Rafael Carvalho Falleiros

20322135

Memes and Puns in Translation: an MT Evaluation from English into Brazilian Portuguese

Trinity College Dublin
MPhil in Literary Translation

2021

Supervised by Dr. Sheila Castilho
I declare that this dissertation has not been submitted as an exercise for a degree at this or any other university and that it is entirely my own work.

I agree that the Library may lend or copy this dissertation on request.

Signed: Rafael Conceição Belluom Date: 07/09/2021
This work aims to solve a gap in research in machine translation (MT) of humorous pieces, specifically memes and puns from English into Brazilian Portuguese. The reason for this study is the importance that humor has played in social media, particularly during the COVID-19 pandemic as a form of comic relief – memes are continuously produced in massive numbers, and puns are historically relevant for Humor Studies. Despite the number of linguistic and psychological studies on humor, the scientific literature of the field indicates a lack of solid implementations in MT and humor, both in academia and in industry. Thus, this research translated nine memes and six puns using freely available MT systems, i.e. Google Translate, Bing Microsoft Translator, and DeepL Translate, in order to identify if and how humor is conveyed by these applications. To carry out this research, a User-Centered Translation (UCT) approach was chosen alongside Multidimensional Quality Metrics (MQM) definitions. Therefore, two surveys based on UCT were filled out by thirty volunteers. Furthermore, a translation quality analysis was performed by an invited professional translator using MQM as a basis. The results of both analyses indicate that, in general, Google Translate had a better performance in conveying humor, despite its error in register. It was followed by a tied overall functioning of Bing Translator and DeepL Translate, the latter presenting more errors regarding the wrong variety of Portuguese in the translations, and the first showing flaws on untranslated terms and mistranslations. Interestingly we found that images are vital for memes’ understanding and the source text also influences users’ perception. Hence, we concluded that humor can be conveyed in a basic MT outputs, which can be useful for non-native speakers or users who want to have a basic idea of the meme or pun’s semantics.
Acknowledgements

First and foremost, I would like to thank my parents, Adriana Lobo de Carvalho and Mauro de Paula Falleiros, for supporting me with an unconditional love. Without their motivation, I would not be able to finish this work. Also, I would like to thank my fiancée, Anna Beatriz Dimas Furtado, who also joined me in this adventure of studying a master’s degree abroad in the middle of this very challenging moment we are living in.

Secondly, I would like to thank my supervisor, Dr. Sheila Castilho, who has been patient and very helpful throughout this dissertation and my portfolio. I could not ask for a better guidance during this period. My colleagues, who became friends, are also responsible for making me laugh and help me throughout this whole process. Even though most classes were virtual, and we were not able to see each other often, I am really grateful for the connections we made during this course. Furthermore, I would like to thank Dr. James Hadley for being always so helpful, organized, and supportive. I learned a lot about being succinct, scientific, and responsible in my writing because of him.

Finally, I would like to acknowledge the importance of all of my friends and colleagues in Brazil who took part in this research directly or indirectly. Thank you for the thirty participants who filled out the survey, and for João Camargo, who helped me with the translation analysis of the pieces presented here. I could not forget to mention Patricia Cardoso Moreira, one of my professors in my undergraduate course in Brazil, who inspired and helped me with her humor and research.

We have not been living in one of the easiest years of the past decades. COVID-19 showed the importance of caring about each other, and of respecting my own limitations. Thus, I would also like to thank all of the health professionals around the world who have been fighting for almost three years now to preserve life the best way they can. To those who lost someone during this tragic period, I hope you can find comfort and hope with time.
# Contents

Introduction .......................................................................................................................... 1

Chapter 2 – Literature Review ............................................................................................. 4
  2.1. Machine Translation ................................................................................................. 4
  2.2. Machine Translation Evaluation ............................................................................... 6
  2.3. Overview of Humor Studies ...................................................................................... 8
  2.3.1 Memes and Puns .................................................................................................. 13
  2.4. Machine Translation and Humor Studies ................................................................ 14

Chapter 3 – Methodology ................................................................................................. 18
  3.1. Justification ............................................................................................................. 20
  3.2. Materials ................................................................................................................. 21
  3.2.1. Memes and Puns ................................................................................................. 21
  3.2.2. Ranking and Questionnaire – Surveys ................................................................. 26
  3.2.3. Participants .......................................................................................................... 31
  3.2.4. Translators Analysis and MQM ............................................................................ 33

Chapter 4 – Finding and Analysis ....................................................................................... 37
  4.1. MT Quality via Translator’s Assessment ................................................................. 37
  4.2. End-User Reception ................................................................................................ 41
    4.2.1. Survey 1 – ST ........................................................................................................ 41
    4.2.2. Survey 2 – No ST .................................................................................................. 49
    4.2.3 Average Answers for Both Surveys ....................................................................... 57

Chapter 5 – Discussion ...................................................................................................... 58

Conclusion ......................................................................................................................... 61

Limitations ....................................................................................................................... 61

Future Research .............................................................................................................. 63
List of Figures

Figure 1 ............................................................................................................................. 25
Figure 2 ............................................................................................................................. 25
Figure 3 ............................................................................................................................. 25
Figure 4 ............................................................................................................................. 26
Figure 5 ............................................................................................................................. 27
Figure 6 ............................................................................................................................. 28
Figure 7 ............................................................................................................................. 29
Figure 8 ............................................................................................................................. 31
Figure 9 ............................................................................................................................. 32
Figure 10 ......................................................................................................................... 43
Figure 11 ......................................................................................................................... 45
Figure 12 ......................................................................................................................... 45
Figure 13 ......................................................................................................................... 45
Figure 14 ......................................................................................................................... 46
Figure 15 ......................................................................................................................... 46
Figure 16 ......................................................................................................................... 47
Figure 17 ......................................................................................................................... 50
Figure 18 ......................................................................................................................... 51
Figure 19 ......................................................................................................................... 51
Figure 20 ......................................................................................................................... 52
Figure 21 ......................................................................................................................... 52
Figure 22 ......................................................................................................................... 52
Figure 23 ......................................................................................................................... 53
Figure 24 ......................................................................................................................... 53
Figure 25 ......................................................................................................................... 53
Figure 26 ......................................................................................................................... 54
Figure 27 ......................................................................................................................... 54
Figure 28 ......................................................................................................................... 55
Figure 29 ......................................................................................................................... 55
List of Tables

Table 1 .............................................................................................................................. 23
Table 2 .............................................................................................................................. 24
Table 3 .............................................................................................................................. 35
Table 4 .............................................................................................................................. 36
Table 5 .............................................................................................................................. 38
Table 6 .............................................................................................................................. 38
Table 7 .............................................................................................................................. 42
Table 8 .............................................................................................................................. 48
Table 9 .............................................................................................................................. 56
Table 10 ............................................................................................................................ 57
**Abbreviations**

AI – Artificial Intelligence

BLEU – Bilingual Evaluation Understudy

EN-US – American English

ICLR – International Conference on Learning Representations

MQM – Multidimensional Quality Metrics

MT – Machine Translation

NMT – Neural Machine Translation

NBF – Non-bona fide

OED – Oxford English Dictionary

PT-BR – Brazilian Portuguese

PT-EU – European Portuguese

RQ – Research Question

SMT – Statistical Machine Translation

ST – Source Text

TER – Translation Error Rate

UCT – User-Centered Translation

US – United States

UX – User Experience

TQA – Translation Quality Assessment

WER – Word Error Rate
Introduction

This work aims at answering an overarching research question (RQ), namely ‘Would humor still be perceived in nine memes and six puns after being translated from American English (EN-US) into Brazilian Portuguese (PT-BR) by Machine Translation (MT) systems?’.

Moreover, this RQ resulted in four sub-research questions (see Chapter 3). The systems chosen were freely available translation software, i.e. Google Translate, Bing Microsoft Translator, and DeepL Translate. Therefore, three fields of research are necessary to encompass a possible answer to this RQ. That is MT, Translation Quality Assessment (TQA), and Humor Studies.

To carry out this work, the primary methodology chosen was the User-Centered Translation (UCT) approach, formulated by Suojanen, Koskinen, and Tuominen (2015) and applied in TQA for MT by Castilho (2016) and Castilho et al. (2017a, 2018). In addition, this research also draws significantly on Multidimensional Quality Metrics (MQM) definitions, which were used as the main parameters of analysis.

As a result, the nine memes and six puns selected were analyzed through two main categories of participants: (1) by an invited translator who assessed each translation using MQM’s concepts of accuracy, fluency, and register – this study was aimed at assessing MT quality; and (2) by thirty volunteers who partook two surveys with a ranking and a questionnaire to measure the end-user reception of the translations presented.

Three memes were selected (i.e. Drake, Distracted Boyfriend, and Confession Bear), alongside three different captions for each one – thus, resulting in nine pieces. They were chosen to reduce bias. In other words, these images and captions would be neither totally unknown nor totally unfamiliar.

Moreover, six puns were selected and divided into two sets according to their textual length. The first three were textually longer, with more contextual and elaborate wordplay. The reason for that is to test MT quality again and how they would function with more contextual pieces of text. The second set of three puns were shorter, with a more direct wordplay, requiring less contextual knowledge, which is another parameter for analyzing MT performance.
Chapter 2 is the literature review, where an overview of the historical background of the theories selected (i.e. MT, TQA, Humor Studies, and their relation to translation and MT) for this work are introduced. This chapter traces the state-of-the-art technology used during the twentieth century (Statistical Machine Translation) and how it has switched in the twenty-first century (Neural Machine Translation). Furthermore, the theories of humor of Attardo and Raskin (2017) are described for a further understanding of how linguistics views humorous texts. At last, we identify a gap in the research of memes and puns in MT mainly through the work of Miller (2019).

Furthermore, Chapter 3 contains the methodology (UCT and MQM), where the justification and the materials (e.g. memes, puns, surveys, and translator’s analysis spreadsheet) are presented.

Chapter 4 focuses on the results of the surveys and the translator’s analysis and on answering each of the four sub-research questions accordingly. We found issues in the three MT systems, Google Translate, Bing Translator, and DeepL Translate, mainly regarding register – the high level of formality of the translations was acknowledged by the invited translator and the participants. Nonetheless, untranslated terms and mistranslations were also pointed out as obstacles for users’ humor perception. Even so, the translations were considered a reasonably good source of basic understanding of memes and puns, which could be helpful for a non-native speaker. In the end, Google Translate was generally chosen, both by the invited translator and by the participants, as the system which provided the most satisfactory translations. Bing Translator was placed in second place by the invited translator and in third place by the participants. And DeepL came in last, according to the translator; on the other hand, it was considered the second best by the participants.

Our final chapter, Chapter 5, presents a discussion about the results to answer the overarching RQ. From the results and the answers to the four sub-research questions, we could conclude that it is somewhat possible to convey basic humor in memes and puns translated by the aforementioned MT systems from EN-US into PT-BR, despite the numerous improvements that those systems still can benefit from.
The conclusion of this work elaborates on its limitations, namely the constricted number of participants and potential future research opportunities. For instance, a new MT system could be built explicitly for humor research in translation. Furthermore, appendixes with the surveys, the spreadsheet, and the gloss translations are attached to this dissertation's last pages.
Chapter 2 – Literature Review

Overview – This chapter focuses on the literature used to base this dissertation’s research. First, it describes the state-of-the-art Machine Translation systems and evaluation methods. It then presents a brief background of Humor Studies, their relation to Machine Translation, and the latest research on the topic.

2.1. Machine Translation

Machine Translation (MT) has been defined as the automated translation of texts written in natural languages with the support of software (Hutchins 2003, 483). The studies in this area started in the 1950s when MT was thought of as a coding and decoding problem (Poibeau 2017, 11.24), which could allegedly be solved with a coded system.

One of the first approaches to machine translation was the direct approach, which used bilingual dictionaries alongside translation/grammar rules to avoid ambiguity and translate sentences on a word-for-word basis (ibid).

The ALPAC report, published in 1966, was a landmark for MT funding and relevance in the United States since it significantly devalued the importance, the effectiveness, and the cost that translations done by machines could offer. It concluded that human translators were better and cheaper than developing this kind of technology because the post-editing effort was too great to be worth it (ibid): ‘There is no emergency in the field of translation. The problem is not to meet some non-existent need through non-existent machine translation.’ (ALPAC Report, 16). As a consequence of the ALPAC report, the MT field in the US paused until the 1990s, continuing only in a few universities (e.g. Wayne State University and the University of Texas) and in companies (Poibeau 2017, 12.14). Meanwhile, MT funding continued in Europe and Canada, mainly because of multilingual contexts (ibid). Again, Canada is a relevant example, as Thierry Poibeau (2017, 12.19) points out: ‘The need to produce a large quantity of official documents in English and in French led to high costs, which created a strong incentive to launch research in the field.’

Available at: https://www.nap.edu/read/9547/chapter/12.
In that context, Systran, one of the first relevant automatic translation systems developed by a private company, was created in 1968 by Peter Toma. And with the evolution of computers, technology, and data collection, brand-new systems arrived in the field. Using bilingual parallel texts (i.e., corpora), methods were developed to either have the examples as a basis for the system to translate using an analogy – or copying the examples given (Example-Based MT)\(^2\), or statistically predict how a text should be translated (Statistical MT (SMT)). Those approaches are encompassed in the Corpus-Based Machine Translation Approach because it primarily uses corpora methodology for extracting meaning and performing the translations (Okpor 2014; Rivera-Trigueros 2021). Furthermore, SMT was one of the significant factors that raised governmental and private funding back to the field (Hutchins 2007, 17). Thanks to the technological development of GPUs and computer science in general (Koehn 2020; Poibeau 2017), SMT was gradually adopted by notable names in the industry e.g. Google, Microsoft, SYSTRAN (ibid). These developments enabled SMT to become and stay relevant for years, being used by numerous companies, governments, translation industries, and end-users (Koehn and Knowles 2017; Koehn 2020; Poibeau 2017).

Neural Machine Translation (NMT) is the name given to MT systems that use artificial neural networks to learn how to best translate a text through deep learning and artificial intelligence (AI). It was first (re-)introduced by Bahdanau et al. (2015), who presented at the International Conference on Learning Representations (ICLR), with results highlighting the studies that enabled the developments in the field (Kalchbrenner and Blunsom 2013; Sutskever et al. 2014; Cho et al. 2014b), while also proposing an improved approach of Cho et al. (2014a) and Sutskever et al. (2014), the RNN Encoder-Decoder method (Bahdanau et al. 2015, 2). NMT uses a more considerable collection of data strengthened by the Internet and companies such as Google, Facebook, and Wikipedia (Koehn 2020, 37–39).

Additionally, even though NMT is still a somewhat new field, requiring more research to enhance the technology to be fully launched in the market (Castilho et al. 2017b, 118; Castilho et al. 2019), it has quickly become the new state-of-the-art MT system (ibid

\(^2\) See more about the different types of systems and the history of MT at Poibeau (2017) and Koehn (2020).
2017b, 109), being implemented by relevant translation companies, e.g. Google, Microsoft, DeepL, and Systran (Crego et al. 2016; Junczys-Dowmunt et al. 2016; Koehn and Knowles 2017; Rivera-Trigueros 2021; Wu et al. 2016). An example of how NMT has been predominant in the industry is Google Translate’s significant mark of processing 143 billion words a day, showing the influence of this technology in users’ daily lives.3

In this work, we used three online NMT systems, namely Google Translate,4 Bing Microsoft Translator,5 and DeepL Translate,6 because of their ease of use and free availability to end-users who want to translate from languages they are not fully fluent in.

2.2. Machine Translation Evaluation

To maintain a quality standard and keep track of the innovations in the field, MT evaluation is an essential practice carried out by scholars and the industry’s specialists. However, there is no consensus in that area, which varies from university to university – having a theoretical focus –, and from company to company – bearing more direct, user-focused, and product-focused evaluation approaches (Castilho et al. 2018; Rivera-Trigueros 2021).

Currently, there are essentially two types of MT evaluation: automated and human evaluation metrics. As the name suggests, the first is done mainly by a software program,7 which compares the MT’s output translation with a couple of reference translations (Castilho et al. 2018; Han 2016; Rivera-Trigueros 2021); the second is carried out by professional translators, researchers, crowd workers, or simply bilingual people. Furthermore, automated and human metrics use different indicators or metrics as a baseline for the evaluation (Castilho et al. 2018; Rivera-Trigueros 2021). Throughout the

3 Data made available by Google’s Senior Vice President of Hardware, Rick Osterloh, at the ‘Made by Google’ technology event. Retrieved from: https://www.androidpolice.com/2018/10/09/google-translate-processes-143-billion-words-every-day/#:~:text=Google%20Translate%20is%20one%20of,single%20day%2C%20across%20100%20languages.
4 Available at: https://translate.google.com/.
5 Available at: https://www.bing.com/translator.
6 Available at: https://www.deepl.com/en/translator.
7 Since the reference translations used by the system to compare the output text are primarily translated by humans, it is not considered a ‘totally automated’ evaluation (Castilho et al. 2018; Rivera-Trigueros 2021).
years, the primary automated metrics developed were the following: Word Error Rate (WER), Bilingual Evaluation Understudy (BLEU), Translation Error Rate (TER).

In human evaluation, one of the most used approaches is to have the evaluator look for adequacy and fluency indicators (Rivera-Trigueros 2021), usually using a Likert-scale (Likert 1932). Adequacy evaluates semantic quality: whether the information was accurately transferred or not. This assessment requires comparison with reference translations (monolingual) or with the original text (bilingual). Fluency, on the other hand, evaluates syntactic quality. In this situation, comparison with the original text is unnecessary (but can be displayed), and the evaluation carried out is monolingual. Scales, rankings, error annotation, gap-filling, annotation, and classification are other translation quality evaluation methods used (Castilho et al. 2018; Chatzikoumi 2020; Rivera-Trigueros 2021).

For this dissertation, error annotation was employed by the invited translator by looking into errors that affect adequacy, fluency, and register by two translators. The error categories are taken from the Multidimensional Quality Metrics (MQM), which was created to define ‘a model to declare multiple metrics rather than one single metric’ (Lommel et al. 2014, 459). MQM measures the quality of a text-based on multiple and flexible quality indicators built on textual characteristics or on textual errors. For instance, accuracy (adequacy) and fluency are already present there. Still, ramified metrics stem from them (e.g. mistranslations, omission, over-translation, under-translation, cohesion, coherence, inconsistency, spelling, and untranslation). Further, the style was another indicator considered with its register ramification. The definition, retrieved from MQM, for each term is described below:

- **Accuracy**: ‘The target text does not accurately reflect the source text, allowing for any differences authorized by specifications. The extent to which the translation transfers the meaning of the source text translation unit into the target.’

---

8 Other precision-focused metrics include NIST (Doddington, 2002), ROUGE (Lin and Hovy, 2003), F-measure (Turian et al., 2003), and METEOR (Banerjee and Lavie 2005). See more at Rivera-Trigueros (2021), ‘Machine Translation Systems and Quality Assessment’.

9 Adequacy is often known as accuracy. The latter will be the main term used in this work.

• **Fluency**: ‘Issues related to the form or content of a text, irrespective as to whether it is a translation or not. The flow and naturalness of the translation in the context of the target audience and its linguistic and sociocultural norms in the given context.’

• **Register**: ‘The text uses a level of formality higher or lower than required by the specifications or general language conventions.’

Moreover, a principal methodology for this dissertation is the **user-centered translation** (UCT) approach. It was described by Suojanen, Koskinen, and Tuominen (2015) and applied by Castilho (2016) and Castilho et al. (2017a, 2018), who used this methodology in MT evaluation. UCT is a method derived from user-centered design (UCD), a technique used in various fields, e.g., user-experience design, user-interface design, design thinking, and many more (Castilho 2016; Suojanen, Koskinen and Tuominen 2015), which has the user as a main theoretical and practical participant and goal in the production and/or evaluation of a product; in this case, translation (ibid). This implemented by applying surveys, questionnaires, interviews, and other tools to assess the user experience in each stage of the product’s development (ibid).

In this dissertation, UCT is employed using rankings and questionnaires to verify if end-users of freely available online MT systems (Google Translate, Bing Microsoft Translator, and DeepL Translate) can perceive humor in memes and puns translated by those systems. The study was conducted by asking them to compare three different MT translations of nine memes and six puns. They will evaluate each of them according to their satisfaction\(^{11}\) and to their technical knowledge in translation or the source text’s (ST) language (English).

### 2.3. Overview of Humor Studies

As we know it today, Humor has modern roots, being traced no further back than the 20th century (Larkin-Galiñanes 2017, 4). Nonetheless, the concept has ancient origins, being seen described as comedy/satire by the ancient Greeks, who seemed to perceive the

---

\(^{11}\) Satisfaction is defined in the ISO 9241-11:1998 as ‘freedom from discomfort, and positive attitudes towards the use of the product.’ Available at: [https://iso.org](https://iso.org).
laugher as a vulgar\textsuperscript{12} activity, preaching ‘for its restriction rather than to study it in any sort of detail.’ (ibid). For Plato, excessive laughter was a false superiority that one would feel. Aristotle complemented that thought by differentiating the buffoonery, diminishing (lower) humor from the wordplay, language, incongruity (superior, refined) humor (ibid).

According to Amy Carrell (2008, 314–15), ‘humor resides with the audience; and thus, nothing is inherently humorous, or funny. Some joke texts will succeed for one audience and fail to fire for another.’ Not only the audience is essential, though. The relationship between audience, environment, comedian, and joke must be in harmony, working together. Otherwise, humor is not achieved (ibid).

Moreover, humor is situational, depending not only on the audience’s previous knowledge of the themes presented but also on the society’s cultural, social, and moral values of the period analyzed (Meyer 2000). Therefore, what appears highly amusing to one audience may seem irrelevant or just intriguing in another (Carrell 1992; Duncan 1982; Meyer 2000; Winick 1976). Humor may also be considered more or less acceptable in communicating in specific contexts (Hackman and Barthel-Hackman 1993; Meyer 2000). This is one of the reasons why the participants of this study were somewhat homogeneous in their age and their average amount of humor consumption.

Humor has been studied in philosophy, psychology, linguistics, sociology, and even computer science (Attardo 2017). Thus, some relevant theories repeatedly mentioned throughout the years by scholars can be highlighted:\textsuperscript{13}

\begin{itemize}
  \item \textbf{Relief Theory}: this theory attempts to explain the psychological reasons why people laugh. According to it, people laugh and enjoy comedy because they perceive their tension has been eased in some way (Berlyne 1972; Morreall 1983; Shurcliff 1968). The physiological ‘symptoms’ of humor are relevant to this theory, claiming that humor is derived from the relief experienced when tensions are induced and eliminated from the spectator (Meyer 2000).
\end{itemize}

\textsuperscript{12} This is due to their perception of laughter as an emotion that should be refrained by Reason. Additionally, being an emotion, it should be reserved for the lower-class in the comedies and satires. When you laugh, you feel a superiority that does not exist, making you look ridiculous (Larkin-Galiñanes 2017, 5).
\textsuperscript{13} See in more detail at Cristina Larkin-Galiñanes (2017, 4–16) and John Meyer (2000, 310–331).
• **Incongruity Theory:** this theory stems from Cognitive Studies. It states that people laugh at what surprises them, something unexpected or strange, in a nonthreatening way (Berger 1976; Deckers and Divine 1981; McGhee 1979). Thus, an accepted pattern is broken, or a change is noticed—close enough to the norm to be unthreatening, yet far enough from it to be noticeable (Meyer 2000).

• **Superiority Theory:** As discussed previously, this theory dates to ancient Greece, being mentioned by Plato, Aristotle, and other philosophers (Larkin-Galiñanes 2017). It tries to discover the reason for laughter by suggesting that people laugh at others because they feel triumphant over them or superior to them in some way (Feinberg 1978; Grotjahn 1957; Gruner 1997, 1978; Meyer 2000; Morreall 1983; Rapp 1951; Ziv 1984).

Furthermore, scholars have been theorizing about humor’s different functions has for the audience (e.g. educational, social, satirical). Some of the major ones will be listed above, but it is important to note that functions are not static (Meyer 2000, 323). All functions can act together in pairs or trios to fulfill the different objectives that the comedian/joke has. Moreover, their definitions seem to intertwine in meaning, bearing very similar uses by communicators/comedians. All this, of course, varies depending, again, on the context, the audience, and their previous knowledge (ibid):

• **Identification:** Humor can help generate support by identifying communicators/comedians with their listeners, increasing the speaker’s credibility (Chang and Gruner 1987; Gruner 1967,1985; Malone 1980; Meyer 2000), and rising group cohesiveness (Graham et al. 1992; Meyer 2000). Feelings that would otherwise be suppressed due to a lack of social acceptance can, because of this function, now be securely shared through humor (Meyer 2000; Winick 1976).

• **Clarification:** Humor is also used by communicators/comedians to state their points of view into memorable phrases or brief anecdotes, resulting in the explanation of topics or stances. This method not only increases audience recollection of the event
(Goldstein 1976; Gruner 1967, 1985), but it also gives a focus for media coverage in today's era of brief television announcements (Meyer 2000).14

- **Enforcement**: Humor allows a communicator/comedian to discreetly impose rules by leveling criticism while preserving some amount of listener identification (Graham et al. 1992). Politicians widely use this to reinforce their agendas while criticizing the opposition without looking like a ‘bitter, angry critic’ (Meyer 1990, 2000).

- **Differentiation**: Finally, communicators frequently employ differentiation to contrast themselves with their opponents, their opinions with those of others, and their social group with that of others. Both alliances and distinctions are made using humor. Instead of outrage, anger, or violence, one might criticize something or someone with humor by ridiculing the adversary through laughing (Meyer 2000; Volpe 1977). While highlighting flaws and disagreements with others, comic criticism can also sustain identification and political solidarity among members of one group (Meyer 2000; Schutz 1977).

As a conclusion to this overview, it is relevant to add the three major linguistic theories developed to understand verbal humor:15

- **Semantic-Script Theory of Humor (SSTH)**: according to this theory, any joke is entirely or partially compatible with two different scripts, each of which is opposed in some way. Scripts are defined as logically organized bits of information about the world (Raskin 2017).

- **Ontological Semantic Technology (OST)**: it is a theory, methodology, and technology that describes natural language meaning, automatic transposition of a text into formatted text-meaning representation (TMR). Also, researchers apply further manipulation with TMRs for reasoning and inferencing, both theoretically and in practice (Nirenburg and Raskin 2004; Raskin 2017). The OST ontology, a built model of reality, a world theory, provides information about the world (Raskin 2017). 'It is a highly structured system of concepts covering the processes and

---

14 In the current social media era, humor seems to be even more vital for communicators, since smaller and quicker-to-consume types of videos and texts are being offered to spectators (e.g. TikTok, Twitter).

15 See Victor Raskin (2017, 109–125) for more details in each theory.
objects, linked by properties in all of their pertinent complex relations, to the grain size determined by an application or considerations of computational complexity’ (ibid). This model uses hierarchies as the main framework of textual analysis. Namely, the significant ramifications are ‘event,’ ‘object,’ and ‘property,’ categories that also branch into several others (ibid).

- **Ontological Semantic Theory of Humor (OSTH):** this theory seeks to model and depict human understanding of the text as a whole, rather than attempting to avoid it by analyzing some non-semantic features of text, such as hidden statistics in machine learning (Raskin 2017). It introduced for comprehending meaning representation in texts, using complete TMRs for each sentence. However, due to the current limitations (and continuing progress) of cognitive computation i.e. AI, Deep Learning, there is still plenty of room for innovation in the OSTH (ibid).

- **General Theory of Verbal Humor (GTVH):** this theory expands Raskin’s (1985) SSTH, as well as complements Attardo’s taxonomy of textual connections (Attardo and Raskin 2017). In it, jokes (or any other form of text) contain relationships with different texts based on textual similarities and concerning other texts (Attardo 2017). The GTVH, unlike the purely semantic SSTH, included phonological, morphological, and additional information, i.e. text-level, sociological information, cognitive information, and more. Moreover, according to Attardo (2017), SSTH missed two crucial points: (1) the differentiation between puns and non-puns and (2) the identification of similar jokes or scripts. Thus, it was named ‘General’ theory for generalizing Raskin’s SSTH aspects (ibid).

Essentially, the study of humor is a broad field, branching across various other areas of expertise. There are definitions, types of humor, and theories attempting to describe its function and its effect on people. For this study, only the major topics were covered. The previously described approaches are essential for this dissertation since they have been historically used as frameworks by translators of humorous texts (Chiaro 2017). Thus, they might also be used to build NMT systems (Taylor 2017).
2.3.1 Memes and Puns

In this work, we will be focusing on translating memes and puns. The first will be studied for their wide-range dissemination on the Internet (Zannettou et al. 2018), highlighting at least a significant presence in people’s entertainment choices. The second was chosen due to its academic relevance for Humor Studies (Aarons 2017), being a relevant source of analysis and application for the SSTH and the GTVH (ibid).

Firstly, a pun can be defined as a joke in which one sound sequence (e.g. a word) has two meanings. The similarity in sound establishes a relationship between the two meanings, resulting in humor (Aarons 2017; Hempelmann 2014; Hempelmann and Miller 2017). Thus, the phonological feature of puns seems to be a relevant asset to its humor. Nonetheless, ambiguity and incongruity have to follow to fulfill its success as a humorous piece (Aarons 2017; Hempelmann 2003).

As defined in the Oxford English Dictionary (OED), ambiguity is ‘the fact or quality of having different possible meanings; capacity for being interpreted in more than one way; (also) lack of specificity or exactness.’ It is a common phenomenon in human language, being observed whenever the context is removed from an utterance and apparently being only easily identifiable by humans (Aarons 2017; Boeckx 2010; Attardo 1994). However, ambiguity alone is not sufficient to build the pun’s humorous aspect since it can only mean a grammatical error is being corrected by the receiver of the message (Aarons 2017, 81). Incongruity, defined in the OED as a ‘disagreement in character or qualities; want of accordance or harmony; discrepancy, inconsistency,’ also plays a significant part in the humor generated by the pun. Furthermore, the union of the phonological, the ambiguous, and the incongruent elements alone do not compose humor; the agreement between speakers to engage in a humorous dialogue also accounts for that.17

On the other hand, memes can be preliminarily defined as any piece of information that can be duplicated through imitation with a somewhat successful rate of propagation

---

16 Available at: https://oed.com.
17 This dialogue was defined by Raskin (1985) as non-bona fide (NBF) communication. It is commonly initiated with phrases such as ‘let me tell you a joke,’ or ‘have you heard this funny story about’ (Aarons 2017, 81).
Expanding on this definition, Shifman (2014, 39) sets two parameters that complement it to establish her own concept of internet memes: firstly, scholars can analyze memes as if they contain various memetic dimensions—that is, several characteristics that one might imitate. Secondly, memes can be framed as groups of content units with shared characteristics rather than single, dispersed content propagating successfully (ibid). This analysis results in the following definition: internet memes are a collection of digital items with similar content, form, or stance produced with the awareness of one another and then distributed, replicated, or altered by a large number of people over the Internet (ibid).

This dissertation aims to identify if humor is perceived after MT systems translate puns (textual-only pieces) and memes (in this case, composed by images and texts, a genre of internet memes described by Shifman (2014, 113–118) as Rage Comics). Bearing in mind the theory of humor described in Section 2.3, alongside the definitions of puns and memes, we can identify how humor in memes and puns works linguistically and how they are observed in current NMT systems.

2.4. Machine Translation and Humor Studies

Humor Studies have somewhat recent progress in the computational field in general – arising in the 1990s (Taylor 2017, 457). However, when it comes to MT, these studies are even more contemporary. Initially, scholars would focus on the study of puns, attempting to make computers create or detect them with the aid of templates provided by the SSTH and the GTVH (Raskin and Attardo 1994; Taylor 2017):

As Raskin & Attardo (1994) argued, it is possible to create a template for a well-defined set of jokes, such as light bulb jokes, and create a table with columns used for substitution of some of the fields. A computer would then assemble information that is in the table with the template,
and the result would be a joke produced by a computer (extremely low on the autonomous score of the manual/autonomous dimension).

Nowadays, the focus remains quite similar. Researchers are teaching AI to create humor and analyze if a text is humorous or not (ibid). As mentioned in Section 2.4, this field is divided between a theoretical point-of-view, i.e. the essence of humor, and a more practical, direct point-of-view, i.e. the relationship between people and humor (Ritchie 2015; Taylor 2017).

To justify the ‘usefulness’ of Humor Studies to get research funding, scholars argue that humor would enhance the human-computer relationship by bringing naturality to a bot/Al speech or textual pattern (Binsted et al. 2006; Taylor 2017). There is also the potential to evolve our knowledge on the human brain; thus, being an addition to the neuroscience/cognitive studies field (Taylor 2017, 458):

From its inception, several researchers argued that computational framework is a great mechanism for testing the understanding of human humor phenomenon—a computational humor model that accurately predicts a behavior of one human with known characteristics, but not another, will show that we still don’t know about humor perception / understanding / appreciation (Raskin 1996). In a similar vein of reasoning, arguments have been made that gaining understanding of humor will make a step closer to understanding how the human brain works (Ritchie 2001; Binsted et al. 2006).

Humor generators have already been created. According to Taylor (2017, 458), very similar to an MT system, those generators should be able to:

- Identify words.
- Represent knowledge about the world and be able to access it.
- Be able to reason about this knowledge.

---


- Understand grammar and apply it appropriately.

- Identify what can be funny or not.

Nonetheless, the majority of current systems are still mainly based on pre-set templates for a computer to identify or create humor, which is still not an ideal performance, since the pieces are made/determined randomly, without actual processing or learning by the machine (ibid, 460–462). Significant landmarks in this field have been found chiefly in the analysis of one-liners, puns, and cartoons (ibid).

Moreover, with the advancements in AI, Deep Learning, and statistical models, this field is rapidly changing. For instance, Kao et al. (2013) built a probabilistic system of sentence comprehension to formalize incongruity and test its relationship to humor in puns. They do this by using two metrics: ambiguity and distinctiveness (Taylor 2017). The other example regarding creative translation (in this case, poems) is Stein et al. (2014) acrostic generator. Even though these are comparatively impressive developments, there is still a gap regarding Machine Translation since the examples provided generate text in only one language. All of this effort with developing and identifying creative uses of language could, thus, be also reverted into MT, humor, and its evaluation (Miller 2019).

There is a clear gap in the research in this field (ibid) as there are not many methodologies or experiments developed to study humor and MT, even less when we narrow the scope down to memes and puns (wordplay). For example, according to Miller (2019, 62), only Stede and Tidhar (1999) and Farwell and Helmrreich (2006) wrote relevant papers on wordplay but with no actual implementation. On the other hand, Low (2011) is one step towards the development of a solid study of how human translators can translate puns with SSTH or GTVH’s scripts model; applying this to brand-new NLP and MT technologies could perhaps accelerate the discoveries in the field (Miller 2019, 62).

Finally, this dissertation aims to identify whether online MT systems convey a similar or different humor present by memes and puns. Hence, it used the state-of-the-art human evaluation metrics (error annotation based on the MQM framework), carried out by the author and an invited professional translator, and an end-user evaluation (based on a survey with a ranking of systems and a questionnaire) to assess humor awareness in
pieces (i.e. memes and puns) translated by Google Translate, Bing Microsoft Translator, and DeepL Translate, three significant NMT systems.

This work attempts to fill the gap of research of Machine Translation Evaluation specifically related to Humor Studies comprising memes and puns for the Brazilian-Portuguese language.
Chapter 3 – Methodology

This chapter focuses on the methodology, justification, and materials used to consolidate this work. In sum, four research questions, stemming from a broader query, conducted the research. As a result, two surveys and an analysis spreadsheet were used to assess Machine Translation quality in translating memes and puns from American English (EN-US) into Brazilian Portuguese (PT-BR). Both surveys started with the Trinity College Dublin’s Consent Form, a mandatory step to proceed in the task.

This dissertation’s experiments attempt to answer the overarching research question (RQ) by analyzing whether humor is perceived after translating nine memes and six puns from EN-US into PT-BR by three MT systems, i.e. Google Translate, Microsoft Bing Translator, and DeepL Translate. This primary research question branched into two specific ones:

RQ1) Which MT system performs better in the proposed task (i.e., translating humor in this language pair)?

RQ2) How is the user experience perceived with the different translations provided?

To answer RQ1, we used the participants’ responses to the survey. A more in-depth analysis was carried out by a professional translator and the researcher to evaluate the MT translations, according to the Multidimensional Quality Metrics (MQM). RQ2 was analyzed with the overall experience of the participants during the survey.

During the research period, two additional questions which influenced our choice of memes and puns were added. Those are sub-research questions that mainly relate to the understanding of humor and meaning in the translated text itself by users:

RQ3) Does the presence of the image, in the case of the memes, affect user perception/understanding and, thus, their evaluation of the MT output?

RQ4) Does the presence of the ST, for memes and puns, affect user perception/understanding and, thus, their evaluation of the MT output?

22 For the spreadsheet, see Appendix 1.
RQ3 was investigated with specific questions in the survey. They asked: ‘Did the image influence your perception of the translation in the memes?’ which was formulated as a ‘yes’ or ‘no’ type of question, and a more descriptive question, ‘How?’, where participants could write their impressions about the first question. Furthermore, they were asked to rank which meme’s image influenced them more and why, and if the image influenced their proposed translation and their evaluation of the translations. The influence of the image was studied because of the multimodal nature of memes. We wanted to investigate how much the image influences humor perception and if it would alter the translations’ comprehension for better or for worse.

Additionally, RQ4 was queried with the following questions: ‘has the lack of the original text (source text) affected your perception of the meme?’ This ‘yes’ or ‘no’ question was followed by the descriptive ‘How?’. The influence of the ST was studied because we wanted to see how the translations would stand by themselves, thus assessing how well they conveyed humor, mainly when complex wordplay or context were applied. Half of the participants had access to the ST, and the other half did not. Both groups offered us a comparison on humor perception. Therefore, to investigate the sub-research question RQ4, two separate surveys were created: one contained the ST, alongside its translations, and appropriate questions to investigate its influence on perception; on the other hand, only the MT outputs were presented in the second survey.

Hence, this work adapted the methodology initially proposed by Castilho (2016), who used Suojanen, Koskinen, and Tuominen’s (2015) User-Centered Translation (UCT) approach. Consequently, this task was carried out by firstly having thirty Brazilian volunteers (with various proficiency levels in English),23 whose data was anonymized, fill out two surveys on Microsoft Forms containing an analysis activity, alongside a ranking and post-task questionnaire. All participants were later randomly divided equally between the two surveys.

To avoid bias when filling the surveys, the volunteers were not fully informed about the translations’ content, nor that they were translated by MT systems. Moreover, two answers from participants who first tested the surveys were maintained, while the third one, for having been more informed about the research than the others, was excluded from the results.
The memes selected were retrieved from one of the biggest memes-gathering websites, Imgur.com. On the other hand, the puns were gathered from the article written by Liles (2020) on Parade.com and from Evans’s World’s Funniest Jokes (2015). To avoid bias in the survey, the pieces were chosen for being neither too well-known nor too unknown by the audience. This was done by analyzing if they had very culturally specific jokes or if the images were influencing the perception of the joke excessively. Moreover, to test how MT systems performed, two sets of puns were chosen: the ones retrieved from the website are simpler and shorter. In contrast, the ones selected from the book are complex jokes, which are heavily based on wordplay and the English language phonology.

3.1. Justification

Humor has been paramount in the entertainment industry for decades now, especially when we analyze past TV Series, e.g. ‘The Sopranos,’ ‘Friends,’ or ‘The Big Bang Theory’ (Costabile 2021). Nonetheless, with the internet and social media, user-generated content, i.e. memes, has been increasingly produced (Zannettou et al. 2018). This production might signal a significant source of entertainment for nowadays, especially considering the connection that memes have with historical events, as we can see in Dynel’s (2021) article about users creating covid-19/face coverings memes. Hence, studying humor - an essential source of comedy relief for readers - is an important research topic in this pandemic context. Also, since the MT systems analyzed here are freely available, users might use them to consume memes in other languages.

On the other hand, puns are a very traditional type of wordplay, and even more ancient kind of humor, being studied by the founders, Raskin and Attardo, of the current linguistic theories about the field (e.g., SSHT, GVTH). Bearing in mind that this pandemic moment makes people stay more at home than usual, memes and puns were selected for this work for being potentially significant in humor consumption by internet and social media users.

24 Available at: https://imgur.com/.
25 Available at: https://parade.com/1024249/marynliles/}


Another reason for choosing these formats of content is their potential to be resourceful in NMT development. As raised in Chapter 2, there is a gap in the research of MT performance and evaluation of humor, even more for memes. Thus, this work might be framed as a head start to fill this gap.

Finally, the choice for these formats was intended to investigate if the multimodal aspect of memes, i.e. their images, influenced users’ perception of the translations and their humorous aspects. Therefore, it was feasible to select puns that follow a similar comic construction – usually, they are relatively short, being rapidly consumed by readers.

3.2. Materials

This dissertation used the STs, which were retrieved from the websites and books referred to in the previous section, the MT systems (i.e. Google Translate, Bing Microsoft Translator, and DeepL Translate), and the answers provided by the volunteers in the two Microsoft Forms Surveys (Appendix 2 and Appendix 3) and by the invited professional translator in the Analysis Spreadsheet (Appendix 1).

3.2.1. Memes and Puns

As previously mentioned, we collected nine memes and six puns. We chose to repeat the same meme template three times to study how each caption would be affected or perceived according to the main image. Thus, we have three ‘Drake’ memes with three different STs, three ‘Distracted Boyfriend’ memes with three distinct STs, and three ‘Confession Bear’ memes with three different STs – thus, arriving at a total of nine memes. The six puns, on the other hand, are all different from each other. The only difference is that the three first puns were chosen to challenge the MT systems, being more contextually dependent puns, with a more complex wordplay; the last three puns, therefore, are shorter and have a more direct wordplay.

Table 1 shows the memes’ captions with their STs and the three MT systems outputs (Google Translate, Bing Translator, and DeepL Translate). Figure 1, figure 2, and figure 3 illustrate how the participants saw each set of memes, i.e. ‘Drake,’ ‘Distracted Boyfriend,’
and ‘Confession Bear’ in Survey 1 (with ST). Table 2 shows the six puns with their STs and the three MT outputs. For the complete gloss translation of all the texts, see Appendix 4.

<table>
<thead>
<tr>
<th>Selection of Memes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source Text</strong></td>
</tr>
</tbody>
</table>
| *1 – Drake – Alarms*
Schedule appointments on a calendar / Set a billion phone alarms | Agende compromissos em um calendário / Defina um bilhão de alarmes de telefone | Agende compromissos em um calendário / Defina um bilhão de alarmes telefónicos | Agendamento de compromissos em um calendário / Estabelecer um bilhão de alarmes telefónicos |
| *2 – Drake – Kids’ food*
My kids reaction to food at the store / My kids reaction to the same food at home | A reação dos meus filhos à comida na loja / Reação dos meus filhos à mesma comida em casa" | Meus filhos reações à comida na loja / Meus filhos reações à mesma comida em casa | Os meus filhos reagem à comida na loja / Os meus filhos reagem à mesma comida em casa |
| *3 – Drake – Shoelaces*
Buy new shoelaces / Take shoelaces from shoes I don’t wear anymore | Compre cadarços novos / Pegue cadarços de sapatos que não uso mais | Compre cadarços novos / Pegue cadarços de sapatos que eu não uso mais | Comprar novos atacadores / Tire os atacadores dos sapatos que eu não uso mais |
| *4 – Distracted Boyfriend – People Walking*
Wide open sidewalk / People walking / The street | Calçada aberta / Pessoas caminhando / A rua | Calçada aberta / Pessoas andando / A rua | Calçada bem aberta / Pessoas a andar / A rua |
| *5 – Distracted Boyfriend – Back to Work*
Literally anything else/ Me / Work / First day back at work... | Literalmente qualquer outra coisa / Mim / Trabalhos / Primeiro dia de volta ao trabalho... | Literalmente qualquer outra coisa / Me / Trabalho / Primeiro dia de volta ao trabalho... | Literalmente qualquer outra coisa / Eu / Trabalho / Primeiro dia de volta ao trabalho... |
| *6 – Distracted Boyfriend – Microwave*
Every other button on the microwave / Me / +30 second button | Todos os outros botões do microondas / Eu / +30 botão segundos | Todos os outros botões no microondas / Me / +30 segundos botão | Cada outro botão no microondas / Eu / +30 segundo botão |
| *7 – Confession Bear – Drinking*
**Table 1 – Memes selected for this work alongside their respective translations from each MT system (Google Translate, Bing Translator, and DeepL Translate).**

<table>
<thead>
<tr>
<th>Selection of Puns</th>
<th>Source Text</th>
<th>Google Translate</th>
<th>Bing Translator</th>
<th>DeepL Translate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 – Granddad Frog Joke</strong></td>
<td>A little girl asks her granddad, &quot;Would you make a frog noise for me?&quot;</td>
<td>Uma garotinha pergunta ao avô: &quot;Você faria barulho de sapo por mim?&quot;</td>
<td>O avô, confuso, pergunta: &quot;Por quê?&quot;</td>
<td>A menina responde: &quot;Pai diz que, quando você coxar, vamos todos para Disneyland&quot;.</td>
</tr>
<tr>
<td></td>
<td>The granddad, confused asks, &quot;Why?&quot;</td>
<td>Uma garotinha pergunta ao avô: &quot;Você faria um barulho de sapo por mim?&quot;</td>
<td>O avô, confuso pergunta: &quot;Por quê?&quot;</td>
<td>A garotinha responde: &quot;O pai diz que quando você morrer, vamos todos para a Disneyland&quot;.</td>
</tr>
<tr>
<td></td>
<td>The little girl replies, &quot;Dad says when you croak, we are all going to Disneyland&quot;.</td>
<td>Uma garotinha pergunta ao avô: &quot;Você faria um barulho de sapo por mim?&quot;</td>
<td>O avô, confuso, pergunta: &quot;Por quê?&quot;.</td>
<td>A menina responde: &quot;O pai diz que quando você coxar, vamos todos à Disneyland&quot;.</td>
</tr>
<tr>
<td><strong>8 – Confession Bear – People Skills</strong></td>
<td>Hired for my &quot;people skills&quot; / Starting to hate &quot;people&quot;</td>
<td>Contratado por minhas &quot;habilidades pessoais&quot; / Começando a odiar &quot;pessoas&quot;</td>
<td>Contratado para minhas &quot;habilidades pessoais&quot; / Começando a odiar &quot;pessoas&quot;</td>
<td>Contratado para minhas &quot;habilidades pessoais&quot; / Começando a odiar &quot;pessoas&quot;</td>
</tr>
<tr>
<td><strong>9 – Confession Bear – Pandemics</strong></td>
<td>People have been complaining since the pandemic started how much they miss being able to touch people / I personally have actually been really enjoying how less touchy-feely everyone is right now.</td>
<td>Desde o início da pandemia, as pessoas reclamam do quanto sentem falta de tocar nas pessoas. / Eu, pessoalmente, tenho realmente gostado de como todos estão menos melindrosos agora.</td>
<td>As pessoas têm reclamado desde que a pandemia começou o quanto sentem falta de poder tocar as pessoas/ Eu pessoalmente tenho realmente gostado de como todos são menos sensíveis agora.</td>
<td>As pessoas têm reclamado desde o início da pandemia o quanto sentem falta de poder tocar as pessoas./ Eu, pessoalmente, tenho realmente gostado muito de como todos estão menos sensíveis ao toque agora.</td>
</tr>
<tr>
<td>2 – Four-chin Teller</td>
<td>O que você chama de médium muito gordo?</td>
<td>Como se chama um vidente gordo?</td>
<td>O que você chama de um médium realmente gordo?</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>What do you call a really fat psychic?</td>
<td>Um contador de quatro queixos</td>
<td>Um caixa de quatro queixos</td>
<td>Um caixa de quatro moedas</td>
<td></td>
</tr>
<tr>
<td>A four-chin teller</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 – Too Many Shifts</th>
<th>Por que o funcionário da unidade de fabricação de teclados foi demitido?</th>
<th>Por que o trabalhador da unidade de fabricação de teclados foi demitido?</th>
<th>Por que o trabalhador da unidade de fabricação de teclados foi demitido?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why did the worker at the keyboard manufacturing unit get fired?</td>
<td>Por colocar muitos turnos</td>
<td>Por colocar em muitos turnos</td>
<td>Por fazer demasiados turnos</td>
</tr>
<tr>
<td>For putting in too many shifts</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 – The Guy Who Invented Zero</th>
<th>Para o cara que inventou o zero, obrigado por nada.</th>
<th>Ao cara que inventou zero, obrigado por nada.</th>
<th>Para o cara que inventou o zero, obrigado por nada.</th>
</tr>
</thead>
<tbody>
<tr>
<td>To the guy who invented zero, thanks for nothing.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5 – Restaurant on the Moon</th>
<th>Você ouviu sobre o restaurante na lua? Ouvi dizer que a comida era boa, mas não tinha atmosfera.</th>
<th>Você ouviu sobre o restaurante na lua? Ouvi dizer que a comida era boa, mas não tinha atmosfera.</th>
<th>Você ouviu falar sobre o restaurante na lua? Ouvi dizer que a comida era boa, mas não tinha atmosfera.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you hear about the restaurant on the moon? I heard the food was good, but it had no atmosphere.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6 – Irrelephant</th>
<th>Por que Dumbo estava triste? Ele se sentiu irrelephant.</th>
<th>Por que Dumbo estava triste? Ele se sentiu irrelephant.</th>
<th>Por que o Dumbo ficou triste? Ele se sentiu irrefletido.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why was Dumbo sad? He felt irrelephant.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 – Puns selected for this work alongside their respective translations from each MT system (Google Translate, Bing Translator, and DeepL Translate).
Figure 1 – An example of how the set of memes with the ‘Drake’ meme template was presented to the participants in the survey with ST.

Figure 2 – An example of how the set of memes with the ‘Distracted Boyfriend’ meme template was presented to the participants in the survey with ST.

Figure 3 – An example of how the set of memes with the ‘Confession Bear’ meme template was presented to the participants in the survey with ST.
3.2.2. Ranking and Questionnaire – Surveys

As previously mentioned, two different surveys were designed to address RQ4 (how having access to the original text influences humor perception by users). While Survey 1 contained the ST, Survey 2 did not include the ST to serve as a comparative analysis. Figure 4 and figure 5 show screenshots of how Survey 1 (with ST) was designed for memes and puns, while figure 6 and figure 7 present screenshots of Survey 2 (without ST).

Figure 4 – Screenshots of the memes section of the Survey 1 (with the ST). Note that the three translations are anonymized and randomized, presented and labeled as ‘1,’ ‘2,’ and ‘3.’
Puns (trocadilhos) - Frog noise

ST
A little girl asks her granddad, "Would you make a frog noise for me?"
The granddad, confused asks, "Why?"
The little girl replies, "Dad says when you croak we are all going to Disneyland".

Tradução 1
Uma garotinha pergunta ao avô: "Você faria um barulho de sapo por mim?"
O avô, confuso pergunta: "Por quê?"
A garotinha responde: "Papai diz que quando você morrer, todos nós vamos para a Disneylândia".

Tradução 2
Uma garotinha pergunta ao avô: "Você faria barulho de sapo para mim?"
O avô, confuso, pergunta: "Por quê?"
A menina responde: "Papai diz que, quando você coaxar, todos nós vamos para a Disneylândia".

Tradução 3
Uma menina pergunta ao seu avô: "Você faria um barulho de sapo para mim?"
O avô, confuso, pergunta: "Por quê?"
A menina responde: "O pai diz que quando você coaxar, vamos todos à Disneylândia".

Figure 5 – A screenshot of the puns section of the Survey 1 (with the ST). Note that the three translations are anonymized and randomized, presented and labeled as ‘1,’ ‘2,’ and ‘3.’
Meme - Drake - A Billion Phone Alarms

Caso não consiga enxergar, clique com o botão direito do mouse na imagem e acesse o menu "abrir imagem em nova guia", onde conseguirá ampliá-la.

Transcrição:

1 = Agende compromissos em um calendário / Defina um bilhão de alarmes telefônico

2 = Agende compromissos em um calendário / Defina um bilhão de alarmes de telefone

3 = Agendamento de compromissos em um calendário / Estabelecer um bilhão de alarmes telefônicos

Figure 6 – A screenshot of the memes and puns sections of Survey 2 (No ST). Only the three translations, anonymized and randomized, are presented and labeled as ‘1,’ ‘2,’ and ‘3.’
Before starting the survey, participants answered a pre-task questionnaire which aimed to explore the volunteers’ background: their preferred types of humor; whether they knew what memes and puns were; their proficiency level in English; their experience with translation (if any); if they used MT systems in their own translations; and if they created memes or puns themselves.

Both surveys had the ranking and the post-task questionnaire divided by each meme and pun. Thus, first, a humorous piece was presented alongside its three translations (see Figures 2 and 3), followed by a ranking, which enabled them to choose what the best and worst translations for them were. Participants were also asked whether they understood the humor in the piece and what caused humor (e.g. if it was caused by the translations’ errors or successes, or by the jokes themselves). Then, in Survey 1 (with ST), participants were also asked to translate the original meme/pun to identify further which
aspects of the MT they disagreed with. In Survey 2 (No ST), only the three translations were presented.

Moreover, we did not disclose with the participants which translation was from each MT system. Instead, they were anonymized and randomized, labeled as ‘Translation 1,’ ‘Translation 2,’ and ‘Translation 3.’ In the end, Survey 1 had 161 questions, taking an average of 45–60 minutes for participants to answer. Survey 2 totaled 113 questions, taking 35–50 minutes to answer (see Appendix 2 and Appendix 3).

Both surveys contained a questionnaire section (after the memes and puns sections ended) which analyzed the sub-research questions RQ3 (e.g. if, in the case of memes having the image would affect the participants’ perception of the humor) and RQ4 (whether having the ST would affect their assessment). It is important to note that the translator’s analysis and the surveys complemented each other in all four sub-research questions. However, the analysis directly aimed at answering RQ1 due to being more specific throughout the research. The surveys were designed to answer RQ2, RQ3, and RQ4 through particular questions.

Finally, after the participants had finished the survey, they were asked to answer a final and general post-task questionnaire to assess if they noticed that they were presented with translations and which elements made them realize it. They were also asked if the translations were helpful for them to understand the humor in a language other than their native language.

A limitation on the survey we observed later on was positioning the memes in the first section of the analysis. This decision might have biased the participants, who got used to the visual context to aid their assessment of the translations, especially in Survey 2 (No ST). When the second section, with the puns, began, their understanding seemed to have been harmed by the lack of visuals.

This UCT methodology will complement the following section’s analysis to answer our main research question – which MT system performs better for this language pair to translate memes and puns – besides all the four sub-research questions mentioned at the beginning of this chapter.
3.2.3. Participants

Thirty participants were recruited for this experiment who were all Brazilians, ranging from 18-34 years old, with various proficiency levels in English\textsuperscript{26} (see figure 8 and figure 9 for the results on the proficiency levels in Survey 1 and 2, respectively). They were contacted via social media (i.e. WhatsApp, Facebook, and Instagram) to participate in the study. Afterward, the link to fill out the surveys were sent either by email or by the social media’s messages function. Moreover, by summing up the results from the pre-task questionnaire from Survey 1 and Survey 2, it was found that they were mainly undergraduates on language or linguistics-related courses, and 63.3% of all participants had experience in translation.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure8.png}
\caption{Results from the pre-task questionnaire in Survey 1 (with ST) showing the proficiency levels of the fifteen participants.}
\end{figure}

\textsuperscript{26} The proficiency levels used in this dissertation were retrieved from the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): ‘The CEFR organises language proficiency in six levels, A1 to C2, which can be regrouped into three broad levels: Basic User, Independent User and Proficient User, and that can be further subdivided according to the needs of the local context.’ The participants were provided with the link for the description of each level. The website also provides a self-assessment grid. Available at: \url{https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions}. 

31
Figure 9 – Results from the pre-task questionnaire in Survey 2 (without ST) showing the proficiency levels of the fifteen participants.
3.2.4. Translators Analysis and MQM

The translators’ analysis was carried out in the Google Spreadsheet software to investigate RQ1 in more detail. Moreover, a file was created and shared with the professional translator.\textsuperscript{27} It contained all the fifteen pieces (i.e. memes and puns) with their respective translations. This time, we disclosed each MT system according to each translation in the spreadsheet.

The spreadsheet contained a tab with the ‘Guidelines’ which explained each MQM definition that would be used in the analysis (i.e. accuracy, fluency, register, and their ramifications), a column ‘How to assess’ explaining to the translator what he should look for/comment on, besides asking him to define which translation was the best and why, if the transfer of humor was possible from English to Brazilian Portuguese and why, and if he had additional comments on the understanding of humor by humans and by machines (see Appendix 1). Table 3 shows a sample of analysis of the evaluation sheet, while table 4 presents the guidelines that the invited translator followed to perform the assessment. Finally, the results of this analysis are discussed in Chapter 4, Section 3.1, where a summary of the translator’s review was presented and where I draw conclusions to the RQ1 with its support.

\textsuperscript{27} It is important to note that the translator also participated in the surveys. To reduce the potential bias in the analysis, the spreadsheet was provided to him a month later.
<table>
<thead>
<tr>
<th>SOURCE TEXT</th>
<th>MT Systems</th>
<th>Translations</th>
<th>Accuracy</th>
<th>Fluency</th>
<th>Register</th>
<th>Best? Why?</th>
<th>Possible to transfer the humor?</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEMES</td>
<td>Google</td>
<td>Agende compromissos em um calendário / Defina um bilhão de alarmes de telefone</td>
<td>Google's MT is accurate, barring the verbs, which should be infinitive, and not imperative.</td>
<td>The verbs should be on infinitive, and not imperative, in the translation. &quot;em um&quot; should be &quot;no&quot; due to shortness/brevity and register appropriateness.</td>
<td>The register appears to be more formal than necessary. &quot;Em um&quot; should become &quot;no&quot;. &quot;Definir&quot; appears to be formal UI/UX talk, it could be &quot;Coloque&quot; to be more informal.</td>
<td>I believe the first one is the closest one to be the best, however, no option reflects on an adequate, informal option.</td>
<td>The transfer is possible, but it's awkward.</td>
</tr>
<tr>
<td>Drake - Alarms</td>
<td>Bing</td>
<td>Agende compromissos em um calendário / Defina um bilhão de alarmes telefónicos</td>
<td>Bing's MT is accurate, barring the verbs, which should be infinitive, and not imperative.</td>
<td>The verbs should be on infinitive, and not imperative, in the translation. &quot;em um&quot; should be &quot;no&quot; due to shortness/brevity and register appropriateness.</td>
<td>The register appears to be more formal than necessary. &quot;Em um&quot; should become &quot;no&quot;. &quot;Definir&quot; appears to be formal UI/UX talk, it could be &quot;Coloque&quot; to be more informal.</td>
<td>The transfer is possible, but it's awkward.</td>
<td>The transfer is possible, but it's awkward.</td>
</tr>
<tr>
<td></td>
<td>DeepL</td>
<td>Agendamento de compromissos em um calendário / Estabelecer um bilhão de alarmes telefónicos</td>
<td>DeepL's MT is accurate, barring the verbs, which should be infinitive, and not imperative.</td>
<td>The verbs should be on infinitive, and not imperative, in the translation. &quot;em um&quot; should become &quot;no&quot;.</td>
<td>The register appears to be more formal than necessary. &quot;Em um&quot; should become &quot;no&quot;.</td>
<td>The transfer is possible, but it's awkward.</td>
<td>The transfer is possible, but it's awkward.</td>
</tr>
</tbody>
</table>
Table 3 – A sample of the analysis’ spreadsheet shared with the invited professional translator. Respectively, an analysis of one of the translations is presented above according to the MQM’s definitions of accuracy, fluency, and register.

<table>
<thead>
<tr>
<th>Definition (MQM)</th>
<th>How to assess</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>The target text does not accurately reflect the source text, allowing for any differences authorized by specifications. The extent to which the translation transfers the meaning of the source text translation unit into the target.</td>
</tr>
<tr>
<td></td>
<td>Look for/comment on mistranslations, omission, over-translation, under-translation and/or untranslated terms that affect the meaning of the humour in the source in the translation.</td>
</tr>
<tr>
<td><strong>Fluency</strong></td>
<td>Issues related to the form or content of a text, irrespective as to whether it is a translation or not. The flow and naturalness of the translation in the context of the target audience and its linguistic and sociocultural norms in the given context.</td>
</tr>
<tr>
<td></td>
<td>Look for/comment on word order, cohesion, coherence, inconsistency, spelling and/or parallelism on the translated sentence (disregard the source). The grammar errors that hinder the intelligibility of the translated sentence and the humour was lost.</td>
</tr>
<tr>
<td><strong>Register</strong></td>
<td>The text uses a level of formality higher or lower than required by the specifications or general language conventions.</td>
</tr>
<tr>
<td></td>
<td>Look for/comment on the degree of formality (e.g. formal or informal) and how that affected the translation of the humour.</td>
</tr>
<tr>
<td><strong>Best?</strong></td>
<td>Choose your preferred translation and comment on why.</td>
</tr>
<tr>
<td>Possible to transfer the Humour?</td>
<td>In your opinion, given the source sentence, could the humour be transferred from English into Portuguese? Why/Why not? Comment on why the translations failed/succeeded and, if failed, suggest a better translation and comment on why your translation is better.</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Other Comments</td>
<td>Comment anything else you think might be helpful for the understanding of the translation of the humour by humans and machines.</td>
</tr>
</tbody>
</table>

Table 4 - A sample of the analysis’ spreadsheet shared with the invited professional translator. The guidelines for the translator’s analysis are described in this tab of the spreadsheet. Also, the MQM’s definitions of accuracy, fluency, and register are also presented.
Chapter 4 – Finding and Analysis

This chapter focuses on the results of the MQM (Section 4.1.) with the invited translator’s assessment and the UCT (Section 4.2) research methods. Both findings provide complementary answers to all RQ1, RQ2, RQ3, and RQ4. Nonetheless, the translator’s analysis delivers a more extensive study of the translations, serving as the primary source of research regarding RQ1. The surveys, nevertheless, were tailored to assess RQ2, RQ3, and RQ4 more in-depth, with specific questions for each sub-research question. Still, as mentioned before, both methods complement each other and aided this work to conclude a final answer for the overarching RQ.

4.1. MT Quality via Translator’s Assessment

<table>
<thead>
<tr>
<th>Best MT system</th>
<th>Meme</th>
<th>Google</th>
<th>Bing</th>
<th>DeepL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drake – Set off alarms</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Drake – Kids’ food</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drake – Shoelaces</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Distracted Boyfriend – People Walking</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distracted Boyfriend – Back to Work</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Distracted Boyfriend – Microwave</td>
<td>✓*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confession Bear – Drinking</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confession Bear – People Skills</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confession Bear – Pandemic</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Pun</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granddad Frog Joke</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Four-chin Teller</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Too many shifts</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>The guy who invented zero</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
The translator’s analysis brought exciting insights. From the results in Table 5, we can conclude that **Google was chosen as the best performing MT system** according to the criteria selected (MQM’s accuracy, fluency, and register), for the memes and chosen puns, and for this language pair designated (EN into PT-BR), **partially answering RQ1**. DeepL performed slightly better than Bing. It was selected as the best MT system five times, while Bing was elected four times. Furthermore, it is essential to note that the translator pointed out the difficulty of analyzing the MT systems’ performances in the case of memes without the support or context of the images, which contributes to answering our RQ3 (see more in Section 4.2.1).

<table>
<thead>
<tr>
<th>Levels of Accuracy, Fluency, and Register</th>
<th>Google</th>
<th>Bing</th>
<th>DeepL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy memes</td>
<td>66.6%</td>
<td>33.3%</td>
<td>44.4%</td>
</tr>
<tr>
<td>Accuracy puns</td>
<td>83.3%</td>
<td>50%</td>
<td>66.6%</td>
</tr>
<tr>
<td>Fluency memes</td>
<td>33.3%</td>
<td>33.3%</td>
<td>44.4%</td>
</tr>
<tr>
<td>Fluency puns</td>
<td>83.3%</td>
<td>50%</td>
<td>66.6%</td>
</tr>
<tr>
<td>Register memes</td>
<td>44.4%</td>
<td>77.7%</td>
<td>33.3%</td>
</tr>
<tr>
<td>Register puns</td>
<td>83.3%</td>
<td>66.6%</td>
<td>50%</td>
</tr>
</tbody>
</table>

As seen in Table 6, puns had an overall better performance than memes, i.e. puns reached 83.3% of translation being accurate, fluent, and with the correct register), it is possible to conclude that the answer to our RQ3 is that images indeed influence the perception and evaluation of these translations, being an essential feature for contextual comprehension. It is important to note that results in Table 6 were calculated based on how many times the translator described the translation as ‘accurate,’ ‘fluent,’ or ‘with an adequate register’.

---

28 In both systems, there was a relevant number of ties with Google Translate. Thus, they are fairly tied according to the invited translator’s assessment.
divided by the total number of puns/memes.’ Results such as ‘somewhat accurate’ or ‘somewhat fluent’ were discarded from this calculation for simplicity.

Moreover, we can deepen the discussion around the answer to the RQ1, regarding the MT quality, by noticing that even though Google Translate, according to table 5, generally outperforms Bing and DeepL, table 6 shows us a more detailed insight. For instance, Google was considered the best only in accuracy, fluency in puns, and register in puns; Bing was the best in register for memes and tied with Google in fluency in memes, and DeepL was the best in fluency in memes. Thus, despite Google Translate being chosen as the overall best in table 5, table 6 highlights the strengths and flaws for each MT system, in these definitions and circumstances, showing that Google does not always perform better.

Furtherly, when we analyze the ‘Possible to transfer the humor?’ tab, we see that the translator evaluates almost all translations as possible to transfer humor (five translations out of the total fifteen were assessed in that way), meaning that a human translator could create a suitable MT output in Brazilian Portuguese; thus, the MT system was not hindered by the impossibility of humor transference and should be able to perform the translation task.

Nonetheless, there were few exceptional examples where the translation was surprisingly good. The MT systems were very accurate, fluent, and they chose the adequate register, leading the translator even to suspect the system was trained with that example by users before. The translations presented from Google Translate, Bing Translator, and DeepL Translate have a few issues on register (see the examples on table 3 and the first pun ‘Granddad Frog Joke’ on table 2), bearing a very formal tone. However, they were classified as mostly adequate, i.e. Google presents 83% adequate register for puns, and Bing shows a 77% good register on memes. See the gloss translation example above (Appendix 4 for the complete gloss translation table):

---

29 DeepL’s translation of the ‘Distracted Boyfriend – Back to Work’ meme, Google’s translations of the ‘Irrelephant’ pun, and the ‘Confession Bear – Drinking’ meme were highlighted for their accurate translation regarding MQM’s parameters. See Appendix 1.
**ST:** Schedule appointments on a calendar / Set a billion phone alarms

**Google:** Agende compromissos em um calendário /
Defina um bilhão de alarmes de telefone [Schedule appointments on a calendar / Define a billion of alarms of telephone]

**Bing:** Agende compromissos em um calendário /
Defina um bilhão de alarmes telefônicos [Schedule appointments on a calendar / Define a billion telephonic alarms]

**DeepL:** Agendamento de compromissos em um calendário /
Estabelecer um bilhão de alarmes telefônicos

(Scheduling of appointments on a calendar/Establish a billion of telephonic alarms)

Besides that, **untranslated** and **mistranslated** terms (especially colloquial ones) presented a challenge in perception too. Moreover, the more complex/phonological puns suffered even more, having wordplays erased completely and not replaced in the translations.

Additionally, DeepL presented errors regarding the variety of Portuguese selected in the software. Even though Brazilian Portuguese was chosen, most accuracy, fluency, and register issues identified by the translator were related to adequacy of terms to other varieties of Portuguese (i.e. European Portuguese) rather than the Brazilian one. This result indicates that DeepL engines are trained with a more significant number of European Portuguese data.

Finally, the translator brings complementary comments to the discussion that helps us gain insights regarding the influence of the ST on humor perception (RQ4). For instance, he mentions that the term ‘tradukibe,’ in Brazilian Portuguese, is used when a meme is plainly viewed as translated by an MT system, which is perceived as negative and unfunny. Thus, this comment serves as a preliminary answer for RQ4, showing that the user’s knowledge or awareness of the original caption and the MT output, being a word-for-word, or a mistranslated text, affects the transfer of humor (see Appendix 1 for the translator’s full comment).
In summary, in our RQ1, regarding MT quality, we note that, overall, **Google has performed better in the translation of memes and puns**, from EN-US into PT-BR, followed by DeepL and Bing. These results were tied according to the translator’s assessment in the tab ‘Best? Why?’ (see table 5 and Appendix 1). However, from the percentages described in table 6, DeepL outperforms Bing in accuracy and fluency. Nonetheless, Bing exceeds in register for memes (77%). DeepL is the best in fluency for memes, both outperforming Google in those metrics (see Appendix 1 for the detailed analysis).

4.2. End-User Reception

This section will display the results from Survey 1 (with the ST) and Survey 2 (No ST). These findings will help us answer research questions (RQ2, RQ3, and RQ4) and complement RQ1.

4.2.1. Survey 1 – ST

Survey 1, which displayed the original caption (the ST), was divided among fifteen random participants of the experiment. It contains 161 questions to investigate nine memes and six puns translated by MT systems. Overall, 36 pieces are analyzed in the memes section (the ST and the three MT outputs), and 24 samples are studied in the puns section (the ST and the three MT outputs).³⁰

Most of the respondents (nine out of fifteen) of this survey had experience with translation, and all of them consumed humor on Instagram and Facebook. As a result, all volunteers had already seen memes, but only eleven of them had the habit of reading puns. Nonetheless, most of them classified most of the original memes and puns in English as humorous. On the other hand, most responses ranked the translations as not funny (see table 7).

---

³⁰ This is due to the fact that each meme/pun counts as one piece. The further three translations results in four pieces to be analyzed for each humorous text/image.


| Table 7 – Results from the humor perception questions. In order to see if the translation’s evaluation is not biased by the lack of humor perception of the original, we asked the volunteers to assess humor in the original and in the translation.

It is important to note that most of the participants had language/translation courses as background (see Section 3.2.3), which might have tailored the answers to the humor in the translations since the errors or mistranslations were not found to be funny for most participants. The hypothesis that the MT output errors could have brought humor to the pieces was discarded from the translator’s analysis and the survey results. This insight can partially answer our RQ4 since, as described in Section 4.2.2, the original caption’s presence

---

31 The volunteers were asked to answer the following question: ‘For what reason you found this translation humorous or not?’ They had to choose between some options – (1) because of the translation’s errors, (2) because of the joke itself, (3) because of the translation’s correctness, and (4) other. The errors were rarely pointed out as one of the main reasons.
alongside the users' expertise background might influence their humor perception of the pieces provided.

Furthermore, the register seemed to be the central issue identified by the volunteers, in contrast with the translator's assessment. For them, the high level of formality diminished the potential humor contained in the translations.

Regarding RQ3 (influence of the image on humor comprehension), most participants agreed that the image influenced their perception of humor in the translations presented (fourteen answered 'yes'), as shown in figure 4. According to them, context and the multimodal nature of memes cannot be disconnected from each other; otherwise, the texts by themselves would not make sense. Therefore, this answers positively our RQ3, related to the image's influence on their perception of humor in the translations.

This perception was complemented by additional questions such as ‘Do you think the images affected your evaluation?’; ‘Would you change your evaluation if images were not included?’, and ‘Do you think the images served as complementary humorous features?’. All of them were answered affirmatively by most participants (approximately 73.3% and 80%, respectively).

However, regarding the puns, we notice a more divided response to RQ3. For instance, 73.3% of the participants answered ‘no’ to the question ‘Do you think these puns would be better with the aid of an image?’. This was followed by a ‘Why?’ question, where insights varied from puns being inherently funny and purely textual, without the need of an image, to having visual aids as an additional help for understanding. Nonetheless, even
the positive comments in favor of the image in puns reaffirmed the genre’s independence from them, being just an interesting addition to illustrate, but not to understand. The next question was ‘Would images alter your humor perception of the translations?’ Again, an expressive 60% of the participants answered ‘yes.’ They again argued that images would not be necessary and that images would probably add a humorous aspect to puns they found unfunny or did not understand completely. We can then conclude that the preliminary answer to our RQ3 in this survey is that images are essential for memes’ understanding, but not for puns, being solely an additional resource to illustrate humor.

Results in figure 14 show us a preference for the simpler, shorter puns, which also had a more direct wordplay. This preference is an addition to the analysis of our RQ2 regarding user perception of MT. It shows that the systems tested in this work performed better in the puns mentioned above than in puns with more extensive sentences and bigger contextualization before the wordplay.

Furthermore, figure 15 shows that an expressive number of participants (33.3%) did not share their humor preferences with the memes and puns selected, leading to a possible influence in their answers regarding the translations. This result means that if participants do not find the original caption humorous, they will not find the translation humorous either. Moreover, figure 13 showcases a 50% division on the participants’ evaluation of the translations presented and a 53.3% approval on the ‘Do you think the translations helped you understand the humor?’ question. However, the memes and puns were considered challenging to translate (80% in figure 11), and MT systems were not seemed as beneficial for them as translators (73.3% of negative responses in figure 12), but more beneficial for the general public or as a base for translators (50% in figure 13). Additionally, some participants commented that MT systems could be helpful for non-native speakers to have a basic idea of what was being said (figure 16). Nonetheless, they were unsure if humor would be transferred successfully.
Figure 11 – Results for the question ‘Have you found difficulties when translating memes and puns? Yes or no’ displayed at the post-task questionnaire.

Figure 12 – Results for the question ‘Would you translate memes and puns with MT? Yes or no’ displayed at the post-task questionnaire.

Figure 13 – Results for the question ‘Do you think MT aids the translation process with memes and puns? Yes or no’ displayed at the post-task questionnaire. The answers to the follow-up ‘Why?’ question varied from MT being useful for non-native speakers, and as a base translation for professional translators.
Figure 14 – Results for the question ‘What were the best puns?’ displayed at the end of the puns section. In the option ‘outra’ [other], participants elucidated which pun they were talking about. The answers to the follow-up ‘Why?’ question varied from the pun’s own creative wordplay to the challenge the translator would face.

Figure 15 – Results from the post-task questionnaire, the section where the final and general comments about the whole activity were gathered from. Questions 147, 148, can be translated as the following: ‘The pieces chosen (memes and puns) reflect the type of humor you consume? Yes or no,’ ‘Did you like the translations presented? Yes or no.’
Figure 16 – Results from the post-task questionnaire, the section where the final and general comments about the whole activity were gathered from. Question 150 can be translated as the following: ‘Do you think the translations helped you understand the humor? Yes or no.’

Regarding the participants’ preferences for a translation, table 8 below shows that the translations by Google Translate were chosen the best translations 5 out of 9 times for the memes and 3 out of 6 times for the puns. DeepL’s outputs were chosen the best 2 out of 9 times for the memes and 3 out 6 times for the puns. Finally, translations made by Bing were placed in the first place 2 out of 9 times for the memes, and 1 out of 6 times for the puns. As a result, in Survey 1, Google Translate performed better overall, according to the participants, followed by DeepL Translate and Bing Translator.

<table>
<thead>
<tr>
<th>Memes</th>
<th>Google</th>
<th>Bing</th>
<th>DeepL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drake – Alarms</td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
</tr>
<tr>
<td>Drake – Kids’ food</td>
<td>1st</td>
<td>3rd</td>
<td>2nd</td>
</tr>
<tr>
<td>Drake – Shoelaces</td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
</tr>
<tr>
<td>Distracted Boyfriend</td>
<td>2nd</td>
<td>1st</td>
<td>3rd</td>
</tr>
<tr>
<td>– People Walking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distracted Boyfriend</td>
<td>2nd</td>
<td>3rd</td>
<td>1st</td>
</tr>
<tr>
<td>– Back to Work</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distracted Boyfriend</td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
</tr>
<tr>
<td>– Microwave</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confession Bear –</td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
</tr>
<tr>
<td>Drinking</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8 – Results from the ranking done by the participants for each set of translations on Survey 1 (with ST). The items marked with (*) were tied because they are identical translations.

| Confession Bear – People Skills | 3rd | 2nd | 1st |
| Confession Bear – Pandemics | 2nd | 1st | 3rd |
| Puns | Google | Bing | DeepL |
| Granddad Frog Joke | 2nd | 1st | 3rd |
| Four-chin Teller | 1st | 2nd | 3rd |
| Too Many Shifts | 2nd | 3rd | 1st |
| The Guy Who Invented Zero | 1st* | 2nd | 1st* |
| Restaurant on the Moon | 2nd* | 2nd* | 1st |
| Irrelephant | 1st | 2nd | 3rd |

In summary, this section concludes that **RQ2 can be preliminary answered relatively negatively**, as the participants of Survey 1 did not have a general good evaluation/perception of humor in the translations provided. Nevertheless, they were not informed of which system was translating it. Nor did they know it was translated by an MT system (this was only disclosed in the post-task questionnaire).

On the other hand, **RQ3 can be answered positively**: the presence of images, in the case of memes, indeed influences users’ perception/evaluation. From the survey and the translator analysis presented in Section 4.1, participants stated that the image is essential for humor comprehension due to its contextualization of the joke.

Finally, **RQ1 can also be complemented by this section because of the ranking provided by the volunteers. Google Translate was again placed as the best MT system** for translating memes and puns from EN-US into PT-BR, as seen in table 8. DeepL Translate followed in the second position, and Bing Translator was positioned as the third. Comments from participants criticizing register32 and the translations’ literalness are helpful for us to

---

32 DeepL translations were again pointed out as resembling other varieties of Portuguese different from Brazilian Portuguese.
reaffirm that there is still a gap in NMT systems to identify and translate humor in different styles and registers.

4.2.2. Survey 2 – No ST

Survey 2, which did not display the ST, was divided among fifteen random participants of the experiment. It contained 113 questions to investigate nine memes and six puns translated by MT systems. Thus, there are 27 pieces to be analyzed in the memes section (three MT outputs per meme) and 18 pieces to be studied in the puns section (three MT outputs per pun).

In this survey’s pre-task questionnaire, most of the participants (66.6%) had experience with translation, and all of them consumed humor online, as mentioned in Section 3.2.3. In addition, all volunteers had already seen memes, and 80% of them had contact with puns before. Therefore, the majority of them understood the humor of the memes and puns presented.

Contrasting with the results from Survey 1, the translations were classified as humorous and helpful to understand the humor in most cases (66.6%). This data is an indication that, regarding the influence of the ST (RQ4), the outputs created by MT systems seem to be a valuable aid to humor comprehension for users. In fact, participants highlighted this support in their explanatory answers, stating that if they were provided with the original captions for the different translations, it would probably be easier for them to recognize the original’s humor.

Again, participants highlighted the translations’ literalness, level of formality (i.e. register), and Portuguese variety as translation problems. Moreover, they could understand the jokes without the ST. Evidence of this can be found when they respond, ‘for what reason is there humor in the text?’ Most answered, ‘because of the joke itself, and not ‘because of the translations’ correctness,’ ‘because of the translations’ mistakes,’ or ‘other’ (see figure 17).
An interesting observation is that humor in the most complex puns, because of their crucial dependency on context (the first three presented: ‘Granddad Frog Joke,’ ‘Four-chin Teller,’ and ‘Too Many Shifts’) was harder to be understood in this Survey 2 (with no ST) than in the previous one. Moreover, participants highlighted that the presence of images would probably be helpful for them to have more context and thoroughly understand the joke’s humor. This fact can be significant evidence of the importance of the ST (RQ3) and the relevance of images as an aid for humor comprehension (RQ4).

As displayed in figure 18, 86.6% of the participants thought images affected their perception of the memes. In the following ‘In what ways?’ descriptive question, they pointed out again that images were essential for context and even for them to find the joke presented in the text as funny. Nonetheless, a surprising 73.3% of them answered that not having the ST did not influence their perception of the memes (figure 19). Instead, they considered that the various translations presented were valuable for them to comprehend the joke.

Finally, despite initial support provided by the MT systems’ translations, participants still prefer having both the ST and the images to aid their assessments of the translations for memes (see figures 18, 19, and 20), as it would be optimal to have all the resources to judge a translation as good or not entirely. They would likely change their evaluation if the image were not provided (figure 21).

Regarding the puns, participants had many more difficulties understanding the humor. According to the ‘Analyzing the Activity’ at the end of the puns section, 80% of the participants said they would rather have the ST (figure 23), and 60% thought they would
likely change their evaluation if they had the ST (figure 25). Regarding having the aid of an image, 53.3% affirmed that the puns would not work better with an image (figure 22). The remaining 46.6% stated that images would probably support their comprehension of the humor, especially in the most complex puns, besides adding an extra layer of humor (RQ4). Interestingly, their votes were split in 50% on whether they would likely change or not their evaluation if there was an image (figure 24).

57. A imagem influenciou sua percepção da tradução no caso dos memes?

![Figure 18](image18.png)

*Figure 18 – Results from the ‘Analyzing the Activity’ from the memes section of Survey 2. Question 57 can be translated as: ‘Did the image influence your perception of the translations in the case of memes? Yes or no.’*

59. Não ter o texto original (Source Text) afetou sua percepção do meme?

![Figure 19](image19.png)

*Figure 19 – Results from the ‘Analyzing the Activity’ from the memes section of Survey 2. Question 59, can be translated as: ‘Not having the ST affected your perception of the meme? Yes or no.’*
61. Preferiria ter visto o texto em inglês?

- Sim: 12
- Não: 3

Figure 20 – Results from the ‘Analyzing the Activity’ from the memes section of Survey 2. Question 61 can be translated as: ‘Would rather have the ST? Yes or no.’

69. Você mudaria sua avaliação se as imagens não estivessem inclusas?

- Sim: 13
- Não: 2

Figure 21 – Results from the ‘Analyzing the Activity’ from the memes section of Survey 2. Question 69 can be translated as: ‘Would you change your evaluation if images were not included? Yes or no.’

95. Você acha que esses puns (trocadilhos) funcionariam melhor com o auxílio de uma imagem?

- Sim: 7
- Não: 8

Figure 22 – Results from the ‘Analyzing the Activity’ from the puns sections of Survey 2. Question 95 can be translated as: ‘Do you think that these puns would work better with the aid of an image? Yes or no.’
Figure 23 – Results from the ‘Analyzing the Activity’ from the puns sections of Survey 2. Question 99 can be translated as: ‘Would you rather have the ST? Yes or no.’

Figure 24 – Results from the ‘Analyzing the Activity’ from the puns sections of Survey 2. Question 103 can be translated as: ‘Would you change your evaluation if images were included? Yes or no.’

Figure 25 – Results from the ‘Analyzing the Activity’ from the puns sections of Survey 2. Question 105 can be translated as: ‘Would you change your evaluation if the ST was included? Yes or no.’

From figure 26, we can analyze the difficulty that participants had to comprehend humor in the most complex puns (i.e. Granddad Frog Joke (Frog noise), Four-chin Teller (Fat psychic), Too Many Shifts (Shifts)). This difficulty is even more apparent when we compare it to figure 14 from the previous section. This additional evidence further concludes that having the ST is paramount for humor comprehension (RQ3), especially in puns.
From the final section of Survey 2, the post-task questionnaire, we could see a significant difference from Section 4.2.1. For example, 66.6% of the participants affirmed, in question 110 (figure 28), that they liked the translations, opposed to the 50% division from the previous survey. Moreover, more volunteers (66.6%) found the translations helped humor perception (against the former 53.3% in Survey 1) – see figure 29. However, on the other hand, their taste (figure 27) remained relatively the same: 66.6% said the selected pieces reflected their usual consumption of memes and puns against the previous 60% in Survey 1.
Figure 28 – Results from the post-task questionnaire of Survey 2. Question 110 can be translated as: ‘Did you like the translations presented? Yes or no.’

Figure 29 – Results from the post-task questionnaire of Survey 2. Questions 112 can be translated as: ‘Do you think the translations helped you understand humor? Yes or no.’

Regarding the participants’ preferences for a translation, table 9 below shows that the translation by Google Translate was chosen the best output 4 out of 9 times for the memes and 3 out of 6 times for the puns. DeepL provided the best translations 4 out of 9 times in memes and 3 out of 6 times in puns. Bing was ranked first 1 out of 9 times in memes and 1 out of 6 times in puns. These results indicate a significant difference from table 8 in Survey 1, where Google Translate’s output was placed as first best 5 out of 9 times in memes and 3 out of 6 times in puns. In Survey 2 (table 9), Google is tied with DeepL Translate with 4 out of 9 in memes and 3 out of 6 in puns. Bing was again placed as the third best provider. These results answers our RQ1, regarding MT quality from the end-user point-of-view, and our RQ2, about end-user reception: Google Translate and DeepL Translate presented the best translations regarding quality (RQ1) and comprehension for end-users (RQ2).
<table>
<thead>
<tr>
<th>Memes</th>
<th>Google</th>
<th>Bing</th>
<th>DeepL</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Drake – Alarms</em></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Drake – Kids’ food</em></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Drake – Shoelaces</em></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Distracted Boyfriend – People Walking</em></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Distracted Boyfriend – Back to Work</em></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Distracted Boyfriend – Microwave</em></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Confession Bear – Drinking</em></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Confession Bear – People Skills</em></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Confession Bear – Pandemics</em></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Puns</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Granddad Frog Joke</em></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Four-chin Teller</em></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Too Many Shifts</em></td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>The Guy Who Invented Zero</em></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Restaurant on the Moon</em></td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>Irrelephant</em></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Table 9 – Results from the ranking in Survey 2 done by the participants for each set of translations. The items marked with (*) were tied because they are identical translations.

Notwithstanding, the answer to our RQ2 is intertwined with the findings in RQ3 and RQ4. **Images were considered essential for humor perception in memes and as a helpful addition/aid in puns (RQ3).** Furthermore, **the presence of the ST (RQ4) significantly affected the results.** Translations were seen as more helpful, and the ST was more required, and, as a contrast, humor was less perceived or lost without the aid of the ST for the respondents of Survey 2.
4.2.3 Average Answers for Both Surveys

By combining the results of both surveys, we can see that Google Translate was perceived as the best MT system, followed by DeepL Translate, and finally by Bing Translator (RQ2) as presented in table 10. Hence, it is safe to assume that humor perception by our end-users, our RQ2, was less affected by Google’s translations and more explicitly identified by Bing’s outputs.

<table>
<thead>
<tr>
<th>Results from the Ranking of Translations – Surveys 1 and 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Table 10 – Results from the ranking in Surveys 1 and 2 done by the participants for each set of translations. This is the sum of how many times each MT system was classified as first, second, and third.

This study can make us conclude that the translations offered by these NMT systems can be helpful as a base for humor understanding, especially for non-native speakers. However, even though they performed surprisingly well, with fluency and accuracy, i.e. DeepL in Distracted Boyfriend – Back to Work, Google in Confession Bear – Drinking, and Google in Irrelephant (see tables 1 and 2 for the texts and Appendix 1 for the full translator’s analysis), they still have flaws in register and in basically identifying jokes and wordplays (i.e. the opposing scripts described by Raskin (2017) in his STH).
Chapter 5 – Discussion

In this chapter, the results detailed presented previously are discussed to answer our overarching RQ: ‘Is humor perceived after the translation of nine memes and six puns from EN-US into PT-BR by three MT systems?’. Since this work attempted to analyze how and if humor in memes and puns can be conveyed from EN-US to PT-BR by MT systems, we can use the four complementary sub-research questions to answer this question.

The main goal of the present work is to fill a gap in the research of humor in MT. This is a relevant topic since creating humorous pieces such as memes is an ever-growing activity by users around social media. Thus, in an interconnected world, translation is continuously essential. Moreover, memes can and are used as a comedy relief tool, which this COVID-19 moment evoked even more for being a global health problem. That said, the translation of wordplay and jokes seems to be significant. Consequently, having research that fills the gap in that area might contribute to improvements in freely available MT systems. Additionally, comments of the surveys’ participants pointing out issues on register and literalness of the translations help us reaffirm that there is still a gap in NMT systems for identifying and translating humor in diverse styles and registers.

The experiments described in the present research were carried out using two primary methodologies as a basis: the User-Centered Translation (UCT), via the two surveys and the thirty participants, and the MQM’s definitions of accuracy, fluency, and register, which aided the invited translator’s analysis in the spreadsheet.

Regarding RQ1, which investigates what MT system can deliver the best translation for each piece of comic text, we have demonstrated that Google Translate’s outputs are ranked in first place most of the time by end-users (see Section 4.2.1 and Section 4.2.2). Despite its flaws, it is also considered the best MT system for the pieces presented according to the invited translator, even though it does not perform equally well in all the definitions analyzed (see table 5 and table 6 in Section 4.1 and Appendix 1). It was followed by DeepL Translate, in second place, and Bing Translator in third, according to participants and the translator (see table 10 with the concatenated results in the surveys, and tables 5 and 6 for the translator’s assessment).
Regarding RQ2, which analyzes the user experience with the translations provided, we have seen, in Sections 4.2.1 and 4.2.2, that the users were more inclined to like or disliked the translations based on whether they had access to the ST or not (Figures 15 and 27). Nonetheless, according to their comments, the translations were generally not satisfying as a solid final MT output. Instead, they were regarded as helpful essential translations for non-native speakers and a base for translators to work on.

RQ3 was investigated in the survey with specific questions such as ‘Did the image influence your perception of the translation in the memes?’, which was formulated as a ‘yes’ or ‘no’ type of question, and a more descriptive question, ‘How?’, where participants could write their impressions about the first question. They were also asked to rank which meme’s image impacted them the most and why and if the image influenced their suggested translation and their opinion of the translations. Because memes are multimodal, the impact of the image was investigated. We wanted to see how much the image affects humor perception and whether it affects translation understanding for the better or for the worse.

Furthermore, we have shown that images are vital for memes’ contextual comprehension (RQ3), proving that the captions of the memes by themselves are not enough for humor understanding. This notion was complemented by the participants’ comments and by the translator’s assessment that the translations for memes were harder to evaluate without the images (see Section 4.1, the translator’s analysis, and Sections 4.2.1 and 4.2.2 for the data on the surveys). On the other hand, puns could stand without images, but they would be a helpful addition, according to the participants, for users who could not understand the wordplay.

Regarding RQ4, the following questions were asked: ‘Has the lack of the original text (source text) affected your perception of the meme?’ This yes/no question was followed by a descriptive ‘How?’. Then, in Sections 4.2.1, 4.2.2, and 4.2.3, we investigated the effect of the ST because we wanted to examine how the translations would stand on their own, judging how well they conveyed comedy, mainly when intricate wordplay or context were used. In addition, half of the participants had access to the ST, whereas the other half did not so that we could compare humor perception.
As a result, we have demonstrated that the ST is vital for end-user experience on humor perception (figure 18). Users seem to become much more critical about the translations provided after seeing the ST (as mentioned in Sections 4.2.2 and 4.2.3). Furthermore, RQ4 was also answered with the translator’s analysis, showing that what he calls ‘tradukibe,’ a non-translated meme or a ‘very obvious’ translation, affects the transfer of humor. Because of that, users might find the joke unfunny due to the influence of the ST in their perception.

Therefore, the overarching RQ is answered positively: humor can be perceived in MT outputs when translating memes and puns. Answers to RQ1 and RQ2 illustrated that a basic understanding of humor is possible according to the users and the translator’s analysis. As mentioned in Chapter 4, the MT systems (Google Translate and DeepL Translate) provided surprisingly accurate, fluent, and adequate register translations in some exceptional cases. Thus, even though numerous improvements are to be implemented (i.e. in register and in mistranslations), the MT systems studied here can convey humor in memes and puns from EN-US into PT-BR.
Conclusion

During this dissertation, we aimed at answering a research question: Is humor still perceived after nine memes and six puns are translated by MT systems (i.e. Google Translate, Bing Microsoft Translator, and DeepL Translate) from EN-US into PT-BR? To answer it, we applied the User-Centered Translation (UCT) approach, via two surveys for thirty volunteers, alongside Multidimensional Quality Metrics (MQM) definitions of accuracy, fluency, and register, which were analyzed in a spreadsheet by an invited professional translator.

This research topic was chosen because of an identified gap in the study of MT for humorous pieces. From what we presented in Chapter 2, Section 2.4, only a few papers directly tackle this issue, especially in the case of memes, a brand-new type of humor that emerged with the internet.

Our results revealed translation problems in all the MT systems, both according to MQM’s definitions and to users’ perceptions. The professional translator indicated the main drawbacks of mistranslations, and untranslated terms, and the survey participants predominantly highlighted issues regarding the register. Nonetheless, participants were able to perceive humor at a basic level, proving that NMT systems and AI will still benefit from many improvements, but they already are translating functioning texts.

Limitations

One of the limitations of this study was the number of participants in the survey and in the analysis. Gathering a much larger number could have given us better insights into the problems faced by the MT systems studies. Also, for the MQM analysis, this work could benefit from more professional translators who could complement and add new perspectives to the analysis. Unfortunately, due to time and space constrictions, we were not able to fulfill such a task.

Another limitation is the multimodal nature of memes, which significantly affects the image linked to them. In this case, a way to judge the quality of the translation of multimodal content like memes with more ecological validity would be to use a multimodal
MT system that employs images in the training process (Specia et al 2016). In that way, a corpus with memes could be used to train a multimodal MT system that would take the images into consideration. However, to use a specific trained MT would not fit the RQ in this research – which was to see if the general public, end-user, could still use freely available MT system to translate humor – as multimodal MT systems are not yet deployed by the companies for free online.

**Contributions**

This work contributes to the MT and Humor Studies fields of research in the sense that it slightly fills the gap in research, and it opens the door for further analysis on the topic of memes and puns in MT. Furthermore, our findings indicate that the MT systems tested are of good quality for basic comprehension of humor. Overall, they were accurate and fluent in most cases, being the errors in the register, Portuguese variation, and mistranslation being their major limitation.

Because of the surveys, we discovered how users perceive humor in memes and puns translated by MT. Despite indicating that this field of research still has obstacles to overcome, our data and analysis demonstrate the current benefit of MTs for non-native speakers.

Moreover, our sub-research questions confirmed that memes are inherently dependent on their visual components, essential for contextual comprehension. On the other hand, puns complemented by images could be beneficial, especially for non-native speakers.

Finally, the surveys taught us that the users’ perception of humor in the MT outputs changes depending on the presence or not of the ST. This factor affects their evaluation of the jokes and the quality of their translations.
Future Research

This work offers some potential research opportunities, such as developing an NMT system directed at translating memes and puns, which could bring further knowledge to the field and humor comprehension by AI. As mentioned before, in Chapter 5, from the surveys, we could gather the information that the end-user identifies potential improvements in register and humor's literalness. Using multimodal MT to encompass translation of memes alongside their images could be a future improvement and a research opportunity.

Also, as mentioned in the Limitations Section, replicating this study on a larger scale could harvest new helpful ideas and discoveries for improving MT technologies. Thus, this study may be a valuable basis for further similar experiments.
Bibliography


Discovery and Data Mining. Paris, France.

https://doi.org/10.1145/2783258.2783388.


Appendix

Appendix 1 – Translator’s Analysis Spreadsheet

<table>
<thead>
<tr>
<th>SOURCE TEXT</th>
<th>MT System</th>
<th>Translations</th>
<th>Accuracy</th>
<th>Fluency</th>
<th>Register</th>
<th>Best? Why?</th>
<th>Possible to transfer humor?</th>
<th>Other comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drake - Kids' food</strong></td>
<td><strong>Google</strong></td>
<td>A reação dos meus filhos à comida na loja / Reação dos meus filhos à mesma comida em casa</td>
<td>Google's MT for some reason omits the article in the second sentence (or simply adds for no reason in the first sentence, depending on your perspective), which proves to be inconsistent.</td>
<td>Barring the &quot;a reação&quot; (the article in the beginning of the sentence), it's rather fluent.</td>
<td>The register appears to be quite adequate, barring the &quot;à&quot; which could become &quot;para&quot; as it would reflect a more oral/informal register.</td>
<td>Google's MT is the closest one to be the best. It does not conjugates verbs for no reason, it maintains a similar structure to the original and it is rather fluent.</td>
<td>The transfer is possible.</td>
<td>Just to note, when a meme is clearly seen as translated by a MT system, we call it &quot;tradukibe&quot;,</td>
</tr>
<tr>
<td><strong>Bing</strong></td>
<td>Meus filhos reações à comida na loja / Meus filhos reações à mesma comida em casa</td>
<td>Bing's MT has problems with nominal groups (meus filhos reações instead of reações dos meus filhos)</td>
<td>This has word order/nominal group issues.</td>
<td>The register is not adequate, as it transforms the sentence structure to not correspond to the text genre 'meme'. (It doesn't necessarily have to</td>
<td></td>
<td>The transfer is rather awkward, it could be fun but it would be easily visible that it was translated.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MEMES
<table>
<thead>
<tr>
<th></th>
<th>DeepL</th>
<th>Google</th>
<th>Bing</th>
<th>DRAM</th>
</tr>
</thead>
</table>
| **Os meus filhos reagem à comida na loja** / Os meus filhos reagem à mesma comida em casa | DeepL's MT present a structure that would be not used in Brazilian Portuguese memes. | Google's MT is not accurate due to the first verb being used in the imperative. | Bing's MT is not accurate due to | The register is adequate as it transforms the sentence structure to not correspond to the text genre 'meme'. (It doesn’t necessarily have to be subject -> verb -> object) | It's a tie between Google's and Bing's. Both retain a good level of register while maintaining the structure. Even if they're not perfect.  
| **Deepl's MT** conjugates “to reagir” as opposed to leaving it as a noun (reação), which is a fluency issue. | The verbs conjugated as the imperative make this meme not fluent. | The verbs conjugated as the | The register is adequate | Yes, it is possible.  
| **The register is not adequate, as it transforms the sentence structure to not correspond to the text genre 'meme'.** (It doesn’t necessarily have to be subject -> verb -> object) | The register is adequate | The register is adequate | The register is adequate | Yes, it is possible.  
| **The transfer is rather awkward, it would be easily visible that it was translated.** | The transfer is rather awkward, it would be easily visible that it was translated. | The transfer is rather awkward, it would be easily visible that it was translated. | Yes, it is possible. | Yes, it is possible.  
| **Drake - Shoelaces** | **Google** | **Bing** | **Google** | **Bing** | Yes, it is possible.  
| Buy new shoelaces / Take shoelaces from shoes I | Compre cadarços novos / Tirar cadarços de sapatos que não uso mais | Compre cadarços novos / | Compre cadarços novos / | Compre cadarços novos / | Yes, it is possible.  
| **Deepl's MT** conjugates “to reagir” as opposed to leaving it as a noun (reação), which is a fluency issue. | The verbs conjugated as the imperative make this meme not fluent. | The verbs conjugated as the | The verbs conjugated as the | The verbs conjugated as the | Yes, it is possible.  
| **The register is not adequate, as it transforms the sentence structure to not correspond to the text genre 'meme'.** (It doesn’t necessarily have to be subject -> verb -> object) | The register is adequate | The register is adequate | The register is adequate | Yes, it is possible.  
| **The transfer is rather awkward, it would be easily visible that it was translated.** | The transfer is rather awkward, it would be easily visible that it was translated. | The transfer is rather awkward, it would be easily visible that it was translated. | Yes, it is possible. | Yes, it is possible.  
| **Yes, it is possible.** | **Yes, it is possible.** | **Yes, it is possible.** | **Yes, it is possible.** | **Yes, it is possible.**  

**Note:** The translations are based on the assumption that the original text was quickly translated by an MT for clout, which is seen as negative and unfunny. That is how I'm basing my subjective justification for the following two "possible to transfer the humour" comments.
<p>| <strong>Don't wear anymore</strong> | Pegue caderços de sapatos que eu não uso mais | The verbs &quot;comprar&quot; and &quot;pegar&quot; being conjugated in the imperative. | Imperative make this meme not fluent. | (informal), even if it's not fluent/accurate. |   |   |
|------------------------|---------------------------------------------|-------------------------------------------------|----------------------------------|------------------------------------------|   |   |
| <strong>DeepL</strong>              | Comprar novos atacadores / Tire os atacadores dos sapatos que eu não uso mais | DeepL's MT is not accurate due to the verb &quot;retirar/tirar&quot; being conjugated in the imperative.| The verbs conjugated as the imperative make this meme not fluent. | The register might be adequate (informal) in other varieties of Portuguese, but it does not appear adequate for Brazilian Portuguese. | Maybe in other varieties of Portuguese, but not Brazilian Portuguese. |   |
| <strong>Distracted Boyfriend - People Walking</strong> | Calçada aberta / Pessoas caminhando / A rua | This is accurate. | This is fluent. | The register is adequate (informal). |   | Yes, it is possible. |
| <strong>Google</strong>             | Calçada aberta / Pessoas andando / A rua | This is accurate. | This is fluent. | The register is adequate (informal). |   | Yes, it is possible. |
| <strong>Bing</strong>               | Calçada bem aberta / Pessoas a andar / A rua | This is accurate. | This is fluent. | The register appears to be accurate in other varieties of Portuguese, but it's not really used in Brazilian Portuguese regarding &quot;a andar&quot; ao invés de &quot;andando&quot; | Yes, it is possible. |
| <strong>Distracted Boyfriend - Back to work</strong> | Literalmente qualquer outra coisa / Mim / | This is accurate | This brings problems in fluency with &quot;mim&quot; being | The register is mostly adequate, barring the use of &quot;mim&quot;. | DeepL is able to be the best system despite its different intended output (another variety | Yes, it is possible. |</p>
<table>
<thead>
<tr>
<th>Source</th>
<th>Translation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Literally anything else / Me / Work / First day back at work...</strong></td>
<td>Trabalhos / Primeiro dia de volta ao trabalho...</td>
<td>used instead of &quot;eu&quot;. It also puts &quot;work&quot; in plural for no reason.</td>
</tr>
<tr>
<td><strong>Bing</strong></td>
<td>Literalmente qualquer outra coisa / Me / Trabalho / Primeiro dia de volta ao trabalho...</td>
<td>This is accurate</td>
</tr>
<tr>
<td><strong>DeepL</strong></td>
<td>Literalmente qualquer outra coisa / Eu / Trabalho / Primeiro dia de volta ao trabalho...</td>
<td>This is accurate</td>
</tr>
<tr>
<td><strong>Distracted Boyfriend - Microwave</strong></td>
<td>Todos os outros botões do microondas / Eu / +30 botão segundos</td>
<td>This is accurate.</td>
</tr>
<tr>
<td><strong>Google</strong></td>
<td>Todos os outros botões do microondas / Eu / +30 botão segundos</td>
<td>This is accurate.</td>
</tr>
<tr>
<td><strong>Bing</strong></td>
<td>Todos os outros botões no micro- ondas / Me /</td>
<td>This is on the accurate side, but it has words not translated such</td>
</tr>
</tbody>
</table>

The register is adequate, barring the mistranslation.

This is fluent. Surprisingly, despite being a system aimed at another variety of Portuguese, the register is adequate for Brazilian Portuguese.

Barring the aforementioned accuracy and fluency issues, they are all on equal ground. The issue is that the problem with "botão de 30 segundos" not being translated doesn't really bring a good candidate.

Yes, it is possible.
<table>
<thead>
<tr>
<th><strong>Confession Bear - Drinking</strong></th>
<th>+30 segundos button</th>
<th>as &quot;me&quot; and &quot;button&quot;</th>
<th>instead of botão de 30 segundos</th>
<th>and the third one could be better in other varieties of Portuguese.</th>
<th>Yes, it is possible.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepL</td>
<td>Cada outro botão no microondas / Eu / +30 segundo botão</td>
<td>This might be accurate? I can't access properly due to register differences in DeepL's MT system. (I'm mostly thinking about &quot;cada outro&quot; instead of &quot;todos os outros&quot;)</td>
<td>This one has issued with word order and agreement (+30 segundo botão instead of botão de 30 segundos)</td>
<td>The register might be adequate for other varieties of Portuguese, but not that much for Brazilian Portuguese.</td>
<td>Yes, it is possible.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Confession Bear - People Skills</strong></th>
<th>+30 segundos button</th>
<th>as &quot;me&quot; and &quot;button&quot;</th>
<th>instead of botão de 30 segundos</th>
<th>and the third one could be better in other varieties of Portuguese.</th>
<th>Yes, it is possible.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>Eu não deveria estar bebendo. / To bebendo</td>
<td>This is accurate.</td>
<td>This is mostly fluent, but it has one orthographical problem &quot;to&quot; instead of &quot;tô&quot;</td>
<td>The register here is perfect. &quot;Tô&quot; is the shortened version of &quot;estou&quot; which most adequately encapsulates the informal register of Brazilian Portuguese.</td>
<td>Yes, it is possible.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Confession Bear - People Skills</strong></th>
<th>+30 segundos button</th>
<th>as &quot;me&quot; and &quot;button&quot;</th>
<th>instead of botão de 30 segundos</th>
<th>and the third one could be better in other varieties of Portuguese.</th>
<th>Yes, it is possible.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bing</td>
<td>Eu não deveria estar bebendo. / Estou bebendo.</td>
<td>This is accurate.</td>
<td>This is fluent.</td>
<td>The register here is okay.</td>
<td>Yes, it is possible.</td>
</tr>
<tr>
<td>DeepL</td>
<td>Eu não deveria estar bebendo. / Eu estou bebendo.</td>
<td>This is accurate.</td>
<td>This is fluent.</td>
<td>The register here is okay.</td>
<td>Yes, it is possible.</td>
</tr>
<tr>
<td>Starting to hate &quot;people&quot;</td>
<td>Bing</td>
<td>DeepL</td>
<td>Google</td>
<td>Confession Bear - Pandemics</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Bing</strong></td>
<td>Contratado para minhas &quot;habilidades de pessoas&quot; / Começando a odiar &quot;pessoas&quot;</td>
<td>This is mostly accurate. The preposition change of &quot;para&quot; instead of &quot;por&quot; changes the meaning a bit.</td>
<td>This is mostly fluent. The reason is related to the one pointed out in accuracy.</td>
<td>The register is adequate (informal).</td>
<td></td>
</tr>
<tr>
<td><strong>DeepL</strong></td>
<td>Contratado para minhas &quot;habilidades pessoais&quot; / Começando a odiar as &quot;pessoas&quot;</td>
<td>This is mostly accurate. The preposition change of &quot;para&quot; instead of &quot;por&quot; changes the meaning a bit.</td>
<td>This is mostly fluent. The reason is related to the one pointed out in accuracy.</td>
<td>The register is adequate (informal).</td>
<td></td>
</tr>
<tr>
<td><strong>Confession Bear - Pandemics</strong></td>
<td>People have been complaining since the pandemic started how much they miss being able to touch people / I personally have actually been really enjoying</td>
<td>This is mostly accurate. Melindroso does not correspond to the meaning of touchy-feely. Melindroso is like threatening or irritable.</td>
<td>This is fluent.</td>
<td>The register is adequate, barring the &quot;melindroso&quot; word which is not exactly informal.</td>
<td></td>
</tr>
<tr>
<td><strong>Google</strong></td>
<td>Desde o início da pandemia, as pessoas reclamam do quanto sentem falta de tocar nas pessoas. / Eu, pessoalmente, tenho realmente gostado de como todos estão menos</td>
<td>This is fluent.</td>
<td>The third one is better, because it is able to keep the meaning of touchy-feely.</td>
<td>Yes, it is possible. Although I would change &quot;melindroso&quot; to something else.</td>
<td></td>
</tr>
<tr>
<td>how less touchy-feely everyone is right now.</td>
<td>melindrosos agora.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bing</strong></td>
<td>As pessoas têm reclamado desde que a pandemia começou o quanto sentem falta de poder tocar as pessoas / Eu pessoalmente tenho realmente gostado de como todos são menos sensíveis agora.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>This is mostly accurate. I feel like just &quot;sensível&quot; loses the &quot;touchy&quot; aspect of &quot;touchy-feely&quot;.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>This is fluent.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The register is adequate.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **DeepL** | As pessoas têm reclamado desde o início da pandemia o quanto sentem falta de poder tocar as pessoas. / Eu, pessoalmente, tenho realmente gostado muito de como todos estão menos sensíveis ao toque agora. |
|  | This is accurate. |
|  | This is fluent. |
|  | The register is adequate. |

**PUNS**

<table>
<thead>
<tr>
<th><strong>Granddad Frog Joke</strong></th>
<th><strong>Google</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>A little girl asks</td>
<td>Uma garotinha pergunta ao avô: &quot;Você faria barulho de sapo</td>
</tr>
<tr>
<td></td>
<td>This is accurate.</td>
</tr>
<tr>
<td></td>
<td>This is fluent.</td>
</tr>
<tr>
<td></td>
<td>Puns are tricky for me as the register either needs to be really formal or</td>
</tr>
<tr>
<td></td>
<td>The first one and the third one are a tie, but something tells me the first is closer to</td>
</tr>
</tbody>
</table>

Yes. There's not much of a punchline, but the effect is there.
<table>
<thead>
<tr>
<th>Bing</th>
<th>DeepL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uma garotinha pergunta ao avô: &quot;Você faria um barulho de sapo por mim?&quot;</td>
<td>Uma menina pergunta ao seu avô: &quot;Você faria um barulho de sapo para mim?&quot;</td>
</tr>
<tr>
<td>This is mostly accurate. The issue is that there's nothing indicating granddad's death!!</td>
<td>This is accurate.</td>
</tr>
<tr>
<td>This is rather fluent, but something tells me that it should be &quot;disse&quot; instead of &quot;diz&quot;.</td>
<td>This is rather fluent, but something tells me that it should</td>
</tr>
<tr>
<td>The register is adequate, see reasons above.</td>
<td>The register is adequate, see reasons above.</td>
</tr>
<tr>
<td>Yes, but for other reasons. Here, it's funny because of the mistranslation of &quot;croak&quot;.</td>
<td>Yes. There's not much of a punchline, but the effect is there.</td>
</tr>
<tr>
<td>Four-chin Teller</td>
<td>Google</td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
</tr>
<tr>
<td>O avô, confuso, pergunta: &quot;Por quê?&quot;. A menina responde: &quot;O pai diz que quando você coxar, vamos todos à Disneylandia&quot;.</td>
<td>be &quot;disse&quot; instead of &quot;diz&quot;.</td>
</tr>
<tr>
<td>The last sentence contributes to low accuracy.</td>
<td>The register appears to be adequate.</td>
</tr>
<tr>
<td>O que você chama de médium muito gordo? Um contador de quatro queixo</td>
<td>In order to better convey humour in translation, this cannot be done by a MT as it requires a good degree of freedom and departure from the literal meaning.</td>
</tr>
</tbody>
</table>

I would like to note that while the solution I propose is "funny", it may wander into inappropriate territory as it may be fatphobic, so proceed with caution. But theoretically, it is possible to convey humour.
<table>
<thead>
<tr>
<th></th>
<th>DeepL</th>
<th>Google</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Too many shifts</strong></td>
<td>O que você chama de um médium realmente gordo? Um caixa de quatro moedas</td>
<td>Por que o funcionário da unidade de fabricação de teclados foi demitido? Por colocar muitos turnos</td>
<td></td>
</tr>
<tr>
<td>(Manufacturer)</td>
<td>The last sentence contributes to low accuracy.</td>
<td>This is an accurate translation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disregarding the translation choices, it is fluent.</td>
<td>This is a fluent translation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The register appears to be adequate.</td>
<td>The register appears to be adequate.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>freedom and departure from the literal meaning.</td>
<td>freedom and departure from the literal meaning.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I’d recommend &quot;Como você chama um médium muito gordo? Um grandium&quot;</td>
<td>I’m not sure if vidente could work here.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>In order to better convey humour in translation, this cannot be done by a MT as it requires a good degree of freedom and departure from the literal meaning.</td>
<td>I’d recommend &quot;Como você chama um médium muito gordo? Um grandium&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>It can be either one, as there’s not really a best solution, considering humour was inevitably lost.</td>
<td>The key (no pun intended) in all of them is transforming the translation of &quot;shift&quot; into something that is used in keyboards in Brazilian</td>
<td></td>
</tr>
<tr>
<td>For putting in too many shifts</td>
<td>Bing</td>
<td>DeepL</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>Por que o trabalhador da unidade de fabricação de teclados foi demitido? Por colocar em muitos turnos</td>
<td>This is an accurate translation. This is a fluent translation. The register appears to be adequate.</td>
<td>The key in all of them is transforming the translation of &quot;shift&quot; into something that is used in keyboards in Brazilian Portuguese as well. Maybe we could say &quot;Por colocar muito espaço nos seus intervalos&quot;, to reference the spacebar.</td>
</tr>
<tr>
<td></td>
<td>Google</td>
<td>Bing</td>
<td>DeepL</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>The guy who invented zero</strong></td>
<td><strong>Para o cara que inventou o zero, obrigado por nada.</strong></td>
<td><strong>Ao cara que inventou zero, obrigado por nada.</strong></td>
<td><strong>Para o cara que inventou o zero, obrigado por nada.</strong></td>
</tr>
<tr>
<td><strong>To the guy who invented zero, thanks for nothing.</strong></td>
<td><strong>This is an accurate translation.</strong></td>
<td><strong>This is a fluent translation.</strong></td>
<td><strong>This is a fluent translation.</strong></td>
</tr>
<tr>
<td><strong>Restaurant on the moon</strong></td>
<td><strong>Você ouviu sobre o restaurante na lua? Ouvi dizer que a comida era boa, mas não tinha atmosfera.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Did you hear about the restaurant on the moon? I heard</strong></td>
<td><strong>This is an accurate translation.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>Portuguese</td>
<td>Translation</td>
<td>Register</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td>the food was good but it had no atmosphere.</td>
<td>Você ouviu sobre o restaurante na lua? Ouvi dizer que a comida era boa, mas não tinha atmosfera.</td>
<td>This is an accurate translation.</td>
<td>The register appears to be adequate, barring the use of &quot;atmosfera&quot;</td>
</tr>
<tr>
<td>Bing</td>
<td>Você ouviu falar sobre o restaurante na lua? Ouvi dizer que a comida era boa, mas não tinha atmosfera.</td>
<td>This is an accurate translation.</td>
<td>The register appears to be adequate, I'm not sure about the use of &quot;atmosfera&quot;, though.</td>
</tr>
<tr>
<td>DeepL</td>
<td>Você ouviu falar sobre o restaurante na lua? Ouvi dizer que a comida era boa, mas não tinha atmosfera.</td>
<td>This is a fluent translation.</td>
<td></td>
</tr>
<tr>
<td>Irrelephant</td>
<td>Por que Dumbo estava triste? Ele se sentiu irrelephant.</td>
<td>This is surprising accurate.</td>
<td>The register appears to be adequate and understandable.</td>
</tr>
<tr>
<td>Google</td>
<td>Por que Dumbo estava triste? Ele se sentiu irrelephant.</td>
<td>This is fluent.</td>
<td></td>
</tr>
<tr>
<td>Bing</td>
<td>Por que Dumbo estava triste? Ele se sentiu irrelephant.</td>
<td>This is rather accurate, the MT system does get lost in &quot;irrelephant&quot;, as it does not get translated.</td>
<td>The register appears to be adequate, kind of.</td>
</tr>
<tr>
<td>DeepL</td>
<td>Por que o Dumbo ficou triste? Ele se sentiu irrefletido.</td>
<td>This one is not really accurate, it misses the mark with &quot;irrefletido&quot;.</td>
<td>I'm not sure about the register on this one, I don't think this even works for other varieties of Portuguese, but it might be wrong.</td>
</tr>
</tbody>
</table>
Appendix 2 – Survey 1 (with ST)

For the full survey in Brazilian Portuguese, access the following link: https://drive.google.com/file/d/1RiKISpY_I6vA zfexCcaEY3PTtCf6goMI/view?usp=sharing.

This is a summarized version of the Survey 1. The memes and puns were discarded and only the questions were maintained. Since various questions were repeated for each meme and pun, the repeated sets were also discarded.

Pre-Task Questionnaire

4 – Gender → (1) Masculine, (2) Feminine, (3) Other.
5 – Age → (1) 12-17, (2) 18-24, (3) 25-34, (4) 35-44, (5) 45-54, (6) 55-64, (7) 65-74, (8) 75+.
6 – Native Language → (1) Portuguese, (2) English, (3) Other.
8 – Do you have any experience in translation? (1) Yes, (2) No.
9 – Please, briefly describe your experience in the field of translation.
11 – Do you consume humor online? In social media, humor websites? (1) Yes, (2) No.
12 – In which websites or social media? (1) Instagram, (2) 9GAG, (3) Twitter, (4) Facebook, (5) Imgur, (6) TikTok, (7) Other.
13 – In which language do you consume humor the most? (1) Portuguese, (2) English, (3) Other.
14 – Do you know what a meme is?
15 – Do you know what a pun is?
16 – What type of meme do you like? Does it involve a specific topic or a specific type of joke?
17 – What type of puns do you like? Does it involve a specific topic or a specific type of joke?
18 – Do you write humorous texts or memes? (1) Yes, (2) No.
19 – Which type of humorous piece do you write? (1) Puns, (2) Memes, (3) Other.
20 – Do you use any tool to translate or to aid your translation process? (1) MT systems, (2) CAT Tools (e.g. Trados, SmartCat, memoQ), (3) I translate manually, (4) Other.

Memes and Puns Questions

21 – Did you understand the humor? (1) Yes, (2) No.
22 – Was there humor in the original text in English? (1) Yes, (2) No, (3) I do not know.

23 – Rank the translations according to your preference → Translation 1, Translation 2, Translation 3.

24 – Did you find the translations funny? (1) Yes, (2) No.

25 – For which reason? (1) Because of the joke itself, (2) Because of the translations’ mistakes, (3) Because of the translations’ adequacy, (4) Other.

26 – Generally, what do are your thoughts on Translations 1, 2, and 3?

27 – Suggest your own translation for this meme.

Analyzing the Activity – Memes

84 – Did the image influence your perception about the translation in the case of the memes? (1) Yes, (2) No.

85 – How?

86 – Would you rather have seen only the ST? (1) Yes, (2) No.

87 – Why?

88 – For which meme do you think the image influenced the most? (1) Drake, (2) Distracted Boyfriend, (3) Confession Bear, (4) None, (5) Other.

89 – Why? Describe what caught your attention in the meme and/or in the translation.

90 – Do you think the image ended up affecting your evaluation of the translation? (1) Yes, (2) No.

91 – Why?

92 – Do you think the images served as a humoristic complement? (1) Yes, (2) No.

93 – Why?

94 – Would you change your evaluation if images were not included? (1) Yes, (2) No.

95 – Why?

Analyzing the Activity – Puns

138 – Do you think these puns would work better with the aid of an image? Transforming these puns into memes would benefit your understanding of the humor? (1) Yes, (2) No.

139 – Why?

140 – Do you think the presence of images would affect your humor perception of the translations?

141 – Why?
142 – Would you rather have access only to the ST? (1) Yes, (2) No.
144 – Why?
145 – Would you change your evaluation if images were included? (1) Yes, (2) No.
146 – Why?

Post-Task Questionnaire
147 – Do the pieces presented (memes and puns) resonate with the type of humor you consume? (1) Yes, (2) No.
148 – Did you like the translations provided? (1) Yes, (2) No.
149 – What is your opinion about the translations provided?
150 – Do you think the translations helped you understand humor? (1) Yes, (2) No.
151 – In what ways?
152 – Did you find any difficulties when translating memes and puns? (1) Yes, (2) No.
153 – Which ones?
154 – Would you translate memes or puns with MT? (1) Yes, (2) No.
155 – Why?
156 – Do you think MT is an aid for the translation process of memes and puns? (1) Yes, (2) No.
157 – In what ways it aids?
158 – In what ways it does not?
159 – Did the translations provided aided you with your own suggested translations? (1) Yes, (2) No.
160 – How did you use the translations provided?
161 – How was your experience with this survey?
Appendix 3 – Survey 2 (No ST)

For the full survey in Brazilian Portuguese, access the following link: https://drive.google.com/file/d/1_mTJo2KwBK0qnMMkZrbuaHyRFPXg6So/view?usp=sharing.

This is a summarized version of the Survey 2. The memes and puns were discarded and only the questions were maintained. Since various questions were repeated for each meme and pun, the repeated sets were also discarded.

Pre-Task Questionnaire

4 – Gender → (1) Masculine, (2) Feminine, (3) Other.

5 – Age → (1) 12-17, (2) 18-24, (3) 25-34, (4) 35-44, (5) 45-54, (6) 55-64, (7) 65-74, (8) 75+.

6 – Native Language → (1) Portuguese, (2) English, (3) Other.


8 – Do you have any experience in translation? (1) Yes, (2) No.

9 – Please, briefly describe your experience in the field of translation.


11 – Do you consume humor online? In social media, humor websites? (1) Yes, (2) No.

12 – In which websites or social media? (1) Instagram, (2) 9GAG, (3) Twitter, (4) Facebook, (5) Imgur, (6) TikTok, (7) Other.

13 – In which language do you consume humor the most? (1) Portuguese, (2) English, (3) Other.

14 – Do you know what a meme is?

15 – Do you know what a pun is?

16 – What type of meme do you like? Does it involve a specific topic or a specific type of joke?

17 – What type of puns do you like? Does it involve a specific topic or a specific type of joke?

18 – Do you write humorous texts or memes? (1) Yes, (2) No.

19 – Which type of humorous piece do you write? (1) Puns, (2) Memes, (3) Other.

20 – Do you use any tool to translate or to aid your translation process? (1) MT systems, (2) CAT Tools (e.g. Trados, SmartCat, memoQ), (3) I translate manually, (4) Other.

Memes and Puns Questions
21 – Rank the translations according to your preference → Translation 1, Translation 2, Translation 3.

22 – Did you understand the humor? (1) Yes, (2) No.

23 – For which reason? (1) Because of the joke itself, (2) Because of the translations’ mistakes, (3) Because of the translations’ adequacy, (4) Other.

24 – Generally, what do are your thoughts on Translations 1, 2, and 3?

Analyzing the Activity – Memes

57 – Did the image influence your perception about the translation in the case of the memes? (1) Yes, (2) No.

58 – In which way?

59 – Did not having the ST affect your perception about the meme?

60 – In which way?

61 – Would you rather have seen the ST? (1) Yes, (2) No.

62 – Why?

63 – For which meme do you think the image influenced the most? (1) Drake, (2) Distracted Boyfriend, (3) Confession Bear, (4) None, (5) Other.

64 – Why? Describe what caught your attention in the meme and/or in the translation.

65 – Do you think the image ended up affecting your evaluation of the translation? (1) Yes, (2) No.

66 – Why?

67 – Do you think the images served as a humoristic complement? (1) Yes, (2) No.

68 – Why?

69 – Would you change your evaluation if images were not included? (1) Yes, (2) No.

70 – Why?

Analyzing the Activity – Puns

95 – Do you think these puns would work better with the aid of an image? Transforming these puns into memes would benefit your understanding of the humor? (1) Yes, (2) No.

96 – Why?

97 – Do you think the presence of images would affect your humor perception of the translations?

98 – Why?
99 – Would you rather have access to the ST? (1) Yes, (2) No.

100 – Why?


102 – Why?

103 – Would you change your evaluation if images were included? (1) Yes, (2) No.

104 – Why?

**Post-Task Questionnaire**

107 – Do the pieces presented (memes and puns) resonate with the type of humor you consume? (1) Yes, (2) No.

108 – Did you realize they were translations in the beginning of the experiment? (1) Yes, (2) No.

109 – What elements made you realize?

110 – Did you like the translations provided? (1) Yes, (2) No.

111 – What is your opinion about the translations provided?

112 – Do you think the translations helped you understand humor? (1) Yes, (2) No.

113 – In what ways?

114 – How was your experience with this survey?
## Appendix 4 – Gloss Translations

<table>
<thead>
<tr>
<th>Selection of Memes</th>
<th>Source Text</th>
<th>Google Translate</th>
<th>Bing Translator</th>
<th>DeepL Translate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 – Drake – Alarms</strong></td>
<td>Schedule appointments on a calendar / Set a billion phone alarms</td>
<td>Agende compromissos em um calendário / Defina um bilhão de alarmes de telefone</td>
<td>Agende compromissos em um calendário / Defina um bilhão de alarmes telefônicos</td>
<td>Agendamento de compromissos em um calendário / Estabelecer um bilhão de alarmes telefônicos</td>
</tr>
<tr>
<td><strong>Gloss Translation</strong></td>
<td>Schedule appointments on a calendar / Define a billion of alarms of telephone</td>
<td>Schedule appointments on a calendar / Define a billion telephonic alarms</td>
<td>Scheduling of appointments on a calendar/ To establish a billion of telephonic alarms</td>
<td></td>
</tr>
<tr>
<td><strong>2 – Drake – Kids’ food</strong></td>
<td>My kids reaction to food at the store / My kids reaction to the same food at home</td>
<td>A reação dos meus filhos à comida na loja / Reação dos meus filhos à mesma comida em casa</td>
<td>Meus filhos reações à comida na loja / Meus filhos reações à mesma comida em casa</td>
<td>Os meus filhos reagem à comida na loja / Os meus filhos reagem à mesma comida em casa</td>
</tr>
<tr>
<td><strong>Gloss Translation</strong></td>
<td>My kids reaction to food at the store / My kids reaction to the same food at home</td>
<td>Reactions my kids to the food at the store / My kids reactions at the same food at home</td>
<td>My kids react to the food at the store/ My kids react to the same food at home</td>
<td></td>
</tr>
<tr>
<td><strong>3 – Drake – Shoelaces</strong></td>
<td>Buy new shoelaces / Take shoelaces from shoes I don’t wear anymore</td>
<td>Compre cadarços novos / Pegue cadarços de sapatos que não uso mais</td>
<td>Compre cadarços novos / Pegue cadarços de sapatos que eu não uso mais</td>
<td>Comprar novos atacadores / Tire os atacadores dos sapatos que eu não use mais</td>
</tr>
<tr>
<td><strong>Gloss Translation</strong></td>
<td>Buy new shoelaces / To take shoelaces from shoes that I don’t use anymore</td>
<td>Buy new shoelaces / Fetch shoelaces from shoes that I don’t use anymore</td>
<td>To buy new fasteners (PT-EU) / Take off the fasteners (PT-EU) from the shoes I don’t use anymore</td>
<td></td>
</tr>
<tr>
<td><strong>4 – Distracted Boyfriend – People Walking</strong></td>
<td>Wide open sidewalk / People walking / The street</td>
<td>Calçada aberta / Pessoas caminhando / A rua</td>
<td>Calçada aberta / Pessoas andando / A rua</td>
<td>Calçada bem aberta / Pessoas a andar / A rua</td>
</tr>
<tr>
<td><strong>Gloss Translation</strong></td>
<td>Open sidewalk (mistranslation) / People commuting / The street</td>
<td>Open sidewalk (mistranslation) / People walking / The street</td>
<td>Very open sidewalk (mistranslation) / People to walk (PT-EU) / The street</td>
<td></td>
</tr>
<tr>
<td>5 – <em>Distracted Boyfriend – Back to Work</em></td>
<td>Literamente qualquer outra coisa / Mim / Trabalhos / Primeiro dia de volta ao trabalho...</td>
<td>Literamente qualquer outra coisa / Me / Trabalho / Primeiro dia de volta ao trabalho...</td>
<td>Literamente qualquer outra coisa / Eu / Trabalho / Primeiro dia de volta ao trabalho...</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Gloss Translation</strong></td>
<td>Literally anything else / My / Work / First day back at work...</td>
<td>Literally anything else / Me (Untranslated) / Work / First day back at work...</td>
<td>Literally anything else / I / Work / First day back at work...</td>
<td></td>
</tr>
<tr>
<td>6 – <em>Distracted Boyfriend – Microwave</em></td>
<td>Todos os outros botões do microondas / Eu / +30 botão segundos</td>
<td>Todos os outros botões no microondas / Me / +30 segundos button</td>
<td>Cada outro botão no microondas / Eu / +30 segundo botão</td>
<td></td>
</tr>
<tr>
<td><strong>Gloss Translation</strong></td>
<td>All the other buttons of the microwave / I / +30 button seconds</td>
<td>All the other buttons at the microwave / Me (Untranslated) / +30 seconds button (untranslated)</td>
<td>Every other button at the microwave (PT-EU) / I / +30 second button</td>
<td></td>
</tr>
<tr>
<td><strong>Gloss Translation</strong></td>
<td>I should not be drinking. / Me drinkin’</td>
<td>I should not be drinking. / I’m drinking.</td>
<td>I should not be drinking. / I am drinking.</td>
<td></td>
</tr>
<tr>
<td>8 – <em>Confession Bear – People Skills</em></td>
<td>Contratado por minhas &quot;habilidades pessoais&quot; / Começando a odiar &quot;pessoas&quot;</td>
<td>Contratado para minhas &quot;habilidades de pessoas&quot; / Começando a odiar &quot;pessoas&quot;</td>
<td>Contratado para minhas &quot;habilidades pessoais&quot; / Começando a odiar as &quot;pessoas&quot;</td>
<td></td>
</tr>
<tr>
<td><strong>Gloss Translation</strong></td>
<td>Hired for my “personal abilities” / Starting to hate “people”</td>
<td>Hired to my “abilities of people” / Starting to hate “people”</td>
<td>Hired to my “personal abilities” / Starting to hate the “people”</td>
<td></td>
</tr>
<tr>
<td>9 – <em>Confession Bear – Pandemics</em></td>
<td>Desde o início da pandemia, as pessoas reclamam do quanto sentem falta de tocar nas pessoas. / Eu, pessoalmente, tenho realmente gostado de como todos estão menos melindrosos agora.</td>
<td>As pessoas têm reclamado desde que a pandemia começou o quanto sentem falta de poder tocar as pessoas / Eu pessoalmente tenho realmente gostado de como</td>
<td>As pessoas têm reclamado desde o início da pandemia o quanto sentem falta de poder tocar as pessoas / Eu, pessoalmente, tenho realmente gostado muito de como todos estão</td>
<td></td>
</tr>
</tbody>
</table>

People have been complaining since the pandemic started how much they miss being able to touch people / I personally have actually been really
<table>
<thead>
<tr>
<th>Source Text</th>
<th>Google Translate</th>
<th>Bing Translator</th>
<th>DeepL Translate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – <em>Granddad Frog Joke</em></td>
<td>Uma garotinha pergunta ao avô: &quot;Você faria um barulho de sapo para mim?&quot;</td>
<td>Uma garotinha pergunta ao avô: &quot;Você faria um barulho de sapo para mim?&quot;</td>
<td>&quot;Uma menina pergunta ao seu avô: &quot;Você faria um barulho de sapo para mim?&quot;</td>
</tr>
<tr>
<td>A little girl asks her granddad, &quot;Would you make a frog noise for me?&quot;</td>
<td>O avô, confuso, pergunta: &quot;Por quê?&quot;</td>
<td>O avô, confuso, pergunta: &quot;Por quê?&quot;</td>
<td>O avô, confuso, pergunta: &quot;Por quê?&quot;</td>
</tr>
<tr>
<td>The granddad, confused asks, &quot;Why?&quot;</td>
<td>A menina responde: &quot;Papai diz que, quando você coxar, todos nós vamos para a Disneylândia&quot;</td>
<td>A garotinha responde: &quot;Papai diz que quando você morrer, todos nós vamos para a Disneylândia&quot;</td>
<td>A menina responde: &quot;O pai diz que quando você coxar, vamos todos à Disneylândia&quot;</td>
</tr>
<tr>
<td>The little girl replies, &quot;Dad says when you croak, we are all going to Disneyland&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Selection of Puns**

<table>
<thead>
<tr>
<th>Source Text</th>
<th>Google Translate</th>
<th>Bing Translator</th>
<th>DeepL Translate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gloss Translation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enjoying how less touchy-feely everyone is right now.</td>
<td>todos são menos sensíveis agora.</td>
<td>menos sensíveis ao toque agora.</td>
<td></td>
</tr>
<tr>
<td>The grandfather, confused, asks: “Why?”</td>
<td>The girl answers: “Daddy says that, when you croak, all of us will go to the Disneyland.”</td>
<td>The grandfather, confused, asks: “Why?”</td>
<td>The girl answers: “Daddy says that, when you die, all of us will go to the Disneyland.”</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>The grandfather, confused asks: “Why?”</td>
<td>The girl answers: “Daddy says that, when you croak, all of us will go to the Disneyland.”</td>
<td>The grandfather, confused asks: “Why?”</td>
<td>The girl answers: “Daddy says that, when you die, all of us will go to the Disneyland.”</td>
</tr>
</tbody>
</table>

### 2 – Four-chin Teller

What do you call a really fat psychic?

A four-chin teller

<table>
<thead>
<tr>
<th>O que você chama de médium muito gordo?</th>
<th>Como se chama um vidente gordo?</th>
<th>O que você chama de um médium realmente gordo?</th>
<th>Um caixa de quatro queixos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Um contador de quatro queixo</td>
<td>Um caixa de quatro queixos</td>
<td>Um caixa de quatro moedas</td>
<td></td>
</tr>
</tbody>
</table>

### Gloss Translation

What do you call a very fat medium?

An accountant of four chin

<table>
<thead>
<tr>
<th>What do you call a very fat medium?</th>
<th>How do you call a fat fortune teller?</th>
<th>What do you call a really fat medium?</th>
</tr>
</thead>
<tbody>
<tr>
<td>An accountant of four chin</td>
<td>A box of four chins</td>
<td>A box of four coins</td>
</tr>
</tbody>
</table>

### 3 – Too Many Shifts

Why did the worker at the keyboard manufacturing unit get fired?

For putting in too many shifts

<table>
<thead>
<tr>
<th>Por que o funcionário da unidade de fabricação de teclados foi demitido?</th>
<th>Por que o trabalhador da unidade de fabricação de teclados foi demitido?</th>
<th>Por que o trabalhador da unidade de fabricação de teclados foi demitido?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Por colocar muitos turnos</td>
<td>Por colocar em muitos turnos</td>
<td>Por fazer demasiados turnos</td>
</tr>
</tbody>
</table>

### Gloss Translation

Why did the employee of the keyboard fabrication unit get fired?

Why did the worker of the keyboard fabrication unit get fired?

<table>
<thead>
<tr>
<th>Why did the employee of the keyboard fabrication unit get fired?</th>
<th>Why did the worker of the keyboard fabrication unit get fired?</th>
<th>Why did the worked of the keyboard fabrication unit get fired?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why did the employee of the keyboard fabrication unit get fired?</td>
<td>Why did the worker of the keyboard fabrication unit get fired?</td>
<td>Why did the worked of the keyboard fabrication unit get fired?</td>
</tr>
<tr>
<td>Row</td>
<td>English</td>
<td>Portuguese</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>4</td>
<td><strong>The Guy Who Invented Zero</strong>&lt;br&gt;For putting many shifts</td>
<td>Para o cara que inventou o zero, obrigado por nada.</td>
</tr>
<tr>
<td>5</td>
<td><strong>Restaurant on the Moon</strong>&lt;br&gt;Did you hear about the restaurant on the moon? I heard the food was good, but it had no atmosphere.</td>
<td>Você ouviu sobre o restaurante na lua? Ouvi dizer que a comida era boa, mas não tinha atmosfera.</td>
</tr>
<tr>
<td>6</td>
<td><strong>Irrelephant</strong>&lt;br&gt;Why was Dumbo sad? He felt irrelephant.</td>
<td>Por que Dumbo estava triste? Ele se sentiu irrelephant.</td>
</tr>
</tbody>
</table>