RiQuA: A Corpus of Rich Quotation Annotation for English Literary Text

Sean Papay and Sebastian Padó

University of Stuttgart, Institute for Natural Language Processing
Pfaffenwaldring 5b, 70569 Stuttgart, Germany
{sean.papay, sebastian.pado}@ims.uni-stuttgart.de

Abstract

We introduce RiQuA (RIch QUotation Annotations), a corpus that provides quotations, including their interpersonal structure (speakers and addressees) for English literary text. The corpus comprises 11 works of 19th-century literature that were manually doubly annotated for direct and indirect quotations. For each quotation, its span, speaker, addressee, and cue are identified (if present). This provides a rich view of dialogue structures not available from other available corpora. We detail the process of creating this dataset, discuss the annotation guidelines, and analyze the resulting corpus in terms of inter-annotator agreement and its properties. RiQuA, along with its annotations guidelines and associated scripts, are publicly available for use, modification, and experimentation.

Keywords: quotations, interpersonal structure, literary text

1. Introduction

In literature, spoken interactions between characters are often of central importance to the narrative. Reported speech, or *quotations*, provide readers a direct window into the world of the narrative, where the author's own presence within the work becomes minimally intrusive (Page, 1988). Authors can use quoted speech for storytelling in order to describe complex verbal interactions between characters without explicit exposition. In dialogue-heavy works, quoted speech can often exceed 50% of a novel's text (Elson and McKeown, 2010). Therefore, quotations are an important structural component of literary texts.

When analyzing the information contained in quotations, both content and context are of fundamental importance. A quotation conveys information about *what* was said, *how* it was said, *who* said it, and *to whom*. Of these data, only the first one is generally provided by a quotation's content, while broader context is needed for the other three. As a concrete example, take the following sentence from Charles Dickens' *A Christmas Carol*:

> ["What's to-day?"]$_{\text{direct quotation}}$ [cried]$_{\text{Cue}}$ [Scrooge]$_{\text{Speaker}}$, calling downward to [a boy in Sunday clothes]$_{\text{Addressee}}$, who perhaps had loitered in to look about him. (1)

The quote itself (["What's to-day?"]$_{\text{direct quotation}}$) provides the content of the utterance, and the type of the quotation (direct speech, as opposed to indirect speech) tells us that this is a verbatim rendering of the utterance. However, the quotation tells us relatively little about the interpersonal structure. The reader requires knowledge of the full sentence to understand which character is speaking ([Scrooge]$_{\text{Speaker}}$), whom he is addressing ([a boy in Sunday clothes]$_{\text{Addressee}}$), and the manner of speech (he [cried]$_{\text{Cue}}$), indicating anxiety. Often, even a full sentence is not enough, and the interpersonal structure must be reconstructed from the larger context.

There exist a substantial number of studies which have shown that knowledge about quotations can be useful in tasks such as extracting social networks (Elson et al., 2010), modeling discourse structure (Redeker and Egg, 2006), formality (Faruqui and Padó, 2012), affect (Nalisnick and Baird,

2013; Iosif and Mishra, 2014), or even plain co-reference resolution (Almeida et al., 2014).[1] In many of these tasks, the interpersonal structure is as important as the content level, or arguably more so. This mirrors a general tendency in other areas of semantics towards structured (relational) annotations that link states and events with the relevant actors and entities. For example, in sentiment annotation, additional structure can be annotated by linking subjective phrases with relevant aspects (Liu, 2012) and in emotion annotation, emotion phrases in text can be linked with experiencers and causes (Kim and Klinger, 2018).

At the same time, existing corpora with quotation annotation vary greatly in whether, how, and to what extent they cover the interpersonal structure (i.e., who communicates what to whom – see Section 2. for details). In particular, we are not aware of any large-scale, publicly available corpora which mark both speakers and addressees for quotations in literary text. This resource gap makes investigation of conversation structures in text difficult. For example, Elson et al. (2010), lacking full addressee information, must assume that speakers are addressing one another when their quotations occur within a window of 300 words. Such assumptions, as we will show, have many exceptions, and introduce noise into analyses of quotation structures.

To address this deficiency, and provide the research community with a more complete dataset of quotations and their structure, we introduce RiQuA. This corpus consists of 11 English-language works of 19th-century literature, manually annotated with rich quotation structure. Within each work, all quotations, direct and indirect, and identified. Furthermore, for each quotation, we identify speakers, addressees, and cue words, when these are present. This corpus is freely available to download[2].

2. Related Work

Compared to other phenomena, resources with quotation annotation are still relatively scarce. Those resources that do

---

Corpus	Language	Genre	# Quotations	Indirect	Addressees
CQSAC	English	Literature	3,176	✗	✗
PARC 3	English	Newspaper	19,712	✓	✗
STOP	English	Fiction, Biography, & Newspaper	13,237	✓	✗
ACDS	English	Bible	1,245	✗	✓
RWG	German	Narrative, Magazine, & Newspaper	9,451	✓	✗

Table 1: A summary of some prominent existing corpora for quotations

exist exhibit large amounts of variance in their scope, conceptual assumptions, and structural assumptions (Papay and Padó, 2019). In this section, we describe existing datasets which fulfill a similar purpose to RiQuA and contrast their attributes. Table 1 provides a summary of existing similar corpora.

CQSAC. The Columbia Quoted Speech Attribution Corpus (CQSAC) was one of the first corpora to provide annotation of quotations (Elson and McKeown, 2010). The corpus identifies quotation spans in literary text, and for each quotation identifies a noun phrase in the vicinity as a speaker for that quotation, if one exists. Unfortunately, this corpus only considers direct quotations, and not indirect ones. Additionally, quotation spans are detected automatically and not manually annotated, which makes them unreliable. Finally, no addressee information exists.

PARC 3. PARC 3 (Pareti, 2016) provides quotation annotation for the more than 2000 news articles in the Wall Street Journal portion of the Penn Treebank. In each article, both direct and indirect quotations are annotated, together with cues and speakers. The annotation assumes that each quotation is introduced by a cue, which is reasonable for newswire, but not for literary text. Again, no addressees are provided.

STOP. STOP is a corpus of "speech, thought, and writing presentation" (Semino and Short, 2004), primarily developed to empirically illustrate the authors' taxonomy of types of language reporting. The corpus comprises English texts across three genres: fiction, biography, and newswire. The corpus marks spans which correspond to speech, thought, and writing separately, and further categorizes each such span as either direct, indirect, free-indirect, or reported. This categorization yields a product of twelve distinct span types which are annotated. Each such span is also marked with a speaker (or medium) where possible, but not at the textual level, but relative to a (manually compiled) list of relevant entities. Addressees are not marked.

ACDS. The annotated corpus of directed speech (Lee and Yeung, 2016) consists of the four gospels of the New Testament annotated for direct speech. For each quotation, the corpus provides speech verb (cue), speaker, and listener (addressee), when applicable. Speakers and addressees are identified as spans of text, and coreferent speakers and addressees are also marked. The main limitations of this corpus are the use of a single

work, and a single type of quotation. Also, as far as we know, this corpus has not been made publicly available.

RWG. The Redewiedergabe ('reported speech') corpus (Brunner, 2013), henceforth RWG, is a German-language corpus for quotations in narrative text. The corpus consists of a short snippets of samples of German fiction and non-fiction text, published between 1840 and 1920. Within each text, spans are classified in the same 3×4 categories as in STOP. Additionally, for each identified quotation, a span of text is also identified as a speaker. No addressees are given, and the use of snippets precludes full-document analyses.

3. Corpus Creation

3.1. Source Texts

We decided to annotate the same set of texts covered by the Columbia speech attribution corpus (Elson and McKeown, 2010). These texts cover a range of 11 works of fiction by various authors, all written in the 19th century. Our main motivation for this choice was to stay close to an established standard in quotation detection, while extending the coverage of the annotation by (a) providing fully manual annotation and (b) extending the scope to include indirect quotations as well as including addressees.

In cases where Elson and McKeown (2010) used excerpts from the full texts (specifically, for the three novels), we use the same excerpts. While all annotated works are in English, one novel is a contemporary translation from the original French, and two short stories and a novella are contemporary translations from Russian. Table 2 lists the properties of the texts. The texts all contain large amounts of quoted speech, and the quotations contained are highly varied in properties such as length, directness, and the presence or absence of speakers and addressees. Section 4. presents a detailed statistical analysis of the texts and the quotations they contain.

3.2. Annotation Schema and Guidelines

As sketched above, our goal was to create a corpus which exemplifies a wide range of types of reported speech events (direct quotations, indirect quotations), the speech events (**cues**) themselves where realized linguistically, as well as the **speakers** and **addressees** (again, where appropriate).

Figure 1 illustrates the annotation of a simple speech event. Each of these concepts needs to be defined for annotation to proceed reliably, as the experiences from the creation of previous corpora show (cf. Section 2.). To this end, we developed detailed annotation guidelines that combined
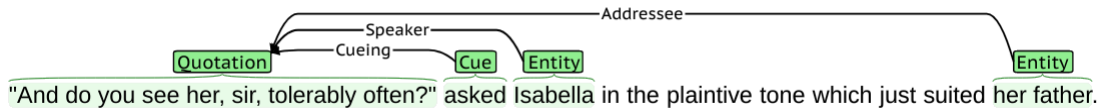
Figure 1: An annotated speech event with a direct quotation, speaker, cue, and addressee

Work	Published	# tokens
Jane Austen		
Emma[†]	1815	69k
Charles Dickens		
A Christmas Carol	1843	36k
Gustave Flaubert		
Madame Bovary*[†]	1856	139k
Mark Twain		
The Adventures of Tom Sawyer[†]	1876	52k
Sir Arthur Conan Doyle		
"A Scandal in Bohemia"	1891	10k
"The Red-Headed League"	1891	11k
"A Case of Identity"	1891	8k
"The Boscombe Valley Mystery"	1891	11k
Anton Chekhov		
The Steppe*	1888	47k
"The Black Monk"*	1894	15k
"The Lady with the Dog"*	1899	8k

Table 2: The source texts used for RiQuA. * indicates that the English text is a translation, and [†] indicates that excerpts were used.

definitions with examples for each concept and address problems and borderline cases explicitly. Below, we outline the most important definitions and side conditions that we ended up with after several revision cycles. The full guidelines will be made available together with the corpus.

Speech event. Speech events are descriptions of verbal communication within the text. Our guidelines' formal definition of a speech event is inspired by the Communication and Statement frames in frame semantics as realized in Berkeley FrameNet (Fillmore et al., 2003). We avoid structural constraints on speech events as far as possible to capture as many of them as possible. The default case of a speech event is triggered by a verbal communication predicate, but we also include nouns ('Jacob's *interruption*') and multi-word expressions ('Jacob *hit the nail on the head*'). Further afield, we also cover types of events that imply communication with a high degree of certainty ('The butler *was afraid* he could be of little help').

This liberal definition of speech event includes many instances of communication that do not fall into the purview of existing quotation corpora. However, structural constraints (Is there a cue? Is the speech event a single communication verb or noun?) should make it possible to identify the "canonical" cases.

Semantic and Discourse Structure. We broadly disre-

gard considerations regarding the surrounding semantic and discourse structure: We annotate speech events irrespective of factuality or modality, and at all narrative levels (e.g., within other quotations). This corresponds to standard practice in semantic role annotation (Fillmore et al., 2003).

Cues. The words and phrases showed in italics in the previous paragraph are examples of cues, that is, the events that express or imply speech. Note that cues are not necessary for speech events. In particular, literary renderings of dialogue can omit cues and just let quotations stand: "How are you?" – "Good, and yourself?".

Quotation spans. The quotation span is defined as the linguistic unit (word, phrase, sentence) which provides the content of a speech event. In the case of **direct quotations**, this span will start and end with qutoation marks, and will consist of a verbatim transcription of the contents of the speech event. For **indirect quotations**, this span will not be surrounded by quotation marks, and will generally correspond to an indirect description of what was spoken. With multi-word cues, there may be grey areas between cue and quotation: In 'Peter rejected the claim', is 'reject the claim' an MWE cue, or is it a cue ('reject') combined with a quotation span ('claim')?

Speaker. Speakers are characters that utter quotations. Documents or communication channels ('letter','TV') do not qualify. The main problem with speakers is that they are not always realized locally. When they are not, their identification is often defeasible, as is the case with unrealized semantic roles (Fillmore, 1986). Even when the speaker is identifiable at the level of entities, there may be multiple phrases in the text that refer to this entity, which forms a problem for annotation. Our guidelines provide heuristics to resolve such ambiguities. Importantly, we decided that speakers should always be outside the quotation, and more specifically, on the same narrative level as the speech event.

Addressee. Addressees are the recipients of a speech event. They are typically even more pragmatically determined than speakers: They may be realized locally, but often are not (see Section 4.). If they are not, they can often be recovered from context. We include such cases in the annotation, but only annotate intended, not accidental, addresses.

3.3. Annotation Process

As user interface for the annotation process, we used brat (Stenetorp et al., 2012), a relatively light-weight web-based
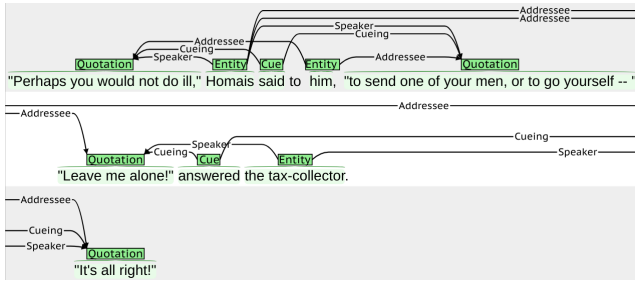
Figure 2: Complex interpersonal structure of quotation

platform which supports the annotation of spans and relations between spans on running text. Crucially, it is able to sensibly visualize very long spans and many relations, a requirement where almost all other tools we tested fail (see Figure 2).

The annotation took place over the course of approximately one year. All texts received two independent annotations by two native speakers (one with a linguistics background, one without) and then merged by a third, more senior annotator. Before the 10-month "production" phase started, a 1-month "training phase" served to train the annotators and refine the guidelines. During the production phase, each annotator worked approximately 400 hours, corresponding to a speed of roughly 54 sentences or 15 speech events per hour (with substantial variance across texts).

The final merging was carried out using a semi-automated approach: an automatic script[3] was used to resolve easily-resolvable disagreements (such as off-by-one errors and disagreements regarding punctuation), and manual annotation was used to resolve the remaining differences.

4. Corpus Analysis

4.1. Descriptive Statistics

Quotations. The resulting corpus, RiQuA, comprises a total of 5963 quotations across all works. 93.6% of these are direct quotations – that is, they were surrounded with quotation marks – while the remaining 6.4% are indirect. Almost all quotations (99.7%) have marked speakers, 95.3% have marked addresses, and 68.3% have marked cues. These observations validate that having no speaker is an exception for quotations, that having an addressee is standard almost to the same degree, but that the role of cues to introduce quotations is frequently taken over by other linguistic indicators. Notable in literature are dialogic passages where utterances of two (or sometimes now) interlocutors simply alternate, such as in the following example from Mark Twain's *Tom*

3The script will be made available with the corpus.

Sawyer:

> Presently Tom seized his comrade's arm and
> [said]$_{Cue}$:
> ["Sh!"]$_{Quotation}$
> ["What is it, Tom?"]$_{Quotation}$ And the two clung together with beating hearts.
> ["Sh! There 'tis again! Didn't you hear it?"]$_{Quotation}$
> ["I—"]$_{Quotation}$
> ["There! Now you hear it."]$_{Quotation}$
> ["Lord, Tom, they're coming! They're coming, sure. What'll we do?"]$_{Quotation}$

(2)

On the other hand, there are many split quotations sharing cues, so that there is a significant discrepancy between the number of spans identified as cues, and the number of quotations with cues. On average, each cue span is associated with about 1.4 quotations. A total of 2884 text spans are identified as cues, but these correspond to only 277 distinct strings. The most common, by far, is the word "said," which occurred as a cue 1199 times. The next most common cues are "asked," "cried," "replied," "answered," and "went on." 170 cues occur only once.

Quotations	5963	
Direct	5580	93.6%
Indirect	383	6.4%
With Cue	4074	68.3%
With Speaker	5946	99.7%
With Addressee	5685	95.3%
Nested	150	2.5%
Speakers	3274	
Pronouns	1307	40.0%
Addressees	2881	
Pronouns	995	34.5%

Table 3: Descriptive Statistics of Annotations

Quotation Length. We observed significant variance in quotation length. While the average quotation was about 26 tokens long, quotations of a single token were identified, and the longest quotation was nearly 1500 tokens long. Figure 3 shows the distribution of quotation lengths found in the corpus.

Speakers and Addressees. RiQuA contains 3274 spans labeled as speakers, and 2881 labeled as addressees. These include 1846 spans which act as both speakers and as addressees. 40% of all speakers and 34.5% of all addressees were pronouns, while the remainder were noun phrases or proper nouns.

Figure 4 illustrates where cues, speakers, and addressees are found relative to their corresponding quotations. We find a strong bias across all relation types to favor selections which occur before the quotation in the text. This is consistent with a linear reading order of the texts, and with the instructions in our annotation guidelines. Comparing the distributions of
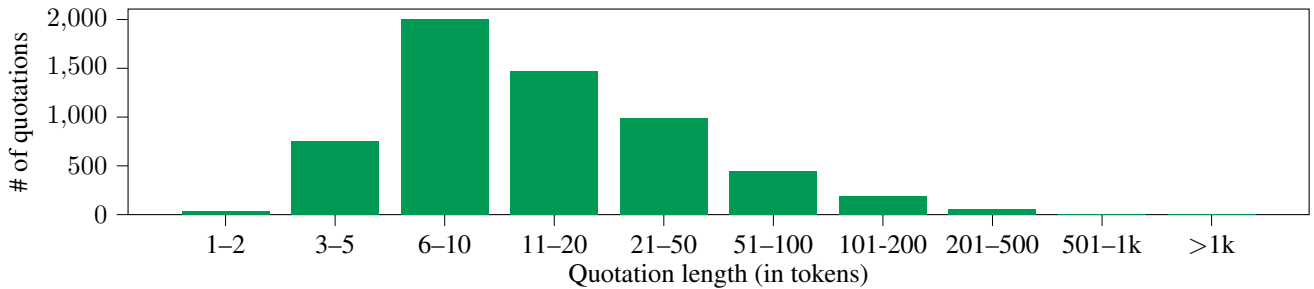
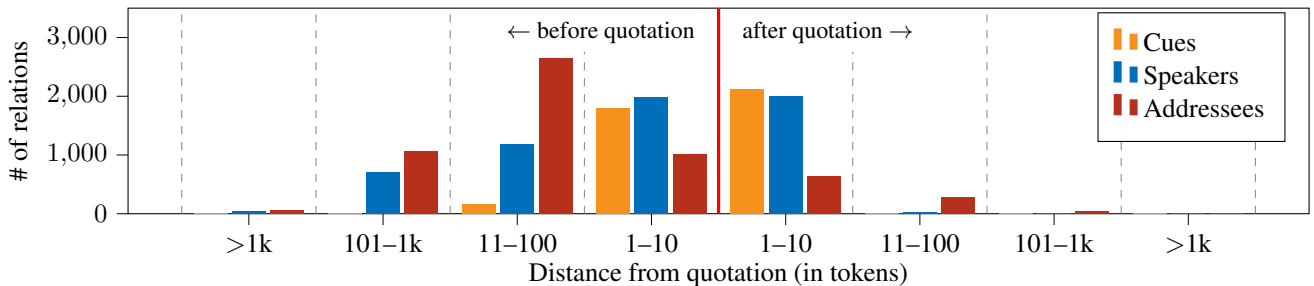Figure 3: Distribution of quotation lengths in RiQuA.

Figure 4: Distances between quotations and their corresponding cues, speakers, and addressees.

the relation types, we find that cues are highly concentrated near their quotations (as expected). Speakers also occur mostly directly before or after the quotation, but addressees tend to occur at a larger distance to their corresponding quotations than speakers.

We believe this is due to the fact that speakers are generally realized more prominently in syntactic terms (e.g., they are often subjects of communication verbs) while addressees are often realized only optionally (as prepositional objects) or must be recovered from the discourse context. Example 1 illustrates this case, where the addressee is an argument of a verb (*call*) different from the primary speech event (*cry*). Here, *call* must be understood by the reader as a further specification of *cry*.

Embedded Quotations. 150 of the quotations identified (2.52%) were nested quotations (quotations within quotations). An example of an embedded quotation occurs in the following passage passage from Anton Chekhov's *The Steppe*:

["And I said to him, ['God bless your compressed air!']$_{\text{Quotation}}$"]$_{\text{Quotation}}$ he brought out through his laughter, waving both hands. (3)

While not disallowed by our annotation guidelines, we observed no deeper quotation embeddings.

4.2. Inter-Annotator Agreement

We now assess the quality of the resulting corpus. We quantify agreement with the family of F_1 measures (precision, recall, F_1), computed between sets of spans in an 'exact match' setup that is often applied to sequence classification tasks like quotation detection – i.e., there is no partial credit for partial match. Note that this evaluation choice does not easily support chance correlation as is usual in classification

		P	**R**	**F$_1$**
Quotation Spans	Inter-annotator	–	–	91.27
	Annotator 1	95.72	92.24	93.94
	Annotator 2	97.79	94.40	96.07
Cues	Inter-annotator	–	–	91.07
	Annotator 1	92.34	93.24	92.79
	Annotator 2	97.66	95.51	96.57
Speakers	Inter-annotator	–	–	84.26
	Annotator 1	91.82	91.52	91.67
	Annotator 2	91.66	91.77	91.71
Addressees	Inter-annotator	–	–	61.44
	Annotator 1	80.01	75.52	77.70
	Annotator 2	82.29	80.93	81.60

Table 4: Inter-annotator agreement, and agreements between each annotator and the final, merged annotations, measured as precision, recall, and F_1. For cues, speakers, and addressees, only those cases are considered where there was agreement for the corresponding quotation span.

(Artstein and Poesio, 2008), since it is not obvious what chance agreement on spans is supposed to be.

More specifically, we measure agreement between the two initial annotations, and between each initial annotation and the final (merged) annotation. In the first case, we treat one set of initial annotations as a prediction and another as a ground truth. As the F_1 measure is symmetric, the choice of which annotator is treated as the ground truth is inconsequential. In the second case, we report precision, recall, and F_1 measure, treating the merged annotation as the ground truth.

Table 4 shows the results of all levels of our annotation: quotation spans, and those quotations' cues, speakers, and

	FP	FN	PM	Cor
Quotation Spans	16	44	35	5
Cues	10	47	34	9

	CoRef	Err	PM	Cor
Speakers	41	46	13	0
Addressees	57	33	10	0

Table 5: A sample of annotator errors for different annotation levels, classified manually. Quotation spans and cues are classified as either false positives (FP), false negatives (FN), partial matches (PM), or correct (Cor) (for when the merged corpus was in error). Speakers and addressees are classified as either Coreferent (when the annotator's selected span is distinct from, but coreferent to, the final corpus), plain errors (Err), partial matches (PM), or correct (Cor).

addressees. We only evaluate the interaction-structure levels of cues, speakers, and addressees for those speech events where there is agreement on the quotation span in the first place. We observe that quotation spans and cues are relatively easy to annotate, speakers are more challenging, and addressees are hard. This trend is not surprising – quotation spans and cues are defined mainly by surface-level, syntactic properties of the text, while speakers and addressees require more text understanding to identify correctly. These results also correlate well with the distances discussed above (cf. Figure 4).

4.3. Annotator Errors

To better understand the causes of annotator disagreement, we manually examined a sample of cases where individual annotators disagreed with the merged gold standard, and classified these samples. We examined 100 such cases each for disagreements in quotation spans, speakers, cues, and addressees – 50 for each annotator. Our findings are summarized in Table 5.

For quotation spans and cues, we observe that false negatives are much more common than false positives. This indicates that, as is often the case in manual linguistic analysis, it is challenging for annotators for stay alert and catch every instance of a given phenomenon. While this trend is already somewhat apparent in our precision and recall results, it is brought out much more clearly in the manual analysis. This is likely due to the fact that, when manually classifying mismatches, we can distinguish partial matches from "real" mismatches, while the automatic evaluation would treat such a partial match as a combination of a false positive and a false negative.

Partial matches also account for about one third of all disagreements, indicating that the guidelines regarding the choice of boundaries when annotating quotation spans still leave too much room for interpretation. Finally, we found a small number of cases where the original annotations matched the guidelines better than the merged ones, i.e., merging errors.

For speakers and addressees, a new class of potential errors arise: not only do annotators have to identify their spans correctly, but they are also supposed to pick the same men-

tion from the document. To assess the difficulty of this task, we added a new error category, CoRef, which indicates that the that the annotator picked a different span of text which nevertheless referred to the same referent as the gold standard span. The following example, with the quotation and gold-standard speaker span marked, illustrates the difference between a CoRef error and a plain error (Err) for speaker annotation:

[He]$_{Err}$ felt [the Spirit]$_{CoRef}$'s glance, and stopped. ["What is the matter?"]$_{Quotation}$ asked [the Ghost]$_{Speaker}$.   (4)

Indeed, we find that CoRef cases constitute a majority of the errors for addressees, and a significant minority of the errors for speakers. This means that annotators were actually rather successful in identifying the proper *character* in the narrative, but that the guidelines were not specific enough to enforce agreement on the same *span*.

These observations also provide a new angle on the agreement results in Table 4: it is not necessarily the case that addressees are *conceptually* more difficult to annotate than speakers: In fact, the speaker annotation shows more plain errors (46 vs. 33). Rather, for addressees it is specifically harder to pick the right mention of the reference – which is not surprising given the observations about the distances between quotation span and addressee from above (Section 4.1.).

In a scenario in which a more comprehensive computational analysis of literary texts, coreference resolution for characters is available, the choice of exact mention becomes irrelevant, because all mentions are mapped onto the entity. In this case, the category of CoRef errors vanishes, and surprisingly, addressee annotation might become even simpler than speaker annotation.

5. Conclusion and Future Work

In this paper, we introduced the RiQuA corpus which provides around 6,000 quotations, including rich interpersonal structure, for 11 literary 19th century texts. The corpus provides gold annotations, resulting from full double annotation, for quotation spans (both direct and indirect) with cues, speakers and addressees, filling an important resource gap for quotation analysis. In the future, we plan to augment the structure of our corpus by adding coreference information for speakers and addressees, keeping in mind the challenges of the genre (Roesiger et al., 2018). We hope that our corpus will prove to be a valuable resource for future work in quotation analysis, social network extraction, and discourse structure modeling.

6. Acknowledgments

7. Bibliographical References

Almeida, M. S. C., Almeida, M. B., and Martins, A. F. T. (2014). A joint model for quotation attribution and coreference resolution. In *Proceedings of EACL*, pages 39–48, Gothenburg, Sweden.

Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Brunner, A. (2013). Automatic recognition of speech, thought, and writing representation in German narrative texts. *Literary and Linguistic Computing*, 28(4):563–575.

Elson, D. K. and McKeown, K. R. (2010). Automatic attribution of quoted speech in literary narrative. In *Proceedings of AAAI*, pages 1013–1019, Atlanta, GA.

Elson, D. K., Dames, N., and McKeown, K. R. (2010). Extracting social networks from literary fiction. In *Proceedings of ACL*, pages 138–147, Uppsala, Sweden.

Faruqui, M. and Padó, S. (2012). Towards a model of formal and informal address in English. In *Proceedings of EACL*, pages 623–633, Avignon, France.

Fillmore, C. J., Johnson, C. R., and Petruck, M. R. (2003). Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250.

Fillmore, C. J. (1986). Pragmatically controlled zero anaphora. In V Nikiforidou, et al., editors, *Proceedings of the Berkeley Linguistics Society*, volume 12, pages 95–107.

Iosif, E. and Mishra, T. (2014). From speaker identification to affective analysis: a multi-step system for analyzing children's stories. In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature*, pages 40–49, Gothenburg, Sweden.

Kim, E. and Klinger, R. (2018). Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In *Proceedings of COLING*, pages 1345–1359, Santa Fe, NM.

Lee, J. and Yeung, C. Y. (2016). An annotated corpus of direct speech. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pages 1059–1063, Portorož, Slovenia.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Nalisnick, E. T. and Baird, H. S. (2013). Character-to-character sentiment analysis in Shakespeare's plays. In *Proceedings of ACL*, pages 479–483, Sofia, Bulgaria.

Page, N. (1988). *Speech in the English novel*. Springer.

Papay, S. and Padó, S. (2019). Quotation detection and classification with a corpus-agnostic model. In *Proceedings of RANLP*, Varna, Bulgaria.

Pareti, S. (2016). Parc 3.0: A corpus of attribution relations. In *Proceedings of LREC*, pages 3914–3920, Portorož, Slovenia.

Redeker, G. and Egg, M. (2006). Says who? On the treatment of speech attributions in discourse structure. In *Proceedings of the Workshop on Constraints in Discourse*, pages 141–147, Maynooth, Ireland.

Roesiger, I., Schulz, S., and Reiter, N. (2018). Towards coreference for literary text: Analyzing domain-specific phenomena. In *Proceedings of the Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 129–138, Santa Fe, NM.

Semino, E. and Short, M. (2004). *Corpus Stylistics: Speech, Writing And Thought Presentation In A Corpus Of English Writing*. Routledge Advances In Corpus Linguistics. Routledge, London.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of EACL System Demonstrations*, pages 102–107, Avignon, France.