How good can an automatic translation of Pokémon names be?

Léa Talec-Bernard[a]

[a] *University of Western Brittany (UBO), 20 Duquesne Street, Brest, 29490, France*

Abstract

For those of you who are not familiar with the successful franchise of Pokémon, it revolves around imaginary creatures often representing animals mixed with objects, plants, etc. Their names reflect these characteristics and are, most of the time, wordplays. For this paper some Pokémon names were automatically translated by using the translation model T5 with the use of Python. The aim of this paper is to evaluate the overall quality of such translation. The results were very diverse, most source sentences stayed unchanged but some others were surprising and interesting to point out for diverse reasons.

Keywords

Automatic translation; Wordplay; Humour; Pokémon names

1. Introduction

For those of you who are not familiar with the successful franchise of Pokémon, it revolves around imaginary creatures often representing animals mixed with objects, plants, etc. Their names reflect these characteristics and are, most of the time, wordplays. For this paper the T5 model [1] was used to automatically translate short puns, many of which were Pokémon names.

Wordplays and puns are amongst the trickiest things to translate as they often revolve on linguistic and cultural aspects specific to the source language. The translator has to make sure that the translation is understandable to the target language all the while keeping its humoristic value. [2] Therefore, as one can imagine, automatically translating puns can prove very difficult as machines lack the human discernment of humor. This paper aims at analyzing the quality of automatic translations made with the use of the T5 model.

The JOKER projet [3] aims at unifying the scientific community working on the automatic translation of humor and puns. The Handbook of Translation Studies [4] also features a section about humor in translation in which the author addresses the challenges faced with translating humor.

2. Task

The task performed in order to analyze the use of the T5 model in translating humor is the JOKER lab [5] task 2 named "Translate a given punning construction in a proper noun or a neologism from English into French.". This lab was organized as a part of the CLEF-2022 conference. The goal of this task is to automatically translate single-word or short puns from English to French, which is particularly tricky due to the lack of context surrounding each word for the model to use. The data includes 1,161 examples of names of videogames or comics characters containing puns. Many of these names being Pokémon names, this paper will focus on these examples.

3. Methodology description

The Google T5 Model was executed with the use of Python to perform these tasks with the help of the Simple T5 library[1]. This model is one of the most performant models existing today. It allows its users to perform many different NLP (Natural Language Processing) tasks which include translating, summarizing or simplifying documents amongst others. Its performance comes from the fact that it is pre-trained with a very diverse corpus called C4 (Colossal Clean Crawled Corpus) which gathers data from the Common Crawl, an open corpus created from many documents found on the net especially for projects requiring a lot of data. The data from the C4 corpus has been filtered in order to obtain the best result possible. In addition to this data, it is possible to train the T5 model with our own data similar to the document we want to modify. The user also has the possibility to change some of its parameters, allowing a more personalized result. Some of these parameters include the number of times the model will train on the training data or the creativity of the model.

4. Results

The final results were not submitted to the CLEF organizers and were, therefore, not evaluated by them. Some results were interesting enough to point out.

The Pokémon named "Wartorlte" in English was automatically translated into "Brutadou" in French. The English name is composed of the word "war" and the word "turtle". We can see that the T5 model translated the idea of war by the prefix "brut", which keeps the idea of brutality. However, the "turtle" aspect was lost. One interesting thing to note is the end of the name "Brutadou" which makes it fit well with the universe of Pokémon by sounding similar to some other Pokémon names.

Another clever translation was that of "Morelull", a Pokémon name mixing the words "morel" and "lullaby". This name was translated into "Mélulli". Similarly to the case of "Brutadou", the translation only kept one of the two aspects of the original name. The first half of the name "mél", refers to the French word "mélodie" which we could translate by "melody" and conveys a similar idea to "lullaby" in the original name. The second half of the name also conveys this idea but was not translated into French, instead, the "lull" part was kept but and "i" was added at the end, making the name sound more like French Pokémon name.

The T5 model was however not as successful for every translation. In fact, in many cases, the names were not translated at all. In some other cases, the model misinterpreted some aspects of a name. For example, the Pokémon named "Zangoose" was translated into "Gazouille" a French word evoking the song of a bird. The model surely interpreted the end of the name "zangoose" as the bird "goose", in which case the translation is clever, however, "Zangoose" is a pun using the words "zankon" ("Scar" in Japanese) and "mangoose" instead of "goose".

5. Conclusion

Although the model T5 often did not translate the names it was given to translate, it still created some very interesting translations. Some of which could be used as such, some other could spark hints of better ideas in the corrector. With some tweaking and improvement, the T5 model could become, in the future, a great tool in helping translators with their work.

6. Acknowledgements

---

[1] https://github.com/Shivanandroy/simpleT5

[Tapez ici]

from the University of Western Brittany, for her precious help in setting up the T5 model, for mentioning this CLEF event as well as for hosting the SEA EU. I would also like to thank my classmates Nina Španović (University of Split, Croatia), Julliette Le Berrigot (University of Western Brittany, France) and Mikołaj Bondaryk (University of Gdańsk, Poland) for their collaboration on this project during the SEA EU class.

7. References

[1] Delia Chiaro, Translation, Humour and Literature, Volume 1, Continuum Advances in Translation, 2010.

[2] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J. Liu, (2019), Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer

[3] CLEF Workshop JOKER: Automatic Wordplay and Humour Translation, Liana Ermakova, Tristan Miller, Orlane Puchalski, Fabio Regattin, Élise Mathurin, Sílvia Araújo, Anne-Gwenn Bosser, Claudine Borg, Monika Bokiniec, Gaelle Le Corre, Benoît Jeanjean, Radia Hannachi, Ġorġ Mallia, Gordan Matas, and Mohamed Saki, 2022

[4] Jeroen Vandaele, Humor in translation, in: Handbook of Translation Studies Volume 1, Edited by Yves Gambier and Luc van Doorslaer (p.157-162)

[5] Ermakova, L., Miller, T., Regattin, F., Bosser, A.-G., Mathurin, É., Corre, G. L., Araújo, S., Boccou, J., Digue, A., Damoy, A., Campen, P., & Jeanjean, B. (2022). Overview of JOKER@CLEF 2022: Automatic Wordplay and Humour Translation workshop. In A. Barrón-Cedeño, G. Da San Martino, M. Degli Esposti, F. Sebastiani, C. Macdonald, G. Pasi, A. Hanbury, M. Potthast, G. Faggioli, & N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022) (p. 25).

[Tapez ici]