INTERNATIONAL JOURNAL OF SOCIETY, CULTURE & LANGUAGE

Operationalising the Culturemic Framework in Translation: AWWTrans and the Turn from Description to Explanation

Hussain Al Sharoufi[1*]
[1]*Gulf University for Science and Technology, Kuwait*

Abstract Fluent machine translation from Arabic can obscure the loss of culturemes, culture-specific expressions such as divine names and toponyms. This study operationalized the Culturemic Framework, a five-layer model linking cultural schemata to surface form, in the AWWTrans Pro platform and tests its central design commitment: importing source culturemes as transliterated anchors rather than replacing them with target-language equivalents. Sūrat al-Fātiḥa, Sūrat al-Ghāshiya, and the Muʿallaqa opening were generated three times and compared with dated Google Translate captures. The twelve archived outputs were coded item by item for cultureme treatment, register, and structural integrity. Two culturemes consistently separate them: Allāh, retained where the baseline substitutes God, and the toponym Siqṭ al-Liwā, preserved where the baseline paraphrases it away. The Culturemic Fidelity Index (CFI) shows near parity on the Qur'anic texts and an advantage on the poetic text. The contribution is an item-level procedure for measuring cultural loss that string-overlap metrics cannot register; systematic anchoring of the full inventory remains a development objective.

Keywords: *ADAR, Culturemic fidelity index, Translation, Language, Culture*

1. Introduction

In the last 20 years, neural machine translation has yielded outputs that are sometimes difficult to distinguish from a good human translation on the surface. The big question of whether the target sentence will read as well-formed English is now answered by public-facing systems like Google Translate, DeepL, and, more recently, Large-Language-Model (LLM) translators. If one has ever had to work seriously with the Arabic religious, poetic, or legal texts, then one knows that this fluency is a trap. The string decodes; the sense that elicited the original speech has seeped out underfoot. A sentence is given to the reader. The community of practice that wrote the source gets virtually nothing.

The failure looks at first like an engineering problem awaiting more training data or a larger model. It is better read, as Section 2 argues, as the downstream

***Corresponding Author:**
Hussain Al Sharoufi
alsharoufih@gust.edu.kw

Received: August 12, 2026
Revised: October12, 2026
Accepted: October 25, 2026
First Published: November 1, 2026

© 2026 Al Sharoufi.
This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

https://doi.org/10.22034/ijscl.2026.2089679.4697

consequence of a linguistic theory that has long excluded culture from its object of study. The Culturemic Framework (Al Sharoufi & Al Sharoufi, 2026) was developed to address that absence: it models meaning as a causal descent through five interlocking layers, cultural schemata, speech acts, pragmemes, practs, and culturemes, in which the surface form is the terminal manifestation of a deeper cultural process.

The problem the study addresses has two sides. On the empirical side, evaluations of Arabic-English machine translation document persistent failure at the cultural level, fluent output whose religious, poetic, and legal meaning is flattened (Almahasees, 2021; Al Sharoufi & Al-Fadhli, 2025; Zakraoui et al., 2021); these assessments, however, evaluate at sentence level with holistic scales, so the treatment of individual culture-specific items remains unexamined and the cultural loss stays folded inside a global score. On the design side, the framework commits the platform to importing source culturemes as transliterated anchors rather than replacing them with target-language equivalents, and whether the implementation realizes that commitment, realizes it in part, or does not realize it has not been tested on documented, reproducible outputs against a public system on the same texts. If cultural misfires follow from a culture-excluding model, then inserting explicit upper-layer analysis before generation should produce measurable differences in how a system treats culture-specific items; whether it does is the question.

The purpose of the study is accordingly twofold. The first objective is to evaluate the AWWTrans (a translation platform; https://awwtrans.org/) implementation on its own terms, against the framework's design commitments, by generating three culturally dense Arabic texts (Sūrat al-Fātiḥa, Sūrat al-Ghāshiya, and the opening of the Muʿallaqa of Imruʾ al-Qays) on the live platform under a documented protocol, archiving every output verbatim, and comparing the outputs with dated captures from a named public system, Google Translate. The second objective is to supply the item-level instrument that the evaluation literature lacks: coding of culturemе treatment, register, and structural integrity on archived outputs, summarized by the Culturemic Fidelity Index (CFI), a layer-weighted score of how much of the source's cultural profile survives in a rendering, reported also in a completeness-adjusted form (CFI-C) so that truncated output cannot score as if it were complete. Accordingly, two research questions organize the study, formulated so that each is answerable from the coded data.

1. When the outputs of a framework-guided platform and a general-purpose public system are coded item by item against a fixed culturemе inventory, together with register and structural integrity, how do the two systems treat the culture-specific content of the three texts, and on which items do they diverge?
2. What proportion of each text's culturemе inventory does the implemented AWWTrans import as transliterated anchors, and what do the observed points of convergence and divergence with the baseline indicate about the framework's present contribution and its development priorities?

The theoretical orientation remains explanatory rather than descriptive, asking why a community of speakers made an utterance in a particular form and how it may be reconstructed in a second cultural system. The empirical contribution is deliberately more modest than the theoretical one, and the two are kept separate throughout. The corpus is a focused pilot, so the findings characterise the implementation's behavior on these three texts and are not generalizable to Arabic-English translation as a whole.

2. Theoretical Framework

2.1. Culture in Translation Theory and in Arabic-English Machine Translation Evaluation

The cultural turn established culture as constitutive of translational meaning rather than decoration upon it, and Cultural Linguistics formalized the shared mental frameworks, cultural schemata, that allow a community to recognise the meaning of acts that are indeterminate in purely formal terms (Sharifian, 2017; Wierzbicka, 1997). On the applied side, evaluations of Arabic-English machine translation document persistent failure at exactly this level: fluent output whose religious, poetic, and legal meaning is flattened (Almahasees, 2021; Al Sharoufi & Al-Fadhli, 2025; Zakraoui et al., 2021).

There is genuine agreement across functionalist, descriptive, cultural, and computational traditions that context is constitutive of translational meaning (Almahasees, 2021; Hatim & Mason, 1990; Gutt, 2000; House, 2006, 2015; Katan, 2014); there is much less agreement about what context is, and almost no traction on how to integrate the cultural substrate with the surface form in a single traceable chain. The Culturemic Framework, and AWWTrans as its operational instrument, are offered as one principled answer to that second, harder problem, not as a verdict on the first.

On the empirical side, three findings recur. Assessments of neural systems report fluency that outruns adequacy (Castilho et al., 2017; Laubli et al., 2018), and the same question has been pressed on literary text in particular (Toral & Way, 2018). For Arabic-English, Almahasees (2021) compares Google Translate, Microsoft Translator, and Sakhr using holistic adequacy and fluency ratings assigned at sentence level, Zakraoui et al. (2021) survey the open challenges of Arabic machine translation, and Al Sharoufi and Al-Fadhli (2025) locate the recurrent failures at the pragmatic and cultural level. The strategies available at that level are themselves theorised: dynamic equivalence licenses receptor-oriented substitution (Nida, 1964; Nida & Taber, 1969), foreignising importation keeps the reader in contact with the source (Venuti, 1995), and the functionalist tradition makes the choice between them a matter of declared purpose, or skopos (Nord, 2018; Reiss & Vermeer, 2014; Vermeer, 1989). What none of these studies supplies is a procedure that records, item by item, which strategy a system actually applied to each culture-specific expression.

2.2. From the Formalist Inheritance to the Culturemic Framework

Traditional and minimalist generative linguistics have always viewed language as a self-contained formal system, whose focus is the syntactic competence of the idealized speaker (Chomsky, 1965, 1995). Culture and interactional aspects of language use, the aspects that the translator is required to deal with in real working situations every day, were relegated to an extra, and left outside the explanatory programme. The methodological implication is perhaps the most obvious: A model where culture does not belong in the explanatory framework will not produce the cultural aspect of meaning at its termination. The applied offspring of this inheritance are the most consistent systems, neural machine translation (NMT).

They are extraordinarily good at the layer the formal tradition cared about, and they are predictably poor at the layers it did not. Functional and socio-cultural linguistics have, of course, long resisted this restriction. Halliday and Hasan (1976) define language as a social semiotic in which meaning is not intrinsic to linguistic forms but is constructed through their deployment in social and cultural situations. Tomasello (2003) treats language acquisition as primarily an act of cultural learning rather than the maturation of an innate module; Evans and Levinson (2009) press the case from typological evidence. Cultural Linguistics, in the form developed by Sharifian (2017), takes the further step of formalizing shared mental frameworks, cultural schemata, as the substrate that allows speakers of the same community to recognize the meaning of acts that are, in purely formal terms, indeterminate.

What these approaches share is a willingness to treat culture as a basic feature of linguistic meaning rather than decoration on top of it. What they have generally lacked, in my reading, is a vertically integrated model that connects the cultural substrate to the surface lexical-syntactic form in a single traceable causal chain. That gap is what the Culturemic Framework tries to fill. The relevance of this inheritance to the present study is direct. Neural machine translation systems are the most consistent applied descendants of the formalist programme: their objective functions reward string correspondence, so the cultural substrate enters training only insofar as it is statistically reflected in surface co-occurrence. AWWTrans is designed as a test of the opposite ordering: an explicit upper-layer analysis, schema, speech act, pragmeme, and cultureme detection conditioned by a declared skopos, is inserted before generation, and the empirical question of Section 4 is whether that insertion measurably changes the treatment of culture-specific items relative to a system without it. The critique of this subsection is thus not background; it defines the independent variable of the study. What that variable contains, layer by layer, is specified by the Culturemic Framework itself.

The Culturemic Framework (Al Sharoufi & Al Sharoufi, 2026) models meaning as a five-layer causal descent that integrates three classical lines of pragmatic theory into a single vertical chain. Speech-act

theory (Austin, 1975; Leech, 1983) supplies the second layer, in which an utterance is read as the doing of an action under conventional uptake conditions. Mey's (1999, 2010) pragmeme and Capone's (2005) development of the same construct supply the third, situational layer, in which a family of speech acts is licensed by a culturally recognised situational prototype. The framework's original contribution, already foreshadowed in Al Sharoufi (2020) and consolidated in Al Sharoufi and Al Sharoufi (2026), is to place cultural schemata above this pragmatic apparatus and the bidirectional cultureme below it, so that the five layers form a productive chain in which each layer constrains the form of the next (Table 1).

Table 1
The Five Layers of the Culturemic Framework and Their Roles in Translation

Layer	Definition	Role in translation
Cultural schemata	Pre-conscious, culturally distributed conceptual frameworks that organize perception and interpretation.	Define the cognitive horizon inside which the source utterance is intelligible.
Speech acts	Communicative actions (asserting, supplicating, warning, boasting) recognized by the community.	Determine the illocutionary intent that the translation must preserve.
Pragmemes	Situational prototypes that license a family of speech acts in a given context.	Anchor the act to its situation of use.
Practs	Specific surface realisations of a pragmeme in actual discourse.	Provide the concrete linguistic instantiation.
Culturemes	Minimal units of cultural meaning, simultaneously the output of the chain and the imported anchors that re-feed it.	Carry the dense cultural information that surface translation tends to strip.

Note. Layer definitions and translation roles summarized from Al Sharoufi and Al Sharoufi (2026).

Read bottom-up, every observable surface choice can in principle be traced to a culturally distributed cognitive frame; read top-down, every cultural frame leaves a measurable fingerprint in the surface form. The Advanced Discourse Analysis Research (ADAR) engine was designed to make this reading empirically tractable on real corpora, and AWWTrans extends the same architecture to inter-systemic reconstruction. These two directions of reading are what the study operationalises: the top-down direction defines the item-level coding of Section 3.4.2, and the bottom-up direction defines the anchoring commitment tested in Section 4.

A refinement developed in recent work is the bidirectional or embracing theory of culturemes: culturemes are not only the output of the descent but importable anchors that, carried into a target text, can re-feed the target community's schemata. This motivates the design commitment tested throughout this paper: importation of source culturemes as transliterated anchors rather than replacement by target-language equivalents. Bidirectionality itself is a hypothesis of the framework rather than an established property; the present study does not test it.

2.3. Operationalising the Framework in AWWTrans

ADAR operationalises the framework as a diagnostic instrument. In the form used in this study, the Cultural Analysis engine embedded in the AWWTrans platform, ADAR accepts a text in Arabic or English, applies genre-conditioned analysis, and returns detected cultural schemata, pragmemes, and culturemes, each with a confidence value and an interpretive annotation. Its detection component is implemented over a large-language-model backend prompted with the framework's categories; it operates at the level of cultural concepts realized in phrases rather than at the level of individual lexical anchors, and it has no published validation history against independent expert annotation. These properties set expectations for the comparative pilot in Section 4.4, which empirically examines what the instrument can and cannot currently measure.

IN PRESS

The mapping from the framework's concepts to the implemented pipeline is as follows, stated at the level of what the platform demonstrably does. Cultural schemata enter at two points: the genre profile selected at project creation, which activates a domain terminology base, and the free-text cultural-context field of the skopos brief, which conditions generation; the pre-generation analysis additionally reports detected schemata for the source text. Speech acts and pragmemes are handled by the analysis stages that identify the illocutionary and situational functions of source spans, ritual praise, supplication, rhetorical questioning, and these identifications inform the generation prompt. Practs, the concrete realizations, are steered through the register and tone controls of the skopos brief. Culturemes are detected and annotated by the cultureme stage; their importation as transliterated anchors is, in the current implementation, requested through the skopos text rather than enforced by a hard lexical constraint, which is why the evaluation of Section 4 treats anchoring as an empirical outcome to be measured rather than a guaranteed behavior. Two boundaries of this operationalization should be stated plainly. First, detection at every layer is performed by the platform's language-model components conditioned by the framework's categories, not by hand-built symbolic rules; the framework supplies the category system and the conditioning, the model supplies the recognition. Second, there is no explicit cross-linguistic mapping module: the transfer of an identified speech act or pragmeme into a target realization, and the selection among candidate realizations, are delegated to the generation model operating under the analysis-conditioned prompt and the best-of-N selection described in Section 3.3. The framework's influence on translation is therefore exerted through conditioning rather than through rule-governed transfer, and this is precisely why the study's evaluation is behavioral, measuring what the conditioned system does on coded outputs, rather than mechanism-level. This behavioral stance also defines what an adequate evaluation must be able to measure, and it is on exactly this point that the empirical literature falls short.

One methodological property of the empirical literature reviewed in Section 2.1 defines the gap the present study takes up. Almahasees (2021) establishes that the failures occur, but his evaluation design, holistic adequacy and fluency ratings assigned at sentence level, cannot localize a failure to the item or the layer at which it happens: a sentence that translates Allāh as God and one that garbles the syntax lose points on the same scale for different reasons, and the cultural loss remains folded inside a global score. What the literature therefore documents is the existence of cultural failure; what it does not provide is an instrument that identifies which culture-specific items are lost, in which layer, and under which system behavior, nor any test of whether a framework-conditioned system such as AWWTrans treats those items differently from a general-purpose one. Thus, the purpose of this study is to supply both: a fixed cultureme inventory coded item by item on archived outputs, and a documented comparison of the implemented AWWTrans with a public baseline on three texts, from which the two research questions stated in Section 1 are answered.

3. Methodology

3.1. Research Design

The framework's full design specifies a seven-stage, coordinate-bound, deterministic translation process in which culturemes are imported as bidirectional anchors. That design remains the project's horizon; this study reports on the implementation as it exists, states which design components are not yet realized, and assigns them to the development roadmap.

Methodologically, the study is a framework-internal, descriptive, comparative evaluation. Two systems, the implemented AWWTrans Pro and the public Google Translate baseline, are compared on the same three Arabic source texts. Because AWWTrans generation is stochastic, each text was generated three times on AWWTrans and captured once on the baseline, giving twelve archived outputs, and every run is reported rather than a selected best output. Each output is coded on three dimensions: cultureme treatment against a fixed inventory, register, and structural integrity, and the coding is summarized by the CFI with and without its completeness factor. The design tests the implementation against the framework's own design commitments, in particular the importation of source culturemes as transliterated anchors, and does not test the framework's theoretical claims, such as bidirectionality, which remain hypotheses.

3.2. Corpus

Three Arabic texts were selected on stated criteria: Sūrat al-Fātiḥa (liturgical prayer; divine-name culturemes), Sūrat al-Ghāshiya (eschatological narrative; doctrinal lexicon), and the opening of the Muʿallaqa of Imruʾ al-Qays (pre-Islamic poetry; toponyms of the aṭlāl convention). Each is the canonical exemplar of its genre, with a documented translation history that makes departures from source anchoring recognizable, and together they vary the cultureme type at stake while remaining small enough for exhaustive item-level coding of every archived output. The corpus is deliberately a focused pilot: the findings characterise the implementation's behavior on these three texts and are not generalizable to Arabic-English translation as a whole. The Arabic source texts, reproduced at the head of each case in Section 4, are the Qur'anic texts in standard printed orthography and the Muʿallaqa opening in its standard recension.

3.3. Instruments

AWWTrans Pro is a web-based environment organised around translation projects. A project fixes the language pair and a genre profile that activates a domain terminology base (Al-Mawrid, Hans Wehr, and ISESCO resources for the Religious and Theological profile; style-preservation components for Literary and Creative); a four-step Skopos Profiler then records purpose, audience, cultural context, register, and tone, and the completed brief conditions all generation. Translation is produced by an AI generation component presented as a five-stage framework-guided process (skopos, cultural schema, LCT cohesion, pragmeme, cultureme); generation is best-of-N sampling from a stochastic large-language-model backend, displayed with an internal score and screened by a phrase-level collocation check. The study's software instruments were the platform itself, its embedded ADAR engine, the public Google Translate interface for baseline capture, and a structured workbook holding the verbatim output archive and coding record. Three design components are not realized in the implementation: coordinate-bound schema packs, deterministic rendering, and systematic culturemic re-anchoring.

3.4. Procedure

The procedure comprised the generation and capture of the outputs and their coding and analysis (see Figure 1 for the summary of the workflow).

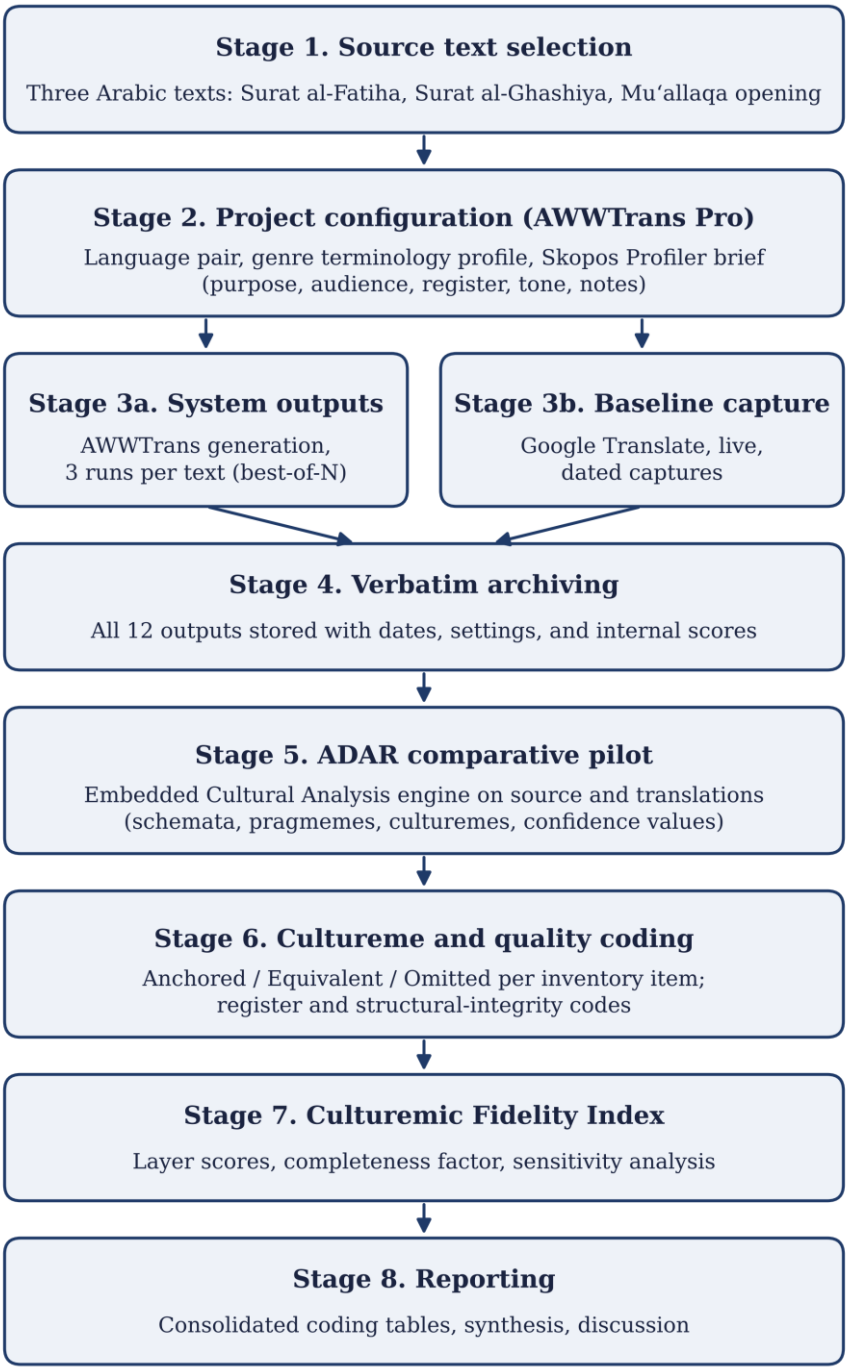
3.4.1. Data Collection

All system outputs were generated on the live platform between 13 July and 5 August 2026 under the protocol summarized in Figure 1. For each text, a dedicated project was created (Religious and Theological profile for the sūras; Literary and Creative for the Muʿallaqa), Arabic to English, with a written skopos requesting reconstruction of the text's cultural-pragmatic meaning, preservation of named culturemes as transliterated anchors, and an elevated register; the register was set to formal and the tone to authoritative throughout. Each text was generated three times in immediate succession; all settings were identical across runs, so the observed variation originates in the stochastic best-of-N backend, not in the protocol. Every output was archived verbatim with its analysis panel and internal score; because generation is stochastic, all runs are reported, with the first structurally complete output of each triplet serving as the reference rendering for excerpting. On the evidence of Section 4.5, variability in wording and register falls within an acceptable range for the study's purposes, since the coded behaviors of interest are stable across runs and every run is reported; structural truncation does not, and it is identified as the implementation's gating engineering issue.

The comparison baseline is Google Translate, captured live with the identical source strings on 5 August 2026 for all three texts. No simulated or reconstructed baseline output is used anywhere in the evaluation: the "legacy NMT" exhibits of earlier drafts, whose generation record could not be documented, are withdrawn and replaced by the captured outputs.

IN PRESS

Figure 1
Workflow of the Study, From Source Selection to Reporting



3.4.2. Data Analysis

The evaluation is framework-internal and descriptive. For each source text, a target cultureme inventory was fixed in advance from the framework's analysis of the text (for the Fātiḥa: Allāh, al-Raḥmān, al-Raḥīm, Rabb, Dīn, al-ṣirāṭ al-mustaqīm). Every output, each AWWTrans run and each baseline capture, was coded item by item on a three-way scheme: A (anchored) when the transliterated form of the item, with or without diacritics, appears in the output; E (equivalent) when the item's span is rendered by a target-language equivalent; O (omitted) when the span corresponding to the item is absent. This portion of the scheme is deterministic, turning on the presence or absence

of specified surface forms, and its application can be verified directly against the verbatim outputs in the Appendix. Register is coded elevated-archaic when archaic address forms or inflections (thee, thou, hath, verb forms in -eth) or scriptural inversion predominate; modern when contemporary address predominates; and mixed when both occur within a single output. Structural integrity is coded complete when every source verse or line is rendered in order in the main output body; truncated when the output terminates before the final source unit; and malformed when rendered content is displaced, misnumbered, or omitted from the body in a non-terminal position. Coding was performed by the research team in a single joint pass, so no independent disagreements arose; the one borderline case encountered, an output combining a modern-register body with an archaic-register footnote, was resolved by the mixed category defined above rather than by ad hoc judgment. No independent second coder was engaged, and no inter-coder agreement statistic is therefore reported. The embedded ADAR engine was additionally applied in a comparative pilot to the Fātiḥa source and both of its translations, to establish the instrument's current scope.

4. Results

Analyses are reported individually for each case, and then combined. The target cultureme inventory, coding of each archived run and baseline capture, and the register and integrity codes are provided for each case. A verbatim output of the full results can be found in the Appendix.

4.1. Case 1: Sūrat al-Fātiḥa

The Arabic source text analyzed is:

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ. الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ. الرَّحْمَنِ الرَّحِيمِ. مَالِكِ يَوْمِ الدِّينِ. إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ. اهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ. صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ.

Three generations of AWWTrans (in each case, 5 August 2026, with an internal score of 90/100) and one Google Translate capture were coded against a six-item inventory. The result of this is reported in Table 2.

Table 2
Cultureme Coding for Sūrat al-Fātiḥa (A = Anchored, E = Equivalent, O = Omitted)

Cultureme	AWW run 1	AWW run 2	AWW run 3	Google
Allāh	A (Allāh)	A (Allāh)	A (Allāh)	E (God)
al-Raḥmān	E	E	E	E
al-Raḥīm	E	E	E	E
Rabb	E (Lord)	E (Lord)	E (Lord)	E (Lord)
Dīn	E (Day of Judgment)	E (Day of Judgment)	E (Day of Judgment)	E (Day of Judgment)
al-ṣirāṭ al-mustaqīm	E (straight path)	E (straight path)	E (straight path)	E (straight path)

Three AWWTrans generations (5 August 2026; internal score 90/100 in each case) and one Google Translate capture were coded against the six-item inventory (Table 2). Anchoring is confined to Allāh, retained in all three runs where the baseline substitutes God; every other item is rendered by an equivalent in all four outputs. Register alternates between elevated (runs 1-2) and modern (run 3), two runs truncate before the final clause, and wording differs across all three runs on identical input, confirming run-level stochasticity.

4.2. Case 2: Sūrat al-Ghāshiya

The Arabic source text analyzed is:

بسم الله الرحمن الرحيم هَلْ أَتَىٰكَ حَدِيثُ ٱلْغَٰشِيَةِ (١) وُجُوهٌ يَوْمَئِذٍ خَٰشِعَةٌ (٢) عَامِلَةٌ نَّاصِبَةٌ (٣) تَصْلَىٰ نَارًا حَامِيَةً (٤) تُسْقَىٰ مِنْ عَيْنٍ ءَانِيَةٍ (٥) لَّيْسَ لَهُمْ طَعَامٌ إِلَّا مِن ضَرِيعٍ (٦) لَّا يُسْمِنُ وَلَا يُغْنِي مِن جُوعٍ (٧) وُجُوهٌ يَوْمَئِذٍ نَّاعِمَةٌ (٨) لِّسَعْيِهَا رَاضِيَةٌ (٩) فِي جَنَّةٍ

عَالِيَةٍ (١٠) لَّا تَسْمَعُ فِيهَا لَٰغِيَةً (١١) فِيهَا عَيْنٌ جَارِيَةٌ (١٢) فِيهَا سُرُرٌ مَّرْفُوعَةٌ (١٣) وَأَكْوَابٌ مَّوْضُوعَةٌ (١٤) وَنَمَارِقُ مَصْفُوفَةٌ (١٥) وَزَرَابِيُّ مَبْثُوثَةٌ (١٦) أَفَلَا يَنظُرُونَ إِلَى ٱلْإِبِلِ كَيْفَ خُلِقَتْ (١٧) وَإِلَى ٱلسَّمَآءِ كَيْفَ رُفِعَتْ (١٨) وَإِلَى ٱلْجِبَالِ كَيْفَ نُصِبَتْ (١٩) وَإِلَى ٱلْأَرْضِ كَيْفَ سُطِحَتْ (٢٠) فَذَكِّرْ إِنَّمَآ أَنتَ مُذَكِّرٌ (٢١) لَّسْتَ عَلَيْهِم بِمُصَيْطِرٍ (٢٢) إِلَّا مَن تَوَلَّىٰ وَكَفَرَ (٢٣) فَيُعَذِّبُهُ ٱللَّهُ ٱلْعَذَابَ ٱلْأَكْبَرَ (٢٤) إِنَّ إِلَيْنَآ إِيَابَهُمْ (٢٥) ثُمَّ إِنَّ عَلَيْنَا حِسَابَهُم (٢٦)

Three AWWTrans generations (5 August 2026; internal scores 80, 75, 80) and one Google Translate capture were coded against a four-cultureme inventory at five positions, with Allāh occurring at two positions (the basmala and verse 24). Table 3 reports the outcome.

Table 3
Cultureme Coding for Sūrat al-Ghāshiya

Cultureme	AWW run 1	AWW run 2	AWW run 3	Google
Allāh (basmala)	E (God)	A (Allāh)	A (Allāh)	E (God)
Allāh (v. 24)	A (Allāh)	A (Allāh)	O (verse missing)	A (Allāh)
ḍarīʿ	E (a bitter plant)	E (bitter thorns)	E (a bitter thorn)	E (a bitter, thorny plant)
Jannah	E (a high paradise)	E (a high garden)	E (a lofty paradise)	E (a lofty Paradise)
al-ʿadhāb al-akbar	E	E	O (verse missing)	E

Three generations (internal scores 80, 75, 80) and one capture were coded at the five positions of the inventory (Table 3). This case adds the structural dimension: run 1 is malformed (most verses displaced to a misnumbered notes list, two verses absent), run 2 is complete, run 3 truncates at verse 22, while the baseline capture is complete and well-formed. The same referent is rendered God in run 1's basmala and Allāh elsewhere, and the baseline capture shows the identical split, a pattern analyzed across systems in Section 4.5; lexically the two systems converge ("lofty paradise" in both). Register is at its least stable in this case, modern in the complete run and mixed across and within the others, including the output whose numbered body is modern while its appended footnote is archaic. The platform's collocation check reported no issues for all three outputs, including the defective ones, and its internal score ranked the malformed run above the complete one; neither internal signal reflects structural quality.

4.3. Case 3: The Muʿallaqa of Imruʾ al-Qays

The Arabic source text analyzed is:

قِفَا نَبْكِ مِنْ ذِكْرَى حَبِيبٍ وَمَنْزِلِ. بِسِقْطِ اللِّوَى بَيْنَ الدَّخُولِ فَحَوْمَلِ. فَتُوضِحَ فَالمِقْراةِ لَمْ يَعْفُ رَسْمُها. لِما نَسَجَتْها مِنْ جَنُوبٍ وَشَمْأَلِ.

Three AWWTrans generations (Literary and Creative profile, 5 August 2026; internal score 85/100 in each case) and one Google Translate capture were coded against the toponymic inventory of the nasīb opening. Table 4 reports the outcome.

Table 4
Cultureme Coding for the Muʿallaqa Opening

Cultureme	AWW run 1	AWW run 2	AWW run 3	Google
Siqṭ al-Liwā	A	A	A	E (the edge of the sand dunes)
al-Dakhūl	A	A	A	A (Ad-Dakhul)
Ḥawmal	A	A	A	A
Tawḍiḥ / al-Miqrāt	A	A	A	A (Tawdih, Al-Miqra)

Three generations (Literary and Creative profile; internal score 85/100 in each case) and one capture were coded against the toponymic inventory (Table 4). The case yields the best culturemic result and

the clearest structural failure. Most toponyms are transliterated by both systems, which is standard practice and non-differentiating, but the baseline dissolves Siqṭ al-Liwā into "the edge of the sand dunes" while AWWTrans anchors it in all three runs; both systems anchor the Tawḍiḥ/al-Miqrāt pair, and AWWTrans renders the dual imperative as an address to companions, with the study's most stable elevated register. All three runs, however, truncate the final hemistich, so no run delivers the complete opening; the baseline capture is complete.

4.4. ADAR Comparative Pilot

The embedded ADAR engine was applied, under the Religious and Theological genre setting, to the Fātiḥa source and both of its translations; Table 5 reports the profiles. The pilot establishes the instrument's current scope in two findings. Detection operates at concept level, crediting the invocation cultureme to "In the name of Allāh" and "In the name of God" alike, so the anchored-equivalent distinction on which the framework's translation theory turns is below the instrument's resolution; and both translations receive richer profiles than the source, which is counter-intuitive if profile size is read as cultural content: the English renderings unpack the compact morphology of the Arabic into a larger number of discrete phrases, each of which can host a detection, and the annotator is more productive on English input, so profile size tracks phrasal segmentation and language sensitivity rather than cultural content, and subtractive loss measurement is not supported in the instrument's present form. ADAR's contribution to the study is accordingly concept-level inventory building and interpretive annotation; the comparative findings rest on the item-level coding of Section 3.4.2, which does register the distinction.

Table 5
ADAR Profiles for the Fātiḥa Source and Its Two Translations

Profile element	Arabic source	AWWTrans (run 2)	Google
Culturemes detected	4	6	7
Pragmemes detected	1	2	2
Cultural schemata	2	2	2
Highest confidence	95% (invocation)	95% (invocation)	90% (invocation)

4.5. Cross-Case Synthesis and Answers to the Research Questions

Table 6 consolidates the coding across all texts, systems, and dimensions, so that the patterns invoked below are verifiable in one place; the item-by-item treatment underlying the anchored counts remains in Tables 2-4, of which Table 6 is the aggregation.

Table 6
Consolidated Coding Across Texts, Systems, and Dimensions

Text/system	Anchored items	Register	Structural integrity	Distinct wordings
Fātiḥa, AWWTrans	1 of 6 (all runs)	elevated (runs 1-2), modern (run 3)	1 complete, 2 truncated	3 of 3 runs
Fātiḥa, Google	0 of 6	modern	Complete	single capture
Ghāshiya, AWWTrans	2 of 5 (reference run)	mixed across and within runs	1 complete, 1 truncated, 1 malformed	3 of 3 runs
Ghāshiya, Google	1 of 5	modern	Complete	single capture
Muʿallaqa, AWWTrans	4 of 4 (all runs)	elevated (all runs)	3 truncated	3 of 3 runs
Muʿallaqa, Google	3 of 4	modern	Complete	single capture

IN PRESS

Three patterns organise Table 6. First, divine-name policy drift, the same referent rendered God in one position and Allāh in another, occurs in both systems: within AWWTrans it appears across runs on the Ghāshiya basmala (God in run 1, Allāh in runs 2-3, against a stable Allāh policy in all three Fātiḥa runs), and within a single output in both systems on the Ghāshiya, AWWTrans run 1 and the Google capture each rendering the basmala as God while rendering verse 24 as Allāh. In frequency, the two systems are comparable on this corpus: each produces exactly one within-text drift instance, both on the Ghāshiya, and each is internally consistent on the Fātiḥa. What the drift reveals about enforcement is precise: both AWWTrans projects named Allāh explicitly in the skopos brief, so the drift in one Ghāshiya run demonstrates that prompt-level conditioning reduces the inconsistency, the Fātiḥa policy having held in all three AWWTrans runs (Table 2), but does not guarantee it; consistent anchoring of even a single high-salience cultureme requires the hard lexical constraint identified in the development roadmap, and the framework's present contribution is that its coding scheme makes the drift visible and locates it at the schema layer, where a declared naming policy could govern it. Second, run-to-run variability is not uniform across dimensions: anchoring decisions are the most stable behavior (Allāh anchored in three of three Fātiḥa runs; all four Muʿallaqa toponyms anchored in three of three runs; only the Ghāshiya basmala varies), register is stable in one case of three, structural integrity is defective in seven of nine runs, and surface wording differs in nine of nine runs. The system is, in other words, most consistent exactly where the framework conditions it lexically and least consistent at the surface, which is the profile expected of skopos-conditioned stochastic generation. Third, structural reliability favours the baseline without exception: all three Google captures are complete, and the platform's internal score and collocation check detect none of the seven defective runs.

The research questions are answered directly from Table 6. RQ1: the two systems converge on most of the inventory, rendering doctrinal and theological culturemes by equivalents, and diverge at exactly two items, Allāh in the Fātiḥa and Siqṭ al-Liwā in the Muʿallaqa, both of which AWWTrans anchors in every run and the baseline does not; the systems additionally diverge in register, elevated in most AWWTrans runs and modern in every capture of the baseline, and in structural integrity, where the relation is reversed. RQ2: the implemented AWWTrans anchors one of six inventory items on the Fātiḥa, two of five on the Ghāshiya reference run, and four of four on the Mu͟allaqa, so the design commitment of systematic anchoring is realised for proper names and for the highest-salience divine name, and not realized for the remaining doctrinal lexicon; the divergences indicate that the framework's conditioning measurably shifts anchoring where the item is lexically enforceable, and that the development priorities are an enforced anchor lexicon covering the full inventory, a completeness gate, and a deterministic generation mode.

4.6. The Culturemic Fidelity Index (CFI)

The CFI summarizes the coding as the layer-weighted congruence between the source profile and a target rendering, and its completeness-adjusted form (CFI-C) multiplies that congruence by the fraction of the source rendered:

$$CFI = w_1 \cdot C(\text{Schemata}) + w_2 \cdot C(\text{SpeechActs}) + w_3 \cdot C(\text{Pragmemes}) + w_4 \cdot C(\text{Practs}) + w_5 \cdot C(\text{Culturemes})CFI\text{-}C = \text{Comp} \times CFI$$

where Comp is the fraction of source verses or lines rendered in the output body and the weights are $w_1 = 0.30$, $w_2 = 0.25$, $w_3 = 0.20$, $w_4 = 0.10$, $w_5 = 0.15$. Structural completeness is a core multiplicative component, so an output cannot compensate at the layers it renders for the source it fails to render; the index is reported without the factor (CFI) for layer comparability and with it (CFI-C) as the headline value. The weights encode the framework's causal-descent ordering and are theoretically motivated but empirically uncalibrated; the index is therefore a structured summary of the coding, not a validated quality measure, and its calibration against expert human ratings is a stated requirement of the research agenda. A sensitivity analysis (Table 7) recomputes the index under uniform and cultureme-weighted schemes; the system ranking is unchanged under all three, so the comparative findings do not depend on the particular weights.

Table 7
Sensitivity of the Index to the Weighting Scheme (CFI Without Completeness Factor)

Case/system	Original weights	Uniform weights	Cultureme-weighted
Fātiḥa, AWWTrans	0.58	0.53	0.44
Fātiḥa, Google	0.55	0.50	0.38
Ghāshiya, AWWTrans	0.71	0.68	0.61
Ghāshiya, Google	0.68	0.64	0.53
Muʿallaqa, AWWTrans	0.90	0.80	0.90
Muʿallaqa, Google	0.69	0.75	0.73

Layer scoring is operationalised as follows, with each criterion tied where possible to the codes of Section 3.4.2 so that judgment is minimized. The cultureme layer is the anchored fraction of the case inventory from Tables 2-4, a deterministic count. For the schemata layer, the key anchors of a case are defined in advance as the inventory items that carry the case's governing frame: for the two sūras, the divine-name and divine-attribute set (Allāh, al-Raḥmān, al-Raḥīm, Rabb), and for the Muʿallaqa, the toponym set with Siqṭ al-Liwā as the schema-central item of the aṭlāl frame; appropriate register is determined by the operational register code of Section 3.4.2, not by impression. A rendering scores 1.0 when the majority of the key anchors, including the schema-central item, are anchored, and the register code is elevated; 0.5 when the frame is recoverable but the key anchors are predominantly rendered as equivalents or the register code is modern or mixed; 0.0 when the frame is not recoverable. The speech-act and pragmeme layers are scored against the functions identified per case in Section 4 (1.0 when the function is identifiable in its source form, 0.5 when present but weakened, as when the dual address is flattened to a first-person plural, 0.0 when absent), and the practs layer combines the structural codes (1.0 complete and unmixed in register, 0.5 truncated, malformed, or register-mixed, 0.0 absent). Applied to the reference renderings and baseline captures, these criteria reproduce the scores of Table 8 deterministically at the cultureme and practs layers; residual judgment at the upper layers was exercised by the research team. Table 8 reports the layer scores, completeness, and both index values.

Table 8
CFI Layer Scores, Completeness, and Index Values

Case/system	Schemata	Speech acts	Pragmemes	Practs	Culturemes	Comp	CFI	CFI-C
Fātiḥa, AWWTrans (run 2)	0.50	1.00	0.50	0.50	0.17	1.00	0.58	0.58
Fātiḥa, Google	0.50	1.00	0.50	0.50	0.00	1.00	0.55	0.55
Ghāshiya, AWWTrans (run 2)	0.50	1.00	1.00	0.50	0.40	1.00	0.71	0.71
Ghāshiya, Google	0.50	1.00	1.00	0.50	0.20	1.00	0.68	0.68
Muʿallaqa, AWWTrans (run 1)	1.00	1.00	1.00	0.00	1.00	0.88	0.90	0.79
Muʿallaqa, Google	0.50	0.50	1.00	1.00	0.75	1.00	0.69	0.69

On the Qur'anic texts, the index places AWWTrans marginally above the baseline under all three weightings, with the margin carried almost entirely by the anchored fraction of the cultureme layer. On the poetic text, the completeness factor does the work it was introduced to do: the Muʿallaqa AWWTrans score falls from 0.90 to 0.79 because every run truncates the final hemistich, so the headline value registers the structural defect the unadjusted index concealed, and under uniform

IN PRESS

weights completeness adjustment reverses the unadjusted ranking, confirming that completeness cannot sit outside the index. The larger poetry advantage admits two explanations that the present corpus cannot separate: the framework's conditioning is most enforceable exactly where culturemes are proper names, the toponym-dense aṭlāl opening being the ideal case, and the baseline is independently weakest on the same material, dissolving a rare pre-Islamic toponym into a common-noun paraphrase, plausibly for reasons of training-data sparsity. Both explanations predict the observed pattern; the ablation and larger-corpus designs of Section 5 are required to distinguish them.

5. Discussion

This study set out to operationalize the Culturemic Framework (Al Sharoufi & Al Sharoufi, 2026) in a working translation platform and to evaluate the implementation against the framework's own design commitments under a fully documented and reproducible protocol. Its objective was deliberately double: to test whether an explicit upper-layer analysis, schemata, speech acts, pragmemes, and culturemes, conditioned by a declared skopos (Nord, 2018; Reiss & Vermeer, 2014; Vermeer, 1989), measurably changes the treatment of culture-specific items relative to a general-purpose public system, and to develop an item-level procedure for measuring cultural loss that the standard evaluation repertoire of Arabic-English machine translation research (Almahasees, 2021; Zakraoui et al., 2021) does not provide. Twelve archived outputs across three culturally dense texts, coded exhaustively for cultureme treatment, register, and structural integrity, constitute the evidence base, and every claim advanced below is traceable to the consolidated coding of Table 6. In answer to RQ1, the two systems converge on most of the inventory, rendering doctrinal and theological culturemes as equivalents, and diverge at exactly two items, Allāh in the Fātiḥa and Siqṭ al-Liwā in the Muʿallaqa, which AWWTrans anchors in every run and the baseline does not. They diverge further in register, elevated in most AWWTrans runs and modern in every baseline capture, and in structural integrity, where the baseline is more reliable. In answer to RQ2, the implemented AWWTrans imports one of six inventory items on the Fātiḥa, two of five coded positions on the Ghāshiya reference run, and four of four on the Muʿallaqa, so the anchoring commitment is realized for proper names and for the highest-salience divine name and not for the remaining doctrinal lexicon. The development priorities that follow are an enforced anchor lexicon covering the full inventory, a completeness gate, and a deterministic generation mode.

The first finding the evidence supports is that the layer at which cultural fidelity is decided is invisible to surface evaluation, and this carries a direct implication for evaluation practice. The AWWTrans and Google renderings of the Fātiḥa are close paraphrases at the propositional level; string-overlap metrics of the Bilingual Evaluation Understudy (BLEU) family (Papineni et al., 2002) or character-level measures such as chrF (Popovic, 2015) would score them nearly identically, yet the item-level coding shows they differ at exactly the point, Allāh against God, at which the reader is either anchored in the source community's schematic field or relocated into a generic monotheistic frame. This is the measurable form of a failure long described qualitatively in the literature on culture in translation (Bassnett & Lefevere, 1990; Faiq, 2004; Hatim & Mason, 1997; Katan, 2014) and in the assessments of neural systems whose fluency outruns their adequacy (Castilho et al., 2017; Laubli et al., 2018; Toral & Way, 2018). The conclusion is not that overlap metrics should be abandoned but that they are structurally insufficient for culturally dense text: evaluation of such material requires a cultureme-level layer, of the kind operationalised here through inventory coding, alongside string metrics and the human assessment that quality frameworks demand (House, 2015). The framework's vocabulary (Capone, 2005; Mey, 2010; Sharifian, 2017) is what made the difference codable, and in that respect its diagnostic contribution stands independently of the platform's generation performance. What the coding adds to the understanding of cultural meaning in translation is that, on texts of this kind, the loss is not diffuse: it localises to discrete, enumerable items whose treatment can be read directly off the surface, and that localisability is what makes cultural fidelity measurable at all rather than a matter of holistic impression.

The second finding is that the implementation anchors selectively rather than systematically, and the pattern of that selectivity is informative once alternative explanations are confronted. AWWTrans anchored Allāh in every Fātiḥa run and every Muʿallaqa toponym in every run, while rendering the

remaining doctrinal lexicon by equivalents exactly as the baseline does; anchoring succeeded, in other words, precisely where the culturemeis a proper name that the skopos conditioning can enforce lexically, and failed where importation would require the generator to override its distributional preference for idiomatic equivalents. Classified in these terms, the failure is an engineering limitation rather than a theoretical one: the framework specifies the constraint and the implementation requests it through the prompt instead of enforcing it as a hard lexical rule, with the backend's training-data distribution as the force the missing enforcement would have to override; nothing in the results indicates that the framework's account of what should be anchored is itself deficient. One alternative explanation, that the observed differences reflect training-data distributions rather than the framework's conditioning, cannot be excluded on a three-text corpus and is stated as such; what the archive does show is that the anchoring behavior is the most stable dimension across stochastic runs of the same backend, which is the signature of a conditioning effect rather than of sampling accident, and the decisive test, an ablation of the same platform with and without the skopos analysis, is specified as future work. A second alternative reading, that God for Allāh is a legitimate receptor-oriented choice rather than a failure, is accepted in its own terms: dynamic equivalence licenses it (Nida, 1964; Nida & Taber, 1969), and foreignising importation is one strategy among others (Venuti, 1995). The framework's claim, sharpened by the evidence, is narrower than a verdict on either strategy: the choice between them should be a declared, schema-governed policy, and what the coding actually documents in both systems is the absence of any policy, the same referent drifting between God and Allāh within a single output or across runs. The failure the study measures is ungoverned drift, not domestication.

The implications extend in three directions. For evaluation methodology, the study contributes a reproducible procedure, fixed culturemeinventories, item-level anchored/equivalent/omitted (A/E/O) coding on archived outputs, per-run reporting for stochastic systems, and a completeness-adjusted layer-weighted index that can be applied to any language pair in which culture-specific anchors are at stake, and it demonstrates on a documented case why single-output reporting under-describes stochastic translation systems. For system design, the findings convert directly into engineering requirements: an enforced anchor lexicon applied as a hard generation constraint, a structural completeness gate, since seven of nine runs shipped defective output that the platform's internal quality signals did not detect, and a deterministic or seed-reported generation mode as a precondition for scholarly citation of system output. For translation theory, the study offers a working demonstration that the cultural turn (Munday, 2016; Snell-Hornby, 2006) can be carried past programmatic statement into measurement: the vertical model connecting schema to surface, whatever its ultimate empirical fate, is here shown to generate falsifiable, item-level predictions about system behavior, which is the explanatory standard the framework set for itself.

The limitations mark the boundaries of these claims and define the research agenda: expansion of the corpus across genres and language pairs with double coding and reported inter-coder agreement; blind human evaluation of adequacy, fluency, and cultural fidelity (House, 2015; Laubli et al., 2018), against which the CFI can be calibrated; an ablation study isolating the skopos conditioning from the backend's training distribution; extension of ADAR to anchor-level detection with independent expert validation; comparison with contemporary large-language-model translators under the identical protocol (Castilho et al., 2017; Toral & Way, 2018); and an empirical test of the bidirectionality hypothesis (Kecskes, 2014; Sharifian, 2017). The acceptance test used here, three texts, three runs per text, item-level coding on archived outputs, is the verification instrument for each of these steps. The wider premise of the programme, that translation is better modeled as the reconstruction of a cultural process than as the mapping of strings, closes this study with its diagnostic half supported by documented evidence and its generative half given a defined, testable path.

Acknowledgement

The author used an AI-based assistant in the preparation and editing of this manuscript and in generating Figure 1. The study design, the platform and framework evaluated, the data, and all interpretations and claims are the author's own, and the author takes full responsibility for the content

of the paper. The author would like to express sincere gratitude to Gulf University for Science and Technology for its support and encouragement throughout this research.

Disclosure Statement

The author claims no conflict of interest.

Funding

The research was funded by Gulf University for Science and Technology (No. 228).

Data Availability

All outputs analysed in this study, the three AWWTrans runs and the Google Translate capture for each of the three texts, are reproduced verbatim in the Appendix together with their generation dates and internal scores. The coding derived from them is reported in full in Tables 2-8. No further data were generated.

Ethics Declaration

The study analysed machine-generated translations of published texts and involved no human participants, personal data, or animal subjects. Ethical approval was therefore not required.

References

Almahasees, Z. (2021). *Analysing English--Arabic machine translation: Google Translate, Microsoft Translator and Sakhr*. Routledge.

Al Sharoufi, H. (2020). Culturemes in critical discourse analysis. *Studies in Pragmatics and Discourse Analysis*.

Al Sharoufi, H., & Al-Fadhli, W. S. (2025). Bridging the gap: Pragmatic and cultural challenges in machine translation. *International Journal of Society, Culture & Language*, *13*(2), 29-43. https://doi.org/10.22034/ijscl.2025.2062567.4066

Al Sharoufi, H., & Al Sharoufi, Z. (2026). A conceptual culturemic framework for analyzing cross-cultural miscommunication. *International Journal of Society, Culture & Language*, *14*(2), 75-90. https://doi.org/10.22034/ijscl.2026.2087259.4471

Austin, J. L. (1975). *How to do things with words* (2nd ed.). Oxford University Press.

Bassnett, S., & Lefevere, A. (Eds.). (1990). *Translation, history and culture*. Pinter.

Capone, A. (2005). Pragmemes (a study with reference to English and Italian). *Journal of Pragmatics, 37*(9), 1355-1371. https://doi.org/10.1016/j.pragma.2005.01.013

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, *108*(1), 109-120. https://doi.org/10.1515/pralin-2017-0013

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Chomsky, N. (1995). *The minimalist program*. MIT Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429-448. https://doi.org/10.1017/S0140525X0999094X

Faiq, S. (2004). *Cultural encounters in translation from Arabic*. Multilingual Matters.

Gutt, E. A. (2000). *Translation and relevance: Cognition and context* (2nd ed.). St. Jerome Publishing.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Hatim, B., & Mason, I. (1990). *Discourse and the translator*. Longman.

Hatim, B., & Mason, I. (1997). *The translator as communicator*. Routledge.

House, J. (2006). Text and context in translation. *Journal of Pragmatics*, *38*(3), 338-358. https://doi.org/10.1016/j.pragma.2005.06.021

House, J. (2015). *Translation quality assessment: Past and present*. Routledge.

Katan, D. (2014). *Translating cultures: An introduction for translators, interpreters and mediators* (3rd ed.). Routledge.

Kecskes, I. (2014). *Intercultural pragmatics*. Oxford University Press.

IN PRESS

Laubli, S., Sennrich, R., & Volk, M. (2018). Has machine translation achieved human parity? A case for document-level evaluation. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), *The 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4791-4796). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1512

Leech, G. (1983). *Principles of pragmatics*. Longman.

Mey, J. L. (1999). *When voices clash: A study in literary pragmatics*. Mouton de Gruyter.

Mey, J. L. (2010). Reference and the pragmeme. *Journal of Pragmatics*, *42*(11), 2882-2888. https://doi.org/10.1016/j.pragma.2010.06.009

Munday, J. (2016). *Introducing translation studies: Theories and applications* (4th ed.). Routledge.

Nida, E. A. (1964). *Toward a science of translating: With special reference to principles and procedures involved in Bible translating*. Brill.

Nida, E. A., & Taber, C. R. (1969). *The theory and practice of translation*. Brill.

Nord, C. (2018). *Translating as a purposeful activity* (2nd ed.). Routledge. https://doi.org/10.4324/9781351189354

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. In P. Isabelle, E. Charniak, & D. Lin (Eds.), *The 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311-318). Association for Computational Linguistics. https://doi.org/10.3115/1073083.1073135

Popović, M. (2015). chrF: Character n-gram F-score for automatic MT evaluation. In O. Bojar, R. Chatterjee, C. Federmann, B. Haddow, C. Hokamp, M. Huck, V. Logacheva, & P. Pecina (Eds.), *The Tenth Workshop on Statistical Machine Translation* (pp. 392-395). Association for Computational Linguistics. https://doi.org/10.18653/v1/W15-3049

Reiss, K., & Vermeer, H. J. (2014). *Towards a general theory of translational action: Skopos theory explained* (C. Nord, Trans.). Routledge.

Sharifian, F. (2017). *Cultural linguistics: Cultural conceptualisations and language*. John Benjamins Publishing Company. https://doi.org/10.1075/clscc.8

Snell-Hornby, M. (2006). *The turns of translation studies: New paradigms or shifting viewpoints?* John Benjamins Publishing Company. https://doi.org/10.1075/btl.66

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text? In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment* (pp. 263-287). Springer. https://doi.org/10.1007/978-3-319-91241-7_12

Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.

Vermeer, H. J. (1989). Skopos and commission in translational action. In A. Chesterman (Ed.), *Readings in translation theory* (pp. 173-187). Oy Finn Lectura Ab.

Wierzbicka, A. (1997). *Understanding cultures through their key words*. Oxford University Press.

Zakraoui, J., Saleh, M., Al-Maadeed, S., & Alja'am, J. M. (2021). Arabic machine translation: A survey with challenges and future directions. *IEEE Access*, *9*, 161445-161468. https://doi.org/10.1109/ACCESS.2021.3132488

Appendix

Archived Outputs (Verbatim)

All AWWTrans outputs were generated on the platform on 5 August 2026 under the protocol of Section 3.4.1; Google Translate captures are dated 5 August 2026. Truncation and formatting are reproduced as generated.

1. Fātiḥa, AWWTrans run 1 (score 90/100; truncated)

In the name of Allāh, the Most Gracious, the Most Merciful. Praise be to Allāh, the Lord of all worlds. The Most Gracious, the Most Merciful. Master of the Day of Judgment. Thee alone do we worship, and Thee alone do we ask for help. Guide us to the straight path, the path of those upon whom Thou hast bestowed Thy blessings, not of those who have incurred [Thy] wrath, nor of those who have [output ends]

IJLRES

2. Fātiḥa, AWWTrans run 2 (score 90/100; complete; reference rendering)

In the name of Allāh, the Most Gracious, the Most Merciful. All praise is due to Allāh, the Lord of all worlds, the Most Gracious, the Most Merciful, Master of the Day of Judgment. Thee alone do we worship, and from Thee alone do we seek assistance. Guide us on the straight path, the path of those upon whom Thou hast bestowed Thy blessings, not of those who have incurred Thy wrath, nor of those who have gone astray

3. Fātiḥa, AWWTrans run 3 (score 90/100; truncated)

In the name of Allāh, the Most Gracious, the Most Merciful. Praise be to Allāh, the Lord of all worlds. The Most Gracious, the Most Merciful. Master of the Day of Judgment. It is You alone we worship, and it is Your aid we seek. Guide us on the straight path, the path of those upon whom You have bestowed Your favors, not of those who have incurred Your wrath, nor of those who have [output ends]

4. Fātiḥa, Google Translate

In the name of God, the Most Gracious, the Most Merciful. All praise is due to God, Lord of the worlds. The Most Gracious, the Most Merciful. Master of the Day of Judgment. You alone we worship, and You alone we ask for help. Guide us to the straight path. The path of those upon whom You have bestowed favor, not of those who have incurred Your wrath, nor of those who are astray.

5. Ghāshiya, AWWTrans run 1 (score 80/100; malformed)

Main body: In the name of God, the Most Gracious, the Most Merciful. Have you received the tidings of the Overwhelming? [1] (2) Laboring, weary. [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] (16) Do they not then observe the camels, how they are created? (17) And the sky, how it is raised high? Remaining verses displaced to an appended notes list with offset numbering; verses 18-19 absent.

6. Ghāshiya, AWWTrans run 2 (score 75/100; complete; reference rendering)

In the name of Allāh, the Most Gracious, the Most Merciful. Have you received the news of the Overwhelming Event? Some faces on that Day will be humbled, toiling and weary, as they enter a blazing fire. They will be given to drink from a boiling spring, and there will be no food for them except for bitter thorns, which neither nourish nor satisfy hunger. On that Day, some faces will be joyous, pleased with their endeavors, in a high garden where idle talk will not be heard. Therein is a flowing spring, raised couches, set goblets, arranged cushions, and rich carpets spread out. Do they not observe how the camels are created, how the heaven is raised, how the mountains are firmly set, and how the earth is spread out? So remind, for your task is only to remind; you are not a controller over them. Except for those who turn away and disbelieve, Allāh will punish them with the greatest punishment. Indeed, their return is to Us, and it is for Us to call them to account.

7. Ghāshiya, AWWTrans run 3 (score 80/100; truncated)

In the name of Allāh, the Most Gracious, the Most Merciful. Has the story of the Overwhelming Event reached you? (1) On that Day, faces will be humbled, (2) laboring and weary, (3) scorched by a blazing fire, (4) and given drink from a boiling spring. (5) They will have no food except for a bitter thorn, (6) which will neither nourish nor alleviate hunger. (7) On that Day, faces will be joyous, (8) content with their past efforts, (9) in a lofty paradise, (10) where they will hear no idle talk. (11) In it is a flowing spring, (12) raised couches, (13) goblets set, (14) cushions arrayed, (15) and rich carpets spread out. (16) Will they not then consider the camels, how they are created? (17) And the sky, how it is raised? (18) And the mountains, how they are rooted? (19) And the earth, how it is spread? [20] (21) You are not a controller over them, (22) except for those who [output ends; verse 20 displaced to a footnote; verses 23-26 absent]

8. Ghāshiya, Google Translate

In the name of God, the Most Gracious, the Most Merciful. Has there come to you the account of the overwhelming event? Some faces, that Day, will be downcast, toiling, weary, entering a blazing Fire. They will be given to drink from a boiling spring. They will have no food except from a bitter, thorny plant, which neither nourishes nor satisfies hunger. Some faces, that Day, will be radiant, pleased with their striving. In a lofty Paradise. You will not hear therein any ill speech. Therein is a flowing spring. Therein are thrones raised high. And goblets placed ready. And cushions lined up. And carpets spread out. Do they not look at the camels, how they were created? And at the sky, how it was raised? And at the mountains, how they were firmly set? And to the earth, how it was spread out. So remind, for you are only a reminder. You are not a controller over them. Except for those who turn away and disbelieve, then Allāh will punish them with the greatest punishment. Indeed, to Us is their return. Then indeed, upon Us is their reckoning.

9. Muʿallaqa, AWWTrans run 1 (score 85/100; truncated; reference rendering for coding)

Stand now! Let us weep for the memory of a beloved and a dwelling amidst the sands of Siqt al-Liwā, situated between al-Dakhul and Hawmal. The ruins of Tawdiha and al-Miqrāt still endure, undisturbed by [output ends]

10. Muʿallaqa, AWWTrans run 2 (score 85/100; truncated)

Stand with me; let us weep in remembrance of a beloved and cherished dwelling place, by the lowland of Siqt al-Liwā, between al-Dakhul and Hawmal. The site of Tawdhih and al-Miqrāt still bears undiminished traces, for the [output ends]

11. Muʿallaqa, AWWTrans run 3 (score 85/100; truncated)

Stand with me; let us weep for the memory of a beloved and a dwelling at Siqt al-Liwā, situated between al-Dakhul and Hawmal. The markings at Tawdiha and al-Miqrāt remain intact, as the winds from the south and [output ends]

12. Muʿallaqa, Google Translate

Let us pause and weep for the memory of a beloved and a dwelling place. At the edge of the sand dunes between Ad-Dakhul and Hawmal. Then Tawdih and Al-Miqra, whose traces have not faded. Because of what the south and north winds have woven around them.

IN PRESS