabic diglossia because H

variety upholds social value and official power but L variety serves for regular communication purposes.

The theoretical framework shows that Arabic diglossia functions as one system which affects language processing during courtroom activities when speakers alternate between MSA and colloquial Arabic.

2.1.2 Ferguson's Diglossia (Classical/Narrow Diglossia)

According to Charles Ferguson's classic 1959 definition, diglossia is a very specific sociolinguistic situation with clearly defined characteristics (**Fathi, 2024,2-5**):

1. Two Varieties of the *Same* Language

Diglossia involves two distinct varieties of one language—what Ferguson calls the **High (H)** and **Low (L)** varieties. These are not treated as separate languages, but as divergent forms of a single linguistic system.

Example: Standard Modern Arabic vs. regional spoken Arabics such as Moroccan Arabic.

2. Clear Functional Separation

Each variety has its own well-established role in society.

- The **H variety** is reserved for formal, official, and written contexts—public speeches, education, newscasts, religious texts, and literature.
- The **L variety** is the everyday spoken language and is **not** typically used for formal writing.
Speakers instinctively know which variety belongs in which situation; they rarely confuse the two.

3. Prestige and Learning Path

The H variety carries overt prestige and is usually learned through formal schooling rather than at home. The L variety, despite being every speaker's mother tongue, may be viewed as less "correct," less elegant, or even as a degraded form of the H variety—an attitude that is part of the diglossic system Ferguson describes.

4. Long-Term Stability

Ferguson emphasizes that true diglossia is remarkably stable. Many diglossic communities maintain these patterns for centuries, with both varieties fulfilling their roles without one replacing the other.

5. Classic Examples

Ferguson highlights several well-known cases of diglossia:

- **The Arab world:** Standard Arabic (H) vs. local spoken dialects (L)
- **German-speaking Switzerland:** Standard German (H) vs. Swiss German dialects (L)
- **Haiti:** Standard French (H) vs. Haitian Creole (L).

2.1.3 Fishman's Diglossia (Extended/Broad Diglossia)

Joshua Fishman (1967) developed Ferguson's core idea into a complete methodological framework (**Fathi, 2024,2-5**).

1. The model of Fishman operates with different languages instead of language varieties. The model works for any situation where communities use two or more separate languages which can be either connected or unrelated.
2. The essential element of diglossia according to Fishman involves the distribution of social tasks between languages rather than their linguistic characteristics.
3. Bilingualism Connection: Fishman established direct links between diglossia and bilingualism.
 - A society can exhibit three different patterns of language use. The existence of diglossia does not depend on bilingualism because two social groups communicate through different languages yet they use the same official language for official purposes.
 - Bilingualism exists without diglossia when two languages are used but lack established functional boundaries which results in language transition.
 - The population shows bilingualism and diglossia because most members use two languages to fulfill their social roles.
 - **Examples:** The official language of Paraguay uses Spanish for official duties but Guarani serves as the language for casual interactions. The official language of Peru uses Spanish for official purposes but Quechua

functions as the language for indigenous communities. European colonial languages such as English and French operate as High Varieties for administrative and educational purposes in numerous African and Asian countries while Local Varieties maintain their use in everyday life (Fathi, 2024,2-5).

2.2 Diglossia between the Arabic Language and the Iraqi Dialect

Jabbari (2013, 139) presents a study that looks at the features of diglossia, between Modern Standard Arabic and Iraqi Colloquial Arabic. The study by Jabbari describes the relationship in Iraq between Modern Standard Arabic (MSA) the High Variety (H) and Iraqi Colloquial Arabic (ICA) the Low Variety (L). Diglossia creates a situation where the two distinct varieties of the language live together each, with its own social role. MSA is used in written and public settings, like media, education and official speeches. ICA is the tongue for informal communication. The sound, word and grammar differences make MSA and ICA mostly unintelligible to each other. Jabbari shows the difference, on the levels:

1. Phonological Differences

The sound changes follow rules. The sound changes shape the way the two varieties sound. The sound changes make the two varieties sound very different.

a- Consonant Changes:

- MSA /q/ (ق) becomes /g/ (ك) in ICA (e.g., qamar (moon) → gamar).
- MSA /k/ (ك) becomes /ch/ (چ) in ICA (e.g., kalb (dog) → chalib).
- MSA glottal stop /ʔ/ (ء) becomes /j/ (ي) in ICA (e.g., maʔ (water) → maj).

b- Vowel Changes: Vowels systematically shift, e.g., the definite article /al-/ becomes /il-, and present-tense verb prefixes change (/yu-/ → /yi-/ or /a-/).

c- Syllable Structure: ICA often adds a prosthetic glottal stop (/ʔi-/) to the beginning of words, changing the syllable structure from CV to CVC (e.g., uri:d (I want) → ʔari:d).

d- Final Deletion: ICA tends to delete case and mood endings (/un/, /an/, /in/) that are obligatory in MSA.

2- Lexical and Morphosyntactic Differences

The grammatical systems also diverge, with ICA demonstrating a simplified morphology.

- **Morphology:** ICA frequently deletes the conjunction *ʔan* (that) and often violates complex MSA grammatical rules, such as subject-predicate agreement for case and the obligatory deletion of the /n/ sound in dual and plural nouns in genitive constructions.
- **Syntax:** Word order can differ. For example, the MSA phrase "May God help you" (*sa:ʔadək alla:h*) becomes "God helps you" (*alla:h jisa:ʔadək*) in ICA.
- **Lexical Differences**
This is the most drastic area of difference, leading directly to mutual unintelligibility.
- **Complementary Distribution:** MSA and ICA often use completely different words for the same common concepts. They form a doublet, where each word is reserved for its specific (H or L) domain.

Examples: now: MSA *ʔalʔa:n* → ICA *hassa*; here: MSA *huna:* → ICA *ʔihna:*; what: MSA *ma:ða:* → ICA *ʔinu:*; much: MSA *kaθi:r* → ICA *ʔihwa:ja*.

- **Extensive Borrowing:** ICA's lexicon is enriched with loanwords from neighboring languages and international influences, which are largely absent from MSA.
- **Persian:** *gilas* (cherry), *majz* (table)
- **Turkish:** *xaʔu:g* (spoon), *qu:tʔi:* (box)
- **English:** *tikit* (ticket), *bajsi:kil* (bicycle)
- **French:** *ka:rt* (card), *mitru:* (subway)

The cumulative effect of these differences across all linguistic levels results in the mutual unintelligibility of MSA and ICA. A single concept can be expressed with entirely different words and structures, as shown by the example:

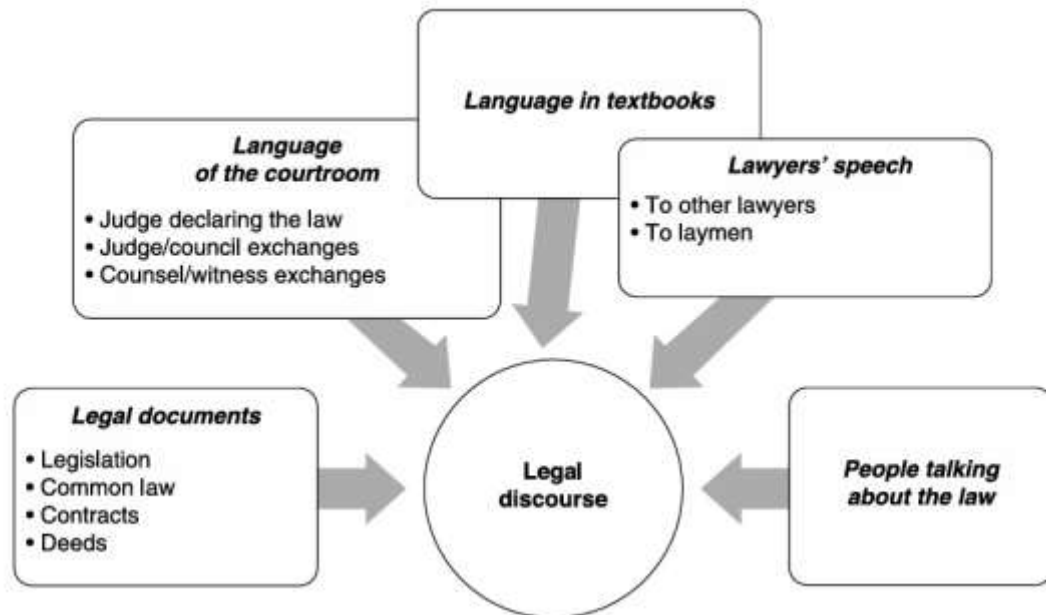
- MSA: *kajfa ʔastaʔi:ʕu-l-istifa:da min barna:maʕ?* (How can I use this program?)
- ICA: *ʔiʕlawm ʔagdar ʔastafid min il-barna:maʕ?*

The research shows that Iraq demonstrates all characteristics of Ferguson's diglossia model. The H and L varieties exist as separate linguistic systems which continue to maintain their social language function. The study shows that children must learn MSA as their second language to succeed in school because it represents a major educational challenge. The author proposes additional research on Arabic dialects beyond this study to achieve complete comprehension of Arabic diglossia.

2.3 Language Use in Iraqi Courtrooms

Anna Trosborg demonstrates through her functional and participant-based classification of legal discourse (El-Farahaty, 2015, 30) that legal language exists as multiple distinct discourses which develop based on their environmental context and the people involved and their intended goals. The schema presents two main categories which include communication methods between spoken and written formats and the participants who take part in the communication process. The spoken discourse section divides into two parts which show how legal professionals and non-professionals communicate with each other. The courtroom contains two essential communication systems which judges use to deliver official statements and lawyers employ strategic dialogue to question witnesses. The courtroom uses distinct language from what lawyers use when speaking to other lawyers through specialized terminology or when speaking to non-lawyers through explanatory or persuasive language. The written discourse section contains official legal documents which include laws created through legislation and contracts that establish formal agreements. The diagram shows how people discuss legal matters through textbooks and media and casual conversations which form part of the complete legal discourse system. The framework enables translation studies to determine suitable translation methods because it demonstrates that different legal discourse subdomains need distinct translation approaches.

Figure1: Trosborg's classification on Legal Discourse



The legal participants in Iraqi court trials which appear on Youtube follow Trosborg's classification system based on available parties and case types. The following list shows all participants who attend Iraqi court trials and their specific roles in the legal process.

The Iraqi legal system combines Civil Law with Islamic Law so the following parties attend trials in Iraqi courtrooms.

- The Judge (Qadi) starts the trial by evaluating evidence and questioning witnesses until he reaches the final decision and punishment.
- The Public Prosecutor (Na'ib 'Am) serves as the state representative who builds the case against defendants and conducts investigations into alleged offenses.
- The Defendant (Muttaham) faces criminal accusations or civil lawsuits as the person who stands accused in court.
- The Defense Lawyer (Muhamme) defends their client through case development and evidence opposition against prosecution evidence.
- The Plaintiff (Mudda'i) starts legal proceedings against another party in civil court cases.
- The Clerk (Katib) performs three main judicial duties which include document management and oath administration and official trial recordkeeping.
- The court accepts testimony from Shuhood who swear to tell the truth about relevant case facts and events.
- The court selects Experts (Khubara') to deliver specialized knowledge about technical matters including forensic science and handwriting analysis.
- The (Shurta) police force maintains courtroom security through their work of transporting defendants and they occasionally act as witnesses in court cases.

The participants in court proceedings depend on specific details from each individual case. The defendant and plaintiff must appear in court for all cases but there are exceptions which allow them to stay absent.

The Iraqi court system demonstrates diglossic characteristics because Modern Standard Arabic (MSA) appears in official legal documents and court decisions, yet Iraqi Arabic dialects dominate witness statements and lawyer-client dialogue and everyday courtroom communication (Radam, 2020, p. 225). The author demonstrates that legal Arabic contains multiple errors because lawyers and judges incorrectly apply Modern Standard Arabic grammar rules and use outdated vocabulary and foreign words from Ottoman and European sources (Radam, 2020, pp. 267–270). The legal system continues to use "طابو" (Tapu) for land registry and "أسبقية" (priority) because these terms stem from historical linguistic patterns which still affect modern legal terminology (p. 270). The study shows that legal professionals and average citizens do not possess enough MSA abilities which produces ambiguous legal texts that result in wrong interpretations (p. 277). The combination of these factors makes it difficult for MST and ChatGPT to achieve accurate legal translation because the tools operate mainly with Modern Standard Arabic while lacking essential knowledge about legal dialects and cultural aspects.

2.4 Microsoft Office and ChatGPT in Focus

The following presentation demonstrates the fundamental principles of MST and ChatGPT systems.

2.4.1. MST in Focus

MST operates as a neural machine translation service which Microsoft developed for use across different platforms that include Windows operating systems and Microsoft Office applications. The system lets users use fundamental multilingual communication features which include instant messaging but it does not work effectively for users who require advanced legal or critical communication needs.

Research studies about MST demonstrate that the system produces acceptable results for official document translation yet requires human assistance to achieve proper translation quality in particular technical domains. The study by Almahasees (2023) showed that MST produced incorrect translations in 52% of institutional documents from the United Nations and the Arab League because it misinterpreted both words and sentence order. The research findings directly apply to courtroom discourse because legal communication needs Modern Standard Arabic formal language and exact legal terminology.

The MST tool demonstrates its most significant flaws when users enter information which does not follow the rules of official language standards. Karosekali et al. (2023) showed that the system relies heavily on literal, word-for-word translation strategies, with minimal use of pragmatic or adaptive techniques. The research analyzed Indonesian folk narratives instead of legal documents yet its results prove useful for methodological purposes because both folk discourse and courtroom speech contain expressions which derive from cultural backgrounds and require knowledge of specific contexts. The system shows a direct link between actual text transfer and its inability to handle discourse which needs both pragmatic inference and contextual modification.

The research results demonstrate that MST achieves its best results when translating simple words and basic sentence structures but it fails to understand intricate discourse patterns. The diglossic nature of Iraqi courtrooms which uses Modern Standard Arabic and Iraqi Colloquial Arabic creates challenges for speakers who need to switch between these two languages while courtroom meaning depends on pragmatic elements and emotional intensity and interruptive speech patterns will most likely lead to incorrect interpretations. The system operates at its best as a summary translation system for legal

settings which use two languages but requires human expert supervision to achieve both accuracy and unbiased courtroom translation results.

Here are the key characteristics found:

- Low use of techniques noticed; that the pragmatic techniques of amplification, generalization and transposition were absent or used little.
- Dependence, on transfer: The tool often used borrowing and calque. Dependence, on transfer shows that the tool relied on source-language forms.
- MST works well for literal translations. MST does not work well with cultural texts such, as folklores. Human translators do better with those texts because human translators have an understanding of context and culture.

2.4.2 ChatGPT in Focus

The translation-focused tool MST operates as a dedicated multilingual processing system but ChatGPT operates as a large language model (LLM) artificial intelligence system. The translation-focused architecture of MST differs from the architecture of ChatGPT which OpenAI developed. The model ChatGPT receives training from a massive internet database which enables it to create text and understand context and produce lengthy written responses. The translation function of ChatGPT develops through its general training process instead of being a specific built-in capability. ChatGPT operates through distinct methods than typical MT systems because it handles complex linguistic operations. ChatGPT functions as a general-purpose tool which enables translation capabilities that traditional MT systems cannot perform but it generates mistakes that expert translators usually avoid. The performance of ChatGPT can be evaluated through multiple recent academic studies which provide essential knowledge for this research (Jiang, et al, 2024). The following studies discussed the use ChatGPT as a translation tool:

- **Jiao et al. (2023), "Is ChatGPT a Good Translator? A Preliminary Study."**

The research provides an initial comparison between ChatGPT and traditional MT systems through its evaluation process. The research shows ChatGPT understands both context and pragmatic elements but it produces fake information and generates inconsistent translations. The research findings from this study match the translation errors which were found in this study.

<https://arxiv.org/abs/2301.08745>

- **Zhang, etal. (2023). “Multilingual large language models are not (yet) code-switchers.”**

The research shows that current multilingual large language models perform poorly when processing code-switched text even though they excel at monolingual tasks. The evaluation of LLMs in zero- and few-shot settings shows that these models achieve worse results than smaller models which received direct fine-tuning on code-switched data for tasks including sentiment analysis and translation and summarization and language identification. The performance difference shows that pretraining multilingual models does not provide them with sufficient ability to handle code-switching tasks. The research results demonstrate a critical weakness in modern multilingual LLMs which requires developers to create specific code-switched resources or training methods or model architectures for effective multilingual processing (Zhu, etal. 2023).

[2023.emnlp-main.774.pdf](#)

- **Khondaker et al. (2023), " GPTAraEval: A Comprehensive Evaluation of ChatGPT on Arabic NLP"**

The introduction of ChatGPT brings about a revolutionary period for NLP which demonstrates its outstanding results on multiple English language evaluation tests. The model shows excellent results on English benchmarks but its performance in different linguistic environments has not been fully explored. The research investigates ChatGPT's Arabic language processing abilities while focusing on its performance with different dialects. The research conducts extensive automated and human assessments of ChatGPT through 44 language understanding and generation tasks across 60+ datasets. The research presents the first detailed assessment of ChatGPT's Arabic NLP performance. The research shows that ChatGPT performs better in English than in Arabic even though smaller models that received Arabic finetuning outperform it. Our research examines the performance difference between ChatGPT and GPT-4 when processing Modern Standard Arabic (MSA) and Dialectal Arabic (DA) to show their weaknesses in Arabic dialect processing compared to MSA. The research investigates GPT-4 as a human evaluation tool but our study contributes to existing research that demonstrates ChatGPT's performance restrictions (Khondaker et al., 2023).

<https://aclanthology.org/2023.emnlp-main.16.pdf>

3. Data Collection and Analysis

3.1 Data Collection

The research study uses speech recordings from Iraqi courtroom videos available on YouTube as its primary data source. The recordings serve as authentic courtroom speech examples because they show natural speaker transitions between Modern Standard Arabic (MSA) and Iraqi Colloquial Arabic (ICA).

The following selection criteria were used to choose the recordings:

1. **Diglossia Presence:** The recordings needed to show natural code-switching between MSA and ICA during authentic courtroom dialogues.
2. **Audio Quality:** The audio recordings needed to have good sound quality to enable proper manual transcription and analysis.
3. **Courtroom Diversity:** The selected recordings needed to show different courtroom participants including judges and witnesses and lawyers and plaintiffs and defendants to display various linguistic functions and speaking approaches to speaking.

The researchers performed manual transcription of all selected recordings. The symbol ($\geq$) with a numerical index (e.g. ≥ 1 , ≥ 2) indicated diglossic shifts in the annotated data. The underlined MSA segments in the text served to distinguish them from ICA segments and to indicate functional language shifts in the discourse.

3.2 Translation Procedure

The evaluation of system performance required both MST (MST) and ChatGPT to process the same input through the following sequence:

- **MST Input:** Researchers play the audio recordings through MST's voice recognition tool to obtain immediate English translations.
- **ChatGPT Input:** The Arabic text from the transcript was fed into ChatGPT to produce an English version of the original content.
- **Data Recording:** The systems produced their results which were used for evaluation purposes.
- **Error Identification:** The researchers conducted a review of their interpretations to detect mistakes that involved language usage and semantic interpretation and pragmatic understanding and diglossic processing.

3.3 Analytical Framework: The research uses qualitative and quantitative error analysis methods which follow established MT evaluation frameworks. The researchers studied each error to determine its effect on legal content because courtroom speech demands exact word choice and unbiased message delivery and clear speaker intention representation.

3.5 Data Analysis

The criteria of analysis is conducted on two levels: **qualitative and quantitative analysis**.

The Qualitative analysis is based on the following criteria:

1. A set of random trials in Iraqi courtrooms available publicly on YouTube are chosen to extract the samples to be studied.
2. The criteria of selecting the samples is based on two points:
 - The sample must contain words not available in the Modern Standard Arabic Dictionaries.
 - The personal linguistic knowledge of the authors in identifying what counts Modern Standard Arabic (MSA) and Iraqi Colloquial Arabic (ICA).
3. Play the audio samples to the models in study, namely ChatGPT and MST.
4. Compare the interpretations of the models to the interpretation of the authors to identify if the models make a full accurate interpretation.
5. Identify the number of the diglossic shifts in the source audio by using the following marks (...<³, <², <¹).
6. Underline the MSA utterances to distinguish them from the ICA.
7. Identify other mistakes if noted.
8. The study is after analyzing the correct interpretation of the shift regardless the accuracy of the interpretation of the whole utterance. That is to say, if an utterance is inaccurately interpreted in general, but the shifts are accurately interpreted; the study counts the shifts as correct. However, a note for the accuracy of each interpretation as whole is added to the end of each analysis.

The Quantitative analysis is based on the following points criteria:

1. Counting the total number of diglossic shifts in the utterances.
2. Dividing the number of the correct shifts interpretations by the total number of shifts to identify the percentage using the following formula:
3. $\text{percentage} = (\text{part} \div \text{whole}) \times 100$
examples: if the result is 45 out of 60
 $\text{percentage} = (45 \div 60) \times 100 = 75\%$

4. Counting the total number of acceptable interpretations for the whole utterance by each model.
5. Generate a comparative chart to compare the results.

Table 1: (4 shifts)

Source Audio	
وطلّعوا المتهم لي بره واني¹ ≤ بلغت الشرطة² ≤ الي موجودين وصار³ ≤ اتصال بمركز الشرطة. هذا على مستوى الحادث⁴ ≤ الي صار. Interpretation: The accused was escorted to outside the building and I reported to the attending police; eventually they reported to the police station. This is on the base of the incident.	
Source Output by ChatGPT	Interpretation
و وطلّعوا ال آ المتهم ريبّر، وأني بلغت الشرطة إليّ موجودين، وصارت اتصالك ب بمركز الشرطة، هذا على مستوى ال ال الحادث إليّ صار.	And they told the accused Reaper, and I informed the police who were there, and you got in touch with the police station, this is at the level of the incident that happened.
Source Output by MST	Interpretation
و وطلّعوا ال آ المتهم آ لبرّا، وأني بلغت الشرطة إليّ موجودين، وصار اتصال بمركز الشرطة، هذا على مستوى ال الحادث إليّ صار	And they took the accused out, and I informed the police who were there, and there was a call to the police station, this is on the level of the incident that happened

Analysis of ChatGPT's output: The ChatGPT failed to capture the source audio properly, and eventually it made some mistakes on producing an accurate interpretation. The first mistake can be noticed in that (لي بره Li Bara) means (to outside), which was interpreted into a name (Reaper) as a name of person. Shifts 1, 2 and 4 were interpreted correctly, but shift 3 was misinterpreted by adding “you got”, which is not stated in the source audio. **Note: the interpretation is not acceptable for the partial failure noted.**

Analysis of MST's output: The MST succeeded to capture all the source audio, and interpreted the whole statement correctly. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

Table 2: (3 shifts)

Source Audio	
ولكن شاء الله أن تتوفي من خلال العملية¹ ≤ ولمن طلعت² ≤ من صالة العمليات هذا المتهم بحضور الناس وجه³ ≤ الي تهديدات كلي انو ما راح تطلعين منا عايشة Interpretation: However, Allah had His will that the patient dies during the surgery. When I got out of the surgery room, this accused man threatened me publically, and told me that I wouldn't leave alive.	
Source output by ChatGPT	Interpretation
ولكن إن شاء الله أن تتوفي من خلال العملية ولما طلعت من صالة العمليات هذا المتهم بحضور الناس وجه إلي تهديدات قللي انو ما راح تطلعين منا عايشة.	But, God willing, you would die during the surgery, and when I came out of the operating room, this accused person, in the presence of people, threatened me, telling me that I would not leave alive
Source Output by MST	Interpretation
ولكن إن شاء الله أنت توفي من خلال العملية، ولمن طلعت من صالة العمليات، هذا المتهم بحضور الناس وجه إلي تهديدات، قال لي إنه ما راح تطلعين من نوع عيشة.	But hopefully, you died during the operation, and when I came out of the operating room, this accused in the presence of people made threats to me, he told me that you are not looking forward to a kind of life.

Analysis of ChatGPT's output: The first underlined section was misunderstood as shown by the table as ChatGPT interpreted it into "God willing, you would die" which is not stated in the source audio. Moreover, shift 1 was also misinterpreted as the person meant changed from 3rd person into 1st person. However, only shift 3 was interpreted correctly. **Note: the interpretation is not acceptable for the partial failure noted.**

Analysis of MST's output: It failed to capture the first underlined section properly, which led to misinterpretation. The second mistake can be spotted in (منا عايشة) (wouldn't leave alive), interpreted as (a kind of life). However, the 3 shifts made in the statements were interpreted correctly. **Note: the interpretation is not acceptable for the partial failure noted.**

Table 3: (4 shifts)

Source Audio	
حضرة القاضي¹ > اني جنت مو بو عيي بذاك الوقت² > فقدت أعز الناس. وطفلة بريئة³ > راحت بهاي الحادثة وجنت كل الي أعرفه⁴ > هم السبب في هذا الشيء Interpretation: Your Honor, I lost my temper at that time as I lost the dearest person to me. An innocent infant died in the incident. All I knew was that they were the cause of this.	
Source output by ChatGPT	Interpretation
حضرة القاضي، أنا كنت مو بو عيي في ذاك الوقت، فقدت أعز الناس، وطفلة بريئة راحت بهاي الحادثة. كنت كل اللي أعرفه هما السبب في هذا الشيء	Your Honor, at that time I was not in my senses. I lost my dearest ones, and an innocent child was lost in this incident. All I knew was that they were the cause of this.
Source Output by MST	Interpretation
حضرة القاضي، أني، كنت مو بو عيي بذاك الوقت، فقدت أعز الناس، و طفلة بريئة راحت بهاي الحادثة، وكنت كل اللي أعرفه هم السبب في هذا الشيء.	Your Honor, Annie, I was conscious at the time, I lost my dearest people, and an innocent child was hurt by this incident, and all I knew were the reason for this.

Analysis of ChatGPT's output: All the shifts in the SL audio were captured and interpreted correctly including the shifts. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

Analysis of MST's output: The first mistake is (اني) which was rendered into a name (Annie). The second section in the source language is negated, but the interpretation made is affirmative, which is incorrect. However, all the shifts were interpreted incorrectly. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

Table 4: (1 shift)

Source Audio	
شنو¹ < جوابك يا صادق؟ هل أنت مذنب أم بريء؟ Interpretation: What is your answer, Sadiq? Are you guilty or innocent?	
Source Output by ChatGPT	Interpretation
شنو جوابك يا صادق؟ هل أنت مذنب أم بريء؟	What is your answer, Sadiq? Are you guilty or innocent?
Source Output by MST	Interpretation
شنو جوابك يا صادق؟ هل أنت مذنب أم بريء؟	What is your answer Sadek? Are you guilty or innocent?

Only 1 shift is identified in the SL audio, however both ChatGPT and MS Translator succeeded to capture and interpret the SL audio properly. **Note: the interpretation is of the whole utterance acceptable for delivering the intended meaning for both models.**

Table 5: (2 shifts)

Source Audio	
بريء سيدي¹ < لأن جنت مو.....² فاقد أعصابي Interpretation: Innocent Your Honor. I was not..., I lost my temper.	
Source output by ChatGPT	Interpretation
بريء سيدي لأنني لم أن فاقد أعصابي	I am innocent Sir, because I did not lose my nerve.
Source Output by MST	Interpretation
بريء سيدي لأن جنت مو فاقد أعصابي	Innocent, Sir, because I am not nervous.

Note: The speaker made a pause (لأن جنت مو.....) and shifted from an incomplete utterance in section 2 to make a new utterance in section 3 as shown in the first interpretation above. This shift requires a special kind of awareness that the speaker wanted to state

something, but paused and made a new statement. The analysis of this case by ChatGPT and MS Translator is as follows:

Both ChatGPT and MS translator failed to capture the shift made in the statement above; eventually led to resulting a negated interpretations (ChatGPT: I did not lose my nerve.) and (MS Translator: I am not nervous) **Note: the interpretation is not acceptable for the partial failure noted.**

Table 6: (1 shift)

Source Audio	
يعني شفته¹ < شهادة عينية بأنه قد سرق؟ Interpretation: Did you witness with your own eyes that he stole?	
Source Output by ChatGPT	Interpretation
فشفته شهادة عينية أنه سرق	I saw an eyewitness testimony that he had stolen.
Source Output by MST	Interpretation
يعني شفته شهدة عينية بأنه قد سرق؟	I mean, did you see him as an eyewitness that he stole?

Note: The statement above is interrogative and requires a recognizing awareness to identify the intonation in the given case.

Analysis of ChatGPT's output: The shift is not captured correctly. Eventually, the interpretation is incorrect. The result is an declarative statement, while the source audio is interrogative. The interpretation is considered inaccurate though it includes the intended information, but in a declarative rather than interrogative one. **Note: the interpretation is not acceptable for the partial failure noted.**

Analysis of MST's output: The shift is correctly interpreted. The interpretation is correct and the MS Translator succeeded to capture the intended tone and meaning; eventually made an accurate interpretation. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

Table 7: (1 shift)

Source Audio	
المشكلة دائما متكررة¹ هي على مود الورث Interpretation: The problem is always recurring. It is about inheritance.	
Source Output by ChatGPT	Interpretation
المشكلة دائما متكررة هي على مود الورث	The problem is always recurring; it is related to inheritance.
Source Output by MST	Interpretation
المشكلة دائما متكررة هي على مود الورث	The problem is always recurring is on the heirloom mode.

Analysis of ChatGPT's output: The interpretation of the shift is accurate. ChatGPT succeeded to capture and interpret the statement in concern. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

Analysis of MST's output: The interpretation of the MS translator is inaccurate, though it captured the source audio correctly, but produced an inaccurate interpretation of the shift in the SL audio. **Note: the interpretation is not acceptable for the partial failure noted.**

Table 8: (1 shift)

Source Audio	
مشكلتهم ¹ .. جت فترة حسين يتأخر هواي عن الحضور الى الدوام فشفتنا وضعه مو تمام	
Interpretation:	
Their problem... for a while, Hussien was always late to work; we realized he is not fine.	
Source Output by ChatGPT	Interpretation
مشكلتهم كانت فترة طويلة يتأخرون باستمرار عن الدوام فشفتنا الوضع مو تمام.	Their issue was that, for a while, they were consistently late to work, so we noticed that the situation wasn't good.
Source Output by MST	Interpretation
مشكلتهم جت فترة حسين يتأخر هواي عن الحضور الى الدوام، فشفتنا وضعه مو تمام.	Their problem came during the period of Hussein, who was very late to attend to work, so we cured his situation.

ChatGPT's output Analysis: in spite of the fact that CatGPT captured this shift correctly and interpreted it accurately, it failed to capture the noun “حسين” in the source audio, and replaced it with “they were”. **Note: the interpretation is not acceptable for the partial failure noted.**

MST's output Analysis: the result shows failure in capturing the shift. In the source audio, there is a pause after the first word (مشكلتهم....), and then the speaker resumes talking (جت فترة حسين يتأخر...). Eventually, misinterpretation occur. **Note: the interpretation is not acceptable for the partial failure noted.**

Table 9: (1 shift)

Source Audio	
سيادة القاضي بعدما اتصلت بالمتهم وبلغني بقدومه في تمام الساعة العاشرة ليلاً ≤ ضليت واكف عالشارع أنتظر بيه، تأخر هواي، ورجعت اتصلت بيه ما جاوبني Interpretation: Your Honor, after I contacted the defendant and he informed me that he would arrive at exactly ten o'clock at night, I remained standing on the street waiting for him. He was very late, and when I called him again, he did not answer.	
Source Output by ChatGPT	Interpretation
سيادة القاضي، بعدما اتصلت بالمتهم وبلغني بقدومه في تمام الساعة العاشرة ليلاً ظلت واقف على الشارع أنتظر بي، تأخره وايه، ورجعت اتصلت بي ما جاوبني	Your Honor, after I called the defendant and he told me he'd arrive at 10 p.m., I waited out on the street for him, but he was very late and then didn't answer my follow-up call.
Source Output by MST	Interpretation
القاضي بعد ما اتصلت بالمتهم وبلغني بقدومه في تمام الساعة 10:00 ليلاً، ظليت واقف ع الشارع أنتظر بيه. تأخر هواي، ورجعت اتصلت بيه ماجاوبني.	After I called the defendant and told me that he was coming at 10:00 p.m., I was standing on the street waiting for him. I was very late, and I called him back and he didn't answer me.

ChatGPT's output Analysis: the Interpretation of the ChatGPT is accurate and delivers the intended meaning. The shift from العاشرة ليلاً to ضليت was captured accurately, and the intended meaning was delivered correctly. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

MST's output Analysis: The interpretation of the shift is correct, and meets the intended meaning. However, the only mistake to be highlighted is the difference of tense between the source and interpretation. The source includes proximity of arrival time in future as seen in the interpretation made by the author, but the tense that MS Translator made is past continuous. Moreover, (تأخر هواي) is (he was very late); instead interpreted into (I was very late), which is incorrect. **Note: the interpretation is not acceptable for the partial failure noted.**

Table 10 (3 Shifts)

Source Audio	
أحنا ننطي فرصة لكل حتى يحجي¹، للمتهم والمدعي ووكلاء الدفاع²، فلا تسولف زايد وخلي الأمور³ تسير بالسياق الصحيح We give everyone the opportunity to speak including the defendant, the claimant, and the defense attorneys; so do not talk excessively and let matters proceed in the proper course.	
Source output by ChatGPT	Interpretation
نحن نعطي فرصة لكل حتى يحكي للمتهم والمدعي ووكلاء الدفاع فلا تسالف زايد وخلي الأمور تصير بالسياق الصحيح	We give everyone a chance to speak—defendant, claimant, and defense attorneys—so don't talk too much and let things proceed in the proper way.
Source Output by MST	Interpretation
إحنا ننطي فرصة لكل حتى يحكي للمتهم والمدعي ووكلاء الدفاع، فلا تسولف زايد وخلي الأمور تصير بالسياق الصحيح.	We give everyone an opportunity to tell the defendant and the prosecutor about the defense attorneys, so don't ask too much and let things fall into the right context.

ChatGPT's output Analysis: All of the shifts were interpreted correctly. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

MST's output Analysis: the MS translator failed in interpreting the first shift, from حتى إلى للمتهم. As the output is incorrectly (to tell the defendant). The second shift, which is from وكلاء الدفاع to فلا تسولف, is also incorrectly interpreted into (the defense attorneys, so don't ask) which is obviously not the intended meaning. However, the MS succeeded to capture and interpret the last shift. **Note: the interpretation is not acceptable for the partial failure noted.**

Table 11: (3 shifts)

Source Audio	
بعد هذا، حدثت مشادة كلامية بيني وبين المتهم بخصوص مبلغ مالي¹، فقام بفشر علي ويغلط هواي واني تنرفزت، والي² <صدر مني أعتقد أي إنسان يحمل ذرة كرامة>³ ما ممكن يتقبل هالشئ After that, a verbal altercation occurred between me and the defendant regarding a sum of money. He began shouting at me and repeatedly insulting me. I became upset, and what resulted from my reaction is something I believe any person with a shred of dignity would not accept	
Source Output by ChatGPT	Interpretation
بعد هذا حدثت مشادة كلامية بيني وبين المتهم بخصوص مبلغ مالي فقام بفشر علي ويغلط هواي واني تنرفزت واللي صدر مني اعتقد اي انسان يحمل ذرة كرامة لا يقبل بهذا الشئ	After that, a verbal altercation happened between me and the defendant about a sum of money, and he started cursing at me and insulting me a lot, so I got angry, and what I did was something I think anyone with a shred of dignity wouldn't accept in that situation.
Source Output by MST	Interpretation
حدثت مشادة كلامية بيني وبين المتهم بخصوص مبلغ مالي، فقام بفشر علي ويغلط هواي، وأني تنرفزت، واللي يصدر مني. أعتقد أي إنسان يحمل ذرة كرامة لا يقبل بهذا الشئ.	There was a verbal argument between me and the defendant about a sum of money, and he started to scold me and make a mistake about me, and that I was nervous, and that I was nervous. I think any human being with an atom of dignity would not accept this thing.

ChatGPT's output Analysis: The ChatGPT successfully captured all the shifts and interpreted them correctly. **Note: the interpretation the whole utterance is acceptable for delivering the intended meaning.**

MST's output Analysis: The MS Translator performed correct interpretations of the shifts in the utterances. However, the interpretation of *يغلط هواي* is inaccurately interpreted as (make a mistake about me) instead of (repeatedly insulting me). **Note: the interpretation is not acceptable for the partial failure noted.**

Table 12: (5 Shifts)

Source Audio	
انت قلت باليوم الي صارت بي الجريمة¹ < ، أنت كنت مع المدعى عليه² <، جنتوا³ في سيارة المدعى عليه باتجاه محافظة ميسان⁴ < . بس لمن شفنا⁵ < الكشف الجنائي، موقع هاتفك مسجل في شركة الاتصال تواجدك في مكان آخر You said that on the day the crime occurred, you were with the defendant, and that you were in the defendant's car heading toward Maysan Province. However, when we reviewed the forensic report, your phone location is marked by the telecommunications company as being in a different location.	
Source Output by ChatGPT	Interpretation
انت قلت باليوم اللي صارت بي جريمة انت كنت مع المدعى عليه كنت في سيارة المدعى عليه باتجاه محافظة ميسان بس لما شفنا الكشف الجنائي موقع هاتفك مسجل في شركة الاتصال تواجدك في مكان آخر	You said that on the day the crime happened, you were with the defendant in his car heading toward Maysan province, but when we checked the forensic report, your phone's location was recorded by the telecom company in a different place."
Source Output by MST	Interpretation
أنت قلت باليوم إللي صارت بيه جريمة، أنت ك أنت كنت مع المدعى عليه؟ في سيارة المدعى عليه باتجاه محافظة ميسان. شفنا الكشف الجنائي. موقع هاتفك مسجل في شركة الاتصال تواجدك في مكان آخر.	You said on the day the crime happened, you were with the defendant? In the defendant's car towards Maysan Governorate. We saw the forensic test. Your phone's location is registered with your carrier and you are located elsewhere.

ChatGPT's output Analysis: The interpretation is accurate and all the shifts and correctly captured. It is to be noted that ChatGPT deleted the repetition in the source (انت) (كنت مع المدعى عليه) and (جنتوا في سيارة المدعى عليه); all merged in one utterance "you were with the defendant in his car); which is acceptable that it does not affect the intended meaning. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

MST's output Analysis: All of the shifts were captured and interpreted correctly. However, the source audio is (أنت كنت مع المدعى عليه.) which is a declarative utterance; but the MS Translator produced an interrogative one (أنت كنت مع المدعى عليه؟) which affects the

direction of the sentence. **Note: the interpretation is not acceptable for the partial failure noted.**

Table 13: (1 shift)

Source Audio	
كل الي طلبته أشوف ابني ، ما أكرر أفهم شلون ¹ < ممكن أب يحرم أم من أطفالها؟	
Source Output by ChatGPT	Interpretation
كل اللي طلبته أشوف ابني ما أقدر أفهم شلون ممكن أب يحرم أم من أطفالها	All I asked for was to see my son. I just can't understand how a father could deprive a mother of her children.
Source Output by MST	Interpretation
كل إللي طلبته أشوف ابني. ما أقدر أفهم شلون ممكن أب يحرم أم من أطفالها؟	All I asked for was to see my son. I can't understand how a father can deprive a mother of her children?

ChatGPT's output Analysis: The shift is interpreted correctly. **Note: the interpretation the whole utterance is acceptable for delivering the intended meaning.**

MST's output Analysis: The shift is interpreted correctly. **Note: the interpretation of the whole utterance is acceptable for delivering the intended meaning.**

4. Key Findings: Results of Data Analysis

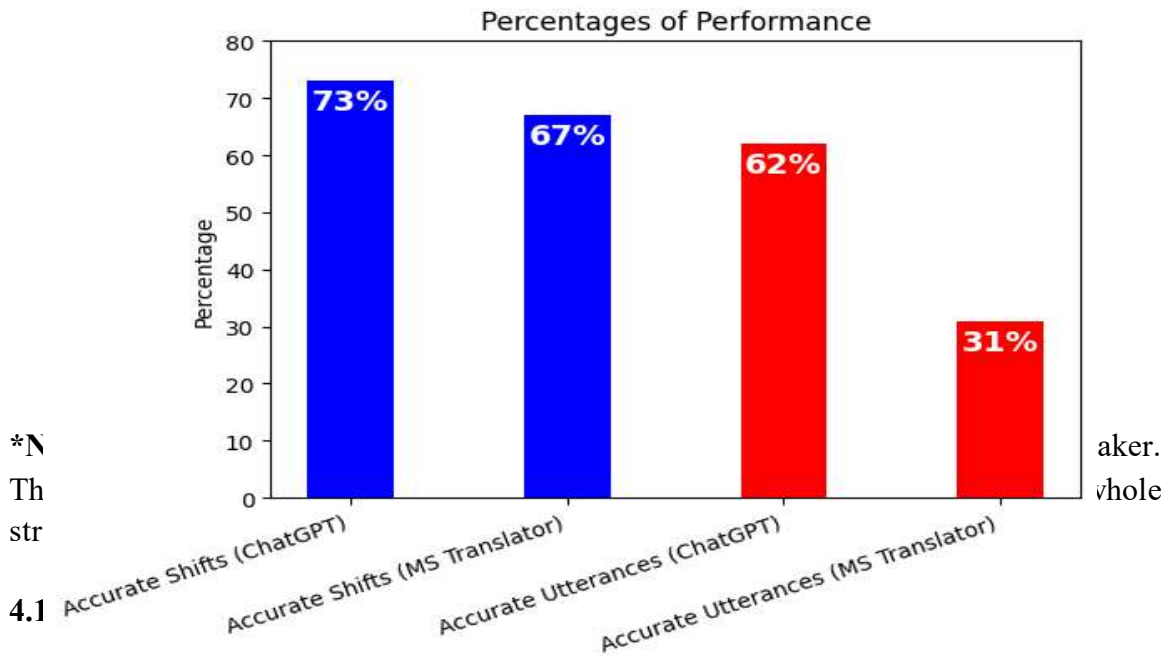
Table1: Statistical Analysis of Diglossic Shifts and Translator Performance

Total No. of Diglossic Shifts and Utterances	No. of Accurate shifts Interpretations (ChatGPT)	No. of Accurate shifts Interpretations (MS Translator)	No. of Accurate Interpretations for Utterances (ChatGPT)	No. of Accurate Interpretations for Utterances (MS Translator)
- Shifts: 30 - Utterances: 13	22	20	8 out of 13	4 out of 13

Table 2: Percentages of Performance

Percentage of Accurate Shifts Interpretations (ChatGPT)	Percentage of Accurate Shifts Interpretations (MS Translator)	Percentage of Accurate Utterances Interpretations (ChatGPT)	Percentage of Accurate Utterances Interpretations (MS Translator)
73%	67%	62%	31%

Figure 2: Percentage of Performance of ChatGPT and Microsoft



ChatGPT:

1. The system failed to recognize speech tone differences between declarative and interrogative statements.
2. The system generated wrong output results when it performed people and pronoun detection on the text input.
3. The system failed to operate correctly when it encountered silent periods and uncompleted statements which happened during spoken dialogues.
4. The system performed a basic translation that did not take into account the current dialogue situation.

MST:

1. The system generated illogical results because it conducted literal character-by-character substitutions.
2. The system failed to detect dialectal expressions which included "مود الورث" and produced "heirloom mode" instead.
3. The system failed to recognize "اني" as "Annie" and it also misread multiple words that share similar features.
4. The system failed to detect both negative statements and statements which were incomplete.

5. Conclusions

- This study investigated the performance of ChatGPT and MST in translating diglossic Iraqi courtroom discourse
- The findings show that Arabic diglossia presents a serious challenge for current AI translation systems, especially in legal settings
- Translation accuracy decreases when speakers shift between Modern Standard Arabic and Iraqi Colloquial Arabic during courtroom interaction
- ChatGPT demonstrated better overall performance than MST in understanding context, speaker intention, and pragmatic meaning
- MST showed relatively better results only when the discourse was formal and close to Modern Standard Arabic
- Both systems showed weaknesses in handling pauses, incomplete utterances, negation, and interrogative versus declarative tone
- Several translation errors observed in the data had the potential to change the legal meaning of witness statements
- These errors pose a serious risk to judicial accuracy and fairness if AI translation tools are used without supervision
- The results confirm that AI translation systems cannot be safely relied upon as independent tools in courtroom environments

- Human translators remain essential to ensure accurate interpretation, legal clarity, and the protection of justice
- Future AI translation systems should be trained on Arabic dialects, code switching, and spoken legal discourse to improve their reliability

References

- Abdul-Mageed, M., Hassan, S., & Elaraby, M. (2024).** A survey of Arabic natural language processing: History, limitations, and opportunities. *Journal of King Saud University – Computer and Information Sciences*, 36(2), 101945. <https://doi.org/10.1016/j.jksuci.2022.101945>
- Almahasees, Z. (2023).** A synchronic assessment of Google Translate, MST and Reverso in English–Arabic translation. *Proceedings of the 3rd International Conference on Language and Education*. <https://doi.org/10.24086/ICLANGEDU2023/paper.931>
- El-Farahaty, H. (2015).** *Arabic–English–Arabic legal translation*. Routledge.
- Fathi, S. (2024).** Dual diglossia in Morocco: A new sociolinguistic situation. *International Journal of Language, Linguistics, Literature and Culture*, 3(4). <https://doi.org/10.59009/jillic.2024.0078>
- Ghawi-Dakwar, O., & Saiegh-Haddad, E. (2025).** Word learning in Arabic diglossia in children with typical language development and developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 68(3), 1533–1551.
- Huang, R., Li, J., & Xiong, D. (2023).** *ChatGPT for speech translation* (arXiv:2309.03409). arXiv. <https://arxiv.org/abs/2309.03409>
- Jabbari, M. J. (2013).** Arabic in Iraq: A diglossic situation. *International Journal of Applied Linguistics and English Literature*, 2(1), 139–150. <https://doi.org/10.7575/ijalel.v.2n.1p.139>
- Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023).** *Is ChatGPT a good translator? A preliminary study* (arXiv:2301.08745). arXiv. <https://arxiv.org/abs/2301.08745>
- Jiang, Z., Lv, Q., Zhang, Z., & Lei, L. (2024).** Convergences and divergences between automatic assessment and human evaluation: Insights from comparing ChatGPT-generated translation and neural machine translation (arXiv:2401.05176v3) arXiv. <https://doi.org/10.48550/arXiv.2401.05176>

- Karosekali, S. C. B., Sipayung, K. T., & Saragi, C. N. (2023). Comparison of translation techniques between human and MST (Windows II) in translating folklores. *Visi Sosial Humaniora*, 5(1), 4–9. <https://ejournal.uhn.ac.id/index.php/humaniora/>
- Khondaker, M. T. I., Waheed, A., Nagoudi, E. B., & Abdul-Mageed, M. (2023). GPTAraEval: A comprehensive evaluation of ChatGPT on Arabic NLP. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 220–247). Association for Computational Linguistics.
- Radam, B. A. H. (2020). A study of linguistic awareness in legal discourse: Iraqi courts as a model. *Journal of the Revival of Arab Scientific Heritage*, (4), 213–295. (Original work published in Arabic)
- Zhang, R., Cahyawijaya, S., Salameh, M., Fung, P., & Abdul-Mageed, M. (2023). Multilingual large language models are not (yet) code-switchers. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.
- Zhu, W., Li, Z., & Hsu, W.-N. (2023). On the code-switching benchmarks of large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 15198–15211). Association for Computational Linguistics. <https://aclanthology.org/2023.emnlp-main.841>