FanfictionNLP: A Text Processing Pipeline for Fanfiction

Michael Miller Yoder*, **Sopan Khosla***, **Qinlan Shen, Aakanksha Naik, Huiming Jin, Hariharan Muralidharan, Carolyn P. Rosé**
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA, USA
{yoder,sopank,qinlans,anaik,huimingj,hmuralid,cprose}@cs.cmu.edu

Abstract

Fanfiction presents an opportunity as a data source for research in NLP, education, and social science. However, answering specific research questions with this data is difficult, since fanfiction contains more diverse writing styles than formal fiction. We present a text processing pipeline for fanfiction, with a focus on identifying text associated with characters. The pipeline includes modules for character identification and coreference, as well as the attribution of quotes and narration to those characters. Additionally, the pipeline contains a novel approach to character coreference that uses knowledge from quote attribution to resolve pronouns within quotes. For each module, we evaluate the effectiveness of various approaches on 10 annotated fanfiction stories. This pipeline outperforms tools developed for formal fiction on the tasks of character coreference and quote attribution.

1 Introduction

A growing number of natural language processing tools and approaches have been developed for fiction (Agarwal et al., 2013; Bamman et al., 2014; Iyyer et al., 2016; Sims et al., 2019). These tools generally focus on published literary works, such as collections of novels. We present an NLP pipeline for processing fanfiction, amateur writing from fans of TV shows, movies, books, games, and comics.

Fanfiction writers creatively change and expand on plots, settings, and characters from original media, an example of "participatory culture" (Jenkins, 1992; Tosenberger, 2008). The community of fanfiction readers and writers, now largely online, has been studied for its mentorship and support for writers (Evans et al., 2017) and for the broad representation of LGBTQ+ characters and relationships in fan-written stories (Lothian et al., 2007; Dym et al., 2019). Fanfiction presents an opportunity as a data source for research in a variety of fields, from those studying learning in online communities to social science analysis of how community norms develop in an LGBTQ-friendly environment. For NLP researchers, fanfiction provides a large source of literary text with metadata, and has already been used in applications such as authorship attribution (Kestemont et al., 2018) and character relationship classification (Kim and Klinger, 2019).

There is an vast amount of fanfiction in online archives. As of March 2021, over 7 million stories were hosted on just one fanfiction website, Archive of Our Own, and there exist other online archives of similar or even larger sizes (Yin et al., 2017). We present a pipeline that enables structured insight into this vast amount of text by identifying sets of characters in fanfiction stories and attributing narration and quotes to these characters.

Knowing who the characters are and what they do and say is essential for understanding story structure (Bruce, 1981; Wall, 1984). Such processing is also useful for researchers in the humanities and social sciences investigating identification with characters and the representation of characters of diverse genders, sexualities, and ethnicities (Green et al., 2004; Kasunic and Kaufman, 2018; Felski, 2020). The presented pipeline, which extracts text related to characters in fanfiction, can assist researchers building NLP tools for literary domains, as well those analyzing characterization in fields such as digital humanities. For example, the pipeline could be used to explore how characters are voiced and described differently when cast in queer versus straight relationships.

The presented pipeline contains three main modules: character coreference resolution, quote attribution, and extraction of "assertions", narration that relates to particular characters. We incorporate new and existing methods into the pipeline that perform well on an annotated set of 10 fanfiction stories. This includes a novel method using

* Denotes equal contribution.

Proceedings of the 3rd Workshop on Narrative Understanding, pages 13–23

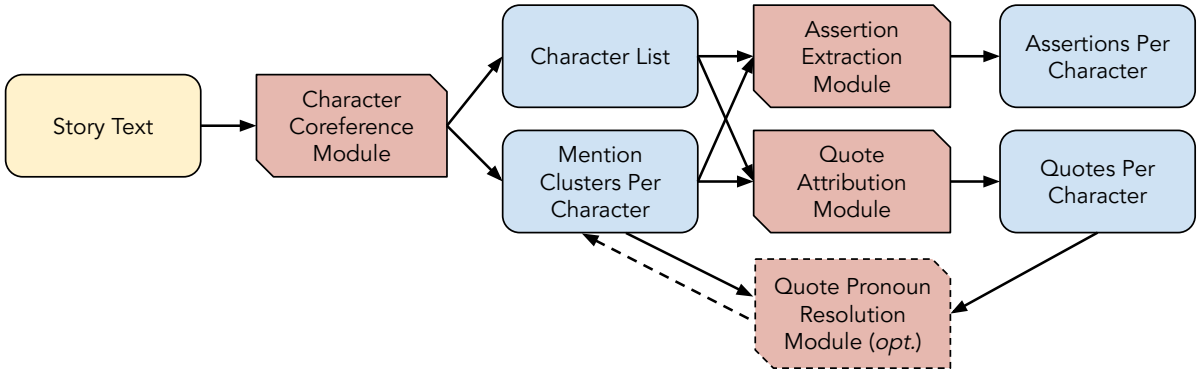

Figure 1: Fanfiction NLP pipeline overview. From the text of a fanfiction story, the pipeline assigns character mentions to character clusters (character coreference). It then attributes assertions and quotes to each character, optionally using the quote attribution output to improve coreference resolution within quotes (see Section 3.3).

quote attribution information to resolve first- and second-person pronouns within quotes.

Fanfiction is written by amateur writers of all ages and education levels worldwide, so it contains much more variety in style and genre than formal fiction. It is not immediately clear that techniques for coreference resolution or quote attribution that perform well on news data or formal fiction will be effective in the informal domain of fanfiction. We demonstrate that this pipeline outperforms existing tools designed for formal fiction on the tasks of character coreference resolution and quote attribution (Bamman et al., 2014).

Contributions. We contribute a fanfiction processing pipeline that outperforms prior work designed for formal fiction. The pipeline includes novel interleaving of coreference and quote attribution to improve the resolution of first- and second-person pronouns within quotes in narrative text. We also introduce an evaluation dataset of 10 fanfiction stories with annotations for character coreference, as well as for quote detection and attribution.

2 Fanfiction and NLP

Data from fanfiction has been used in NLP research for a variety of tasks, including authorship attribution (Kestemont et al., 2018), action prediction (Vilares and Gómez-Rodríguez, 2019), fine-grained entity typing (Chu et al., 2020), and tracing the sources of derivative texts (Shen et al., 2018). Computational work focusing on characterization in fanfiction includes the work of Milli and Bamman (2016), who found that fanfiction writers are more likely to emphasize female and secondary characters. Using data from WattPad, a platform that includes fanfiction along with original fiction, Fast et al. (2016) find that portrayals of gendered characters generally align with mainstream stereotypes.

We are not aware of any text processing system for fanfiction specifically, though BookNLP (Bamman et al., 2014) is commonly used as an NLP system for formal fiction. We evaluate our pipeline's approaches to character coreference resolution and quote attribution against BookNLP, as well as against other task-specific approaches, on an evaluation dataset of fanfiction.

3 Fanfiction Processing Pipeline

We introduce a publicly available pipeline for processing fanfiction.[1] This pipeline is a command-line tool developed in Python. From the text of a fanfiction story, the pipeline extracts a list of characters, each mention of a character, as well as what each character does and says (Figure 1). More specifically, the pipeline first performs character coreference resolution, extracting character mentions and attributing them to character clusters with a single standardized character name (Section 3.1). After coreference, the pipeline outputs quotes uttered by each character using a sieve-based approach from Muzny et al. (2017) (Section 3.2). These quote attribution results are optionally used to aid the resolution of first- and second-person pronouns within quotes to improve coreference output (Section 3.3). In parallel with quote attribution, the pipeline extracts "assertions", topically coherent segments of text that mention a character (Section 3.4).

[1]The pipeline is available at https://github.com/michaelmilleryoder/fanfiction-nlp.

3.1 Character Coreference Module

The story text is first passed through the coreference resolution module, which extracts mentions of characters and attributes them to character clusters. These mentions include alternative forms of names, pronouns, and anaphoric references such as "the bartender". Each cluster is then given a single standardized character name.

Coreference Resolution. We use SpanBERT-base (Joshi et al., 2020), a neural method with state-of-the-art performance on formal text, for coreference resolution. This model uses SpanBERT-base embeddings to create mention representations and employs Lee et al. (2017)'s approach to calculate the coreferent pairs. SpanBERT-base is originally trained on OntoNotes (Pradhan et al., 2012). However, we further fine-tune SpanBERT-base on LitBank (Bamman et al., 2020), a dataset with coreference annotations for works of literature in English, a domain more similar to fanfiction. The model takes the raw story text as input, identifies spans of text that mention characters, and outputs clusters of mentions that refer to the same character.

Character Standardization. We then assign representative character names for each coreference cluster. These names are simply the most frequent capitalized name variant, excluding pronouns and address terms, such as *sir*. If there are no capitalized terms in the cluster or if there are only pronouns and address terms, the most frequent mention is chosen as the name.

Post-processing. SpanBERT-base resolves all entity mentions. In order to focus solely on characters, we post-process the cluster outputs. We remove plural pronouns (*we*, *they*, *us*, *our*, etc.) and noun phrases, demonstrative pronouns (*that*, *this*), as well as *it* mentions. We also remove clusters whose standardized representative names are not named entities and have head words that are not descendants of *person* in WordNet (Miller, 1995). Thus clusters with standardized names such as "the father" are kept (since they are descendants of *person* in WordNet), yet clusters with names such as "his workshop" are removed.

For each character cluster, a standardized name and list of the mentions remaining after post-processing is produced, along with pointers to the position of each mention in the text. This coreference information is then used as input to quote attribution and assertion extraction modules.

3.2 Quote Attribution Module

To extract quotes, we simply extract any spans between quotation marks, a common approach in literary texts (O'Keefe et al., 2012). For the wide variety of fanfiction, we recognize a broader set of quotation marks than are recognized in BookNLP's approach for formal fiction.

The pipeline attributes quotes to characters with the deterministic approach of Muzny et al. (2017), which uses sieves such as looking for character mentions that are the head words of known speech verbs. We use a standalone re-implementation of this approach by Sims and Bamman (2020) that allows using the pipeline's character coreference as input. Muzny et al. (2017)'s approach assigns quotes to character mentions and then to character clusters. We simply assign quotes to the names of these selected character clusters.

3.3 Quote Pronoun Resolution Module

Recent advances in coreference resolution, such as the SpanBERT-base system incorporated in the pipeline, leverage contextualized word embeddings to compute mention representations and to cluster these mentions from pairwise or higher-order comparisons. They also concatenate features such as the distance between the compared mentions to their representations. However, these approaches to not capture the change in point of view caused by quotes within narratives, so they suffer when resolving first- and second-person pronouns within quotes. To alleviate this issue, we introduce an optional step in the pipeline that uses the output from quote attribution to inform the resolution of first- and second-person pronouns within quotes.

Prior work (Almeida et al., 2014) proposed a joint model for entity-level quotation attribution and coreference resolution, exploiting correlations between the two tasks. However, in this work, we propose an interleaved setup that is modular and allows the user of the pipeline to use independent off-the-shelf pre-trained models of their choice for both coreference resolution and quote attribution.

More specifically, once the quote attribution module predicts the position of each quote (q_i) and its associated speaker (s_i), the first-person pronouns within the quote (e.g. *I*, *my*, *mine*, *me*) are resolved to the speaker of that quote, s_i. For second-person pronouns (e.g. *you*, *your*, *yours*), we assume that they point to the addressee of the quote (a_i), which is resolved to be the speaker of the nearest

Fandom	Primary media type(s)
Marvel	Comics, movies
Supernatural	TV show
Harry Potter	Books, movies
DCU	Comics, movies
Sherlock Holmes	Books, TV show
Teen Wolf	TV show
Star Wars	Movies
Doctor Who	TV show
The Lord of the Rings	Books, movies
Dragon Age	Video game

Table 1: The most popular 10 fandoms on Archive of Our Own by number of works, as of September 2018. We annotate 1 story from each fandom to form our test set.

quote before the current quote ($a_i = s_{i-j}$ such that $s_{i-j} \neq s_i$). We only consider the previous 5 quotes to find a_i.

Since there are no sieves for quote attribution that consider pronouns within quotes, the improved coreference within quotes from this optional step does not affect quote attribution. Thus, this "cycle" of character coreference, then quote attribution, then improved character coreference, need only be run once. However, the improved coreference resolution could impact which assertions are associated with characters.

3.4 Assertion Extraction Module

After coreference, the pipeline also extracts what we describe as "assertions", topically coherent segments of text that mention a character. The motivation for this is to identify longer spans of exposition and narrative that relate to characters for building embedding representations for these characters. Parsing these assertions would also facilitate the extraction of descriptive features such as verbs for which characters are subjects and adjectives used to describe characters.

To identify such spans of texts that relate to characters, we first segment the text with a topic segmentation approach called TextTiling (Hearst, 1997). We then assign segments (with quotes removed) to characters if they contain at least one mention of the character within the span. If multiple characters are mentioned, the span is included in extracted assertions for each of the characters.

Evaluation Dataset	
# stories	10
# words	22,283
# character mentions	2,808
# quotes	876

Table 2: Fanfiction evaluation dataset statistics

4 Fanfiction Evaluation Dataset

To evaluate our pipeline, we annotate a dataset of 10 publicly available fanfiction stories for all mentions of characters and quotes attributed to these characters, which is similar in size to the test set used in LitBank (Bamman et al., 2020). We select these stories from Archive of Our Own[2], a large fanfiction archive that is maintained and operated by a fan-centered non-profit organization, the Organization for Transformative Works (Fiesler et al., 2016). To capture a representative range of fanfiction, we choose one story from each of the 10 most popular *fandoms* on Archive of Our Own when we collected data in 2018 (Table 1). *Fandoms* are fan communities organized around a particular original media source. For each fandom, we randomly sampled a story in English that has fewer than 5000 words and does not contain explicit sexual or violent content.

Two of the authors annotated the 10 stories for each of the tasks of character coreference and quote attribution. All annotators were graduate students working in NLP. Statistics on this evaluation dataset and the annotations can be found in Table 2.

These stories illustrate the expanded set of challenges and variety in fanfiction. In one story, all of the characters meet clones of themselves as male if they are female, or female if they are male. This is a variation on the practice of "genderswapping" characters in fanfiction (McClellan, 2014). Coreference systems can struggle to keep up with characters with the same name but different genders. Another story in our test set is a genre of fanfiction called "songfic", which intersperses song lyrics into the narrative. These song lyrics often contain pronouns such as *I* and *you* that do not refer to any character.

For quote attribution, challenges in the test set include a variation of quotation marks, sometimes used inconsistently. There is also great variation in the number of indirect quotes without clear quota-

[2] http://archiveofourown.org/

tives such as “she said”. This can be a source of ambiguity in published fiction as well, but we find a large variety of styles in fanfiction. One fanfiction story in our evaluation dataset, for example, contains many implicit quotes in conversations among three or more characters, which can be difficult for quote attribution.

Annotation details and inter-annotator agreement for this evaluation dataset are described below. An overview of inter-annotator agreement is provided in Table 3.

4.1 Character Coreference Annotation

To annotate character mentions in our evaluation dataset, annotators (two of the authors) were instructed to identify and group all mentions of singular characters, including pronouns, generic phrases that refer to characters such as “the boy", and address terms. Possessive pronouns were also annotated, with nested mentions for phrases such as `<char1><char2>his</char2> sister</char1>`. Determiners and prepositional phrases attached to nouns were annotated, since they can specify characters and contribute to characterization. For an example, `<char1>an old friend of <char2>my</char2> parents</char1>`. Note that “parents" is not annotated in this example since it does not refer to a singular character. Appositives were annotated, while relative clauses (“the woman who sat on the left”) and phrases after copulas (“he was a terrible lawyer”) were not annotated, as we found them to act more as descriptions of characters than mentions.

After extracting character mentions, annotators grouped these mentions into character clusters that refer to the same character in the story. Note that since we focus on characters, we do not annotate other non-person entities usually included in coreference annotations. Full annotation guidelines are available online[3].

To create a unified set of gold annotations, we resolved disagreements between annotators in a second round of annotation. The final test set of 10 annotated stories contains 2,808 annotated character mentions.

In Table 3, we first provide inter-annotator agreement on extracting the same spans of text as character mentions by comparing BIO labeling at the token level. Tokens that begin a mention are labeled B, tokens that are inside or end a mention are labeled I, and all other tokens are labeled O.

	Character Coreference	**Quote Attribution**
Extraction (BIO)	0.95	0.97
Attribution (all)	0.84	0.89
Attribution (agreed)	0.95	0.98

Table 3: Inter-annotator agreement (Cohen’s κ) between two annotators for each task, averaged across 10 fics. Extraction (BIO) is agreement on extracting the same spans of text (not attributing them to characters) with token-level BIO annotation. Attribution (all) refers to attribution of spans to characters where missed spans receive a NULL character attribution. Attribution (agreed) refers to attribution of spans that both annotators marked.

Which mentions are identified affects the agreement of attributing those mentions to characters. For this reason, we provide two attribution agreement scores. First, we calculate agreement on mentions annotated by either annotator, with a NULL character annotation if any annotator did not annotate a mention (Attribution (all) in Table 3). We also calculate agreement only for character mentions annotated by both annotators (Attribution (agreed) in Table 3). Character attribution was labeled as matching if there was significant overlap between primary character names chosen for each cluster by annotators; there were no disagreements on this.

For all these categories, inter-annotator agreement was 0.84 Cohen’s κ or above, “near perfect”, for character coreference (Table 3).

4.2 Quote Attribution Annotation

Two of the authors annotated all quotes that were said aloud or written by a singular character, and attributed them to a list of characters determined from the character coreference annotations. Annotation was designed to focus on characters’ voices as displayed in the stories. Thus characters’ thoughts were not annotated as quotes, nor were imagined or hypothetical utterances. We also chose not to annotate indirectly reported quotes, such as “the friend said I was very strange” since this could be influenced more by the character or narrator reporting the quote than the original character who spoke it. However, we did annotate direct quotes that are reported by other characters.

Inter-annotator agreement on quote attribution

[3] https://github.com/michaelmilleryoder/fanfiction-nlp/blob/main/annotation_guidelines.md

was 0.89 Cohen's κ on the set of all quotes annotated by any annotator (see Table 3). Attribution agreement on the set of quote spans identified by both annotators was very high, 0.98 κ. Token-level BIO agreement for marking spans as quotes was 0.97 κ. The final test set of 10 stories contains 876 annotated quotes.

5 Pipeline Evaluation

We evaluate the pipeline against BookNLP, as well as other state-of-the-art approaches for coreference resolution and quote attribution.

5.1 Character Coreference Evaluation

We evaluate the performance of the character coreference module on our 10 annotated fanfiction stories using the CoNLL metric (Pradhan et al., 2012; the average of MUC, B^3, and CEAFE) and LEA metric (Moosavi and Strube, 2016).

We compare our approach against different state-of-the-art approaches used for coreference resolution in the past. Along with BookNLP's approach, we consider the Stanford CoreNLP deterministic coreference model (CoreNLP (dcoref); Raghunathan et al., 2010; Recasens et al., 2013; Lee et al., 2011) and the CoreNLP statistical model (CoreNLP (coref); Clark and Manning, 2015) as traditional baselines. As a neural baseline, we evaluate the more recently proposed BERT-base model (Joshi et al., 2019), which replaces the original GloVe embeddings (Pennington et al., 2014) with BERT (Devlin et al., 2019) in Lee et al. (2017)'s coreference resolution approach.

Micro-averaged results across the 10 annotated stories are shown in Table 4. The FanfictionNLP approach is SpanBERT-base fine-tuned on LitBank, with the post-hoc removal of non-person and plural mentions and clusters (as described in Section 3.1). Note that these results are without the quote pronoun resolution module described in Section 3.3. Traditional approaches like BookNLP and CoreNLP (dcoref, coref) perform significantly worse than the neural models, especially on recall. Neural models that are further fine-tuned on LitBank (OL) outperform the ones that are only trained on OntoNotes (O). This suggests that further training the model on literary text data does indeed improve its performance on fanfiction narrative. Furthermore, the SpanBERT-base approaches outperform their BERT-base counterparts with an absolute improvement of 4-5 CoNLL F1 percentage points and 6 LEA F1 percentage points. Post-hoc removal of non-person and plural entities improves CoNLL precision on characters by more than 12 percentage points over SpanBERT-base OL.

	CoNLL (Avg.)			LEA
	P	R	F1	F1
BookNLP	67.7	27.4	38.5	28.7
CoreNLP (dcoref)	26.9	49.5	29.6	21.9
CoreNLP (coref)	39.8	47.0	40.5	36.7
BERT-base O	45.8	53.2	49.2	50.9
BERT-base OL	55.0	62.3	58.4	63.1
SpanBERT-base OL	60.3	**71.1**	64.8	69.4
FanfictionNLP	**72.6**	70.1	**71.4**	**73.5**

Table 4: Character coreference performance on CoNLL and LEA metrics. **O**: Model is trained on OntoNotes. **L**: Model is also fine-tuned on LitBank corpus. **FanfictionNLP** is the SpanBERT-base OL model with post-hoc removal of non-person entities. Note that none of the approaches had access to our fanfiction data. These results are without the quote pronoun resolultion module described in Section 3.3.

5.2 Quote Attribution Evaluation

Using our expanded set of quotation marks, we reach 96% recall and 95% precision of extracted quote spans, micro-averaged over the 10 test stories, compared with 25% recall and 55% precision for BookNLP.

For attributing these extracted quotes to characters, we report average F1, precision, and recall under different coreference inputs (Table 5). To determine correct quote attributions, the canonical name for the character cluster attributed by systems to each quote is compared with the gold attribution name for that quote. A match is assigned if a) an assigned name has only one word, which matches any word in the gold cluster name (such as *Tony* and *Tony Stark*), or b) if more than half of the words in the name match between the two character names, excluding titles such as *Ms.* and *Dr.* Name-matching is manually checked to ensure no system is penalized for selecting the wrong name within a correct character cluster. Any quote that a system fails to extract is considered a mis-attribution (an attribution to a NULL character).

As baselines, we consider BookNLP and the approach of He et al. (2013), who train a RankSVM model supervised on annotations from the novel

	With system coreference			*With gold coreference*			*With gold quote extraction*		
	P	R	F1	P	R	F1	P	R	F1
BookNLP	54.6	25.4	34.7	66.8	38.9	49.2	65.0	49.7	56.3
He et al. (2013)	54.0	53.3	53.6	56.5	55.7	56.1	56.7	56.0	56.3
Muzny et al. (2017) **(FanfictionNLP)**	**68.7**	**67.0**	**67.8**	**73.5**	**75.4**	**74.4**	**77.5**	**77.5**	**77.5**

Table 5: Quote attribution evaluation scores. Scores are reported using the respective system's coreference (*system coreference*), with gold character coreference supplied (*gold coreference*) and with gold character and gold quote spans supplied (*gold quote extraction*). Attribution is calculated by a character name match to the gold cluster name. If a quote span is not extracted by a system, it is counted as a mis-attribution. Micro-averages across the 10-story test set are reported. We include Muzny et al. (2017)'s approach in the FanfictionNLP pipeline.

Pride and Prejudice.

The quality of character coreference affects quote attribution. If an entire character is not identified, there is no chance for the system to attribute a quote to that character. If a system attributes a quote to the nearest character mention and that mention is not attributed to the correct character cluster, the quote attribution will likely be incorrect. For this reason, we evaluate quote attribution with different coreference settings. *System coreference* in Table 5 refers to quote attribution performance when using the respective system's coreference. That is, BookNLP's coreference was evaluated with BookNLP's quote attribution and FanfictionNLP's coreference with FanfictionNLP's quote attribution. We test He et al. (2013)'s approach with the same coreference input as FanfictionNLP. Evaluations are also reported with gold character coreference, as well as with gold character coreference and with gold quote extractions, to measure attribution without the effects of differences in quote extraction accuracy.

The deterministic approach of Muzny et al. (2017), incorporated in the pipeline, outperforms both BookNLP and He et al. (2013)'s RankSVM classifier in this informal narrative domain.

5.3 Quote Pronoun Resolution Module Evaluation

We test our approach for resolving pronouns within quotes (Section 3.3) on character coreference on the fanfiction evaluation set. We show results using gold quote attribution as an upper bound of the prospective improvement, and using quote attributions predicted by Muzny et al. (2017)'s approach adopted in the fanfiction pipeline. As shown in Table 6, post-hoc resolution of first-person (*I*) and second-person (*you*) pronouns with perfect quote annotation information (Gold QuA) substantially improves the overall performance of coreference resolution across both CoNLL and LEA F1 scores (by 1.6 and 3.5 percentage points respectively).

	CoNLL			**LEA**
	P	R	F1	F1
FanfictionNLP	72.6	70.1	71.4	73.5
+ I (Muzny QuA)	72.9	70.2	71.6	74.4
+ I + You (Muzny QuA)	73.1	70.2	**71.7**	**74.5**
+ I (Gold QuA)	73.9	71.2	72.5	76.0
+ I + You (Gold QuA)	74.6	71.6	**73.1**	**77.2**

Table 6: Quote Pronoun Resolution evaluation scores. Coreference resolution scores on the 10 fanfiction evaluation stories are reported. Improvements gained from changing the attribution of *I* and *you* within quotes are shown, with both the Muzny et al. (2017) quotation attribution system used in the FanfictionNLP pipeline, as well as the upper bound of improvement with gold quote annotation predictions.

Similarly, coreference resolution using information from a state-of-the-art quote attribution system (Muzny et al., 2017) also results in statistically significant, although smaller, improvements across both metrics (by 0.3 percentage points and 0.8 percentage points respectively) on the 10 fanfiction stories. These results suggest that our approach is able to leverage the quote attribution outputs (speaker information) to resolve the first and second-person pronouns within quotations. It does so by assuming that the text within a quote is from the point of view of the speaker of the quote, as attributed by the quote attribution system.

Table 7 shows the qualitative results on three consecutive quotes from one of the stories in our fanfiction dataset. For the first two quotations, FanfictionNLP incorrectly resolves *your/you* to the char-

Quote	Speaker (Muzny QuA / *Gold QuA*)	Addressee (Muzny QuA / *Gold QuA*)	FanFictionNLP	FanFictionNLP + I + You (Muzny QuA / *Gold QuA*)
"Alright , give me [**your**] phone . These questions are lame ."	Caitlin / *Caitlin*	Cisco / *Cisco*	**your** = Caitlin	**your** = [Cisco / *Cisco*]
"Would [**you**] rather give up showering for a month or the Internet for a month ?"	Caitlin / *Caitlin*	Cisco / *Cisco*	**you** = Caitlin	**you** = [Cisco / *Cisco*]
"[**You**] know what , do n't reply to that one , [**I**] do n't want to know ."	Cisco / *Caitlin*	Caitlin / *Cisco*	**I** = Cisco **You** = Cisco	**I** = [Cisco / *Caitlin*] **You** = [Caitlyn / *Cisco*]

Table 7: Coreference Resolution of first- and second-person pronouns in three consecutive quotes from one of the fanfiction stories in our dataset. Results show the impact of the Quote Attribution predictions on the performance of the algorithm described in Section 3.3.

acter *Caitlin*. However, FanfictionNLP + I + You correctly maps the mentions to *Cisco*. In the third example, we find that FanfictionNLP + I + You (Muzny QuA) does not perform correct resolution as the speaker output by the quote attribution module is incorrect. This shows the dependence of this algorithm on quality quote attribution predictions.

5.4 Assertion Extraction Qualitative Evaluation

There is no counterpart to the pipeline's assertion extraction in BookNLP or other systems. Qualitatively, the spans identified by TextTiling include text that relates to characterization beyond simply selecting sentences that mention characters, and with more precision than selecting whole paragraphs that mention characters.

For example, our approach captured sentences that described how characters were interpreting their environment. In one fanfiction story in our test set, a character "could see stars and planets, constellations and black holes. Everything was distant, yet reachable." Such sentences do not contain character mentions, but certainly contribute to character development and contain useful associations made with characters.

These assertions also capture narration that mentions interactions between characters, but which may not mention any one character individually. In another fanfiction story in which two wizards are dueling, extracted assertions for each character includes, "Their wands out, pointed at each other, each shaking with rage." These associations are important to characterization, but fall outside sentences that contain individual character mentions.

6 Ethics

Though most online fanfiction is publicly available, researchers must consider how users themselves view the reach of their content (Fiesler and Proferes, 2018). Anonymity and privacy are core values of fanfiction communities; this is especially important since many participants identify as LGBTQ+ (Fiesler et al., 2016; Dym et al., 2019). We informed Archive of Our Own, with our contact information, when scraping fanfiction and modified fanfiction examples given in this paper for privacy. We urge researchers who may use the fanfiction pipeline we present to consider how their work engages with fanfiction readers and writers, and to honor the creativity and privacy of the community and individuals behind this "data".

7 Conclusion

We present a text processing pipeline for the domain of fanfiction, stories that are written by fans and inspired by original media. Large archives of fanfiction are available online and present opportunities for researchers interested in community writing practices, narrative structure, fan culture, and online communities. The presented text processing pipeline allows researchers to extract and cluster mentions of characters from fanfiction stories, along with what each character does (assertions) and says (quotes).

We assemble state-of-the-art NLP approaches for each module of this processing pipeline and evaluate them on an annotated test set, outperforming a pipeline developed for formal fiction on character coreference and quote attribution. We also present improvements in character coreference with a post-processing step that uses information from quote attribution to resolve first- and second-

person pronouns within quotes. Our hope is that this pipeline will be a step toward enabling structured analysis of the text of fanfiction stories, which contain more variety than published, formal fiction. The pipeline could also be applied to other formal or informal narratives outside of fanfiction, though we have not evaluated it in other domains.

Acknowledgements

This work was supported in part by NSF grant DRL 1949110. We acknowledge Shefali Garg, Ethan Xuanyue Yang, and Luke Breitfeller for work on an earlier version of this pipeline, and Matthew Sims and David Bamman for their quote attribution re-implementation. We also thank the fanfiction writers on Archive of Our Own whose creative work allowed the creation and evaluation of this pipeline.

References

Apoorv Agarwal, Anup Kotalwar, and Owen Rambow. 2013. Automatic Extraction of Social Networks from Literary Text: A Case Study on Alice in Wonderland. In *International Joint Conference on Natural Language Processing*, October, pages 1202–1208.

Mariana SC Almeida, Miguel B Almeida, and André FT Martins. 2014. A joint model for quotation attribution and coreference resolution. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 39–48.

David Bamman, Olivia Lewke, and Anya Mansoor. 2020. An Annotated Dataset of Coreference in English Literature. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 44–54.

David Bamman, Ted Underwood, and Noah A. Smith. 2014. A Bayesian Mixed Effects Model of Literary Character. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)*, pages 370–379.

Bertram Bruce. 1981. A social interaction model of reading. *Discourse Processes*, 4(4):273–311.

Cuong Xuan Chu, Simon Razniewski, and Gerhard Weikum. 2020. EntyFi: Entity typing in fictional texts. *WSDM 2020 - Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 124–132.

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.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 4171–4186.

Brianna Dym, Jed R. Brubaker, Casey Fiesler, and Bryan Semaan. 2019. "Coming Out Okay": Community Narratives for LGBTQ Identity Recovery Work. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–28.

Sarah Evans, Katie Davis, Abigail Evans, Julie Ann Campbell, David P Randall, Kodlee Yin, and Cecilia Aragon. 2017. More Than Peer Production: Fanfiction Communities as Sites of Distributed Mentoring. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 259–272.

Ethan Fast, Tina Vachovsky, and Michael S Bernstein. 2016. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. In *Proceedings of the 10th International Conference on Web and Social Media (ICWSM)*, pages 112–120.

Rita Felski. 2020. *Hooked: Art and Attachment*. University of Chicago Press.

Casey Fiesler, Shannon Morrison, and Amy S. Bruckman. 2016. An Archive of Their Own. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 2574–2585.

Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. *Social Media and Society*, 4(1).

Melanie C. Green, Timothy C. Brock, and Geoff F. Kaufman. 2004. Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, 14(4):311–327.

Hua He, Denilson Barbosa, and Grzegorz Kondrak. 2013. Identification of speakers in novels. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1312–1320.

Marti A. Hearst. 1997. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguis-*

tics: Human Language Technologies (NAACL-HLT), pages 1534–1544.

Henry Jenkins. 1992. *Textual Poachers: Television Fans and Participatory Culture*. Routledge.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77.

Mandar Joshi, Omer Levy, Luke Zettlemoyer, and Daniel Weld. 2019. BERT for coreference resolution: Baselines and analysis. 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 5803–5808, Hong Kong, China. Association for Computational Linguistics.

Anna Kasunic and Geoff Kaufman. 2018. Learning to Listen: Critically Considering the Role of AI in Human Storytelling and Character Creation. In *Proceedings ofthe First Workshop on Storytelling*, pages 1–13.

Mike Kestemont, Michael Tschuggnall, Efstathios Stamatatos, Walter Daelemans, Günther Specht, Benno Stein, and Martin Potthast. 2018. Overview of the author identification task at PAN-2018: Cross-domain authorship attribution and style change detection. *CEUR Workshop Proceedings*, 2125.

Evgeny Kim and Roman Klinger. 2019. Frowning Frodo, Wincing Leia, and a Seriously Great Friendship: Learning to Classify Emotional Relationships of Fictional Characters. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 647–653.

Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2011. Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In *Proceedings of the 15th Conference on Computational Natural Language Learning: Shared Task*, pages 28–34.

Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 188–197.

Alexis Lothian, Kristina Busse, and Robin Anne Reid. 2007. Yearning void and infinite potential: Online slash fandom as queer female space. *English Language Notes*, 45(2).

Ann McClellan. 2014. Redefining genderswap fan fiction: A Sherlock case study. *Transformative Works & Cultures*, 17.

George A Miller. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

Smitha Milli and David Bamman. 2016. Beyond Canonical Texts : A Computational Analysis of Fanfiction. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP-16)*, pages 2048–2053.

Nafise Sadat Moosavi and Michael Strube. 2016. Which coreference evaluation metric do you trust? a proposal for a link-based entity aware metric. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 632–642, Berlin, Germany. Association for Computational Linguistics.

Felix Muzny, Michael Fang, Angel X. Chang, and Dan Jurafsky. 2017. A two-stage sieve approach for quote attribution. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017)*, volume 1, pages 460–470.

Tim O'Keefe, Silvia Pareti, James R. Curran, Irena Koprinska, and Matthew Honnibal. 2012. A sequence labelling approach to quote attribution. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, (July):790–799.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. Conll-2012 shared task: Modeling multilingual unrestricted coreference in ontonotes. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, page 1–40, USA.

Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky, and Christopher D Manning. 2010. A multi-pass sieve for coreference resolution. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 492–501.

Marta Recasens, Marie-Catherine de Marneffe, and Christopher Potts. 2013. The life and death of discourse entities: Identifying singleton mentions. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 627–633.

Bingyu Shen, Christopher W. Forstall, Anderson De Rezende Rocha, and Walter J. Scheirer. 2018. Practical text phylogeny for real-world settings. *IEEE Access*, 6:41002–41012.

Matthew Sims and David Bamman. 2020. Measuring information propagation in literary social networks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 642–652, Online. Association for Computational Linguistics.

Matthew Sims, Jong Ho Park, and David Bamman. 2019. Literary event detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3623–3634, Florence, Italy. Association for Computational Linguistics.

Catherine Tosenberger. 2008. Homosexuality at the Online Hogwarts: Harry Potter Slash Fanfiction. *Children's Literature*, 36(1):185–207.

David Vilares and Carlos Gómez-Rodríguez. 2019. Harry Potter and the action prediction challenge from natural language. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1:2124–2130.

Anthony Wall. 1984. Characters in Bakhtin's Theory. *Studies in 20th Century Literature*, 9(1):2334–4415.

Kodlee Yin, Cecilia Aragon, Sarah Evans, and Katie Davis. 2017. Where no one has gone before: A meta-dataset of the world's largest fanfiction repository. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 6106–6110.