



P-ISSN : 2074-9554 | E-ISSN: 2663-811

Journal of Al-Farahidi's Arts

available online at : jfa.tu.edu.iq/index.php/jfa



Assessing AI Translation of Humor: challenges, limitations and post editing solutions

Ammar Ahmed Mahmoud Al-Harhi

Zainab Fahmi Ahmed Al-Khayyat

Fartin Hanna Shaba Al-Hanouk

E-Mail: amaar.a.m@uomosul.edu.iq

Keywords:

Key words: humor; post-edit ; translation; Translation Edit Rate.

Article history:

Received 6/8/2025
Received in revised form 15/9/2025
Accepted 19/10/2025
Available online 9/12/2025

E-mail Jaa@tu.edu.iq

©THIS AN OPEN ACCESS ARTICLE UNDER THE CCBY LICENSE

<http://creativecommons.org/licenses/by/4.0>



ABSTRACT

Humor is widely regarded as a compelling aspect of everyday communication, offering nuanced, often playful insights into political, social, and cultural realities. Despite its significance, humor remains a challenging and under-theorized area in the field of machine translation. This paper principally opts for addressing this somewhat under-theorized topic through (1) presenting a thorough account of humor in terms of the definitions, major types, motivations, and communicative functions; and (2) pinpointing and defining the way exploited by AI-generated translation programs in translating humor to show whether the renderings arrived at are appropriate, effective, and to the point or otherwise. It is hypothesized that translating humor poses untold and still unsolved problems for AI-generated translation programs, as they do not possess the proper systematized and refined algorithms that enable them to efficiently process the SL message and adequately transfer their pragmatic content into the TL. In order to test the validity of this hypothesis, a body of authentic humorous texts is chosen, a translation model is adopted to subtly post-edit renderings as presented by the AI-chosen programs, data analysis is delicately conducted, and some reference renderings are recommended. Translation quality is assessed using the Translation Edit Rate (TER) metric, which measures the effort required to correct them.

تقييم ترجمة الذكاء الاصطناعي للفكاهة: التحديات، حدود الأداء وحلول تعديل الترجمة

م. عمار احمد محمود الحريثي /جامعة الموصل/ كلية الآداب/ قسم الترجمة

م. زينب فهمي احمد الخياط /جامعة الموصل/ كلية الآداب/ قسم الترجمة
م. فارتين حنا شبا ال حنوق /جامعة الموصل/ كلية الآداب/ قسم الترجمة
المستخلص:

يُنظر إلى الفكاهة على نطاق واسع على أنها جانب مُقنّع من التواصل اليومي، إذ تُقدم رؤيةً دقيقةً ومُرحةً في كثيرٍ من الأحيان حول الحقائق السياسية والاجتماعية والنقافية. ورغم أهميتها، لا تزال الفكاهة مجالاً صعباً وقليلَ التنظير في مجال الترجمة الآلية. وتميل هذه الورقة البحثية بالأساس إلى تناول هذا الموضوع الذي لم يُنظر له بشكلٍ كافٍ من خلال (1) تقديم وصفٍ شاملٍ للفكاهة من حيث تعريفاتها وأنواعها الرئيسية ودوافعها ووظائفها التواصلية؛ و(2) تحديد وتعريف الطريقة التي تستغلها برامج الترجمة المُولدة بالذكاء الاصطناعي في ترجمة الفكاهة لإظهار ما إذا كانت الترجمات المُتوصّل إليها مناسبةً وفعالةً ومباشرةً أم لا. ويُفترض أن ترجمة الفكاهة تُشكّل مُشكلاتٍ لم تُحل بعد لبرامج الترجمة المُولدة بالذكاء الاصطناعي، إذ لا تمتلك هذه البرامج الخوارزميات المُنظمة والمُحسنة المناسبة التي تُمكنها من معالجة رسالة لغة المصدر بكفاءة ونقل محتواها البراجماتي إلى لغة الهدف بشكلٍ كافٍ. لاختبار صحة هذه الفرضية، تم اختيار مجموعة من النصوص الفكاهية الأصيلة، واعتماد نموذج ترجمة لإجراء تعديلات لاحقة دقيقة على الترجمات المُقدمة من برامج الذكاء الاصطناعي المختارة، وتحليل البيانات بدقة، والتوصية ببعض الترجمات المرجعية. قُيِّمت جودة الترجمة باستخدام مقياس معدل تحرير الترجمة (TER)، الذي يقيس الجهد اللازم لتصحيحها. يحسب TER عدد التعديلات اللازمة لتحويل الترجمات غير الدقيقة إلى ترجمات دقيقة. يوفر هذا أداة تقييم لإحصائيات دقة برامج ترجمة الذكاء الاصطناعي بشكل عام، والجوانب التي تحتاج إلى تحسين في كل برنامج. وتختتم الورقة ببعض الاستنتاجات والتوصيات التي يُفترض أن تكون ذات فائدة كبيرة.

الكلمات المفتاحية : الفكاهة؛ التحرير اللاحق؛ الترجمة؛ معدل تحرير الترجمة

Introduction

On a daily basis, we all giggle at jokes, share amusing stories for information and entertainment, make fun of one another, and trade witty insults for amusement. The daily routine of human beings encompasses actions that are done regularly for no obvious objective. Such actions are done for interpersonal reasons.

Humor is a personal experience, in other words, subjective; therefore, what one person thinks humorous may not be hilarious to another. Because humor depends on the situation, what one person finds humorous may be unpleasant or inappropriate to a different one. When it comes to humor, different cultures and people may have different tastes and sensitivity levels. When employing humor in English or any other language, it is crucial to keep the audience and context in mind.

The main objective of this paper is to try and present successful renderings for humor. This is the reason why issues such as humor's translatability, its cross-linguistic compatibility, and the characteristics of the obstacles must be addressed from both sides in the arena where humor and translation intersect.

The Concept of Humor:

Humor, like language, has an objective, and it is the ultimate result of externalizing human thought as well as conceptualization. This externalization has various meanings, partially as an outlet for certain feelings, partly as a social device, and partially as an intellectual exercise. The active use of this human ability allows some people to make a living by telling jokes. Thus, in a culture, it is possible to broadcast one's own personal view and vision of the world through a series of jokes.

Attardo (1994, p. 3) asserts that humor is considered a fascinating part of everyday speech and that it subtly and gently conveys social, cultural, and political realities. According to Chafe (1994, p. 9), this is an inherent characteristic of Homo sapiens; it is "the essence of human comprehension: the ability to recognize specific experiences as representations of larger embracing systems". The ability to laugh and identify that something is humorous is the characteristic that the Cambridge Dictionary attributes in very simple words to humor. Humor levels the playing field, allowing individuals that are compatible with one another to form social groups. Today, 'humor' refers to anything that people say or do that is thought to be humorous and makes other people

smile or laugh and has a generally positive, socially desirable connotation. It's interesting how recently this broad definition of humor has emerged. The word, which began with a completely different meaning and gradually accumulated new connotations over the centuries, has, in fact, had a very interesting and complex history.

For Feingold & Mazzella (1993, p. 430), humor is a cognitive capability to remember, understand, reproduce, and create jokes. Ruch (2002, p. 34) presents a similar view when he says that humor denotes a cognitive affective style of dealing with situations. And it is eventually, from its original meaning derived from the Latin root '*umor*,' which means fluid, the first meaning of 'Humores' was a term for the fluids of the body. Humor, as much as our everyday life is concerned, is also how often a person smiles, laughs, or otherwise reveals amusement in different situations (Martin & Lafcourt, 1983, p. 1315).

Humor as an Interpersonal Instrument

Humor serves as a powerful tool in interpersonal and intergroup communication, engulfing various social functions that play a vital role in group dynamics and discourse. In this regard, Bell and Pomerantz (2016, p. 35) are of the viewpoint that humor is a collaborative or co-constructed communicative endeavor. The study of language humor in sociolinguistics provides important insights into how language is an interpersonal instrument that reflects and shapes relationships of power and human interaction. Through a sociolinguistic approach, researchers aim to unravel the complex mechanisms of language humor, investigate its societal relevance, and comprehend how it influences communication among various linguistic communities. Sociolinguistics studies the relationship between humor and language use in society. Sociolinguistic approaches examine how humor interacts with social factors rather than just linguistic forms (such as joke structure).

According to Zabalbeascoa (2005, p. 198), social themes such as sex, religion, bodily functions, and dealing with 'lighthearted' topics are at stake in the former scenario. In social interactions, humor can be utilized to establish a comfortable mood, reduce stress, and break the ice. People can help others relax and facilitate opening conversations by sharing jokes, amusing stories, or lighthearted banter. To make their coworkers feel more at ease and promote a sense of friendship, a newly hired employee might, for instance, utilize comedy to introduce

themselves during a team meeting. Humor is essential for establishing interpersonal ties and enhancing the ones that already exist. Within a group, relationships, social cohesion, and a sense of belonging can be developed through shared laughter and inside jokes. For example, friends frequently tease one another with humor in an affectionate way, strengthening their bond and maintaining friendship.

The Universality of Humor

Laughing, according to scientists, is healthy. But what is it that makes something amusing? Prior definitions of humor classified it as a universal human characteristic. Raskin (1985, p. 2) considers humor as part of human behavior, ability, or competence, showing an important social and psychological manifestation of human beings' language. He goes on to say that humor is partly inbuilt and partly acquired.

Degabriele and Walsh (2010, pp. 225–6) assert that humor is a universal human interaction phenomenon that begins in children at extremely young ages. Shifman, (2007, p. 189) is on the same line when he says that, as a universal phenomenon, humor touches on important issues that all civilizations have in common, including politics, religion, race, sex, gender, and age.

Humor in Language 2 Learning

While incorporating general affective humor into the classroom can enhance learning environments in both language and general educational contexts, language teachers benefit from humor much more when they use it as a targeted tool to illustrate the linguistic, discursive, and cultural aspects of the language they are teaching. Crucially, and in view of the modern dominance of structure-based syllabi in language training, comedy provides a perfect medium for the explanation and application of linguistic mechanics. Deneire (1995, p. 289) examines the use of humor in particular in this kind of linguistic context. He suggests that morphological, phonological, syntactic, and lexical differences between a student's native language (L1) and the target language (TL) or within a single language can be effectively explained to students using comedy. Mind the Lexicon example below:

A: "Waiter, do you serve crabs here?" asks a customer.

B: "We serve everybody. Just have a seat at this table, sir."

Four major aspects of humor instruction are discussed by Bell and Pomerantz (2016, chapter 8). They are: identifying humor,

comprehending humor, producing humor, and responding to humor. To elaborate further, humor instruction, which is very significant in language learning and can be integrated to enhance communicative competence among learners, can involve any or all of the following:

1. Recognition: Understanding clues like tone, context, and linguistic techniques that convey a humorous intention is necessary to be able to identify comedy in communication. This ability enables people to easily pick up on jokes, puns, or humorous aspects in texts or conversations. For instance, identifying sarcasm in a statement such as 'Oh, great, another Monday!' necessitates comprehending the tone of the speaker and the discrepancy between the intended meaning and the words.

2. Comprehension: Comprehending humor is more than just identifying it; it's about grasping the wit, wordplay, or incongruity that underlies a joke's amusement. To properly enjoy the humor, this skill necessitates an understanding of context, linguistic subtlety, and cultural references. Understanding a pun like 'I'm reading a book on anti-gravity,' for instance. 'It's difficult to put down!' entails realizing that the word "put down" can refer to both physically setting something down and being unable to put down the book.

3. Production: Creating or delivering jokes, anecdotes, or clever comments to amuse or interest people is known as humor production. To properly create amusing content, one needs imagination, linguistic flexibility, and an awareness of audience preferences. For instance, coming up with a clever one-liner such as 'I advised my wife to accept her errors.' demonstrates the ability to come up with a joke with a surprising twist for humorous impact. "She gave me a hug."

4. Response: Responding to jokes or humorous situations in a way that recognizes and heightens the comedic effect is known as 'responding to humor.' This ability involves maintaining a humorous touch in a conversation by laughing, making a clever reply, or engaging in playful banter. A witty answer such as 'Well played! You've got jokes for days' in response to a friend's joke, for instance, shows an active role in the humorous exchange and promotes more amusing conversation.

Humor categories:

Generally speaking, people encounter humor in a variety of contexts throughout their daily lives, including through the use of visual

and printed media (comedies, cartoons, movies, etc.) and newspapers, comics, etc. respectively, as well as through the verbal sharing of hilarious comments, jokes, and stories. Degabriele and Walsh (2010).

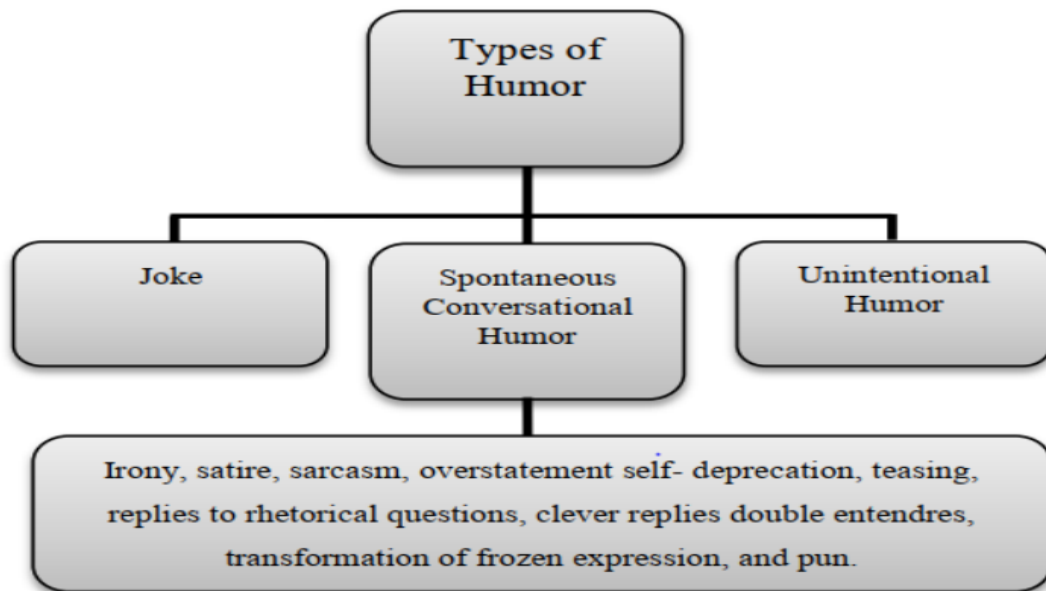
Three components make up a joke, according to Hockett (1972, p. 153): Buildup, pivot, and punch line. The joke is introduced in the first part, and the pivot is a statement that builds to the joke's peak. Lastly, the expression that concludes the joke and influences the intended audience is known as the punch line. As a result, the final chuckle is generated. Here is an example illustrating the idea.

Build-up: A thief broke into my house last night.

Pivot: He started searching for money,

The Punch line: so I woke up and searched with him.

Martin and Ford (2018) present the following classification of types of humor.



Types of humor by Martin and Ford (2018)

Mohammed (2019, pp. 17-8) presents an interesting classification about humor categories, saying that humor can be manifested by the use of

- 1- Wordplay, or pun, is when a term has two meanings: a literal meaning and a metaphorical one. Some writers refer to various forms of linguistic play, including parody, anagrams, spoonerism, and transformed allusion, as wordplay. Wordplay, in specific, is given priority to other categories.

For Zabalbeascoa (2005, p. 195), wordplay, such as in puns, one-liners, limericks, and so forth, can generate humor. Delabastita (1996, p. 128) thinks that wordplay is the umbrella term for a wide range of textual phenomena in which the structural features of the language or languages concerned are manipulated to bring two (or more) linguistic structures with largely similar forms and substantially distinct meanings into confrontation that has communicative significance.

With regard to translation, since wordplay relies on aspects of the source language that often do not have an equivalent in the target language, it has been considered a challenging phenomenon to translate owing to its pragmatic and semantic implications (Delabastita 1994: 223).

The most important aspect of wordplay translation is the retention of the impact of the wordplay. In order to recreate equivalent effects, translators should deviate from the structures of the source text. Paradoxically, Delabastita (1996: 135) suggests that the only way to be faithful to the original text is to be unfaithful to it. In other words, wordplay can be translated without the target text utilizing the same linguistic or formal structures.

For instance, let's talk about rights and lefts. You are right, so I left.

2- Irony is the contradiction between what someone says and what he means and what other people understand, as well as between what is expected or intended and what really occurs. It could be situational, dramatic, or verbal.

For example: A mother walks into her son's room, which seems messy, and says: Honey, your room is so tidy. I am so proud of you!

3- Parody is a purposeful copy of a work, style, genre, or discourse that incorporates comedy and aims to interpret its target in a certain way.

Example:

A young daughter stomps around in her father's large shoes, saying, "I need to make a business call. I am a very busy, very important businessman!"

4- Satire can refer to a work of writing that uses a comedic approach to critique individuals or ideas, often in order to make a political point. Satire is a literary method used by authors to use humor, irony, exaggeration, or mockery to highlight and critique the folly and corruption of a person or a society.

Example: Husband should be like Kleenex: soft, strong, and disposable.

- 5- Spoonerisms are described by Oxford Dictionaries as "a verbal error in which a speaker, often to humorous effect, accidentally transposes the initial sounds or letters of two or more words." These are individuals's inadvertent speech errors that make people laugh in real life and have even been exploited to make people laugh in some television programs.

Example: I fool so feelish.

- 6- A retort is characterized as a humorous or incisive response; in particular, one that contradicts or reverses the initial speaker's assertions. Put another way, humor occurs when the second speaker claims to have misinterpreted the first speaker's earlier statement, so changing the intended meaning. A retort may use a variety of humorous devices, such as ambiguity, aggressiveness, and rhetorical questions.

For instance: what type of flowers grows on a human face?

Answer: the flower tulip. (Two lips).

- 7- Teasing is a common practice in human social life. It is typical for people to playfully tease one another for a variety of reasons (e.g., to socialize, to pass the time, and to mock).

Example: Jill: You manifest the Peter Pan syndrome.

Jack: And you have the Captain Hook syndrome. (Teasing)

Jill: There's no such syndrome.

Jack: Obviously there is. You have it! (Teasing)

- 8- Banter refers to a one-turn between two interlocutors that results in a lengthy, hilarious, and teasing exchange of words. Banter is a quick interchange of amusing remarks focused on a shared issue; however, it is more intended for mutual amusement than for relevant discussion. Positive impoliteness is used by both the speaker and the hearer in this kind of humor, which is associated with social intimacy.

Humor Theories

This section tries to apply ideas of a linguistic nature through the context of the theory of humor as a whole. Here is not the right place to go through a general framework concerning the 'General Theory of Verbal Humor,' which came after a Semantic-Script Theory of

Humor. This section is devoted to placing linguistic theories through the context of humor theory as a whole.

There is no place here to go over the general framework of the General Theory of Verbal Humor, which came after the Semantic-Script Theory of Humor. According to Raskin (1985), diverse theories of humor are classified into three types: incongruity, superiority and release. In our terminology, the release and relief theories only address the emotions and psychological makeup of the listener, while the theories of superiority characterize the attitudes or relationships between the hearer and the speaker.

The three sub-types of humor theories are proposed by Raskin (1985); they are listed below:

- 1- Incongruity Theory: The theory holds that unexpected or discordant events are where humor arises. Laughter is triggered when something defies logic or our expectations. Consider wordplay, puns, or ridiculous scenarios that surprise us. The core of incongruity-based humor is the element of surprise. See (Raskin, 1985, pp. 31-36).
- 2- Superiority or Hostility Theory: According to this theory, humor originates from a feeling of superiority or supremacy. Jokes frequently make fun of someone's shortcomings, bad luck, or ridiculousness in order to make the audience feel better for a short while. One example of the superiority theory in action is when we laugh at a sitcom character who is awkward. See (Raskin, 1985, pp. 36-38).
- 3- Release Theory: According to the release or relief theory, humor acts as a relief mechanism for tension or repressed feelings. As a homeostatic mechanism, laughter lowers psychological stress. We feel relieved from worry or repressed emotions when we laugh. For example, tickling elicits laughter due to the buildup of tension. See (Raskin, 1985, pp. 38-40).

Humor from the cultural standpoint:

Maintaining the overall meaning of the original text while incorporating the humorous touch is the primary objective of the entire translation process in terms of humor. When interpreting humor's cultural

reference, the translator encounters the same challenges. A comprehensive knowledge base is necessary in order to modify these types of references once more for the intended audience. The audience is not familiar with some of these components, and the translator acts as a linguistic and cultural mediator between the two languages.

It is impossible to ignore the importance of the human condition when discussing humor in the context of literature and cultural practice. The manifestations of humor as we know them today, such as laughter, smiling, and distinctive body language, have been with people since the beginning of time, despite the fact that the definition of humor as something "funny" is a relatively recent phenomenon (Morreall, 2009:29).

The value and significance of humor can be found in the earliest human societies, when humor was crucial to their survival. According to John Morreall, some of the earliest uses of humor, and more specifically, the physiological manifestation of laughter, were ways to let go of pent-up energy in response to perceived danger. The benefits of the relief theory and the fact that humor has been a part of the human experience for many centuries are both shown by the fact that when a threat turns out to be a false alarm, the pent-up energy transforms from panic into relief and is frequently released as laughter (Morreall, 2009:28–29).

Humor and AI:

Translation is one of the businesses that has changed as a result of the development of artificial intelligence. The topic of whether AI will soon replace human translators is raised by the notable advancements achieved in eliminating language barriers by AI-powered technologies like Google Translate and DeepL. This paper shows that, even though AI has boosted translation efficiency, AI still finds it not easy to duplicate certain vital components of human translation.

Artificial intelligence (AI) advancements have brought about the development of conversational virtual assistants, some of which have been optimized to interpret and generate amusing responses. AI-driven

humor is presently witnessing dramatic advancements, and it has the potential to improve user interactions and generate humorous content.

The subtle aspects of humor and cultural references, which are crucial for accurate humor translation, are still difficult to capture, even with advancements in AI technology. These issues are addressed by developments in artificial intelligence (AI), which create more inclusive watching experiences via improving comedy translation (Abu-Rayyash, 2024, p. 1). Nonetheless, there is still much space for development in the field of translating comedy using AI models like ChatGPT. Although preliminary findings are encouraging, there are still issues with effectively communicating linguistically complicated jokes, capturing cultural differences, and customizing humor to various audiences. When analyzing neural machine translation systems, such as Google's Neural Machine Translation system, Koehn (2017, p. 90) observes that machine translation quality frequently falls short of human-level language proficiency. This is especially noticeable when discussing extremely contextual humor.

To optimize the efficacy of humor translation, neural machine translation models such as GPT-3 require further advancement in a number of areas. Abu-Rayyash (2024, p. 1) thinks that the requirements are as follows. Firstly, AI's ability to distinguish between linguistic plays on words and cultural references can be improved by refining algorithms using methods like neural networks and reinforcement learning. Secondly, collaborative models that combine AI translation with human expert input can combine sophisticated quality control with technological scale. This one, in specific, is of great relevance to this study. Thirdly, humor can be tailored for better reception and appeal by creating customized AI translations based on audience profiles and preferences.

Humor and Translation:

Translating humor is, first and foremost, a challenging enterprise for the interpreter, who must create and adapt the multiple kinds of jokes as well as humor described earlier by utilizing humankind's strongest ability: creativity. The main goal of the entire translation process in terms of humor is to preserve the total message of the initial version while adding a humorous touch. The translator faces the same difficulties when

translating the cultural reference of humor. To adapt these kinds of references again for the target audience, a broad knowledge base is required. Some of these elements are unfamiliar to the viewer, and the translator serves as a bridge between the two languages, not only linguistically but also culturally. In this regard, Mohammed (2019, p. 19) reckons that translators and interpreters must get past the cross-cultural barriers brought about by the disparities in expectations, conventions, and incongruities between the two cultures in order to convey humor. They are obligated to use adaptations, which they choose based on how similar the actions, words, or visual representations in the source and target cultures are.

Zabalbeascoa (2005, p. 185) has it that there is very little controversy among translation scholars on the notion that translation studies is an interdisciplinary field of study. The same is true of comedy studies, which use sociology, psychology, and linguistics, among other fields, to describe and create their theoretical models. Therefore, the intersection between translation studies and humor is not surprising.

Methodology and Data Analysis

The study features five different English sentences that incorporate wordplay. These sentences are then translated using six different AI programs. The focus of the study is solely on English-to-Arabic translation. In their paper, "A Study of Translation Edit Rate with Targeted Human Annotation," published in the Proceedings of the Association for Machine Translation in the Americas (AMTA) conference in 2006, Snover and colleagues introduced the Translation Edit Rate (TER) metric. This metric is used to evaluate the quality of machine-translated texts. TER is calculated through post-editing efforts and is proposed as a way to assess the quality of machine translation output by comparing it with a reference translation performed by humans. In this study, the researchers are going to apply this model on AI translations.

The number of changes required to convert inaccurate AI translations into an accurate translation is used to determine the TER score. Human translations serve as the basis for the edits. Thus, reference translation refers to the human translation that is carried out by the researchers. The number of words that make up the reference translation

sentence is then divided by the number of modifications. The following formula is used to determine the TER score:

[TER Score = No. of edits / No. of reference translation words]

A translation that is closer to the reference and therefore of higher quality is indicated by a low TER score. Conversely, a high TER score suggests a lower-quality translation and points to the need for additional corrections. However, since we are dealing with wordplay, we expect that any edits will primarily be at the word or phrase level. Therefore, we do not anticipate high TER scores, as the necessary adjustments will likely be limited to individual words or phrases. The researchers provided two reference translations: one in a more formal, classical style and another in a modern, colloquial one, because wordplay often relies on casual, everyday language and has a much stronger impact on the reader that way.

The TER (Translation Edit Rate) model will be used to evaluate the quality of AI-generated translations of English wordplay sentences into Arabic. This evaluation will include two specific criteria suggested by researchers: cultural adaptation and word play preservation, in addition to the typical TER criteria, which include insertions, deletions, substitutions, and shifts. By analyzing whether the translation effectively communicates culturally particular references and subtleties in a way that resonates to Arab audiences, cultural adaptation will be assessed. The final evaluation will thus combine quantitative TER scores with qualitative comments on these two additional criteria, ensuring that the translation not only aligns linguistically, but also retains the cultural and playful essence of the original English lines.

The study attempts to address the following research topics through data analysis:

What level of translation appropriateness may artificial intelligence programs achieve within word play domain? Are AI programs able to understand the word play context of the ST? Which of these artificial intelligence programs is more precise? Are AI programs able to produce cultural adaptation and preserve word play of the ST?

SLT 1: Shortest Horror Story: tomorrow is Monday.

<https://www.facebook.com/groups/768200854795175/posts/1306170860998169/>

Discussion: this expression means that people don't like Monday because it's the return to work day. Gemini, Deep Seek, Sider and Manus see the humor in "Shortest Horror Story: tomorrow is Monday..." comes from the

common feeling of dread that many people have about Monday. It's the day when the weekend ends, and the workweek begins again, often bringing a sense of reluctance or anxiety. By framing the arrival of Monday as a "horror story," the joke playfully exaggerates that universal sentiment—highlighting how many see Monday as the start of their least favorite part of the week. However, they couldn't create an equivalent effect in the target language—except for Deep Seek. That AI system managed to come close to the original effect, even though it relied on colloquial language. DeepL and QuillBot didn't add any explanation to the text; they only render it into Arabic.

The reference translations: أقصر حكاية مرعبة: - أقصر قصة رعب: غدا يوم الاحد
بكرة دوام

As evident in the analysis table, edits often contain the insertions, deletions, and substitutions required to convert the reference translation into the source text, or vice versa, depending on the standard approach used for TER computation. The difficulty with this kind of wordplay lies in how different cultures associate certain days of the week with various meanings and connotations. That's why AI systems often struggle to recognize these cultural nuances. To improve, they need to expand their knowledge of diverse cultural contexts and understand how different societies perceive and relate to specific days.

Although the TER scores are not high in many cases, the translations are still not quite accurate. A low TER score doesn't necessarily indicate an accurate translation, especially since the wordplay in the original text was not preserved. In other words, a low score doesn't guarantee that the essence or humor of the source was maintained.

Table 1. SLT (1) Analysis

No	Translator	TLTs	No. of Edits	TER Score	Cultural adaptation	Word play preservation	Appropriateness
1	Gemini	أقصر قصة رعب: غدا هو يوم الإثنين	2	0.3	-	-	Not appropriate
2	Deep Seek	أقصر حكاية رعب: بكرة يوم شغل	0	0	+	+	+
3	Sider	أقصر قصة رعب: غدا الإثنين	1	0.1	-	-	Not appropriate
4	DeepL	أقصر قصة رعب: غدا هو الإثنين	1	0.1	-	-	Not appropriate
5	QuillBot	أقصر قصة رعب: غدا هو الإثنين	1	0.1	-	-	Not appropriate
6	Manus	أقصر قصة رعب: غدا هو الإثنين	1	0.1	-	-	Not appropriate

SLT 2: What begins with T, ends with T and has T in it?**Teapot**

https://www.classools.net/OR/questions_list.php?fold=32&fname=JT325

Discussion: This expression involves a playful use of the letter "T" — the first two "T"s represent the alphabet letter, while the last "T" stands for "tea". This creates a linguistic challenge for both human and AI translators, as the meaning hinges on the similarity in pronunciation between the letter "T" and the word "tea". Gemini, Deep Seek, Sider, and Manus managed to grasp the core of this wordplay, recognizing that it's about the letter and the beverage since they sound alike. However, they struggled to replicate the effect of the original source, as this trick is deeply tied to the unique features of each language. Deep Seek managed partially to give an equivalent effect although it used colloquial style. Gemini and Manus have appropriate translation with partially preserving the word play in the text while culturally accepted. On the other hand, QuillBot only produced literal translations without capturing the deeper nuance or providing any explanation.

The reference translations: يبدأ بالتاء وينتهي بالتاء ووسطهم رجال كبير.....تابوت
يبدأ أول ما تبدي الحروف وينتهي لما تنتهي الحروف والي بداخله طعمه ينسبك كل الحروف....ابريق الشاي

When comparing the AI translations to the reference versions using the TER formula, it's clear that translating wordplay involving letter tricks is extremely challenging—especially when dealing with two languages that have very different sound systems. This complexity makes it even harder for AI translators to accurately capture the original nuance and humor, highlighting the difficulty of preserving such subtle linguistic features across languages.

The translations by Gemini, Deep Seek, and Manus were generally appropriate, with some minor improvements suggested by researchers. To be fully effective in the target language, some amendments were necessary. Deep Seek tended to use colloquial language, which is why we added "تي" to refer to tea—the word is used colloquially to help with cultural adaptation and to preserve the wordplay.

Table 2. SLT (2) Analysis

No	Translator	TLTs	No. of Edits	TER Score	Cultural adaptation	Word play preservation	Appropriateness
1	Gemini	ما هو الشيء الذي يبدأ بحرف التاء، وينتهي بحرف التاء، وبه حرف التاء؟ توت	4	0.5	+	+	+ With improvements ما هو الشيء الذي يبدأ بحرف التاء، وينتهي بحرف التاء، ومرة سوداء ومرة حمراء؟
2	Deep Seek	إيش يبدأ ب 'ت'، وينتهي ب ' ' 'ت'، وفيه 'ت'؟ تبيوت	2	0.25	+	+	+ With improvements إيش يبدأ ب 'ت'، وينتهي ب ' 'ت'، وفيه 'تي'؟
3	Sider	شيء يبدأ بحرف "ط"، وينتهي بحرف "ط"، وفيه شاي؟ ابريق الشاي	6	0.75	-	-	Not appropriate
4	DeepL	وينتهي T ما الذي يبدأ بحرف ويحتوي على حرف T بحرف T؟إبريق الشاي	6	0.75	-	-	Not appropriate
5	QuillBot	ما يبدأ بحرف التاء، وينتهي بحرف التاء، ويحتوي على حرف التاء؟ إبريق الشاي	5	0.6	-	-	Not appropriate
6	Manus	ماذا يبدأ بحرف التاء، وينتهي بحرف التاء، وبه حرف التاء؟ توت	4	0.5	+	+	+ With improvements ما هو الشيء الذي يبدأ بحرف التاء، وينتهي بحرف التاء، ومرة سوداء ومرة حمراء؟

SLT 3: Why did the math book look sad? Because it had too many problems

<https://www.facebook.com/groups/542942581191814/posts/1178978417588224/>

Discussion: This expression features a clever play on the word "problem," which has two meanings: the first refers to math problems, and the second to personal issues that can make someone sad. Gemini, Deep Seek, Sider, and Manus all understood the double entendre and explained both meanings. However, they struggled to find an Arabic word with a similar dual meaning—except for Deep Seek, who suggested "مسائل," which carries the same double sense as the English word. Deep Seek also added "ملهاش حل" ("no solution"), which, although it's an addition, successfully conveyed the effect of the original punchline. Both DeepL and Manus also used "مسائل," recognizing its double meaning.

QuillBot, on the other hand, translated the phrase literally, without capturing the wordplay.

The reference translations:

لماذا بدا كتاب الرياضيات مكتئب؟.....لأنه ملئ بالمسائل التي لا حل لها!
 ليش كتاب الرياضيات حزين؟.....لأنه التمارين الي بداخله مشكلة!!

By applying the TER formula and evaluating both cultural adaptation and the preservation of wordplay, it's clear that AI systems need to improve their ability to recognize the double meanings of certain words in both languages. They also need to be capable of preserving these puns and wordplays to maintain the original impact when translating from the source language to the target language. Otherwise, the cleverness and effect of the original expression will be lost.

Despite low Translation Error Rate (TER) scores, the accuracy of a translation of a specific text depended on capturing the double meaning of the word "problems." Only the translations by DeepSeek and Manus were successful. DeepSeek used a colloquial style, while Manus used the Arabic word "مسائل," which means both "issues" and "mathematical problems."

Table 3. SLT (3) Analysis

No	Translator	TLTs	No. of Edits	TER Score	Cultural adaptation	Word play preservation	Appropriateness
1	Gemini	لماذا بدا كتاب الرياضيات حزيناً؟ لأنه ملئ بالمسائل	1	0.1	-	-	Not appropriate
2	Deep Seek	ليش كتاب الرياضيات مكتئب؟ لأنه مليان مسائل ملهاش حل	0	0	+	+	+
3	Sider	ليش كتاب الرياضيات حزين؟ عنده مشاكل كثيرة!	2	0.2	-	-	Not appropriate
4	DeepL	لماذا بدا كتاب الرياضيات حزيناً؟ لأنه يحتوي على الكثير من المسائل	1	0.1	+	+	+
5	QuillBot	لماذا بدا كتاب الرياضيات حزيناً؟ لأن لديه الكثير من المشاكل	2	0.2	-	-	Not appropriate
6	Manus	لماذا بدا كتاب الرياضيات حزيناً؟ لأن لديه الكثير من المسائل	1	0.1	+	+	+

SLT 4: Why did the bicycle fall over? Because it was two-tired!

https://www.reddit.com/r/Jokes/comments/2i3m3s/why_did_the_bicycle_fall_over/

Discussion: This expression relies on wordplay with "two-tired," which sounds exactly like "too tired". Gemini, Deep Seek, Sider, and Manus all recognized the double meaning and understood the pun based on the

identical pronunciation. However, effectively conveying the same effect in the target language was challenging because of the unique characteristics of each language. They struggled to find a single word in Arabic that captured both meanings; most either suggested "متعبة" (tired) or "إطارين" (two tires). Only Manus proposed "إطارين متعبين," which successfully preserved the double sound by combining two words that reflect both "two-tired" and "too tired". Deep Seek used the colloquial phrase "اثنين كاوتش" (two tires), but it only conveyed the literal meaning of "two tires" and didn't capture the pun or the double meaning of "too tired." Sider added the unnecessary phrase "تعبانة من التعب" (tired from being tired), which is not an accurate representation of the original joke. Meanwhile, DeepL and QuillBot, as usual, provided only literal translations without any explanation or attempt to preserve the wordplay.

The reference translations: لماذا سقطت الدراجة؟... لأنها ذات إطارين متعبين
ليش وقعت الدراجة؟.... لأنها تعبت من الوقفة على رجليتين!!

Based on the table and the researchers' criteria, translating nuanced linguistic qualities, such as homophones, is difficult for both AI and human translators. AI systems need to improve their ability to handle such linguistic and cultural complexities.

Table 4. SLT (4) Analysis

No	Translator	TLTs	No. of Edits	TER Score	Cultural adaptation	Word play preservation	Appropriateness
1	Gemini	لماذا سقطت الدراجة؟ لأنها كانت مستعجلة	2	0.2	-	-	Not appropriate
2	Deep Seek	ليش وقعت الدراجة وقعت؟ لأنها كانت اثنين كاوتش	2	0.2	-	-	Not appropriate
3	Sider	ليش وقعت الدراجة؟ عشان كانت تعبانة من التعب	4	0.5	-	-	Not appropriate
4	DeepL	ماذا سقطت الدراجة؟ لأنها كانت متعبة!	2	0.2	-	-	Not appropriate
5	QuillBot	لماذا سقطت الدراجة؟ لأنها كانت متعبة جدا!	3	0.4	-	-	Not appropriate
6	Manus	ماذا سقطت الدراجة الهوائية؟ لأنها كانت ذات إطارين متعبين!	0	0	+	+	+

SLT 5: Why eggs don't tell jokes to each other? Because they might crack each other up

https://www.reddit.com/r/Jokes/comments/1455yfe/why_eggs_dont_tell_jokes_to_each_others/

Discussion: This expression contains a wordplay in the phrase "crack each other up," which is an idiom. When used figuratively, it means to make each other laugh heartily, but literally, it suggests that they will break or crack each other. This double meaning creates the humor. Gemini, Deep Seek, Sider, and Manus all easily recognized the pun in the original expression. However, not all of them managed to convey this dual meaning in the target language, due to the unique characteristics of each language. Gemini, Manus, and QuillBot used the words "يفقس" (to hatch) or "ينفجر" (to burst), while Deep Seek used colloquial language and successfully captured the effect of the original. This task was somewhat easier for AI systems because in Arabic, this humor is already expressed with phrases like "يفرقع من الضحك" or "ينفجر من الضحك," which are common idiomatic ways to say someone is laughing very hard.

The reference translations: لماذا لا تروي البيضات نكتا لبعضها؟..... لأنها قد تتكسر من الضحك

ليش البيض ما يحكوا نكت لبعض؟..... لأنه ممكن ينفجروا من الضحك!!

An AI found the expression easy to translate because both languages share the same idiom. This shows that the more similar two languages are, the easier the translation; the more different they are, the more difficult it becomes.

Table 5. SLT (5) Analysis

No	Translator	TLTs	No. of Edits	TER Score	Cultural adaptation	Word play preservation	Appropriateness
1	Gemini	لماذا لا يروي البيض النكات لبعضها؟ لأنهم قد "يفقسون" في الضحك	0	0	+	+	+
2	Deep Seek	ليش البيض ما يحكوا نكت لبعض؟ عشان ممكن يققعوا من الضحك	0	0	+	+	+
3	Sider	لماذا لا تضحك البيوض على بعضها؟ لأنها قد تكسر بعضها!	4	0.5	-	-	Not appropriate
4	DeepL	ماذا لا تروي البيضات النكات لبعضها البعض؟ لأنها قد تكسر بعضها البعض	4	0.5	-	-	Not appropriate
5	QuillBot	لماذا لا تروي البيض النكات لبعضها البعض؟ لأنهم قد ينفجرون من الضحك على بعضهم البعض	0	0	+	+	+
6	Manus	لماذا لا تحكي البيضات النكات؟ ! لأنها قد "تفقس" من الضحك	0	0	+	+	+

Findings

What level of translation appropriateness may artificial intelligence programs achieve within word play domain?

Based on our data, we have 5 texts and 6 AI translators, totaling 30 translations to evaluate. Of these, 9 were appropriate as they accurately conveyed the meaning. An additional 3 translations were suitable but required some improvements by researchers. However, 18 translations were not appropriate because they failed to effectively transfer the wordplay's meaning.

Appropriate	Not Appropriate	Appropriate with improvements	Total
9	18	3	30
30%	60%	10%	100%

Are AI programs able to understand the word play context of the ST?

As shown in the analysis, Gemini, Deep Seek, Sider, and Manus all demonstrated a solid understanding of the context of the source text, including the wordplay and humor present in the original. Although they sometimes failed to fully convey these elements in the target text, they were able to recognize and interpret the wordplay effectively. The AI programs were capable of understanding different types of wordplay, providing explanations for each instance and clarifying for researchers that these were indeed wordplays. In some cases, they also noted that certain wordplays are difficult to translate because they are closely tied to specific language features, such as words with multiple meanings or words that sound similar but are spelled differently.

Which of these artificial intelligence programs is more precise? and Are they able to produce cultural adaptation and preserve word play of the ST?

	Deep Seek	Manus	Gemini	DeepL	QuillBot	Total
Appropriate	3	3	1	1	1	9
Appropriate With improvements	1	1	1			3
Cultural adaptation	4	4	2	1	1	12
Word play preservation	4	4	2	1	1	12

As shown in the table above, the most accurate AI programs in handling wordplay are Deep Seek and Manus. Although Deep Seek primarily used colloquial language, it was still effective in conveying the meaning.

Regarding cultural adaptation and the preservation of wordplay, which are the true essence of our work, it appears that only a few cases—specifically Deep Seek and Manus—managed to preserve the wordplay, while other AI programs failed. This indicates that AI still needs to be improved to better understand and incorporate contextual nuances across different languages. Currently, most AI systems tend to rely on literal translation, especially when they encounter humor or wordplay, as they often see these as just forms of humor rather than linguistic features that require special handling.

Conclusions and Recommendations

Humor is essential to human communication, as a source of entertainment as well as a means of social expression and cohesion. Examining humor from a sociolinguistic perspective demonstrates how intricately it interacts with linguistic, cognitive, and cultural elements.

Translating humor poses particular challenges, because of linguistic subtleties and cultural allusions. Effective decoding of humoristic patterns is necessary for translators to provide the desired humorous impact across languages.

The social functions and psychological implications of comedy are influenced by cultural variances in how it is perceived.

One can recognize humor's transforming power on interpersonal relationships and cultural variety by embracing it within a sociolinguistic framework.

Promoting mental health with culturally appropriate humor interventions requires an understanding of these disparities.

Future studies should concentrate on improving translation techniques and expanding our knowledge of the dynamics of humor across cultural boundaries.

References:

- Abu-Rayyash, H. (2024). AI meets comedy: Viewers' reactions to GPT-4 generated humor translation. *Ampersand* 12 (2024) 100162, Kent State University, 424 Lake Street, Kent, OH, USA
- Attardo, S. (1994). *Linguistic theories of humor*. Berlin: Mouton de Gruyter.
- Bell, N., & Pomerantz, A. (2016). *Humor in the classroom: A guide for language teachers and educational researchers*. Routledge
- Chafe, W. (1994). *Discourse, consciousness and time*. Chicago: University of Chicago Press.
- Degabriele, J. & Walsh, I. P. (2010). "Humour appreciation and comprehension in children with intellectual disability." *Journal of Intellectual Disability Research* 54(6):525-37
- Delabastita, D. (1994). Focus on the Pun: Wordplay as a Special Translation Problem in Translation Studies. *Target* 6: 2, 223–443.
- Delabastita, D. (1996). "Introduction." *The Translator: Studies in intercultural Communication* 2.2.
- Deneire, M. (1995). Humor and Foreign Language Teaching. *Humor*, 8, 285-298.f
- Feingold, A. & Mazzella, R. (1993). Preliminary validation of a multidimensional model of wittiness. *Journal of Personality* 61(3), 439-456. Retrieved June 19, 2010 from Academic Search Premier.
- Koehn, P., (2017). *Neural Machine Translation*. Center for Speech and Language Processing Department of Computer Science Johns Hopkins University ArXiv Preprint ArXiv:1709.07809.
- Hockett, F. (1972) Jokes. In: Estellie Smith (ed.). *Studies in linguistics in honor of George L. Trager* (New York: Mouton), 153–178.
- Martin R. A., Ford T. (2018). *The Psychology of Humor: An Integrative Approach*. Burlington, MA: Elsevier Academic Press.
- Martin, R. & Lefcourt, H. (1983). Sense of humor as a moderator of the relation between stressors and moods. *Journal of Personality and Social Psychology*, 45(6), 1313-1324. Retrieved December 17, 2007 from PsychArticles.
- Mohammed, S. (2019). Translating Humorous Expressions into Arabic with Reference to Loss, Gain, and Compensation. *International Journal of Applied Linguistics & English Literature*, 8(3), pp. 16-22

- Morreall, J. (2009): *Comic Relief: A Comprehensive Philosophy of Humor*.
New Jersey: Wiley-Blackwell,.
- Raskin, V. (1985). *Semantic mechanisms of humor*. Dordrecht, Netherlands: D. Reidel.
- Ruch, W. (2002b). Values is action (Via) Classification of strength. Submitted for publication. Retrieved February 7, 2008 through personal communication with Ruch, W.
- Shifman, L. (2007). Humor in the age of digital reproduction: Continuity and change in Internet-based comic texts, *International Journal of Communication* 1(1),187-209.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers* (pp. 223-231).
- Zabalbeascoa, P. (2005) "Humour and Translation –an interdisciplinary. 'Humour: *International journal of Humour Research* 18.2: 185-207.