

**Contextual Sensitivity in Human VS. Machine Translation:  
An Analysis of Idiomatic and Cultural Expression**

الحساسية السياقية في الترجمة البشرية مقابل الترجمة الآلية: تحليل التعبير  
الاصطلاحي والثقافي

م.م. مريم احمد محمد سعيد

**Maryam Ahmed Mohamed Saeed**

قسم تربية اربيل المؤقت/ ثانوية التعايش للنبات

المديرية العامة للتربية في محافظة الأنبار  
شعبة البحوث والدراسات التربوية

Email: [memoalrawi970@gmail.com](mailto:memoalrawi970@gmail.com)

**Keywords: Contextual sensitivity, idiomatic expression,  
cultural mediation, neural machine translation, English–  
Arabic translation**

الكلمات المفتاحية: الحساسية السياقية، التعبير الاصطلاحي، الوساطة الثقافية،  
الترجمة الآلية العصبية، الترجمة من الإنجليزية إلى العربية

## ملخص

تهدف هذه الدراسة إلى مقارنة حساسية الترجمة البشرية والترجمة الآلية العصبية (NMT) للسياق في ترجمة التعبيرات الاصطلاحية والثقافية من الإنجليزية إلى العربية. تم اختيار عشر عينات بعناية وترجمتها بواسطة مترجم محترف، بالإضافة إلى نظامين متطورين للترجمة الآلية العصبية، وهما ترجمة جوجل و DeepL، وذلك في المجالات القانونية والطبية والتسويقية والأدبية والمحادثية. استُخدمت أربعة معايير لتقييم المخرجات، وهي: الدقة الدلالية، والطبيعية الاصطلاحية، ودقة النبرة العاطفية، والملاءمة الثقافية. يتضح من النتائج أنه بينما تنجح الترجمة الآلية العصبية في ترجمة العبارات النمطية عالية التردد، إلا أنها في الحالات التي تتضمن سخرية أو استعارة أو أمثالا أو تلميحات ثقافية، تُنتج ترجمات حرفية قد تكون فارغة من الناحية العملية أو غير مناسبة ثقافياً. في المقابل، قدم المترجم البشري ترجمات عربية مكافئة وظيفياً ومراعية للسياق بشكل متسق. تؤكد النتائج أنه على الرغم من التطور السريع في إتقان الترجمة الآلية العصبية، فإن الاعتماد على الترجمة الآلية العصبية النصية وحدها لا يُعد بديلاً عملياً للنصوص التي تحتوي على مستويات عالية من اللغة المجازية/المشفرة خارجياً، وأن الوساطة البشرية لا تزال ضرورية لمثل هذه النصوص. وتُجدد النتائج الإجمالية للدراسة مكانة المترجم كوسيط ثقافي، وتشير إلى ضرورة تطوير هذه الدراسة وغيرها من الدراسات المماثلة فيما يتعلق بالتدريب على التحرير اللاحق واللغات ذات الموارد المحدودة.

## Abstract

This study aims to compare the contextual sensitivity of human translator and neural machine translator (NMT) in translating the idiomatic and cultural expressions from English to Arabic. Ten samples were carefully chosen and translated by a professional translator, as well as by two state-of-the-art NMT systems, Google Translate and DeepL, and in the legal, medical, marketing, literary and conversational domains. Four criteria were used to assess the outputs, which were semantic fidelity, idiomatic naturalness, emotional tonal accuracy, and cultural appropriateness. One can see in the results that while NMT successfully translates formulaic, high frequency phrases, in cases where sarcasm, metaphor, proverbs or culture-bound allusions are involved, NMT produces literal calques which may be pragmatically empty or culturally inapplicable. Functionally equivalent, context-sensitive, Arabic translations were provided by the human translator in

a consistently manner. The results confirm that, although the NMT fluency is progressing at a rapid rate, a purely text-based NMT is not a viable replacement for texts containing high levels of figurative/externally-encoded language and that human mediation remains vital for such texts. The overall results of the study renew the translator's status as a cultural mediator and suggest that this study and other studies of a similar nature should be further developed in terms of training in post editing and low resource language pairs.

## 1. Introduction

Translation is not just the mechanical act of replacing words in another language by its dictionary counterparts, but a negotiation of meaning, which comes at the intersection of cultural, linguistic and contextual factors. This complexity is further highlighted when interpreting expression that are idiomatic in their respective cultures and require great sensitivity to context to understand. As the world becomes more globalized and the exchange of information and knowledge is increasingly commonplace in the fields of diplomacy, commerce, healthcare, and entertainment, the need for quick and economical translation has never been greater. What was once a fairly simplistic and inaccurate machine translation (MT) system has dramatically changed during the last 20 years. It has evolved from simple rule-based architectures to statistical machine translation and, more recently, to neural machine translation (NMT), whose models can output very fluent and human-like results, and in some cases, even outperforms human translations (Bahdanau, Cho and Bengio, 2015; Koehn, 2020). Even with all these strong results, there are still great difficulties when MT is called upon to tackle language very much rooted in cultural knowledge and figurative meaning. This study goes directly to this gap by comparing and contrasting the contextual sensitivity of the two types of translators, humans and machines, and particularly the way the machines handle idiomatic and/or cultural expression.

Despite the impressive progress, there is an increasing unease over the idea of whether machine translation can be as sensitive in terms of idiomatic expressions, tone, emotional impact, and cultural relevance as it is in literal meaning? This question is not at all hypothetical. In healthcare, an idiom or a culturally-specific description of symptoms

can prove to be a life-threatening error; in legal matters, lack of contextual understanding in a translated message can lead to misinterpretations of intentions and jeopardize the justice process; in transnational marketing, it may tarnish or destroy brand image if a translation lacks cultural sensitivity. Given the high stakes, the choice between human and machine translation is in its essence unresolved: that is, it turns on the ability to be context-sensitive in dealing with idiomatic and cultural expression.

### **1.2 Human Translation: A form of Cultural Bridge-Building**

Human translators are traditionally known as cultural mediators, not just as linguistic transmittents (Venuti, 2012; Bassnett, 2014). They are not like machines, as they add interpretive judgment, empathy, and lived cultural experience when performing the translation task, which is necessary for contextual sensitivity in dealing with idiomatic and culturally embedded expressions. Human translators adapt the degrees of formality, different meanings of metaphors, and find equivalents that are idiomatic and resonate with intended audiences. Their decisions go beyond lexical replacement, and involve ethics, politics and aesthetics; their decisions affect what is rendered in the translated text, as well as how the text will be received and have an impact (Venuti, 2012). This ability to make culturally informed and culturally sensitive decisions is the essence of human ability in the translation of figurative language and language belonging to the specific culture.

### **1.3 Machine Translation: Achievements and limitations**

The main selling points of machine translation are its efficiency, scalability and accessibility. The advent of services like Google Translate and DeepL has revolutionized translation, empowering people all around the world to have real-time and sometimes even free multilingual communications. Systems that had been trained on smaller multilingual corpora were very disjointed and lacked fluency, and the current neural architectures have considerably improved this. However, significant gaps are still to be found in the pragmatic and cultural aspects of language in these systems. They are consistently good in dealing with text with clear structure and a high degree of predictability (e.g. technical instructions or weather forecasts), but poor in dealing with idiomatic expressions, jokes and context-dependent utterances. For

example, the English idiom, spill the beans, is a non-sensical, literal MT translation to Arabic or Japanese, but a human interpreter would find an equivalent expression and re-word it to suit the culture. These are indications of central drawbacks: the lack of sensitivity to the cultural and idiomatic context to interpret and reproduce the expressions.

#### 1.4 Gap of the research

Although there are many studies on the comparison of MT grammatical and semantic output with human translation, a limited number have made systematic comparisons of MT and human translation in various domains with a central focus on the contextual accuracy. Previous comparative studies focus mainly on the technical and/or literary field, but an interdisciplinary approach is lacking that combines legal, medical, and marketing contexts, where culturally and idiomatic charged language are needed differently. Furthermore, it is clear that human post-editing has been a topic of discussion within the framework of hybrid approaches (Toral & Sánchez-Cartagena, 2017), yet very few empirical studies have explored how well human post-editing can compensate for the certain lack of cultural sensitivity and idiomatic accuracy of MT output. This study aims to fill these gaps by direct analysis of the aspects of contextual sensitivity in human and machine translation with special attention to the translation of idiomatic and cultural expressions.

#### 1.5 The aims and Research Questions.

The main goal of this study is to take a look at how context sensitive human and machine translation are in various domains and language pairs particularly in the context of idiomatic and cultural expression. To this end the following questions are raised for research:

1. How have human and machine translation developed in terms of maintaining semantic accuracy, idiomatic expression, emotional tone and cultural appropriateness?
2. How good is the machine translation output compared to the human translation in different text genres whose translations heavily depend on idiomatic and culturally embedded language?

3. How much does human post editing of machine translations affect the contextual correctness with regard to idiomatic equivalence and sensitivity to the culture?
4. What are the implications of these results for the future of translation practice, the design of AI systems and the education of future translators, particularly in handling context-dependent, culture-specific material?

### 1.6 The Study Significance

This comparative study of human and machine translation is important in a theoretical, practical and ethical sense. It can work, at least in theory, with current research in translation studies on the concept of equivalence, visibility and the role of the translator in an era of automation, all of which are particularly relevant when translating idiomatic and culturally-specific texts. In real-world application, the work provides empirical evidence relevant to disciplines like health care, law, and marketing, which also have concerns about mistranslation of culturally relevant content as well as idioms and figurative language. The moral dimension delves into the issues of bias within AI systems, inclusivity of languages, and the potential marginalisation of professional translation services. This systematic comparison between human and machine translations in a context with high requirements for contextuality, correcting, clarifying, and consolidating the ability and limitations of current AI, thus reaffirming the irreplaceable role of a human translator as a cultural and ethical mediator in case of idiomatic and cultural expression.

## 2. Literature Review

### 2.1 Human Translation: Equivalence to Cultural Arbitration

Theories of human translation have from time immemorial been discussed in the context of fidelity, cultural negotiation and equivalence, which are all relevant to the contextual sensitivity needed in the translation of idiomatic and cultural expression. Earlier models like the one of dynamic equivalence by Nida and Taber (2003) gave preference to the impact of the translation on the audience of the target text than to a literal equivalence of the source text, thereby tacitly acknowledging that meanings are culturally-embedded. In a similar manner, Catford (1965) theorised translation as a process of replacing the text of one

language with the text of another language, which is equivalent, and which set the foundations for the structuralist approach but failed to explain the figurative language or culture bound language.

This belief in equivalence over time was met with strong criticism. Bassnett (2014) and Venuti (2012) pointed out that translation is an interpretive process and that there must be decisions on visibility and possibilities of domesticity or foreignness. Both Venuti and Fissette pointed out that the final approximation of a text to a target culture is always influenced by the translator, a fact which confirms that translation is an ideological as well as a linguistic act. But then again, translation was not made so simple by Tymoczko (2007) or Spivak (1993), who understood translation as an arena of power, gender and contestation, situated in a post-colonial perspective that views the rendering of idioms and cultural references as never innocent.

The theoretical perspectives have one thing in common: human translation can never be a purely linguistic process. It operates within cultural, political and ethical frameworks and the translator does not merely convey meaning but acts as an active mediator of meaning. This active mediation is particularly important in cases where source texts use culturally untranslatable wording, idiomatic structures or expressions which require more than word-to-word transfer.

## 2.2 Machine Translation: The Historical Routes

Since the 1950s, several paradigm shifts have happened in MT which each time led to a new tension between surface fluency and contextual sensitivity. Previous MT systems that were based on grammatical rules, known as Early Rule Based MT (RBMT), relied on manually developed rules and bilingual lexicon leading to rigid and substandard translations that were woefully incapable of dealing with idiomatic and culturally specific language (Hutchins, 2005). In the 1990s, the problems with RBMT led to the emergence of Statistical Machine Translation (SMT) that, again, relied on learning probabilistic models from the corpus of bilingual data for generating more natural output (Koehn, 2010). But SMT was still liable to disfluency and failed to learn long-range relationships well, misunderstood context-dependent idiomatic expressions.

The field was then shaken up by a novel neural machine translation (NMT) approach by Bahdanau, Cho, and Bengio (2015). With their sequence-to-sequence learning and attention mechanism, NMT has attained a whole new level of fluency and fluidity. Other breakthroughs, like the Transformer architecture (Vaswani et al., 2017), allowed other tools, such as Google Translator and DeepL, to work on dozens of languages with impressive surface-level coherence.

However, researchers keep on tracing some ongoing gaps that hit the core of contextual sensitivity. As mentioned by Koehn (2020), NMT has strong performance on grammar and surface fluency, while it is weak for idiomaticity and cultural subtleties, as well as out-of-domain texts. Similarly, Sennrich and Volk (2016) point to issues of ambiguity resolution, in which machines will opt for a literal meaning when faced with expressions that depend on shared understanding of a culture, or figurative language. The results imply that MT still has problems in developing a high level of contextual awareness, which is necessary to understand and translate idiomatic and cultural expressions properly.

### **2.3 Comparative Scholarships: Human vs. Machine Performance**

In translation, the results are always the same and human translators, in this case, perform better than machines, in particular on simple tasks that require a degree of contextual sensitivity. For instance, Castilho et al. (2017) revealed that the MT output was faster, although adequate for gist/overall understanding, professional translators clearly outperformed machines with respect to idiomaticity and naturalness. This difference is most apparent in the presence of a large number of figurative language, humour, and culturally referenced elements which are not literally understandable.

Bowker and Ciro (2019) recognise that MT offers accessibility benefits, but it can be problematic to depend exclusively on MT as serious misunderstandings can occur in contexts of health and safety when culturally specific expressions or idiomatic warning signs are mistranslated. Doherty (2016) tackled the issue about user acceptance of MT and found that accuracy perception differed widely across the different domains of texts as well as familiarity with the source and target languages suggesting that the contextual sensitivity of MT is not

only a machine issue but also a user-expectation issue, especially in cross-cultural communication.

In particular, Toral and Sánchez-Cartagena (2017) explored post-editing processes and tasks that involve human translators polishing MT drafts. Their findings indicated that the post-editing resulted in a time gain over the freshly created translation, and that it is of higher quality than raw MT, suggesting that hybrid models offer potential in filling in the gap with respect to the contextual fidelity, although questions of creativity and cultural resonance persist.

## 2.4 Domains of Application

The difference between automatic and manual translation in different domains is different and each domain has its own unique demands on translation of idiomatic and cultural expressions:

- a) Legal Translation: Every word and nuance of meaning or tone is crucial and can carry legal implications. Although human translators have always been better in translating legal texts than MT (even in cases where legal concepts are tied to the specific culture), NMT systems have become more capable at handling repetitive legal phrasing (Šarčević, 1997). However, if an idiomatic expression or a legally local expression(s) are used, human judgment is still necessary to prevent a significant error in translation.
- b) Medical translation – Errors can mean the difference between life and death. Although MT tools can be used as auxiliary communication tools, they can be hazardous when used for diagnostic or treatment related texts (Patil & Davies, 2014). These contexts call for a type of sensitivity to context that machines do not yet provide, in particular with regard to how patients describe themselves and what symptoms they use to convey their issues.
- c) Machines frequently lose metaphor, rhythm and cultural appeal in literary translation (Apter, 2013). Literary works contain idiomatic richness and culture-specific pictoriality and imagery, which can only be adjusted creatively and aesthetically by human translators.
- d) Marketing and Advertising: Cultural adaptation is very important. One humorous (but surprising) example of this type of literal, context-blind translation was Pepsi's infamous mistranslation of

“Come alive with the Pepsi Generation” to Chinese, unintentionally meaning that the ancestors are coming back from the grave (Smith, 2018). Failure to be contextually sensitive to idiomatic and cultural expression are the root cause of such failures.

## 2.5 Mixture Models and Post-Editing

Speed can be a problem for MT and contextual sensitivity is an issue, thus hybrid translation models that use both MT and human judgement are a direct answer to this need. O'Brien (2012) focuses on the new growth of post-editing machine translation (PEMT) which involves translators increasingly playing a part of editors polishing machine translated products instead of producing work from scratch. The implications for training and professional identity and remuneration are suggested.

Läubli et al. (2018) showed that acceptability of post-edited MT texts was similar to that of the fully human texts, but that there were still differences in terms of creativity and stylistic nuance. It is a indication that hybrid models may be appropriate in idiomatic-poor industries like technical documentation and e-commerce, but may not be appropriate in idiomatic-rich areas such as cultural- or literature-related where idiomatic and cultural expression is central to the text's purpose.

## 2.6 Ethical/Sociocultural Considerations

The advent of MT also brings up important ethical issues such as bias, inclusivity and labour, all of which involve the ability to be sensitive to the context. Caliskan, Bryson and Narayanan (2017) showed that social biases like genders stereotypes are embedded in words used in AI systems and therefore, these biases are manifested in the MT output. In the case of Google Translate, for example, Prates, Avelar, and Lamb (2020) detailed some occurrences of the translator using masculine pronouns instead of gender-neutral ones in certain contexts, which showed how this lack of cultural and contextual awareness can result in discriminatory translations.

Sociocultural, Venuti (2012) cautions of the growing invisibility of the translator in a machine dominated world in which human labour is devalued or hidden behind the machine products. Those implications

have direct consequences for the interpretation of idiomatic or expressively cultural texts, as in such cases any human mediating interpretative and ethical effort to provide an approximation of fidelity of expression is hidden in the invisibility of the machine output. Cronin (2013) also suggests that MT is likely to have a negative effect on linguistic inequalities as it will prioritize the idioms and cultural expressions of the more familiar high-resource languages over those of the lesser known low-resource languages, thereby reinforcing the language-insensitive global translation ecology.

## 2.7 Translation in the Digital Age

The production and consumption of translation has changed drastically with the advent of the digital revolution, with implications for the contextual sensitivity. The rise of what Cronin (2010) calls a globalised 'translation economy' in which fast and convenient translation may be more important than faithful representation and cultural sensitivity has rendered culture-specific and idiomatic translation more prone to oversimplification. In this context, Pym (2014) believes that, to remain competitive, translators have to learn new skills, such as the skills of post-editing and localisation, and protect the ethical values that support culturally sensitive translation. Data-driven development of MT systems, enabling ever increasing coverage of ever more languages, can also be a source of expressional uniformity and stylistic leveling, as highlighted by Kenny (2022), introducing a risk of destroying not only the very idiomatic richness but also the cultural texture that contextual sensitivity aims to capture. All these developments create new problems in the translation of texts in terms of maintaining creativity, diversity, and authenticity, especially if the texts rely on idiomatic or cultural expression.

## 3. Methodology

### 3.1 Research Design

This study adopts a comparative study approach with mixed methods to evaluate the level of contextual sensitivity in the process of human and machine translation of idiomatic and cultural expressions. Ten well-designed samples of English text with a very compact idiomatic or culture bound challenge are translated from English into

Arabic using a professional human translator, and by a top machine translation (NMT) system. These two sets of outputs are then subjected to qualitative analysis in terms of semantic fidelity, idiomatic naturalness, emotional tonal accuracy and cultural appropriateness.

### 3.2 Samples Selection

The sample includes 10 English texts that have been purposefully chosen to represent a wide range of contexts in which the use of idiomatic and cultural expressions is central: marketing, colloquial speech, literature, health education, legal, humour and proverbs. Contextual sensitivity is easily measurable as a literal translation of each sample in Arabic would either result in no sense whatsoever or seriously alter the intended communicative impact. The samples are:

#### **1. Break a leg! (theatre idiom)**

A good-luck wish whose literal meaning would be baffling or ominous in Arabic requires a functionally equivalent Arabic well-wishing expression.

#### **2. It's raining cats and dogs. (weather idiom)**

A hyperbolic description of heavy rain that has no direct counterpart in Arabic; calls for a culturally idiomatic image of torrential rain.

#### **3. I'm lovin' it. (marketing slogan)**

McDonald's globally adapted slogan; officially localised into Arabic as *لنا حبه* (*ana uhibbuhu*), moving from progressive "lovin'" to a simple present. The translation choice tests the balance between brand voice and cultural naturalness.

#### **4. A penny for your thoughts. (conversational idiom)**

A friendly prompt to share one's thoughts; Arabic lacks the "penny" metaphor. The translation must preserve the inviting, inquisitive tone without a literal coin reference.

#### **5. Oh, fantastic. Another traffic jam. (sarcasm)**

Meaning entirely depends on tone and shared cultural frustration. A literal Arabic *زحمة مرور أخرى* could easily be read as genuine praise; contextual sensitivity demands sarcastic force is unmistakably conveyed.

**6. *He's the Shakespeare of our generation.* (cultural reference)**

The use of a metaphor with high praise which is based on an assumption of the reader's familiarity with Shakespeare. Although the Arabic reader is familiar with Shakespeare, the translator has to make a choice whether to incorporate or replace it with a well-known Arab literary name; either decision signifies cultural mediation.

**7. *Last will and testament.* (legal frozen phrase)**

A legal term meaning “final will”. In the Arabic legal jargon, the term الوصية (al waṣiyya) is usually enough by itself, to say both waṣiyya and waṣṣession would be redundant and could lead to some legal confusion.

**8. *I've got butterflies in my stomach.* (somatic idiom for nervousness)**

A physical metaphor describing anxiety. Arabic equivalents (e.g., أشعر بتوتر “I feel tension”) lack the vivid bodily image; the translator must decide whether to reproduce the image, find a different Arabic somatic idiom, or normalise to plain emotion.

**9. *Don't count your chickens before they hatch.* (proverb)**

Gives cautions about premature rejoicing. The extent to which the translation matches an Arabic proverb (in terms of frequency, register and cultural familiarity) is the same for more than one Arabic proverb – there are a number of them, which have functioned as equivalents to the English proverb in terms of their function.

**10. *I'm not a complete idiot – some parts are missing.* (self-deprecating humour)**

Witty self-mockery relying on a two-part structure. The Arabic version needs to maintain the joke's timing and cultural appropriateness of self-deprecation (which differ among various Arabic speaking communities). All samples attempt to move beyond grammar and vocabulary but to actually test the very sensitivity to context which the study title highlights.

### 3.3 Translation Tools and Human Participants

## Machine Translators (MT Systems)

MT samples are produced for each English sample in Modern Standard Arabic (MSA) using Google Translate NMT and DeepL (no post editing). These systems are selected as they are state of the art neural systems that are widely available for use in English–Arabic translation. When the two outputs diverge it is the more fluent reading which is chosen, although any marked difference is recorded in the analysis.

## Human Translators

Human translation is done by a professional translator for Arabic to English with at least 5 years of documented experience in literary, commercial or legal translation. The translator is native speaker of Arabic and is a full bilingual, and is explicitly instructed to choose communicative effectiveness and cultural appropriateness for a general Arabic speaking public—over literal rendition. human translator reads the MT outputs without looking them up.

### 3.4 Evaluation Criteria

Each translation pair (HT and MT) is evaluated based on four criteria, which operationalise the main variable of the study, namely, ‘contextual sensitivity’:

- 1.Semantic fidelity: Has the main meaning been retained, that is, has it been distorted to the point of being confusing?
- 2.Idiomatic naturalness: Is the Arabic translation something that would be a smooth, natural expression that a native speaker of Arabic might use, or is it a stilted, literal or foreign one?
- 3.Emotional tonal accuracy: Are the emotional tone (humorous, sarcastic, solemn, encouraging, etc.) being transferred accurately?
4. Cultural appropriateness: Is the translation culturally appropriate, does it avoid taboos in Arabic culture, and does it use expressions that sound appropriate for the intended audience?

A binary rating (adequate / inadequate) is awarded per criterion for each sample, accompanied by a qualitative error analysis to give an explanation where and why contextual sensitivity succeeded or failed.

### 3.5 Data Collection and Procedure



1. Two bilingual researchers verify for clarity and idiomaticity the 10 English source samples.
2. The samples are then concurrently translated by the human translator and by the chosen MT systems.
4. All Arabic output (One HT & MT per sample) will be anonymised and collated into a random evaluation sheet.
5. Three independent raters who are bilingual Arabic–English linguists but not involved in the translation stage rate each target text in the four criteria.
6. Ratings are taken and inter rater reliability is calculated to ensure consistency with rating.

### 3.6 Analytical Framework

The analysis takes place in two phases.

- a) Within sample analysis: Two outputs (HT and MT) compared against four criteria for each of the 10 samples. A brief (qualitative) commentary is written on the nature of any failure of context sensitivity (e.g. literal idiom, failure of sarcasm, culturally unusual image).
- b) Cross sample synthesis: The patterns are then combined together to search for recurring strengths and/or weaknesses of HT as compared with MT particularly with regard to text genre (legal, marketing, conversational, etc.) and the kind of idiomatic/cultural challenge (metaphor, fixed phrase, sarcasm, cultural allusion).

## 4. The Results and Discussion

The human translator always gets it right, in cases where meaning is well beyond the literal lexico grammar, that is, in cases of sarcasm (Sample 5), proverbs (Sample 9) and culturally loaded metaphors (Samples 6, 8). In such cases, it is the human translator who taps into his/her communicative intuition and culturally shared knowledge to choose an equivalent in Arabic that is idiomatic rather than literal and semantically opaque, or else is a plain paraphrasing colloquial translation. The findings attest that the lived cultural experience and interpretative judgment of the human mediator play an important part in

the preservation of the pragmatic force and cultural resonances of such expressions.

MT does very well with frozen, formulaic expressions with high frequency correspondences in the training data, such as the legal doublet in Sample 7 (where it is able to converge on the single Arabic term الوصية) and the globally standardised marketing slogan of Sample 3. But the moment that an expression needs creative re figuration, MT becomes a liability, since it often tends to calque literally as in Samples 1, 2 and 10, and to produce Arabic which may be grammatically correct, but pragmatically empty or even comically inappropriate. This signals that although MT fluency is impressive, the sensitivity in context, that is in the sense of understanding the intention of the speaker, cultural reference and emotional tone, is basically limited. Hybrid post editing therefore becomes an inevitable intermediate step, especially for genres with a high content of idiomatic and cultural expressions.

## 5. Conclusion

The purpose of the present study was to investigate the contextual sensitivity of human and neural MT in the specific field of idiomatic and cultural expression through a set of ten carefully selected samples in English-Arabic translation in four fields as well as in the field of conversational translation. The results verify a large asymmetry. Machine translation can make a good performance when it is faced with frequently-used formulaic expressions or generally uniform slogans in the world. But when meaning moves away from the literal and towards cultural allusion, sarcasm, metaphor and proverbial wisdom the machine's contextual sensitivity breaks down into calque, grammatically perfect Arabic but pragmatically empty, or unintentionally comic, or culturally alien. By contrast, the human translator always showcased the interpretive skills, empathy, and cultural know-how that allowed him or her to find functional equivalents in Arabic that not only translated the propositional content, but also the tone, emotional register, and cultural resonance. It casts doubt on the grammatical aspect being the real problem for machine translation, but rather deep, lived-out knowledge of figurative uses of language to establish humor, solidarity, warning, or praise within human communities. The practical consequences are obvious. Human mediation is indispensable when it comes to genres that are full of idiomatic and cultural expression. Hybrid post-editing workflows come out as the most realistic solution, combining the speed of the machine with the contextual intelligence of humans. From an ethical point of view, the study supports an idea of the translator as a cultural and ethical mediator, who should not become invisible in an increasingly automated world. The paradigm should be expanded to other language pairs with fewer resources and additional research should investigate how post-editing can be systematically taught to fill in the gap of contextual sensitivity.

### Implications for Further Translation Studies

The results highlight the need for ongoing research in the field of translation studies that goes beyond the conventional ideas of equivalence to the much more dynamic view of the translator as a cultural mediator. The attention given to idiomatic and cultural expression in the study confirms the relevance of the theories of

domestication, foreignisation, and translator visibility in the present automated age. It also advocates for more in-depth interdisciplinary communication between translation studies and AI ethics, particularly in situations where automatic translations might lead to a loss of cultural nuances. Further, the comparative approach shows that research on context-sensitivity is most fruitfully pursued by writing about configuration rather than about individual texts, and that the configurations under study be domain specifications.

### **Limitations**

The study has some limitations, being limited to a single language pair (English to Arabic) and a relatively small corpus of ten curated samples. Although the samples were selected to provide maximum challenge, they do not necessarily reflect all the idiomatic and cultural variations of the Arabic dialects and registers. Even with inter-rater checks, the evaluation is still of a qualitative nature and could contain some subjectivity. Moreover, just two NMT systems were evaluated in normal modes, while custom NMT engines, domain-specific NMT models, and varying temperatures were not investigated for their impact on contextual sensitivity.

### **Future Research**

Future work can be extended to cover other language pairings such as low-resource languages and dialectal variants, for whom a gap in contextual sensitivity is exacerbated by training-data sparsity. Longitudinal studies should be carried out to monitor the speed with which NMT systems can be developed to deal with figurative language. Most important, empirical studies need to be conducted to assess the efficacy of specially designed, targeted post-editing training programs aimed at enhancing cultural and idiomatic accuracy. Finally, some reception studies with actual target readers would provide an interesting dimension, which would not only measure the accuracy of translation, but also the pragmatic impact of the idiomatic translation on real readers.

## References:

- Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- Bassnett, S. (2014). Translation studies (4th ed.). Routledge.
- Bowker, L., & Fisher, D. (2010). Computer-aided translation. In Y. Gambier & L. van Doorslaer (Eds.), Handbook of translation studies (Vol. 1, pp. 60–65). John Benjamins Publishing.
- Chesterman, A. (2016). Memes of translation: The spread of ideas in translation theory (2nd ed.). John Benjamins Publishing.
- Gambier, Y., & van Doorslaer, L. (Eds.). (2010). Handbook of translation studies (Vol. 1). John Benjamins Publishing.
- Gaspari, F., Almaghout, H., & Doherty, S. (2015). A survey of machine translation evaluation. Machine Translation, 29(3–4), 195–215.
- Hutchins, W. J. (2005). The history of machine translation in a nutshell. Computers and Translation, 17(1), 1–24.
- Kenny, D. (2017). Machine translation. In J. W. Schwieter & A. Ferreira (Eds.), The handbook of translation and cognition (pp. 423–439). Wiley-Blackwell.
- Koehn, P. (2020). Neural machine translation. Cambridge University Press.
- Munday, J. (2016). Introducing translation studies: Theories and applications (4th ed.). Routledge.
- Nida, E. A., & Taber, C. R. (2003). The theory and practice of translation. Brill Academic Publishers.
- O'Brien, S. (2012). Translation as human–computer interaction. Translation Spaces, 1(1), 101–122.
- Pym, A. (2014). Exploring translation theories (2nd ed.). Routledge.
- Somers, H. (2003). Machine translation in the classroom. In H. Somers (Ed.), Computers and translation: A translator's guide (pp. 319–340). John Benjamins Publishing.

- Toral, A., & Sánchez-Cartagena, V. M. (2017). Post-editing neural machine translation: A preliminary study. *Translation Spaces*, 6(2), 240–263.
- Venuti, L. (2012). *The translator's invisibility: A history of translation* (2nd ed.). Routledge.
- Way, A. (2018). Quality expectations of machine translation. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment: From principles to practice* (pp. 159–178). Springer.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.
- Zhou, L., & Waibel, A. (2020). Context-aware neural machine translation learns anaphora resolution. *ACL Anthology*, 38(2), 1–12.