



UNIVERSIDADE  
LUSÓFONA

# Annotation Manual for Annotators

*Final Version*

**Luiza Vidal Copolillo Coelho**

*DEISI 89*

**Advisor:** Zuil Pirola, Assistant Professor – Universidade Lusófona

**Co-Advisor:** Manuel Pita, Assistant Professor – Universidade Lusófona

**April 2024**

## Table of Contents

Table of Contents .....	2
1 Introduction.....	3
2 Task Definition .....	4
3 Environment Setup .....	5
3.1 BRAT Installation .....	5
3.2 Accessing BRAT and its features.....	5
3.3 Annotation Rules.....	9
3.4 Final Notes .....	11
4 Guidelines .....	12
4.1 What to annotate?.....	12
4.2 Turns .....	12
4.3 Argumentative Language .....	12
4.4 Argumentative Turns .....	12
4.5 Argument Markers.....	12
5 Additional Resources.....	14

# **1 Introduction**

This manual is crafted to guide the annotators through one of the most important phases of the project, the Annotation Phase. This phase is particularly complex giving the nature of the data we are working with, the process of annotating conversational datasets and identifying whether what has been said contains argumentative language or not has its nuances and complications. The annotated corpus will serve as input for the development and training of a supervised machine learning model that will be capable to understand, analyze and classify unseen data. This document aims to address a comprehensive understanding of the project's annotation task (subsequently explained and defined), annotation methodologies and standards, thereby ensuring consistency, accuracy, and reliability across all annotators, enhancing the Inter-Annotator Agreement.

## 2 Task Definition

The task demands a precise approach, attention to detail and a profound understanding of the characteristics that define argumentative language (section ‘Guidelines’) in conversational texts. The core of our annotation tasks lies in identifying turns within conversations that are characterized by argumentative language.

- ⇒ **Pilot Annotation:** Starting with a pilot annotation, the annotators will be given 5 rooms, each with different topics for discussion among the students. This phase is considered complete once  $\alpha \geq 0,7$ , that will determine the feasibility to proceed to the main annotation phase.
- ⇒ **Main Annotation:** Expanding upon the pilot, the main annotation involves a broader corpus, with more discussion rooms, maintaining the requirement for a high inter-annotator agreement to consider the annotation phase complete and continue with the model training.

## 3 Environment Setup

### 3.1 BRAT Installation

Make sure you have BRAT (Brat Rapid Annotation Tool) correctly installed in your system. If you don't, access <http://brat.nlplab.org> to download and follow installation instructions to continue.

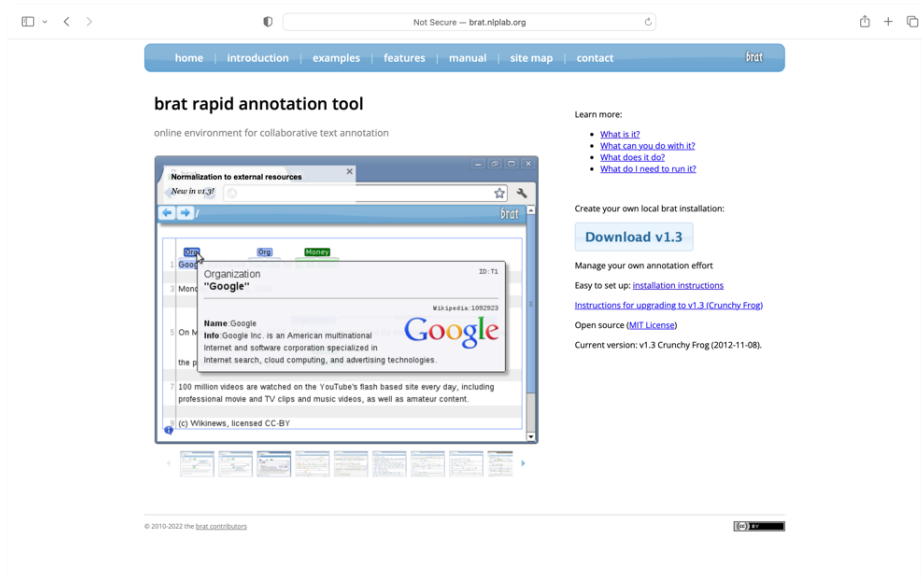


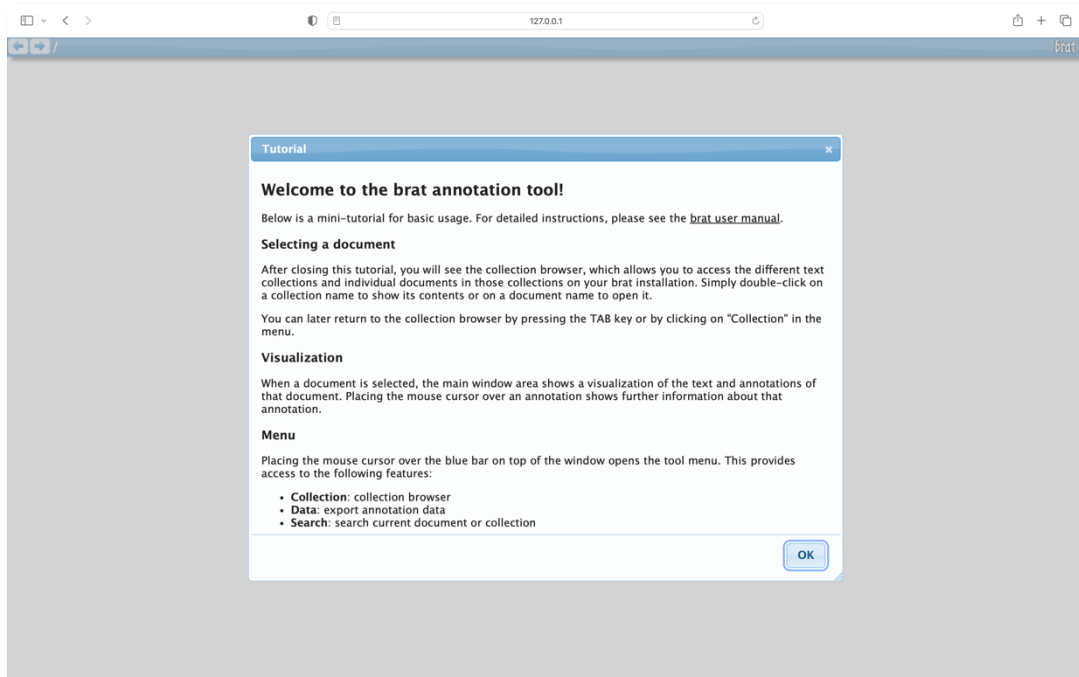
Figure 1 - BRAT website home page.

### 3.2 Accessing BRAT and its features

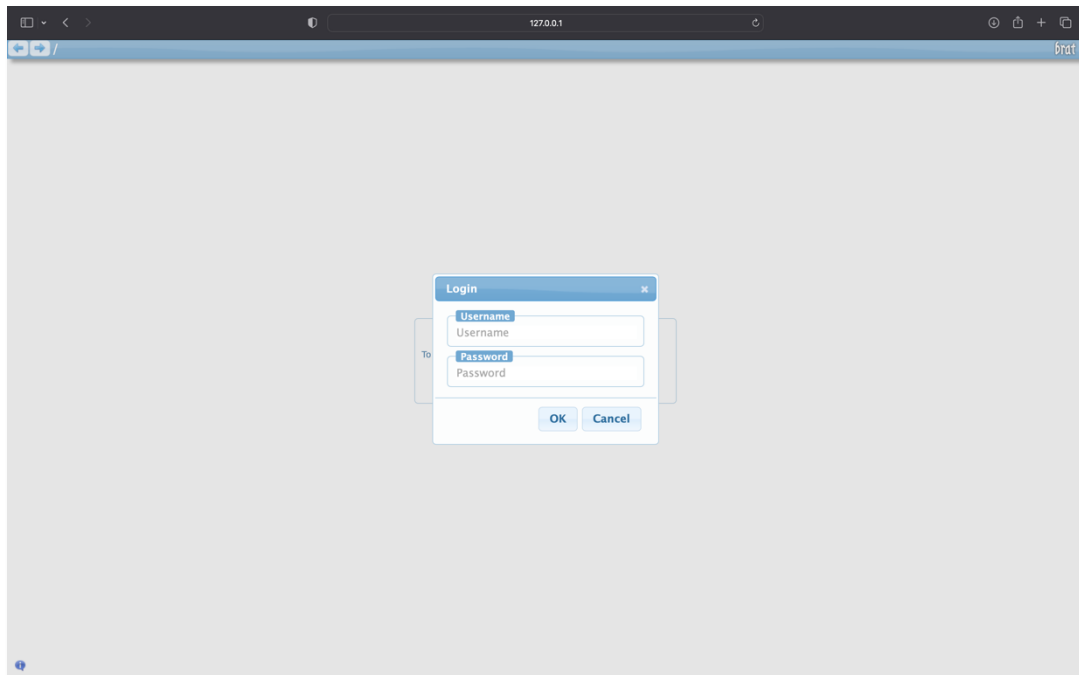
→ Open your terminal, access brat's local directory, and run the following command:

```
Terminal Shell Edit View Window Help
brat-1.3p1 - Python standalone.py - 80x24
luizacopolillo@Luizas-MacBook-Pro ~ % cd Documents
luizacopolillo@Luizas-MacBook-Pro Documents % cd brat-1.3p1
luizacopolillo@Luizas-MacBook-Pro brat-1.3p1 % python2 standalone.py
Serving brat at http://127.0.0.1:8001
```

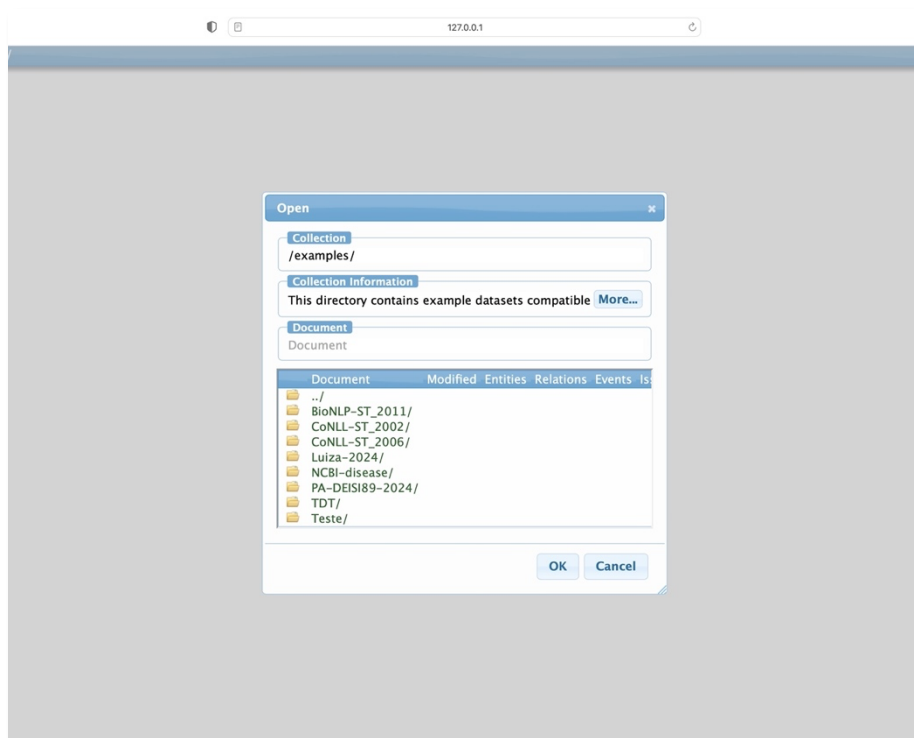
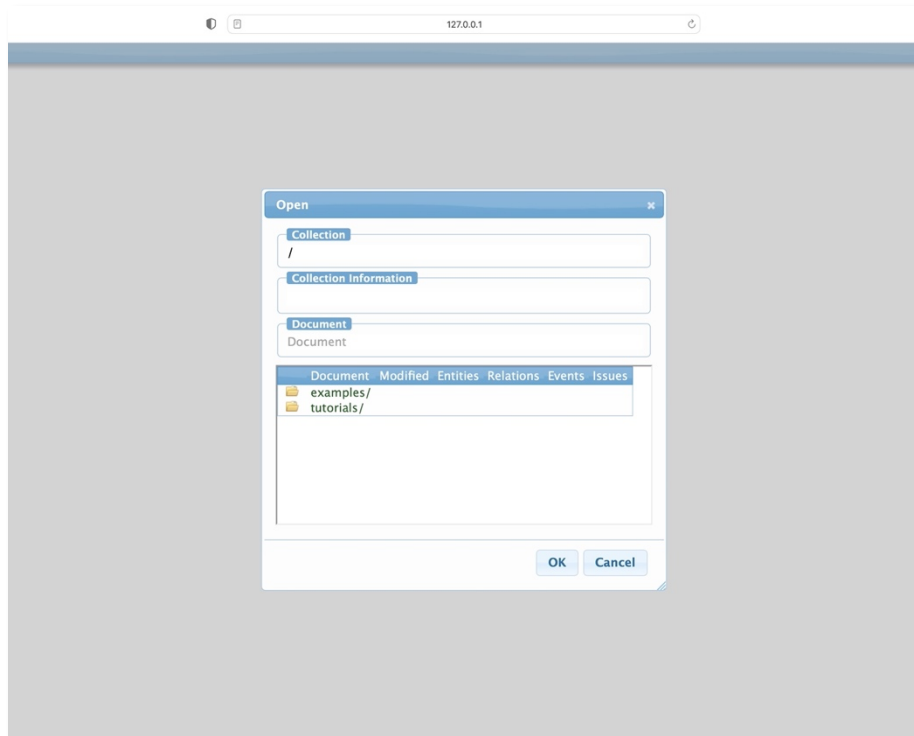
→ By executing the previous command, open the link provided to start brat in your browser.

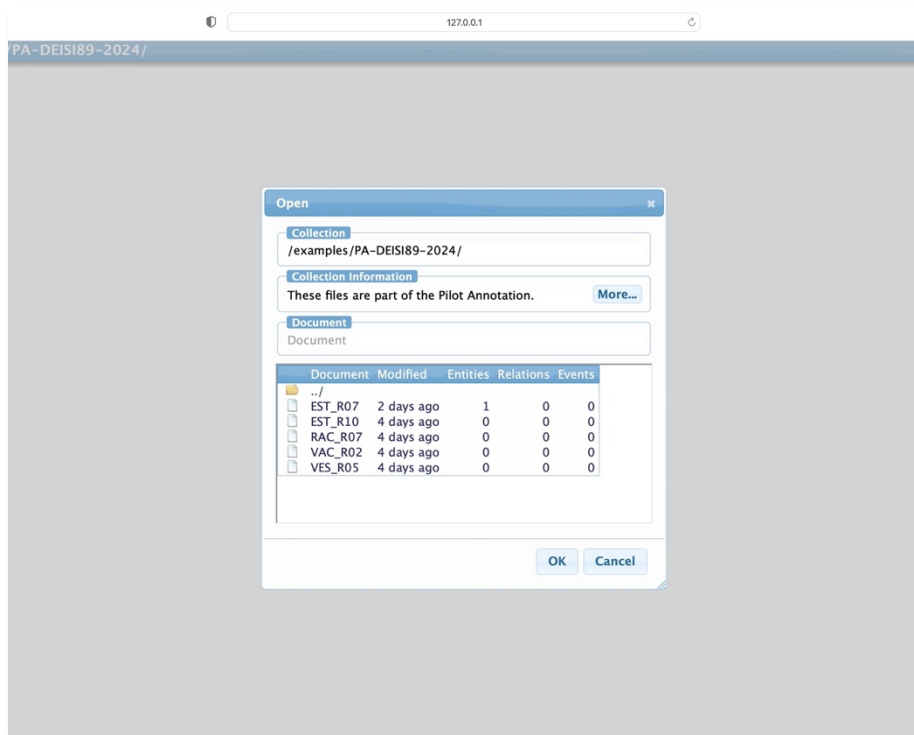


→ In the menu, select 'Login' and use your credentials.

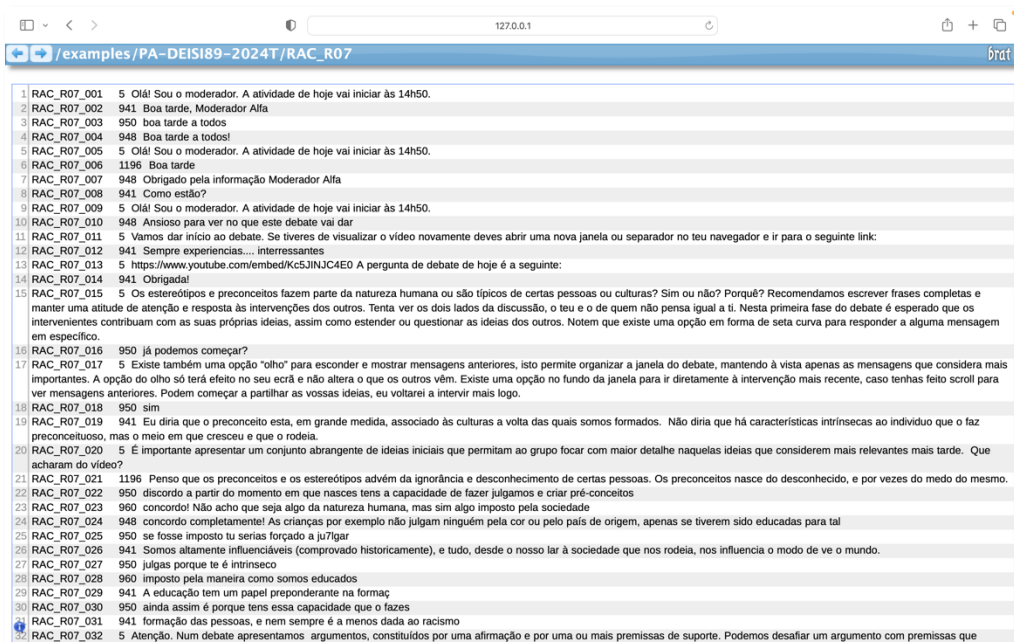


→ In the Menu, select 'Collections'. Then 'examples' > 'PA-DEISI89-2024' > 'file-name.txt'.





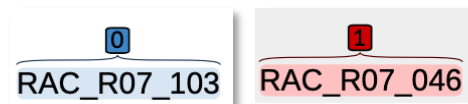
→ When opening one of the files, your interface should look like this:





→ When selecting a text span, you will have this pop-up window:

- To start annotating, select the text span you desire and the type of annotation. In the case of this project, there will be two entity labels for classifying an annotation:
- Label '0': for turns that are argumentative and include a claim but do not include a premise.
  - Label '1': for turns that are argumentative, and include both a claim and a premise, forming a complete argumentative structure.
- Consider the following examples:



### 3.3 Annotation Rules

- **Relevance to the Room Topic:** Annotate only those turns where the statement is directly related to the topic of discussion.
- **Spelling Errors:** If there are spelling errors, but the turn is still comprehensible, proceed with annotation.

941 A educação tem um papel preponderante na formaç

950 este3riotipos são algo que nos deixa confortavel

- **Figures of Speech and Language Vices:** Be cautious with expressions that include figures of speech or language vices, as they can be difficult to interpret and categorize into specific entity labels; such turns should not be annotated.

941 e ideia de "outro" tem uma conotação negativa associada a si.

- **Anaphora:** Anaphora can complicate the determination of whether a claim is being made. If resolving the anaphora depends on prior turns, we should not annotate it.

811 Esse ponto de vista não é válido já que se todos pensarmos dessa maneira, ninguém se vacina.

941 Diria que existe sempre, mas que há, definitivamente, culturas mais racistas que outras.

- **Order of the Argument Components:** The sequence of claim and premise (or premise and claim) can be confusing and hard to classify but should be annotated regardless of the order of components.

941 sim, ate que cesce e se expoe as noticias, a internet, le livros e educa-se de forma diferente. Uma criança fruto de uma familia racista pode nao o ser.

premise claim

- **Turns Containing Only Premises:** Turns that contain only premises and have argument markers starting the sentence should not be annotated, as they do not contain a claim.

941 Porque a educação varia de pais para pais

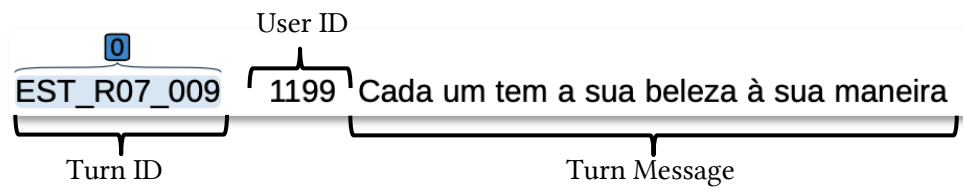
1996 Isso não justifica assediar alguém pela forma como se veste.

- **Non-conventional vs. Common ArgMarks:** Familiarize yourself with the list of common argument markers provided in Section 4.6, as these are typically easy to identify. Be careful with turns that use non-conventional markers to denote arguments, as they can be harder to identify and require cautious handling.

941 por vezes nao. ha paises que formam esteriotipos de outrso (nacionalidades) nao querendo isso dizer que esse esteriotio seja verdadeiro.

### 3.4 Final Notes

→ When annotating, select only the **turn ID**.



→ Do not annotate turns from User '5' and User '1280', these users are the room mediators.

*It is crucial that you follow these steps and the following guidelines, annotations must be consistent across all annotators to ensure the Inter-Annotator Agreement (IAA) is accurately calculated.*

## 4 Guidelines

### 4.1 What to annotate?

As stated in the project report, the core task of our annotation process is to **identify turns within multi-party dialogues that contain argumentative language**. You will focus on distinguishing between argumentative and non-argumentative discourse at the level of conversational turns, rather than breaking down each part of an argument into claims and premises. This approach aims to streamline the annotation process and enhance the Inter-Annotator Agreement (IAA) by providing a clear and manageable task for the annotators.

### 4.2 Turns

In the context we are studying, a turn is defined as a segment where a single student contributes to the dialogue before another participant takes over. Each turn is a potential carrier of argumentative content and should be examined for the presence of such language. **You will label entire turns as argumentative or non-argumentative** based on the criteria set forth, thus capturing the debate's dynamics within the structure of the conversation.

### 4.3 Argumentative Language

Argumentative language involves **constructing and expressing a point of view with the intent to persuade or inform**. It is characterized using claims, sometimes supported by premises and, even, evidence. You must be capable at discerning the intent behind the language used, distinguishing factual statements from opinions, and identifying the persuasive elements within a turn.

### 4.4 Argumentative Turns

With the previous definitions of turns and argumentative language, we can define an argumentative turn as an exchange that involves claims, either alone (forming an incomplete structure) or accompanied by premises (constituting a complete structure). These claims and premises are used in a sequence (no matter the order) aimed at supporting or refuting a central statement, typically through reasoning or presentation of evidence.

It is not merely the expression of an opinion but a structured attempt to sway the audience or interlocutors. **Your task is recognizing these argumentative structures within the flow of conversation and marking the turn accordingly.**

### 4.5 Argument Markers

Argumentation markers or '*ArgMarks*' are linguistic cues that concentrates the rhetorical or logical argumentative role of a turn, such as providing reasons, justifications, counterarguments, or limitations. Below, there are six categories, and next to each category is provided examples of such words or expressions.

1. **Justification:** 'agora que', 'conforme', 'consoante', 'dado que', 'devido a', 'em virtude de', 'graças a', 'já que', 'ora', 'para', 'para que', 'pois', 'pois que', 'por causa', 'porque', 'posto que', 'segundo', 'uma vez que', 'visto que'.
2. **Concessive:** 'ainda assim', 'ainda que', 'apenas', 'embora', 'mesmo que', 'por outro lado', 'se bem que'.
3. **Conclusions:** 'assim', 'assim sendo', 'de modo que', 'derivado de', 'deste modo', 'isto é', 'logo', 'pode ser visto', 'por conseguinte', 'por consequência', 'por isso', 'portanto'.
4. **Contrast:** 'a pesar de', 'contudo', 'em contrapartida', 'enquanto', 'entretanto', 'mas', 'no entanto', 'porém'.

5. **Further-support:** 'adicionalmente', 'além disso', 'além do mais', 'assim como', 'bem como', 'como', 'da mesma forma que', 'igualmente', 'já', 'não só', 'ou', 'por exemplo', 'quanto mais', 'que', 'que mais', 'tal', 'tal que', 'também', 'tanto como', 'tanto que', 'tão que'.
6. **Conditional:** 'a menos que', 'assim que', 'caso, desde', 'desde que', 'logo que', 'qual', 'quando', 'quer', 'se', 'se acaso', 'seja', 'à medida que'.

## 5 Additional Resources

- Stollenwerk, F., Öhman, J., Petrelli, D., Wallerö, E., Olsson, F., Bengtsson, C., ... & Gandler, G. Z. (2023). Text Annotation Handbook: A Practical Guide for Machine Learning Projects. *arXiv preprint arXiv:2310.11780*, <https://arxiv.org/abs/2310.11780>.
- <http://brat.nlplab.org>