

Taking Things Personally: Third Person to First Person Rephrasing

Marcel Granero-Moya

Amazon Alexa AI

Cambridge, UK

marcel.graneromoya@gmail.com

Agis Oikonomou Filandras

Amazon Alexa AI

Cambridge, UK

filandp@amazon.co.uk

Abstract

The recent advancement of digital assistant technologies has opened new possibilities in the experiences they can provide. One of them is the ability to converse with a persona, e.g., celebrities, famous fictional characters, etc. This experience requires that the replies are answered from the point of view of the persona, i.e., the first person. Since the facts about characters are typically found expressed in the third person, there is a need to rephrase them to the first person in order for the assistant not to break character and the experience to remain immersive. However, the automatic solution to such a problem is largely unexplored by the community. In this work, we present a new task for NLP: third person to first person rephrasing. We define the task and analyze its major challenges. We create and publish a novel dataset with 3493 human-annotated pairs of celebrity facts in the third person with their rephrased sentence in the first person. Moreover, we propose a transformer-based pipeline that correctly rephrases 92.8% of sentences compared to 76.2% rephrased by a rule-based baseline system.

1 Introduction

The emergence of digital assistant technologies has opened new possibilities like creating conversational personas that impersonate celebrities or fictional characters and chat with users. When asked about themselves, these personas can respond with answers created manually by experts or taken straight from knowledge bases, which are usually stored in the third person. In order to not break character, we aim to rephrase automatically these third person replies to first person while preserving the original information.

The task can be thought of as a special case of controlled text rephrasing (Logeswaran et al., 2018; Xu et al., 2020) and text style transfer (Fu et al., 2018; Shen et al., 2017). Similarly to the above,

third to first person rephrasing aims to transform a sentence while preserving the content. In our case instead of trying to change the style of the sentence, e.g. formality (Wang et al., 2019), sentiment (Li et al., 2018; Zhang et al., 2018), or politeness (Madaan et al., 2020), we want to change the point of view of the sentence narrator.

We define the task of Third to First Person Rephrasing. The task’s goal is to rephrase a sentence about someone in the third person point of view to the first person point of view as if that person was talking about themselves. The task has a number of challenges. Firstly, we can have text in quotes that should remain unchanged. Secondly, reported statements should be converted to personal statements, e.g., “He said that he loves ...” → “**I** love ...”. Thirdly, there could be multiple subjects in the sentences, e.g., “Samuel and Ron went to ...” → “Ron and **I** went to ...”. Finally, the point of view should be modified only on the mentions to the target persona. Thus, references and co-references have to be identified and rephrased, while keeping unchanged other third person mentions that refer to another subject. This is a known problem called coreference resolution, that aims to find the same entities present in text with a number of proposed solutions in literature (Clark and Manning, 2015, 2016; Lee et al., 2017; Joshi et al., 2019; Xu and Choi, 2020).

For instance, the sentence about Samuel L. Jackson “*His characters often feature the color purple: Jackson chose to have Doyle Gipson wear a purple hat in Changing Lanes; Mace Windu, upon request by Jackson to George Lucas, wielded a purple lightsaber in Star Wars; and Lazarus, his character in Black Snake Moan, plays a purple Gibson guitar.*” should be rephrased to “**My** characters often feature the color purple: **I** chose to have Doyle Gipson wear a purple hat in Changing Lanes; Mace Windu, upon **my** request to George Lucas, **I** wielded a purple lightsaber in Star Wars;

and Lazarus, *my character in Black Snake Moan, plays a purple Gibson guitar*.”. The rephraser needs to be able to disambiguate between the different subjects in complicated sentences and rephrase appropriately.

The contributions of this work are threefold:

- Definition and analysis of the task. We define the goal of rephrasing third person text to first person in order to impersonate a persona, and we identify the main challenges.
- Creation and distribution of a novel supervised dataset specific for the task. The source sentences are scrapped from a website containing facts from celebrities, and the target sentences are annotated with Amazon Mechanical Turk.
- Proposal of a rephrasing model and exploration of several pre-processing techniques. The proposed pipeline establishes a benchmark for future work.

2 Methodology

In this section, we will explain how we created our dataset for the task and present two solutions to the problem, a rule-based one and a deep learning based one.

2.1 Dataset

We built a supervised dataset with third person sentences rephrased to the first person. The source sentences were collected from TheFactSite¹, a website with celebrities facts like “*J.K. Rowling is 54 years old.*” and “*She was born on July 31st, 1965 in Yate, Gloucestershire.*”. The data was collected filtering out groups and bands.

After collecting the facts, they were paired with their first person equivalents. The pairs were created manually, using Amazon Mechanical Turk. We defined microtasks (See Appendix A) containing an original fact, extracted from TheFactSite, to be rephrased to the first person point of view. The target celebrity and a link to the source webpage were provided to help MTurkers with the annotations. MTurkers were asked to change the target celebrity to a first person point of view, to preserve all the initial content in the rephrased sentence, and keep the word order. The last two instructions aimed to reach higher inter-annotator agreement.

¹www.thefactsite.com/celebrities/

Each task was annotated by three MTurkers from the UK or the US, with more than 98% HIT Approval Rate, and more than 5000 HITs approved.

The results from MTurk were post-processed in three different steps to collect the final annotations. First, direct annotation; whenever at least two MTurkers agreed on the transformation, the answer was assumed to be correct. 3352 facts were directly annotated, from which 2307 had full agreement between annotators and 1045 where only two annotators coincided on the rephrasing. At this point, 359 sentences had no agreement between MTurkers. Answers that were not directly annotated in the previous step were simplified by lower-casing, removing final punctuation, and substituting contractions by expanded forms. Thanks to this process, 73 facts were annotated. The remaining 68 facts were annotated by hand. The result is a supervised dataset that consists of 3493 celebrity facts in the third person and the first person point of view.

The dataset contains facts from 117 different celebrities of which 85 are male and 32 are female. There are 2563 pairs for male celebrities and 930 for females. On average, each celebrity has 29.9 facts and each fact has 16.3 words. Table 1 shows an illustrative example of a data pair. It contains the celebrity and the celebrity’s gender, as well as the source and the target fact.

celebrity	Shawn Mendes	gender	male
source	in some of <i>his</i> videos that <i>he has</i> uploaded to youtube, <i>shawn</i> just <i>sings</i> in them and <i>lets</i> other musicians collaborate with <i>him</i> by playing instruments.		
target	in some of <i>my</i> videos that <i>i have</i> uploaded to youtube, <i>i</i> just <i>sing</i> in them and <i>let</i> other musicians collaborate with <i>me</i> by playing instruments.		

Table 1: Example of a fact containing the celebrity name, its gender, the source sentence and the target. In italics, words that differ between source and target.

The most frequently substituted pairs from the original celebrity fact to the transformed first-person fact as expected are personal pronouns. More specifically, the top 5 substituted pairs are the following: his → my, he → I, her → my, she → I, him → me. As the dataset is unbalanced in gender, over two thirds of the substituted personal pronouns are masculine, and one third are feminine.

2.2 Rephrasing Pipeline

To solve the problem, we investigated two possible alternatives. First a rule-based system and second we used a data-driven deep learning approach.

2.2.1 Rule-Based System

We implemented a rule-based system with dictionaries to rephrase text from third person to first person. First, it spots and tags the subject’s name, then, replaces pronouns based on the subject’s gender, and finally, applies substitution dictionaries with over 100 rules (e.g., “He has”→“I have”, “with <subject>”→“with me”) based on the most common substitutions in the dataset.

2.2.2 Deep Learning System

We used GPT-2, a large transformer-based language model with 1.5 billion parameters, trained on a dataset of 8 million web pages, (Radford et al., 2019). Specifically, we finetuned the ‘gpt2’ model from the Hugging Face open source library, (Wolf et al., 2019), on the collected dataset.

Before feeding our inputs to the model, we used a reference recognizer. It adds a tag before and after any of the target celebrity possible names (name, surname, acronym, nickname). This ensures that the model knows, what is the subject whose point of view it is trying to change. Also, we provide the subject’s gender by prepending a gender token <female> or <male> to help the model disambiguate gendered pronouns. Thirdly, we tried another pre-processing step that dealt with the pronoun resolution. NeuralCoref (Clark and Manning, 2015), a neural network that resolves coreference clusters, was used to identify personal pronouns that refer to the target celebrity. Table 2 illustrates the two pre-processing steps, and shows examples of the input to the model.

Finally, the dataset was augmented to provide more training samples to the model, and to decorrelate celebrity names to specific facts. Thus, leveraging the substituted words between the source and target facts, celebrity names were changed in the source while keeping the rest of the fact. Personal pronouns had to change when the original and the new celebrity differed on gender. An example of data augmentation is depicted in Table 3. Male facts were expanded to 4 new male celebrities and 5 new female celebrities, whereas female celebrity facts were only expanded to 4 new female celebrities due to the ambiguity of the word *her*, which

Celebrity	Billie Eilish
Gender	female
Source	Billie Eilish never smiles in photographs because she says it makes her feel " weak and powerless . "
Reference	<cel> Billie Eilish </cel> never smiles in photographs because she says it makes her feel " weak and powerless . "
Coreferences	<cel> Billie Eilish </cel> never smiles in photographs because <ref> she </ref> says it makes <ref> her </ref> feel " weak and powerless . "
Target	I never smile in photographs because it makes me feel " weak and powerless . "

Table 2: Example of reference recognizer and coreference resolution pre-processing steps. Tags <cel> and <ref> stand for celebrity and coreference respectively.

Original	<male> <cel> Post Malone </cel> ’s first single was released during his illness. <cel> Post </cel> says it was thanks to his father.
Augmented A	<male> <cel> Nicholas Hoult </cel> ’s first single was released during his illness. <cel> Nicholas </cel> says it was thanks to his father.
Augmented B	<female> <cel> Ariana Grande </cel> ’s first single was released during her illness. <cel> Ariana </cel> says it was thanks to her father.

Table 3: Example of data augmentation by changing a male’s fact to different male and female celebrities.

can be a personal pronoun replaced by *him* or a possessive replaced by *his*.

3 Experiments & Results

The dataset (see Section 2.1) was split 80% for training, 10% for evaluation and 10% for test. As the number of edits we wanted the model to do in a sentence were few, we evaluated results with a sentence accuracy metric where a rephrased sentence is correct if it perfectly matches the target sentence, and wrong otherwise. Table 5 shows the correct rephrasing for a fact and a possible wrong rephrasing.

Initially, we tried the rule-based system, as discussed in Section 2.2, model #0. We found it got 76.2% accuracy in the test set by handling most straightforward cases, i.e., when the subject was attached to the verb and when all pronouns and possessives referred to the celebrity. Also, it had problems with homographs, e.g., “her” can either be a pronoun or a possessive, so it can be replaced by “me” or “my” respectively.

We investigated using the GPT-2 model to solve

the task. Initially we trained the model without any preprocessing and by finetuning on the non-augmented dataset, model #1. We saw an increase in accuracy by 5.2%. The model could disambiguate between homographs without requiring any explicit input. It still had trouble though when there were multiple possible subjects.

We hypothesized that by providing the model with gender and the name of the subject we wanted it to rephrase, we would see an improvement. Thus, we first prepended a gender token (<female> or <male>) to the source text, model #2, and then, added the reference recognizer pre-processing step as described in section 2.2, model #3. Compared to GPT-2 without any preprocessing step, the accuracy was raised by 6.1% and 8.6% respectively. Our following hypothesis was that by using a coreferencing model, we would be able to help the model with cases where the pronouns were ambiguous, and it wasn’t editing properly the text. By adding the coreference preprocessing step to the GPT-2 along with the reference recognizer and specified gender, model #4 decreased the accuracy compared to model #3. We surmised, this tiny accuracy drop was due to the limited amount of data. We did data augmentation, as discussed in section 2.2, to provide the model with more training examples of those ambiguous cases. By facing multiple similar examples with different subject names and pronouns, the model would have to take into account the subject name and the coreferences provided by the preprocessing steps. Data augmentation worked, models #5 and #6, resulting in an accuracy increase compared to model #3. It is very interesting to note that the deep learning model can achieve 91.4% high accuracy without using any dependency parsing and without the coreferencing model. The best model (#6) reached 92.8% of accuracy by leveraging all the preprocessing techniques. Gender-wise, model (#6) achieves 93.2% accuracy for masculine subjects and 92.0% for feminine ones. Table 4 shows a comparison of all our hypotheses and their results (see appendix C for further finetuned models).

We analyzed our best model’s errors (#7) and categorized them. 24% of mistakes happened due to the removal or addition of an unrelated context words (e.g., predicted “My dog” instead of “Brady, my dog”). Regarding reference disambiguation, other subjects were rephrased 20% of times (“I am Drake’s only child.” instead of “He is my only

Model	Gender	Ref	Coref	Aug	Accuracy
#0					76.2%
#1					81.4%
#2	X				87.5%
#4	X	X	X		90.0%
#3	X	X			90.9%
#5	X	X		X	91.4%
#6	X	X	X	X	92.8%

Table 4: Accuracy results for the baseline (#0) and the different GPT2 models finetuned (#1-#6). An X is marked if the gender token is provided (Gen), reference recognizer used (Ref), coreferences annotated (Coref), and augmented dataset used for training (Aug).

Source	Post credits his father for introducing him to diverse music genres.
Correct prediction	I credit my father for introducing me to diverse music genres.
Wrong prediction	I credits my father for introducing me to diverse music genres.

Table 5: Example of correct prediction and a possible wrong prediction for a given source fact.

child.”), and 23% of times the reference recognition failed because the subject was not identified by the reference recognizer or not rephrased by the model (“PewDiePie is . . .” instead of “I am . . .”), or the subject was referred in plural along with other subjects (“helped me” instead of “helped us”), or the subject was not referred by its name (e.g., “the player” or “the actor”). Finally, there were two kinds of mistakes that could be counted as correct. In 20% of the cases, the subject reported its own statement and the report was omitted (“I like chocolate” instead of “I said I like chocolate”); and 6.8% for synonyms generated by the model (“nearly” instead of “almost”).

4 Conclusions

This work presented a novel NLP task: third person to first person rephrasing. We explored the challenges of the task, when impersonating a celebrity talking about themselves. We presented a supervised corpus of celebrity facts in the first and third person. We also proposed fine-tuning GPT-2 using the dataset and explored a number of techniques that helped the model achieve 92.8% of accuracy in the test split of our dataset, a 16.6% improvement compared to the rule-based baseline. Finally, we leave for future work the exploration of further refinements to our pipeline, extending our solution to other languages and augmenting the dataset.

References

- Kevin Clark and Christopher D Manning. 2015. Entity-centric coreference resolution with model stacking. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1405–1415.
- Kevin Clark and Christopher D Manning. 2016. Deep reinforcement learning for mention-ranking coreference models. *arXiv preprint arXiv:1609.08667*.
- Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. 2018. Style transfer in text: Exploration and evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.
- Mandar Joshi, Omer Levy, Daniel S Weld, and Luke Zettlemoyer. 2019. Bert for coreference resolution: Baselines and analysis. *arXiv preprint arXiv:1908.09091*.
- Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. *arXiv preprint arXiv:1707.07045*.
- Juncen Li, Robin Jia, He He, and Percy Liang. 2018. Delete, retrieve, generate: A simple approach to sentiment and style transfer. *arXiv preprint arXiv:1804.06437*.
- Lajanugen Logeswaran, Honglak Lee, and Samy Bengio. 2018. Content preserving text generation with attribute controls. *arXiv preprint arXiv:1811.01135*.
- Aman Madaan, Amrith Setlur, Tanmay Parekh, Barnabas Póczos, Graham Neubig, Yiming Yang, Ruslan Salakhutdinov, Alan W Black, and Shrimai Prabhumoye. 2020. Politeness transfer: A tag and generate approach. *arXiv preprint arXiv:2004.14257*.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. *arXiv preprint arXiv:1705.09655*.
- Yunli Wang, Yu Wu, Lili Mou, Zhoujun Li, and Wenhao Chao. 2019. Harnessing pre-trained neural networks with rules for formality style transfer. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3564–3569.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface’s transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.
- Lei Xu, Ivan Ramirez, and Kalyan Veeramachaneni. 2020. Rewriting meaningful sentences via conditional bert sampling and an application on fooling text classifiers. *arXiv preprint arXiv:2010.11869*.
- Liyan Xu and Jinho D Choi. 2020. Revealing the myth of higher-order inference in coreference resolution. *arXiv preprint arXiv:2009.12013*.
- Yi Zhang, Jingjing Xu, Pengcheng Yang, and Xu Sun. 2018. Learning sentiment memories for sentiment modification without parallel data. *arXiv preprint arXiv:1808.07311*.

A Data Collection Instructions with MTurk

The dataset was collected via MTurk, a crowdsourcing platform owned by Amazon. In order to help the annotator change the target celebrity from third person to first person, the celebrity name and a link to the celebrity facts webpage were provided. Moreover, two constraints were specified; first, to keep all the content present in the rephrased sentence and ensure no information was deleted. Although the annotations would lack diversity of written form, we specified to keep the same word order in an attempt to reach more inter-annotator agreement which would help to choose the ground truth rephrased sentence.

On the platform, there was a button to open a window with the instructions and illustrative examples showing correct and incorrect annotations. We specified that all HITs (Human Intelligence Tasks) would be reviewed to ensure proper annotations from MTurkers.

MTurk HIT
<p>Instructions</p> <p>Rephrase a celebrity's fact (3rd person) into a personal sentence (1st person).</p> <p>Remember:</p> <ul style="list-style-type: none">• The celebrity must be changed to a 1st person point of view• All the content must be present in the rephrased sentence.• Keep the same word order
<p>ALL HITS WILL BE REVIEWED</p> <p>Example:</p> <p>Facts from: j k rowling facts</p> <p>Celebrity's fact: J.K. Rowling named her daughter, Jessica, after Jessica Mitford who was an English author and one of J.K.s heros.</p> <p>Rephrased sentence: I named my daughter, Jessica, after Jessica Mitford who was an English author and one of my heros.</p>
<p>Your task</p> <p>Rephrase the following fact (3rd person) into a personal sentence (1st person).</p> <p>Fact from: niall horan facts</p> <p>Celebrity's fact: His guitar was the best Christmas present he ever received.</p> <p>Rephrased sentence:</p> <p><Text box for the answer></p> <p><Submit button></p>

Figure 1: MTurk microtask

B Dataset’s Most Frequently Substituted Pairs

In this section, we provide further details about the substituted pairs from the original dataset. Table 6 illustrates the 20 most substituted pairs from the original dataset. Note that masculine pronouns are twice as frequent as feminine pronouns, evidencing again the gender bias. Figure 2 shows the log-log distribution for the substituted pairs.

#	original	transformed	counts
1	his	my	1039
2	he	i	966
3	her	my	464
4	she	i	414
5	he is	i am	138
6	him	me	136
7	he has	i have	123
8	they	we	65
9	she is	i am	54
10	her	me	52
11	she has	i have	51
12	justin	i	35
13	their	our	33
14	has	have	30
15	himself	myself	25
16	he does	i do	21
17	he loves	i love	19
18	he ’s	i ’ve	18
19	shawn	i	17
20	has	i have	16

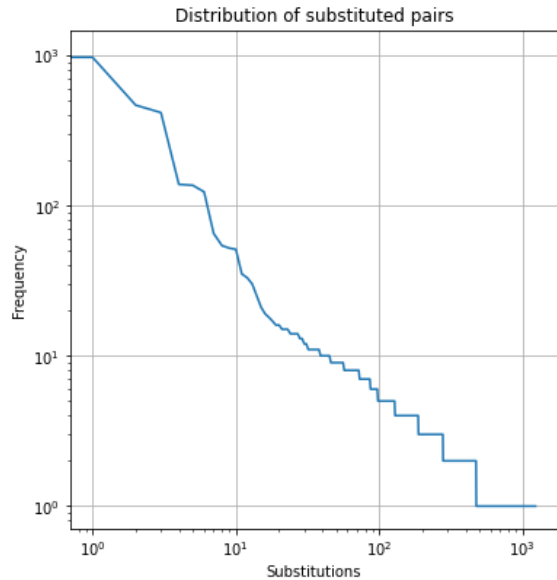


Table 6: Most substituted pairs in the original (non-augmented) dataset.

Figure 2: Log-log distribution of substituted pairs for the original dataset

C Further Model Results

We provide a table with the results for all the models. First, the baseline model (#0), and second, the finetuned GPT2 models which differ on the preprocessing steps applied. The models with a number are explained in section 3. The rest of models provide a wider understanding on how the preprocessing steps help the GPT2 model to rephrase from third to first person. Notice that the coreference resolution has to be used along with the reference recognizer.

Model	Gender token	Reference recognizer	Coreference resolution	Data augmentation	Accuracy
#0					76.2%
#1					81.4%
-				X	82.8%
-		X	X		85.3%
-		X			85.9%
-		X		X	86.1%
-	X			X	86.4%
-		X	X	X	87.0%
#2	X				87.5%
#4	X	X	X		90.0%
#3	X	X			90.9%
#5	X	X		X	91.4%
#6	X	X	X	X	92.8%

Table 7: Accuracy results for the rule-based baseline (#0) and the different finetuned GPT2 models. The GPT2 models differ on the preprocessing steps that were applied to the data before finetuning the GPT2 model. An X means that the column’s preprocessing step was used for that model’s finetuning.