BehancePR: A Punctuation Restoration Dataset for Livestreaming Video Transcripts

Viet Dac Lai[1], **Amir Pouran Ben Veyseh**[1], **Franck Dernoncourt**[2], **and Thien Huu Nguyen**[1]

[1] Dept. of Computer and Information Science, University of Oregon, Eugene, OR, USA
[2] Adobe Research, Seattle, WA, USA
{vietl,apouranb,thien}@cs.uoregon.edu
franck.dernoncourt@adobe.com

Abstract

Given the increasing number of livestreaming videos, automatic speech recognition and post-processing for livestreaming video transcripts are crucial for efficient data management as well as knowledge mining. A key step in this process is punctuation restoration which restores fundamental text structures such as phrase and sentence boundaries from the video transcripts. This work presents a new human-annotated corpus, called BehancePR, for punctuation restoration in livestreaming video transcripts. Our experiments on BehancePR demonstrate the challenges of punctuation restoration for this domain. Furthermore, we show that popular natural language processing toolkits like Stanford Stanza, Spacy, and Trankit underperform on detecting sentence boundary on non-punctuated transcripts of livestreaming videos. The dataset is publicly accessible at http://github.com/nlp-uoregon/behancepr.

1 Introduction

Livestreaming is a powerful broadcasting medium that catches the attention of millions of users. Many video-sharing platforms have supported livestreaming for a wide range of topics such as Twitch for gaming, TikTok for short entertainment videos, Behance for visual creative work, and Youtube/Facebook Live accepting any topics. Among these videos, there are a substantially high number of videos that provide useful knowledge with exceptional visual demonstration. To this end, livestreaming videos are becoming a potential knowledge base waiting for being explored.

Mining videos on video/audio format directly is extremely hard and expensive because of their high data load and complexity in processing images and audio signals. Instead, mining video transcripts, transcribed by either human or machine, is much easier with the existing hardware and software. As such, livestreaming videos should be transcribed at high quality to facilitate future data mining research. As video transcription can be done using existing automatic speech recognition (ASR) systems, a reasonable step to improve the quality of transcribed texts for livestreaming videos involves post-processing produces to remove noises and restore correct language structures and texts from ASR-generated texts.

In this paper, we are particularly interested in punctuation restoration (PR) for livestreaming video transcripts. Punctuation restoration is the task to restore fundamental text structures such as sentences and phrases by inserting punctuation marks into non-punctuated text, e.g. text generated by an automatic speech recognition system for livestreaming videos in our paper. Punctuation restoration is an important post-processing step to improve the readability of ASR texts. Moreover, in natural language processing (NLP), PR is even more important as it enables the use of advanced techniques to process texts at the sentence level to achieve optimal performance for various tasks, e.g., part-of-speech tagging and dependency parsing. Prior studies have shown that with proper sentence split and punctuation, a downstream application can tolerate the word error rate of 25% (Alam et al., 2015), which is extremely high compared to the current state-of-the-art ASR. Figure 1 demonstrates how punctuation restoration improves the readability of ASR-generated texts.

In the literature, PR is considered as a subtask of ASR, in which PR annotation is done as part of the ASR datasets such as the AMI (McCowan et al., 2005) and TED corpus (Federico et al., 2012). However, the speech recorded in these audios involves multi-speaker meetings, as in AMI corpus, or single-speaker talks, as in TED corpus. Our work is different from those work as we consider livestreaming videos that feature many distinctive characteristics that are essential to study. In particular, the number of speakers in livestreaming

Findings of the Association for Computational Linguistics: NAACL 2022, pages 1943 - 1951
July 10-15, 2022 ©2022 Association for Computational Linguistics

use the marquee tool to draw a selection around the empty space on one side then hold shift and add the other areas to the selection too go to edit and fill then change the drop down menu to content aware photoshop should automatically generate a completely new background but it might make a couple of small mistakes

Use the marquee tool to draw a selection around the empty space on one side.
Then hold shift and add the other areas to the selection too.
Go to edit and fill, then change the drop down menu to content aware.
Photoshop should automatically generate a completely new background.
But it might make a couple of small mistakes.

Figure 1: Upper: a ASR-generated text in our dataset. Lower: the corresponding punctuated text in our dataset with greater readability.

videos varies greatly, ranging from one to a few main speakers along with up to thousands of audiences. The audiences might participate in question answering and commenting during the whole duration of the video, hence, constantly changing the information flow of the videos. Furthermore, the speech in livestreaming is much more spontaneous than those in the planned meetings of the AMI corpus and the well-scripted talks of the TED corpus.

An issue with the research of punctuation restoration for livestreaming videos is the lack of a human-annotated dataset for model development and evaluation. This is even more critical when livestreaming has become one of the most powerful communication mediums for not only entertainment but also education purpose. To this end, we introduce a new dataset for Behance Livestreaming Video Punctuation Restoration, called **BehancePR**. The dataset is annotated by skilled transcription annotators for 4 types of punctuation markers. Our experiments reveal the challenges of the BehancePR dataset where the performance of current state-of-the-art models for PR on BehancePR lags far behind those existing PR datasets (e.g., the TED dataset). Our further experiments on cross-domain generation for PR shows that models that are trained on a PR dataset of a different speech scenario perform much worse than those trained on BehancePR even with a much larger training set.

2 Data Annotation

Preparation: The livestreaming videos in this work are collected from the public source of `Behance.net`. Behance is an online platform to showcase and discover creative work such as digital drawing, graphic design, and photo/video editing. In those videos, one or a few creators stream their work on graphic design tools in English, covering a wide range of topics such as design theories, graphical ideas, and tutorials to use graphic design tools. In our dataset, we split the videos into shorter clips of 5 minutes. Next, the clips are transcribed by the Microsoft ASR system. The automatically generated transcripts for each clip (called documents) are then presented to the annotators. To prepare for the PR annotation in livestreaming video transcripts, we inherit the set of three most popular markers, i.e. **period**, **comma**, and **question mark** in prior PR datasets (Federico et al., 2012). In addition, as livestreaming videos of creative works involve a lot of emotional expressions (e.g., excitement), we include **exclamation mark** as a new annotation label to better capture strong feelings and emphasis in this area. Our instruction guidline is presented in Appendix B. To accommodate our annotation budget, we randomly select 2,314 transcribed documents for PR annotation.

Annotation: We recruit 8 annotators from the `Upwork.com` crowdsourcing platform. As Upwork allows its freelancers to submit resumes, we can choose the most experienced annotators with prior experience on audio transcribing. A detailed annotation guideline with many examples is provided to train the annotators. We also develop a customized web-based annotation tool that allows the annotators to work most efficiently with the transcripts and annotation. Appendix A shows the interface and description for our designed annotation tool. After self-practicing on the provided guideline and tool, the annotators are further trained by performing actual PR annotation on transcripts of a 2-hour audio from Behance. Feedback is provided to each annotator in this process to improve the quality. After the training process, the 8 annotators first co-annotate 10% of the documents, leading to the inter-annotation Cohen-Kappa agreement score of 0.59 (i.e., a moderate to substantial agreement level). Afterward, the annotators discuss to resolve the conflicts over the annotated data so far. Finally, the remaining documents are distributed to the 8 annotators for separate annotation to produce a final version of our dataset. To facilitate model development and evaluation, we split the dataset into 3 portions for training/development/test data. Table 1 shows detailed statistics and label distribution of our BehancePR dataset.

	Train	Dev	Test
Statistics			
#Documents	2,174	60	80
#Sentences	115,661	2,969	3,986
#Tokens	1,216,439	34,265	44,224
Label distribution			
#Periods	101,228	2,583	3,229
#Commas	126,739	3,291	4,388
#Questions	7,337	175	437
#Exclamations	7,096	211	320

Table 1: Statistics and label distribution of the BehancePR dataset.

Verbal pause
So, *this is what we got for the site map.*
We talked about six of these being virus killers maps triall assets forum another
So, *those will be that.*
So, *then will also have a footer.*
Um, *so this will be the home page going to,* ***um***, *start grab some assets...*
Duplicate words and phrases
Alright, ***alright***, ***alright***
So, but ***there are all set***
All of ***these are all set***
Incomplete syntax
Um so this will be the homepage ***going to hum*** *start grab some assets I guess image that* ***google "survivor"***
Instructional steps
What if I click it open
2 xbox hub
Inspect
Header
1920 by 1080
Copy
I got it
Word error
Oh *definitely just stubbed my toe.*
And in not very fun pain.
How *just making sure there will be no bleeding.*

Table 2: Examples of noisy texts in transcripts of livestreaming videos. Noisy words are highlighted.

3 Dataset Challenges

Compared to existing PR datasets, e.g., TED (Federico et al., 2012), AMI (McCowan et al., 2005), our dataset BehancePR features several unique challenges. First, as BehancePR's documents are obtained from livestreaming video transcripts, they introduce the unique characteristics of spontaneous speech. This is very different from TED talks, in which the talks are heavily scripted beforehand, and AMI meetings, where the talks are also well prepared. As such, livestreaming video transcripts have a much lower cohesion level as they might present sudden changes of topic and incomplete syntax (among others). Besides, they come with much more verbal pause and repetition of words and phrases, which are the results of hesitation and stutter of the speakers, thus causing a new level of challenges for PR models.

Second, as the documents in BehancePR are generated by an ASR system, it is expected that there is a certain number of word errors (e.g., incorrect transcription, missing words) in the texts. As such, word errors can hinder the language understanding ability, and thus PR performance, for the models on BehancePR. Table 2 shows the examples for different types of noisy texts including verbal pauses, duplicate words and phrases, incomplete syntax, instructional steps, and word errors. In the word error examples, as the streamer has just hurt herself, the ASR system cannot detect the word "*Oww*". Instead, it generates "*Oh*" and "*How*". This error is highly adverse as it might turn a declarative sentence into a WH question starting with the word "*How*".

Third, our introduced exclamation mark is a brand new label that are not captured in existing PR datasets, e.g., the TED and AMI datasets where emotion is rarer. To appropriately restore exclamation mark, a PR model needs to encode not only textual content, but also acoustic features such as frequency and strength of excitement. However, as BehancePR and current PR datasets do not provide access to audio features, the new label for exclamation mark will introduce a new challenging dimension that makes BehancePR an unique PR dataset. We also note that future work can extend BehancePR to include audio features to achieve multi-modal PR.

4 Experiments

Supervised Learning: To reveal the complexity of BehancePR, we evaluate the performance of the state-of-the-art (SOTA) model for PR on this dataset. Similar to prior work, we model PR as a sequence labeling task at the token level that aims to assign one of the five punctuation labels (i.e., 4 designed labels and 1 special labels for non-punctuation) to every space in input texts. In particular, we investigate two major SOTA model architectures for PR: a neural-based model with BiLSTM in (Alam et al., 2020), and graphical-based model with Conditional Random Field in (Makhija et al., 2019). We also investigate the recent advances in data augmentation for PR to automatically produce more training data in (Alam et al., 2020). Applying the data augmentation to two SOTA models leads to four possible model combination as presented in Table 3. We fine-tune the hyper-parameters for the models on the development data of BehancePR. We find that the pre-trained language model RoBERTa (large version) (Liu et al., 2019) delivers the best performance among RoBERTA, AlBERT, and bert-large-uncased version of BERT. This confirms prior results by Alam et al. (2020). The texts are split into sequences of 256 word pieces. The best batch size is 64. The selected learning rate is $3e$-5 for the Adam optimizer. We use a single BiLSTM layer with 200 hidden units for the models. The augmentation rate is set to 0.2 following previous research (Alam et al., 2020).

Table 3(a) presents the performance of four models on the development and the test sets of BehancePR. The first observation from the table is the CRF component can improve the performance of BiLSTM when no data augmentation is applied, thus suggesting the effectiveness of capturing dependencies between labels with CRF for PR. We also observe that data augmentation has zero or little contribution to the performance of the models on BehancePR. In all, the best PR performance on BehancePR is achieved when CRF is applied on top of the BiLSTM model (without using data augmentation). Importantly, we find that the performance of current PR models on BehancePR is far behind that on the TED talk dataset (with F1 score of at least 84%) and perfect performance. It thus indicates the more challenging nature of PR on livestreaming video transcripts with BehancePR and calls for further study on this domain.

Domain Adaptation: To understand the domain difference between BehancePR and current PR datasets, we further explore the cross-domain evaluation setting where the models are trained on a different source domain and evaluated on BehancePR as the target domain. In particular, we choose the TED corpus as the source domain as TED talks are monologues, which is closer to the Behance videos. Table 3(b) presents the the out-of-domain performance of the models. It is clear from the table that the performance of all PR models on BehancePR degrades significantly when they are trained on TED talks. This demonstrates the considerable difference between the domains in TED and BehancePR. It also highlights the benefit of the annotated BehancePR dataset to achieve better PR performance for video transcripts.

Model	**Dev**			**Test**		
	P	**R**	**F**	**P**	**R**	**F**
(a) Behance → Behance						
BiLSTM	63.6	63.1	63.4	62.0	61.4	61.7
+aug	**64.8**	62.2	63.5	**63.8**	60.7	62.2
+CRF	62.8	**65.2**	**64.0**	62.2	**63.5**	**62.9**
+CRF+aug	62.8	64.5	63.7	61.1	62.8	62.0
(b) TED → Behance						
BiLSTM	53.5	59.1	56.2	54.6	59.6	57.0
+aug	55.8	58.0	**56.9**	55.7	58.5	57.1
+CRF	52.7	60.4	56.3	53.2	60.3	56.5
+CRF+aug	56.5	57.3	**56.9**	57.0	57.8	**57.4**

Table 3: Model performance of on BehancePR.

Sentence Splitting: We conduct an additional experiment to demonstrate another benefit of BehancePR on evaluation sentence splitting toolkits. In this task, the models need to predict where the sentences end. As such, we transform the BehanceED dataset by removing comma labels and converting the other labels into a single label of sentence ending. We then train the four models in the supervised learning experiment on the transformed training dataset of BehancePR to detect sentence ending label for sentence splitting. The models are then evaluated on the transformed test set. In addition, we examine the performance of existing NLP toolkits for sentence splitting in this new dataset, including Stanza (Qi et al., 2020), SpaCy (Honnibal and Montani, 2017), and Trankit (Nguyen et al., 2021). The performance of the models and toolkits are presented in Table 4.

As can be seen, existing toolkits perform very poorly on this domain, with the highest F-1 score of only 30.9%. One potential reason for this poor performance is that existing toolkits are trained on perfectly punctuated text (Nivre et al., 2016), making them unfit for our text domain with missing punctuation. As such, the models trained on the transformed BehancePR dataset significantly outperform existing toolkits for sentence splitting with substantial gaps. This demonstrates the ability of

Model	P	R	F
Stanza	70.4	1.4	2.8
Trankit	72.1	7.8	14.0
SpaCy	52.1	21.9	30.9
BiLSTM	70.3	75.6	72.8
+aug	71.5	72.8	72.1
+CRF	73.3	72.0	72.6
+CRF+aug	71.6	73.0	72.3

Table 4: Performance for sentence splitting.

the models to effectively encode contextual information to infer sentence ending. It also suggests the importance of training data for even basic tasks such sentence splitting in challenging domains.

5 Related work

Early studies on PR have explored a wide range of features such as lexical, acoustic, prosodic, and their combination (Gravano et al., 2009; Levy et al., 2012; Xu et al., 2014; Che et al., 2016a; Szaszák and Tündik, 2019). Graphical models such as CRF have been widely used for this task (Lu and Ng, 2010; Zhang et al., 2013) before the emerging of neural networks. Recently, a variety of deep neural network architectures have been explored for PR such as LSTM (Gale and Parthasarathy, 2017), convolutional network (Che et al., 2016b), and transformers (Alam et al., 2020). Corpora for PR are usually created as part of ASR datasets in various domains such as meetings (McCowan et al., 2005), TED talks (Federico et al., 2012), audio books (Panayotov et al., 2015), and film subtitles (Tiedemann, 2016). Among these, the TED corpus is widely used as the benchmark corpus for PR. However, livestreaming video transcripts have not been explored for PR in prior work.

6 Conclusion

We present BehancePR, the first dedicated corpus for punctuation restoration for livestreaming video transcripts. BehancePR is manually annotated for 4 markers and present unique challenges for PR. Our experiments with state-of-the-art models show the challenges of PR for livestreaming videos and call for more research effort in this important area.

Ethical Considerations

In this work we present a dataset on the transcripts of a publicly accessible video-streaming platform, i.e., “*Behance*”[1]. Complying with the discussion presented by Benton et al. (2017), research with human subjects information is exempted from the required full Institutional Review Board (IRB) review if the data is already available from public sources or if the identity of the subjects cannot be recovered. However, to protect the identity of the streamer and any other person whose information are shared in the video transcript, we impose extra processing on the transcribed documents before presenting them to annotators and publicly releasing it later. First, in this dataset, we remove username or any other identity-related information of the streamers in the transcripts to prevent disclosing their identity. Moreover, the proposed dataset only provides textual data (i.e., documents), hence the other content of the videos (e.g., images, audios) are not revealed (to annotators or users) to protect human identity. Finally, to reduce the risk of disclosing the information of the people in the transcripts, in the final version of the dataset, we exclude the transcripts that explicitly or implicitly refer to the identify of the target people.

Acknowledgement

This research has been supported by the Army Research Office (ARO) grant W911NF-21-1-0112 and the NSF grant CNS-1747798 to the IUCRC Center for Big Learning. This research is also based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2019-19051600006 under the Better Extraction from Text Towards Enhanced Retrieval (BETTER) Program. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ARO, ODNI, IARPA, the Department of Defense, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations. We thank the anonymous reviewers and Tracy King for their helpful feedback.

[1] www.behance.net

References

Firoj Alam, Bernardo Magnini, and Roberto Zanoli. 2015. Comparing named entity recognition on transcriptions and written texts. In *Harmonization and Development of Resources and Tools for Italian Natural Language Processing within the PARLI Project*, pages 71–89. Springer.

Tanvirul Alam, Akib Khan, and Firoj Alam. 2020. Punctuation restoration using transformer models for high-and low-resource languages. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 132–142, Online. Association for Computational Linguistics.

Adrian Benton, Glen Coppersmith, and Mark Dredze. 2017. Ethical research protocols for social media health research. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 94–102, Valencia, Spain. Association for Computational Linguistics.

Xiaoyin Che, Sheng Luo, Haojin Yang, and Christoph Meinel. 2016a. Sentence boundary detection based on parallel lexical and acoustic models. In *Interspeech*, pages 2528–2532.

Xiaoyin Che, Cheng Wang, Haojin Yang, and Christoph Meinel. 2016b. Punctuation prediction for unsegmented transcript based on word vector. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 654–658.

Marcello Federico, Sebastian Stüker, Luisa Bentivogli, Michael Paul, Mauro Cettolo, Teresa Herrmann, Jan Niehues, and Giovanni Moretti. 2012. The iwslt 2011 evaluation campaign on automatic talk translation. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).

William Gale and Sarangarajan Parthasarathy. 2017. Experiments in character-level neural network models for punctuation. In *INTERSPEECH*, pages 2794–2798.

Agustin Gravano, Martin Jansche, and Michiel Bacchiani. 2009. Restoring punctuation and capitalization in transcribed speech. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4741–4744. IEEE.

Matthew Honnibal and Ines Montani. 2017. spacy 2: Natural language understanding with bloom embeddings. *convolutional neural networks and incremental parsing*, 7(1).

Tal Levy, Vered Silber-Varod, and Ami Moyal. 2012. The effect of pitch, intensity and pause duration in punctuation detection. In *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, pages 1–4. IEEE.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Wei Lu and Hwee Tou Ng. 2010. Better punctuation prediction with dynamic conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 177–186.

Karan Makhija, Thi-Nga Ho, and Eng-Siong Chng. 2019. Transfer learning for punctuation prediction. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 268–273.

Iain McCowan, Jean Carletta, Wessel Kraaij, Simone Ashby, S Bourban, M Flynn, M Guillemot, Thomas Hain, J Kadlec, Vasilis Karaiskos, et al. 2005. The ami meeting corpus. In *Proceedings of the 5th international conference on methods and techniques in behavioral research*, volume 88, page 100. Citeseer.

Minh Van Nguyen, Viet Dac Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen. 2021. Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 80–90, Online. Association for Computational Linguistics.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666, Portorož, Slovenia. European Language Resources Association (ELRA).

Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. 2015. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5206–5210. IEEE.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

György Szaszák and Máté Akos Tündik. 2019. Leveraging a character, word and prosody triplet for an asr error robust and agglutination friendly punctuation approach. In *INTERSPEECH*, pages 2988–2992.

Jörg Tiedemann. 2016. Finding alternative translations in a large corpus of movie subtitle. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3518–3522.

Chenglin Xu, Lei Xie, Guangpu Huang, Xiong Xiao, Eng Siong Chng, and Haizhou Li. 2014. A deep neural network approach for sentence boundary detection in broadcast news. In *Fifteenth annual conference of the international speech communication association*. Citeseer.

Dongdong Zhang, Shuangzhi Wu, Nan Yang, and Mu Li. 2013. Punctuation prediction with transition-based parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 752–760.

A Annotation Tool

We develop a customized web-based annotation tool for this work. The annotation tool focuses on improving the readability of the annotated text, as the result, improves the annotation quality. Toward this end, we use color coding for punctuation markers. More importantly, whenever a sentence ending marker is assigned, such as period, question mark, and exclamation, it automatically creates a new line to separate sentences. Figure 2 shows the interface and an annotated text using our tool.

B Annotation Guideline

This section summarizes the taxonomy, annotation guideline, and annotation examples. Examples are shown in figure 2.

A **period** is used for:

- Marking the end of a **declarative sentence**.
- Separating independent clauses **without** a conjunction when a semi-colon is usually used (to distinguish with the case that a comma is used when a conjunction presents).

A **question mark** is used for:

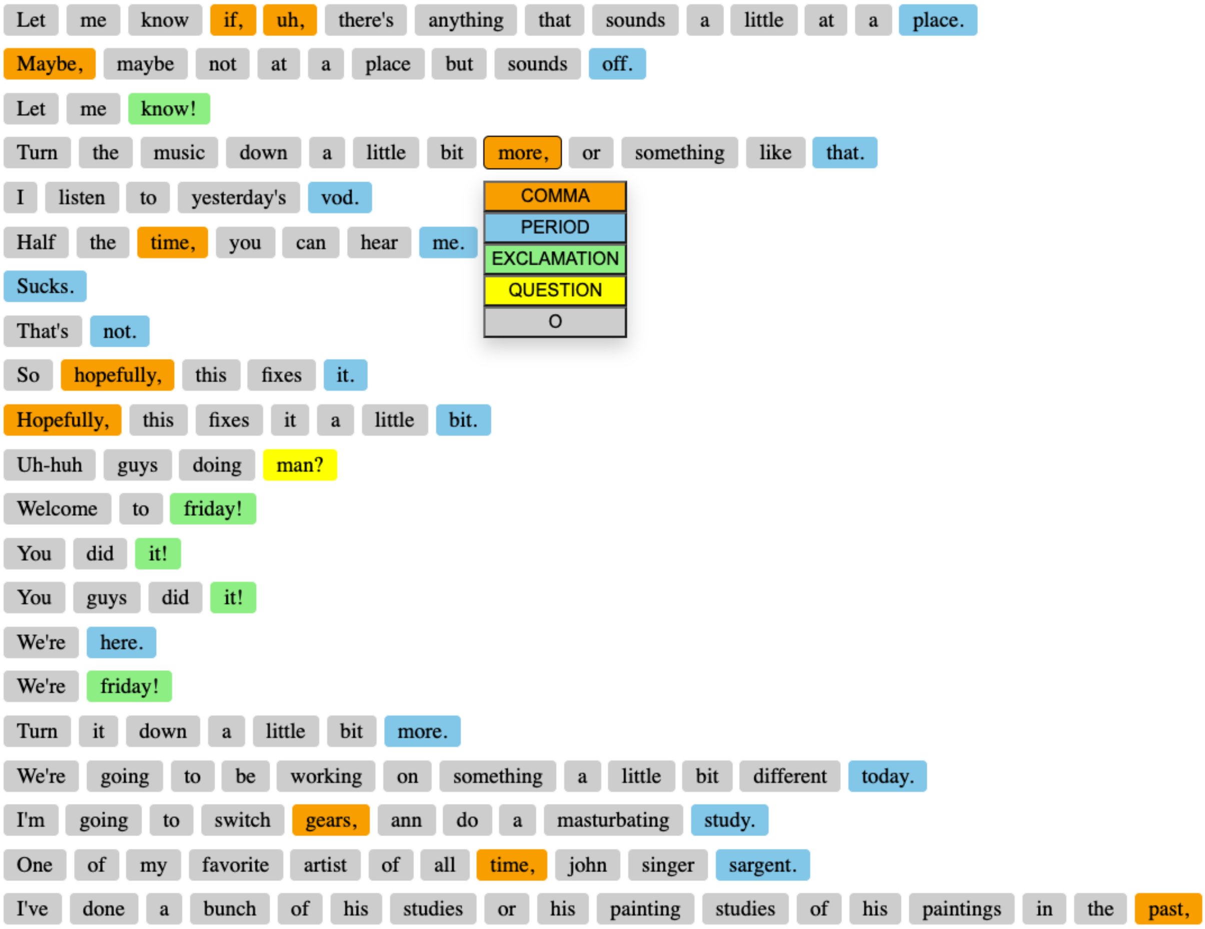
- Marking the end of a **question**.

An **exclamation mark** is used for:

- Exclaiming something. They are commonly used after interjections (words or phrases that are used to exclaim, command, or protest like "wow" or "oh").
- Express the following emotions: excitement, surprise, astonishment, emphasizing a point, and other types of strong emotions.

A **comma** is used for:

- Separating independent clauses when they are joined by any of these seven coordinating conjunctions: and, but, for, or, nor, so, yet.
- Separating introductory clauses, phrases, or words from the main clause.
- Setting off clauses, phrases, and words that are not essential to the meaning of the sentence. Use one comma before to indicate the beginning of the pause and one at the end to indicate the end of the pause.
- Separating three or more words, phrases, or clauses written in a series.
- Separating two or more coordinate adjectives that describe the same noun. Be sure never to add an extra comma between the final adjective and the noun itself or to use commas with non-coordinated adjectives.
- Separating contrasted coordinate elements or to indicate a distinct pause or shift near the end of a sentence.
- Setting off phrases at the end of the sentence that refers back to the beginning or middle of the sentence. Such phrases are free modifiers that can be placed anywhere in the sentence without causing confusion.
- Setting off all geographical names, items in dates (except the month and day), addresses (except the street number and name), and titles in names.
- Shifting between the main discourse and a quotation.
- Preventing possible confusion or misreading.



Let me know if, uh, there's anything that sounds a little at a place.

Maybe, maybe not at a place but sounds off.

Let me know!

Turn the music down a little bit more, or something like that.

I listen to yesterday's vod.

Half the time, you can hear me.

Sucks.

That's not.

So hopefully, this fixes it.

Hopefully, this fixes it a little bit.

Uh-huh guys doing man?

Welcome to friday!

You did it!

You guys did it!

We're here.

We're friday!

Turn it down a little bit more.

We're going to be working on something a little bit different today.

I'm going to switch gears, ann do a masturbating study.

One of my favorite artist of all time, john singer sargent.

I've done a bunch of his studies or his painting studies of his paintings in the past,

Figure 2: The colorful, easy-to-use interface of our annotation tool designed for the annotation of BehancePR. The color codes for comma, period, exclamation mark, and question mark are orange, sky blue, light green, and yellow, respectively.