Wordplay location and interpretation with deep learning methods

Hakima ARROUBAT
École Centrale Marseille

Abstract

Today, all professional translators use machine translation as a basis for understanding the general outline of a text. One of the challenges of translation today is to translate humor, which is often considered untranslatable or difficult to translate. Indeed, the transfer of the humorous side, requires a thorough knowledge of the target language and culture. Humor is often done in a language using on puns that are known by the translation difficulty that arises in the structural ambiguity. Currently, there is still a lot to be done for humor translation, just as there is little research today on machine translation of humor. In the present work, I am interested in the automatic translation of humor from French to English and vice versa, using artificial intelligence techniques. First, I am interested in the data provided for this purpose, I have tried to clean them up. The second task consists in classifying the word sets by providing semantic interpretations.

Keywords

wordplay, humour, Jurassic, machine translation, deep learning

1. Introduction

Translating humor is a complex linguistic exercise because it involves many lexical, stylistic, semantic, phonetic and cultural elements. Indeed, humorous phrases often exploit puns, so translating puns from one language to another requires a mix of rigor and creativity. As a result, the translator must sacrifice either conveying the content, abandoning the pun, or maintaining the pun by replacing the image, deviating from the exact meaning.

Work on pun classification has been fruitful. Bruno de Foucault in his book The Linguistic Structures of the Formation of Wordplays was able to classify wordplays, taxonomically, according to their morphological structure and syntax [1]. A wordplay whose game is understood by humans must be 'audio' and/or 'visual', it exploits many linguistic phenomena such as puns; alliteration, assonance, and consonance... On the other hand, two old problems related to pun patterns must be taken into account while classifying these words: the tendency of alliteration to be flat and the lack of adequate criteria for distinction [2].

The translation of puns is possible by adopting different strategies. Let's take an example: "what the hell" which can be translated by: "c'est quoi ce /chantier/bordel". It is an expression which has been deformed and has become: "what the shell", it is clear that the verbal (semi-homogeneous) and visual (semi-homogeneous) proximity of "shell" to "hell" makes the immediately understandable wordplay for English speakers. Two translation proposals are possible:

- **Keep the meaning distorted:**

For example: “what the shell”: "cé quoi cette coquille" by playing on the double meaning of “shell” in French (shell and misprint).

- **Respect the semantics of the source expression:**

For example: "What the hell" translates a distorted translation as "I'm going to scallop the pellets". This translation proposal expresses the same feeling of exasperation of the speaker, while referring to the distortion of the original in a typo [3].

CLEF 2022 – Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy
EMAIL : hakima.arroubat@centrale-casablanca.ma (H. ARROUBAT)

Today, in the digital age, a wide range of machine translation tools are available to everyone to translate, thanks to artificial intelligence technologies and more particularly to NLP techniques. However, among the challenges of translation automation, we mention the need to maintain the relationship between form and content, the need to preserve in translation the relationship between part and whole, characteristic of the original text. To this end, the automatic translation of humor remains one of the challenges of NLP. Thus, it is essential to design a system that can not only translate the humorous sentence, but also retain the funny effect of the comic texts, by rewriting the puns and similar linguistic quirks for the target culture and language. Additionally, building an ideal system for translating puns and humor requires a quality dataset [2].

Before translating the sentence, we must first discover the pun it creates and try to understand its structure so that it will be easier for us to find its synonym in the target word. Thus, in the present work, my task will be to create a model for the detection of puns in humorous sentences and then to try to find a finer interpretation of it.

2. Method

My first task is to classify the puns by indicating for each of them, the following fields: for singular puns propose an Interpretations (explanation). For humorous phrases, it is first necessary to determine the locations of the puns it contains and then to propose a possible interpretation.

I tried to accomplish this task with more powerful models, including deep learning transformer-based models such as the Jurassic.

2.1. Data Set

We have two databases, one in English and one in French. Both datasets are divided into two categories:

- wordplays based on sentences (wordplay)
- wordplays based on terms (mostly named entities)

I worked with the English dataset, since the French dataset must be cleaned and preprocessed first. Then even patterns will be applied to it. English puns are a collection of offline and online jokes, it contains 3387 standalone English puns with length between 2 and 69 words.

Wordplays instances are stored in a database in Json (or CSV) format with the following fields:
ID: unique pun identifier.
WORDPLAY: singular puns or humorous phrases contain puns
LOCATION: the game of detecting words from the sentence
INTERPRETATION: pun explanation.
HORIZONTAL/VERTICAL: indicate whether the source and the target of the pun are co-present.
MANIPULATION_TYPE: a classification of puns according to similarity, coincidence of source versus target or also Permutation or Abbreviation.
MANIPULATION_LEVEL: a clear sonic similarity or identity.
CULTURAL_REFERENCE: (extra-linguistic factors) takes a binary true or false value.
CONVENTIONAL_FORM: This is a binary category (True/False).
OFFENSIVE (category not rated): some puns are marked as offensive. [4]

The graph in Figure 1 shows the frequency of locations of puns in sentences, we notice that we tend to make puns appear at the end of our jokes. We will see that this fact can have a negative effect on the wordplay classification model, because the built model will also tend to detect the location of wordplay at the end of the sentence, whereas it is not always true.

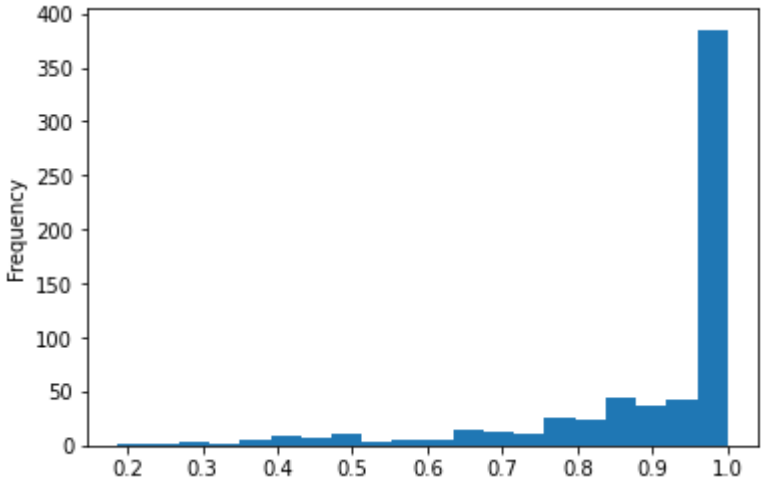

Figure 1: Frequency of locations of puns in sentences

2.2. Jurassic

Jurassic-1 is an autoregressive language model trained on English corpora of 178 B parameters, designed by 10,000 developers, it is a set of reference models, inspired by Open Ai's pioneering work on GPT-3[5]. It allows you to perform huge tasks in NLP such as (summarization, rewriting, text generation, etc.) with great precision. Generally, it can predict text from a wider set of domains (web, academic, legal…) than GPT-3[5].

Jurassic-1 has a complicated architecture which generally differs from that of GPT-3, particularly in terms of the size of the vocabulary and the depth/width ratio of the neural network. Its structure is made up of pairs of auto-regressive language models, from J1-Jumbo with 178B, as well as J1-Large, with parameters 7B. The models use a more efficient architecture and tokenizer, which greatly speeds up the inference. It is based on the decoder module of the Transformer architecture with modifications proposed by Radford et al [6].

The model presents significant advantages either in terms of execution time (speeding up 1.5% per iteration) compared to GPT-3, but also in terms of execution gains in batch inference (7%) and text generation (up to 23%) [5].

3. Result and evaluation

I divided the data into data containing only singular puns (about 32% of the data set) and data containing sentences with puns. I choose among them those that I can give to my model as training data while maintaining the diversity of the examples chosen (about 13% of the data set) and I leave the rest for the testing and evaluation of my model. For the first part of the data containing only the puns, I give as input to my model only the puns and I let it predict the interpretation, on the other hand for the second part I give it the sentence and it predict determine location of puns as well as their interpretations. Finally, I obtained a moderately good match between the predictions and the real interpretations.

Table 1
Interpretations predicted by Jurassic on singular puns on training data

Wordplay	Interpretation	interpretation predicted by Jurassic
Golett	golem + -lett	Gollum (from The Lord of the Rings)
Aegislash	aegis + slash	Aegislash (a character from Final Fantasy)
Brionne	brio + sissonne	Brionne (a character from The Three Musketeers)

We can also see from **Table 1** that Jurassic has well detected the wordplay and has even proposed a definition for it, however, it does not correspond to the result we want. For example, Brionne, we know well that it is a character of the musketeers but we try to understand the meaning of word Brionne.

Table 2
Interpretations predicted by Jurassic on singular puns on test data

Wordplay	Interpretation	interpretation predicted by Jurassic	interpretation predicted by Jurassic better than proposed Interpretation
Peruvian Vipertooth	Viper + tooth	Peruvian viper + tooth	Yes
Squib	squib (damp firecracker)	"squid" = a slang term for "scum"	Yes
Gladrag’s Wizardwear	glad + rags	"glad rags" = fancy clothes	Yes

Other than the training data I have, there is also test data for which I tentatively don’t have the true interpretation values, these are about 148 data with singular word sets. Trying to evaluate the results already given by the Jurassic model, we notice that sometimes Jurassic gives better interpretations than those proposed by the data collectors (**Table 2**). Moreover, out of these 50 manually evaluated examples, 25 (i.e., 50%) of them are well interpreted and can be considered as acceptable interpretations.

Table 3
interpretations predicted by Jurassic on the sentences containing the puns

Wordplay	Location	Location predicted by Jurassic	Interpretation	interpretation predicted by Jurassic
What did the Red light say to the Green light? Don't look at me I'm changing!	changing	changed	change (become different) / change (undress and dress)	look at me / look at me!
"This is where I keep my arrows,"said Tom quiveringly.	Quiveringly	Quiver	quiver (shake) / shake (pouch for arrows)	Arrows (weapons)
An electric company is always looking for high energy employees.	energy	high energy	energy (vitality) / energy (electrical power)	company/employees

We can see from **Table 3** that even if the Jurassic does not give us identical results to the real locations, it does detect the word location in the sentence. On the other hand, he is mistaken in giving a significant interpretation.

4. Conclusion

By way of conclusion, the model developed by Jurassic gives us generally good results. The challenge of this model consists in choosing the prompt (of the training data) to feed it, and it is necessary to ensure their diversity. Another equally important problem in this sense is the evaluation of these models. Indeed, a classical evaluation which consists in smoothing the machine to compare the real result and the interpretation result obtained by the model laterally word by word without taking into account the meaning of the word, is not sufficient and is generally false which requires a human intervention. Otherwise, it is necessary to create an intelligent evaluation model based on artificial intelligence solutions.

The next step in this project is to build models for the translation of these singular puns before passing the entire humorous sentences, keeping in mind that a good translation is one that finds an equivalent in the target language but with keeping this humorous effect.

5. References

[1] Foucault, B. d. " Les structures linguistiques de la genèse des jeux de mots." (1988). https://hal-univ-paris13.archives-ouvertes.fr/hal-01366700/document

[2] Ermakova, Liana and Miller, Tristan and Regattin, Fabio and Bosser, Anne-Gwenn and Mathurin, Élise and Corre, Gaelle Le and Araújo, Sílvia and Boccou, Julien and Digue, Albin and Damoy, Aurianne and Campen, Paul and Jeanjean, Benoît. "Overview of JOKER@CLEF 2022: Automatic Wordplay and Humour Translation workshop" (2022).

[3] Michel, DELARCHE. "Stratégies de traduction des jeux de mots" (2022). https://blogs.mediapart.fr/michel-delarche/blog/170622/strategies-de-traduction-des-jeux-de-mots

[4] Ermakova, L., Miller, T., Puchalski, O., Regattin, F., Mathurin, É., Araújo, S., Bosser, A.-G., Borg, C., Bokiniec, M., Corre, G. L., Jeanjean, B., Hannachi, R., Mallia, Ġ., Matas, G., & Saki, M. "Le jeu de mots et l'humour. " CLEF Workshop JOKER: Automatic Wordplay and Humour Translation. " (2022). https://www.joker-project.com/clef-2022/EN/project

[5] Lieber, Opher, et al. "Jurassic-1: Technical details and evaluation." White Paper. AI21 Labs (2021).https://uploads-ssl.webflow.com/60fd4503684b466578c0d307/61138924626a6981ee09caf6_jurassic_tech_paper.pdf

[6] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. "Language models are unsupervised multitask learners". (2019).