Annotation Guidelines for RɪQuA

Sean Papay, Sebastian Padó

March 3, 2020

Contents

1 Introduction

This document describes the annotation process for the RɪQuA annotation project. This is a living document — it lives a state of modification and addition as the annotation process continues. We expect it to crystallize over time, however. Modifications will become more and more limited to edge cases and fringe cases.

The last section of this document contains a change log.

2 Overview

In the RɪQuA annotation project, we seek to manually annotate information regarding quotations in English literary texts. In particular, we are interested in identifying descriptions of speech events within literary texts, and for these events, identifying

- what was said (the quote)

- the speaker(s)

- the addressee(s)

- and "cue" words in the prose that introduce the quotation.

The RɪQuA annotation process encodes this information by two fundamental types of annotations. The first type of annotation, **span annotations** mark specific, contiguous spans of text as being of interest. Each span annotation has a type, and these annotations can overlap and nest arbitrarily. the RɪQuA annotation project currently recognizes three types of spans: **Quote** spans, which refer to things that are said, **Entity** spans, which correspond to speakers or addressees, and **Cue** spans, which mark words and phrases that directly introduce quotations. These will all be further specified in the Definitions section.

The second type of annotations are **relation annotations**. Relation annotations mark some relation between two marked spans. We use relation annotations to **associate Speakers, Addressees, and Cues with their associated Quotes**. Here is a rule of thumb:

Golden rule of quote annotation. Each speech event[1] corresponds to a single Quote span.

---

[1]The events do not need to have actually occurred. See Section 6.2.

If that communication situation has a well-defined speaker, which is present in the text and marked as an Entity, that span will be associated with the Quote span via the Speaker relation. Addressees and Cues are likewise linked to their corresponding Quotation spans via their respective relation annotations.

2.1 What Is a Speech Event?

Speech events are passages of text that either

- directly describe speech

- imply with a high degree of certainty that some speech occurred.

The **default case** is that speech events are introduced by a communication verbs, the **cue** combines with a syntactically attached **direct or indirect quote** (see Section 4.1):

(1) John *said* "Hello."
(2) Jose *told* the man what he thought.
(3) Nobody *told* him "Good morning."

We also annotate non-default cases.

2.1.1 Non-verb cues

An obvious non-default case are cues that still express communication, but are not verbs. This includes **communication nouns, multi-word expressions such as support verb constructions or idioms**, and potentially other parts of speech:

(4) Jacob's *interruption* that 101 was blocked was met with groans of despair
(5) Jacob *hit the nail on the head*: "This is a disaster."
(6) Jacob *retracted the suggestion* to go out.

2.1.2 Non-communication cues

A second non-default case are **speech events that are not expressed, but only implied** by the cue. Such cues can typically be combined with either direct or indirect quotes.

The cues form a continuum from verbs that clearly imply communication (*greet*) to others that primarily express, for example, attitude or perception, and imply communication only in context:

(7) The waitress asked him what he wanted to drink. He *wanted* a Coke.

(8) The butler *was afraid* that he could be of little help.

(9) In his keynote, Einstein *noted* that Physics made great progress.

2.1.3 Quotations without cues

A third case are **speech events without lexical cues**. This is a special case for direct quotes where the turns in a two-party conversation are given without communication verbs.

(10)   "How are you doing?"
       - "Fine, and yourself'?"
       - "Doing ok."

2.1.4 Cues without quotations

A forth special case are **speech events whose content is not specified**, such as

(11) . . . the wrath which had nearly proved fatal to the pulpit had been excited by the *whispers* of the lady and myself. (Poe: The man that was used up)

The content (quotation) of the whispers is not specified. Nevertheless, this is a communication event, so it should be annotated as such, with the Quotation left unspecified (It also arguably specifies the speaker and addressee.)

Note that sometimes cues refer to complete communication situations rather than individual communication events:

(12) We telegraphed a few signals and then commenced, soto voce, a brisk *tete-a-tete*. "Smith!" said she in reply to my very earnest inquiry: "Smith!- why, not General John A. B. C.?" ... (ibid.)

The "tete-a-tete" (which in this context could be replaced by 'discussion') refers to the whole following passage, so there is no single individual phrase that could be marked as a content. It is generally the case that specific Quotations that do occur in the context have their own cues, as in this case: *"Smith"!* has the cue *said*.

Our general rule is: do not mark Quotations for communication cues that refer to complete situations. We nevertheless mark these cues, as above.

2.2 Narrative Levels and Speech Events

Almost every literary work involves multiple *narrative levels*[2]. For example, there may be a top-level narrative the participants of which tell embedded, and in principle completely independent, stories. Examples of this type include One Thousand and One Nights, or the Canterbury Tales. In modern literature, the different narrative levels and the transitions among them may also be intentionally blurred. For these reasons, **we annotate speech events at all narrative levels**, largely ignoring narrative structure in the process.

Importantly, reports of speech and thought are standard narrative devices that introduce novel narrative levels: what a speaker says or thinks (the embedded level) is independent of the situation in which they speak or think (the embedding level). Therefore, we also annotate speech events that happen within speech:

> Peter **asked**$_{\text{Cue 1}}$: 'Why did you **tell**$_{\text{Cue 2}}$ me to get lost?'

These considerations also motivate the rule formulated below (see Section 5) that the speakers and addresses of a speech even should not be marked within the quote. The reason is that we want to characterize the entities on the level on which the speech event actually takes place, and that is the embedding level, not the embedded level.

3 Background: Communication in Frame Semantics

Our approach to annotation is inspired, in a rather loose way, by the analysis of predicate-argument structure proposed by Charles Fillmore's Frame Semantics[3]. The fundamental idea of Frame Semantics is that predicates (which can be, e.g., verbs, nouns, or adjectives) are understood by reference to ("evoke") abstract situations, so-called frames. The relevant players and props of these situations, the so-called frame elements, are in turn

---

[2]`https://wikis.sub.uni-hamburg.de/lhn/index.php/Narrative_Levels`
[3]See `https://framenet.icsi.berkeley.edu`

supposed to be realizable as arguments of these predicates. Where they cannot be realized, they are supposed to be inferable.

For example, the verb (or noun) *praise* is analysed as evoking the Judgment_Communication frame. In FrameNet, this frame is defined as follows: a "Communicator" expresses a positive or negative judgment of an "Evaluee", possibly with respect to a specific "Topic" or for a specific "Reason". The judgment can be expressed with a body part, the "Expressor", or when it is verbal, through a "Medium".

This frame provides us with an inventory of concepts with which we can analyse the textual instances of *praise*, as well as many other predicates that evoke the same frame, such as *accuse, attack, critical, commendable, commendatory, kudos, rave*, etc.[4]

In its analysis, FrameNet concentrates on the predicate-argument structure, and does not consider factuality: each instance of these predicates can be analysed in terms of the Judgment_Communication situation, no matter where an actual communication took place, or whether it was only intended, or considered, or anything similar.

In RiQuA, we are specifically interested in communication situations – and more specifically, in communication situations where we know what the message is that is being conveyed. This corresponds roughly to the Communication frame. It is described in FrameNet as

> A "Communicator" conveys a "Message" to an "Addressee" or
> A "Communicator" talks about a "Topic" to an "Addressee"

We differ in two important respects from FrameNet.

- We differ in terms of terminology: We call the FrameNet frame element Communicator Speaker in our annotation. Similarly, the FrameNet Message and Topic frame elements are lumped together into the category Quote.

- We differ in that we are somewhat more liberal in what we accept as communication (cf. Section 2.1). In contrast to FrameNet, we are fine if an event only *implies* rather than *explicitly expresses* a speech event.

4 Concrete Criteria for Span Types

Spans mark sections of the text that are of interest to our annotation task. In principle, we are not strictly interested in spans of text themselves. We want to understand the

---

[4]Modulo polysemy – an aspect that we are ignoring here.

things that are said within the work of fiction, and the characters who say these things and who listen to these things. Spans of text are really just a proxy to these utterances and characters we are interested in.

For each span type, we will first describe, at a conceptual level, what sort of "things" we want to identify – either objects, people, or events in the fictional world of the narrative. We will then describe how to map these things to a span of text to be annotated – as unambiguously as possible.

4.1 Quote: What do we annotate as the thing that was said?

A `quote` span represents the content of a speech event – the thing that was said.

Textually, quotes can be realized in one of two ways: direct and indirect quotes.

4.1.1 Direct Quotes

A direct quote occurs between quotation marks, and represents what was said verbatim.

(13)   He said `"I am hungry."`

In these cases, everything between and including the quotation marks should be marked. "Split quotations" should be annotated as if they are separate speech events – that is, as two separate annotations.

(14)   `"So,"` he said, `"it's come to this."`

Note that not everything between quotation marks is a direct quote. In the sentence

(15)   Some "groupies" were following the band.

nothing should be marked as a quote, as nothing is said.

4.1.2 Indirect Quotes

An indirect quote does not use quotation marks, and describes what was said indirectly.

(16) He said `that he was hungry` .

For speech events with no corresponding direct quote, you should try to mark an indirect quote if possible. While, for direct quotes, it is usually obvious what exactly to mark if a quotation occurs, this is not as straightforward for indirect quotes. In general, the phrase describing what was said should be marked, including any introducing words (such as the relative pronoun "that" in the above example). This may either be a dependent clause, or a shorter word or phrase.

(17) The waitress asked him what he wanted to drink. He wanted `a Coke` .

Note that such indirect quotes – in contrast to direct quotes are not necessarily verbatim records on what was said. There is a continuum between almost verbatim records to summaries and further to mere characterizations of the topic. We treat all of these as quotes:[5]

(18) He complained: "I hate this rain. It has been raining since yesterday."
(19) He complained that he hated the rain and that it had been raining since the day before.
(20) He complained about the incessant rain.
(21) He complained about the weather.

In some cases, a speech event will not have any phrase present which could be marked as a quote, either direct or indirect:

(22) Robert cried out in pain.

In these cases, do not make any annotations. Even if the speech event may have other annotatable attributes, such as speakers, addressees, or a cue, do not mark these either.

4.2 Entities: Who Participates in a Speech Event?

An `entity` , in principle, is any character or event, or group thereof, participating in communication as a speaker or addressee. While precise definitions get rather difficult philosophically, this notion can loosely be equated with sentience – any sentient being, or being portrayed as being sentient and capable of engaging in communication, is an

---

[5]In contrast to FrameNet, which distinguishes Messages from Topics, see Section 3.

entity. 99% of the time, an entity is a person, but literature also know talking dogs, cars, or even pancakes.

When marking an entity, the longest possible phrase which refers to the entity that should be annotated.

(23) <mark style="background-color:#00ff00">The man with the incredibly long beard, which nearly reached the floor</mark> , said no.

An entity will often be a noun phrase. However, pronouns can also be entities, including possessive pronouns. This may be relevant for nominal communication cues, among other things:

(24) <mark style="background-color:#00ff00">his</mark> interruption

(25) I addressed <mark style="background-color:#00ff00">him</mark> directly

Sometimes, a set of individuals will participate in a speech event together. This usually occurs when a quotation has a group of people as its addressee. In this case, the group as a whole should be considered an entity, and you should mark a noun phrase which refers to the group as a whole. If no such noun phrase exists, mark nothing — do not just mark a noun phrase that refers to a subset of that group.

(26) The king addressed <mark style="background-color:#00ff00">the public</mark> : "Good morning!".

(27) Alice asked <mark style="background-color:#00ff00">Bob and Carol</mark> for help.

5 Concrete Criteria for Relation Types

Relation types mark how individual spans relate to the speech event. In particular, entities can either be speakers or addressees of a speech act, and predicates of various linguistic forms can be cues of speech acts. Relations should be drawn from the entity spans to the quote span.

5.1 Speaker: Who is Speaking?

For a given quote, the speaker is the entity which does the speaking. Almost all quotes will have a single speaker. In theory, if multiple entities speak in unison, a quote may

have multiple speakers. A quote should always have a speaker at a conceptual level (even when the speech act is hypothetical – see Section 6.2). However, that speaker might not always be realized in the text.

Question 1: Does the Speaker qualify? As we discuss above, we only want to capture speech. Thus, documents do **not** qualify as speakers:

(28) The letter said that we could come.

Communication channels that are not agents do not qualify either.

(29) The radio / telephone / loudspeaker announced the bad news.

Question 2: Which phrase is the most prominent Speaker? When marking a speaker, you will probably have many options for which exact span of text to mark, because entities in a discourse are typically referred to many times by main different phrases on the linguistic surface (a phenomenon called *coreference*). For example, in the following sentence, all possible speakers are shown for the marked quote:

(30) Although `John` said `that` `he` didn't like `his` apartment , `he` had no intention of moving out.

The following rules (listed in order of preference) should be used to pick the most prominent phrase realizing the Speaker. While rules listed first should usually take precedence over those listed later, use some measure of common sense – none of the rules alone should be taken as absolute.

- Never mark a mention within the quote itself as the speaker. For nested quotations (quotes within quotes), you should only mark a speaker that occurs one narrative level up from the corresponding quote.

- If the quote is introduced in a sentence that syntactically specifies the speaker, choose that. Often, this means you should choose the subject of a speech verb such as "said."

- Stay within the general vicinity of the quotation – prefer a reasonable span near the quotation to a perfect span paragraphs away.

- If the quote is part of a dialogue, with other quotes sharing the same speaker, mark the same span for both quotes.

- Prefer descriptions to pronouns, and full names to descriptions. In general, prefer more descriptive phrases to less descriptive ones.

- Prefer mentions within the same sentence as the quote.

- Prefer mentions that come before the quote in the text.

If you can't find any entity to mark, then don't mark a speaker.

5.2 Addressee: Who Gets the Message?

An Addressee is a person or entity being spoken to.

(31)   She told `him` `not to worry`.

(32)   She said `"yes"` to `him`.

Many quotes will have a single addressee. However, sometimes, multiple characters will be addressed at once, and occasionally, something will be said to nobody in particular. You should try to mark a single span of text that describes all addressed characters, using the following preference rules (similar to those for the speaker). If no single span of text describes all addressed characters, then do not mark an addressee.

- Never mark a mention within the quote itself as the addressee. For nested quotations, you should only mark an addressee that occurs one narrative level up from the corresponding quote.

- If the quote is introduced in a sentence that syntactically specifies the addressee, chose that.

- Stay within the general vicinity of the quotation – prefer a reasonable span near the quotation to a perfect span paragraphs away.

- If the quote is part of a dialogue, with other quotes sharing the same addressee, mark the same span for both quotes.

- Prefer descriptions to pronouns, and full names to descriptions. In general, prefer more descriptive phrases to less descriptive ones.

- Prefer mentions within the same sentence as the quote.

- Prefer mentions that come before the quote in the text.

In some cases, there may be ambiguity as to whether a particular character is in fact an addressee of a quotation. The following guidelines should help with the decision process:

- An addressee requires intent on the part of the speaker. An eavesdropper that the speaker is not aware of is not an addressee. A speaker talking in her sleep will have no addressees in the waking world.

- Interjections, such as profanities, can raise some ambiguities about intended addressees. In these ambiguous cases, you should prefer to mark an addressee rather than not marking one. In the following example, John should be marked as the addressee to the quotation "Shit."

  (33) "We need to get out of here in 5 minutes," said John. "Shit," muttered Sam under his breath.

- Sometimes a speaker may be talking to herself, in which case she is also the addressee. In other cases, there may be no addressee, if the speaker is talking to nobody in particular. These two cases can become ambiguous. In general, if there is nothing in the text indicating that the speaker is talking to herself, mark no addressee. If the text explicitly states that the character is speaking with himself, then mark the same character as an addressee, using the rules above to pick the best span to mark.

5.3 Cue: How is the Speech Event Described?

As already discussed in Section 3, `cues` are words or phrases that describe or imply a speech event. Often, there is a syntactic relationship between the cue and the quote. In the easiest case (a verbal cue and an indirect quote), the quote is simply an argument of the cue:

  (34) Peter `said` that [he would like something to eat].

The syntactic relationship can however be more complex, and occasionally the quote is not even in the same sentence:

(35) "Touch my robe!"
      Scrooge did as he was told and held it fast. [Christmas Carol, Stave 3]

We follow a policy of **minimal cues**, that is, we want to identify the smallest expression that denotes the communication event that we annotate. This means that cues should not include adverbs, modifiers, auxiliaries (have) or modal verbs (could, would). In contrast, cues should include particles where appropriate. For example, in the sentence

(36) The suspect suddenly piped up , "It wasn't me, I swear!"

the adverb "suddenly" should not be marked, but the particle "up" should.

Cues can also be multi-word expressions (MWEs). One typical case are so-called *light verb constructions* which, in our case, combine a communication noun with a rather general verbs that can contribute, for example, aspectual information.

(37) Peter launched into a passionate defense : [...]
(38) Peter tried another argument : [...]

For light-verb constructions (i.e., where the noun is a communication event on its own), we only annotate the head noun.

Another type of MWEs is formed by non-compositional metaphors/idioms. These generally do not denote communication events per se, but communication can be inferred from them:

(39) Peter directed his anger at Michael: [...]

Since in these cases, the whole verb phase is necessary to understand that a communication took place (just the noun *anger* does not suffice), we annotate the whole phrase as cue.

Again, we only annotate MWEs as cues when the meaning is non-compositional. Here is an example that we would still see as compositional, since the speech event is essentially introduced by the communication verb:

(40)   All I can `say` is, [I should like to know him too]. [Christmas Carol, Stave 3]

Not all quotes have a cue. If no cue is present, simply don't label anything.

For split quotations, both parts of the split quote should share the same cue. In some cases, it might be ambiguous as to whether two quote spans form a split quotation, which should share a cue, or whether the first is a cued quotation and the second is an uncued quotation. When in doubt, you should assume that the two spans form a split quotation, and have them share a cue.

6 Global Considerations

6.1 The Speech Modality: Speech vs. Thought, Soliloquies

In RıQuA, we decided to annotate only speech events, that is, spoken messages. We do *not* generally annotate thoughts that are not uttered aloud, nor writing. For the same reason, we do not annotate so-called free indirect speech:

(41)   Had he heard something? Yes, he was fairly sure that he had.

This puts soliloquies, where a person speaks with themselves, at the decision boundary. We annotate such cases when it is clear, or at least highly probable that the communication was spoken out loud rather than just thought:

(42)   "Why, it isn't possible," said Scrooge, "that I can have slept through a whole day and far into another night. It isn't possible that anything has happened to the sun, and this is twelve at noon!"

Linguistic markers that provide evidence for speech are:

- Direct quotes as opposed to indirect quotes (weak evidence)

- Manner adverbs: *he said loudly* (strong evidence)

In the following example, it is not clear enough in our opinion that an actual speech event took place, so it should not be annotated:

(43)   Peter asked himself whether the whole thing was a good idea.

6.2 Counterfactuality

In line with FrameNet's analysis, all communication situations should be marked, regardless of their (counter)factuality. As we are annotating works of fiction, nothing is truly "real" anyway, as the people and events are made up by the author. Even within the narrative, there may be different narrative levels (e.g., a character telling a long story, within which further communication takes place).

In addition, language itself provides a number of devices that can make a communication situation 'unreal'. These include negation and modality:

(44) Peter said: "I like it".

(45) Peter did not say: "I like it".

(46) Peter may have said: "I like it".

Note that modality can also be expressed by adverbial modifiers (*maybe, probably*), which introduces a large gray area – another reason not to take factuality into account.

However, note that for this reason we should not claim that we annotate speech events – because these should be factual.

6.3 Mottos, staging directions, etc.

Sometimes, novels contain material that is not directly part of the narrative, such as mottos. Following our considerations regarding narrative levels, these are not to be annotated. Mottos are typically at the beginning of sections/chapters.

The same consideration holds for staging directions:

(47) FRANCISCO: "I think I hear them. Stand, ho! Who's there?"
      [Enter HORATIO and MARCELLUS] (Shakespeare: Hamlet)

The mere presence of square brackets is not sufficient to discount text, though: square brackets can also be used as stylistic devices:

(48) "Never heard!" [This was given in a scream.] "Bless my soul! why, he's the man-"
      (Poe: The man that was used up)

In this case, the text in square brackets is descriptive and can be annotated.

6.4 Foreign language material

The treatment of foreign language text depends on its function in the text. We have not come across foreign language text as part of the main narrative itself. If foreign text serves as (e.g.) a motto, we do not annotate it for that reason. In contrast, Quotations in foreign languages are fair game:

(49)   the writer *shovels in* "haben sind gewesen gehabt haben geworden sein," or words to that effect (Mark Twain)

7 Change log

5-11-18-20-00 Restructured document and added sections on global considerations, background (FrameNet), relation types [SPó]

12-11-18-20-00 Modified precedence rules for speakers and addressees to more highly prioritize names and descritions over pronouns, clarified cues and interjections. [SPy]

30-11-18-21-00 Changed document class parameters. Changed example macro. Clarified modality. Removed term 'speech act'. Moved discussion of what constitutes a speech event to background section. Generally revised the sections about criteria for spans, relations. Added table of contents. [SPó]

6-12-18-20-00 Removed the requirement for "communicational intent" for speech acts. Modified the rules for selecting speakers and addressees. Changed instructions for when there is no markable quote. Added a note on talking to oneself versus talking to nobody.

11-12-18-16-00 Added subsection on narrative levels / speech events. Added discussion of multi-word cues in Cue subsection.

9-1-19-01-00 Modified precedence rules slightly to prefer speakers and addressees that are close to their quotes. Clarified issues with nested quotes, re-defined rules for multiple addressees.

15-1-19-17-00 Clarified rules for cues without quotes (Section 2.1.4); added sections 6.3 and 6.4.