SDRT@DailyDialog Annotation Guideline

Chuyuan Li

January 16, 2022

1 Objective

This is a practical exercise for the course Discourse and Dialogue.

The objective is to use discourse theory SDRT [3] to annotate on DailyDialog dataset [4].

Everyone needs to annotate 3 pilot dialogues (the same ones) and 10 different dialogues individually. There are 2 examples (Section 6) to help you getting used to the annotation process. It is highly recommended that you go through these examples carefully before starting your work.

Bear in mind: annotation work reflects your own reasoning on the understanding of dialogue structures. You should NOT communicate with others during annotation.

2 Hand-ins

Your work is expected to be handed in before **Feb.15th 2022**.

2.1 Package received

Each package is indexed with an number, such as "anno-01". In the package you received, you will find:

- a `txt` file with 13 blocks of dialogues to annotate

- a `html` file which serves as your local annotation platform. You will double click on this file or choose to open it with **Chrome** or **FireFox** web browers for a better presentation. You will learn how to use this platform in Section 3.1.

2.2 Expected return

It should be 13 `json` files. These files will be automatically generated from the platform presented in Section 3.1. You need to put all these files in a `zip` and named with "packagenumber-yourfamilyname-yourfirstname.zip", all lowercase.

3 Annotation Process

3.1 Platform presentation

Figure 1 is your annotation platform. It is consists of 4 parts:

- **Input:** where you put a dialogue text. You will receive a text file in which contains 13 dialogue blocks (3 in common so everyone annotates the same and 10 different ones). For each annotation, copy paste **one block** of text inside this box and press `enter`. The text will be shown automatically in the `Annotation` area. Remember to copy the first line of each block, which is the dialogue ID.

- **Operations:** where you choose actions (ADD EDGE, REMOVE EDGE, or JOIN NODES) and label types. ADD EDGE enables you to add a directed arc between two `EDUs`; REMOVE EDGE enables you to delete an arc; JOIN NODES enables you to combine two `EDUs` into one. See Section 3.2 for details.

- **Annotations:** where you create SDRT graphs, your main working part. You will consider each `EDU` as a node, with speaker ID and `EDU` index at the beginning (e.g.: "A1", "B2"), followed by the text. You will make decisions on how to link these nodes with which label type. At the end of annotation, every `EDU` should be linked with at least one other `EDU` (i.e., no `EDU` should be isolated, so the number of **orphans** should be 0.). And each `EDU` should have an **incoming link**, meaning that it has a **head**. Every pair of `EDUs` should be given one relation type except for the ones with JOIN. You can drag the nodes for a better view. You can also zoom in/out. Click on the blank area and drag can move the graph together.

- **Result:** where shows your annotation result as a `json` file. You don't need to mind this part. When you finish annotating, click on **Download** button to download the `json` file. Please don't modify this file for that it is generated automatically.
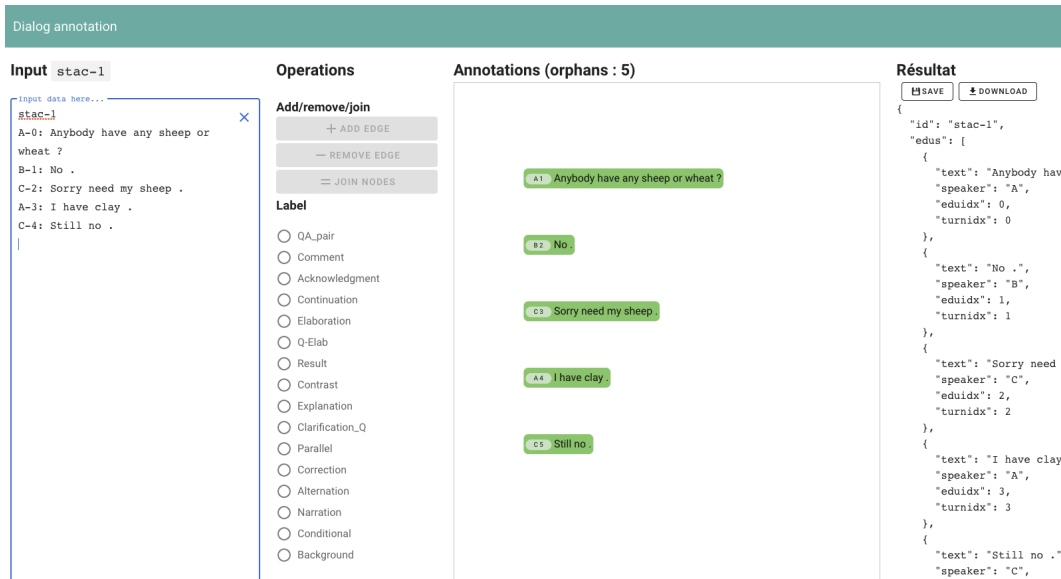

Figure 1: Annotation platform.

3.2 Actions

ADD EDGE: You will follow this order to create an edge: first, click on two nodes (they will be highlighted), then choose one label in the **Label** part, and in the end click on `+ADD EDGE`. As shown in Figure 2.
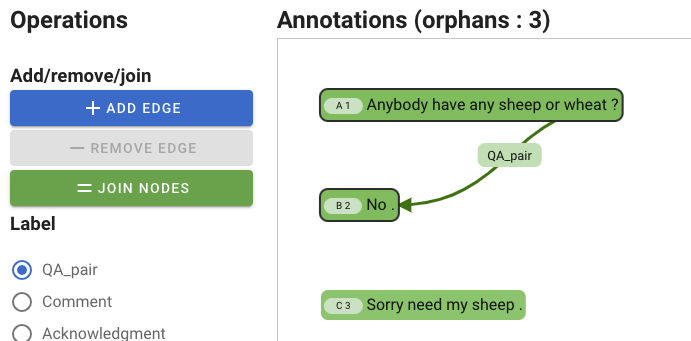

Figure 2: Action: Add an edge.

2

REMOVE EDGE: Whenever you wish to modify your previous annotation, click on the label of two linked edges (as shown here "Comment") and then click on `-REMOVE EDGE`. The link as well as the relation type (if given) will be deleted. As shown in Figure 3. You can also just change the name of the label instead of remove the edge. In this case just click on the label ("Comment") and then choose another label in the left box.
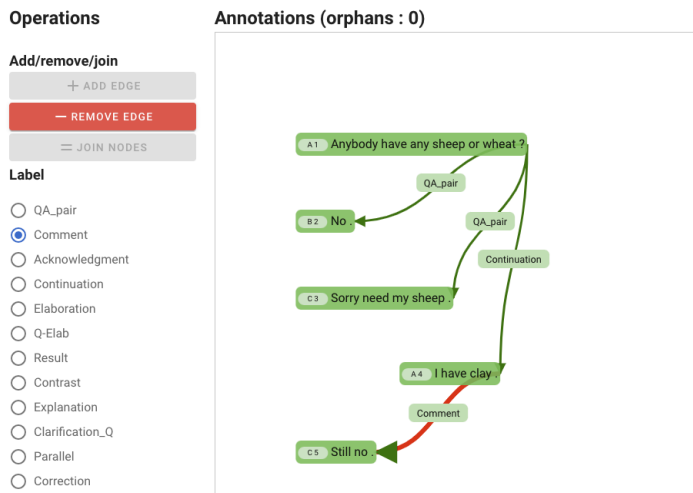

Figure 3: Action: Remove an edge.

JOIN NODES: You can also choose to join two texts where you consider they shouldn't be separated, in that way just click on two `EDUs` and then click on `JOIN NODES`, no need to choose a **label**. As shown in Figure 4.
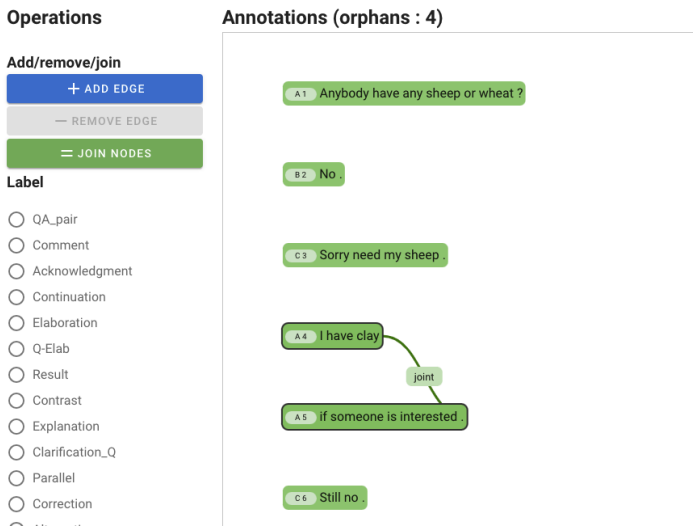

Figure 4: Action: Join nodes.

3.3   Important notes

Link direction: In the same vain of [1], we consider that dialogue moves are sequential, and thus imposing a constraint on the attachment process: in dialogue there are **no "backwards"** rhetorical links. Thus you do not need to worry about the direction of the link when you select two `EDUs`, it will almost pointing from previous `EDU` to the latter one.

Discourse unit (`DU`) simplification: Normally, the annotation process begins by segmenting each text or dialogue into a set of what we will call *elementary discourse units* or `EDUs`, which serve as the basic building blocks of discourse structures. These `EDUs` can also be added to a group of discourse units that collectively provide the argument to a discourse relation. The units that work together to provide a discourse argument form what is called a *complex discourse unit* or `CDU` in SDRT [2]. Discourse units comprise not only `EDUs` but also `CDUs`. To simplify the annotation process, we have already separated **speech turns** (the sentence(s) produced by one speaker before switching to another speaker) into `EDUs`, and we will only consider linking these units.

Join two `EDUs`:   we provide an option to `join` (shown as "=JOIN NODES") to link **two adjacent `EDUs`** whenever you think that the segmentation is incorrect. For instance: if you see these three adjacent `EDUs` from speaker A and B:

   A: I suggest a walk over to the gym
   A: where we can play pingpang and meet some of our friends.
   B: Sounds good.

   You would consider that the first two belong to the same `EDU` because they convey one complete meaning and that there is no relation type in between. **Please use this link with caution.**

4   Evaluation Metric

Following [1], we use an **Exact Match** criterion of success for the 3 pilot dialogues, the inter annotator agreement score is a Kappa score calculated for attachment on structures and on labelling for annotated dialogues, which will be calculated later[1]. You don't need to worry about the calculation. We will use this score to examine the annotation quality.

5   Relation Types

Table 1 gives the 16 relation types and their definition for your reference. For detailed explanation and examples, you can check STAC annotation manual here (starting from section 4.1, page14): `https://www.irit.fr/STAC/stac-annotation-manual.pdf`.

Labels	Description (from target `EDU` to source `EDU`)
QA_pair	gives answer to a question
Comment	provides opinion or evaluation of the first EDU, emotions also indicate comments.
Acknowledgment	signals an understanding/acceptance, such as *OK, right, good, fine, etc.*
Continuation	holds between two EDUs when they both elaborate or provide background to the same segment
Elaboration	provides more information introduced in the first EDU
Q-Elab	provides a question that elaborates
Result	describes the result, *so* is a good marker
Contrast	compares to emphasis the differences, *but, however, nevertheless, etc.* are cues
Explanation	describes the cause, *because, since* are important cues
Clarification_Q	tries to clarify what was said
Parallel	links similar/analogous contents, share a common theme, cues such as *too, also*
Correction	provides a correction/negation
Alternation	provides a disjunction
Narration	describes an event occurs afterwards, *then* is a cue
Conditional	gives hypothesis and consequence, such as *if...then*
Background	provides a stage setting for the event that happens in first EDU

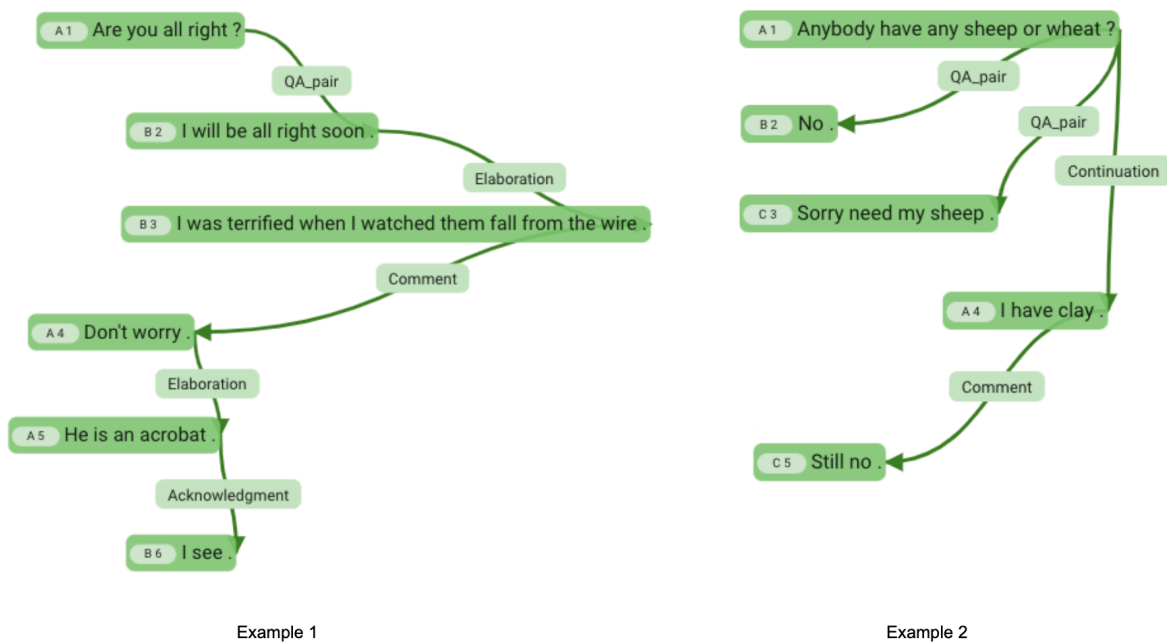Table 1: List of rhetorical relations and their definition.

Figure 5: Complete annotation example 1 and 2.

6 Examples

As shown in the second example in Figure 5, the links are not always adjacent. There could be links across longer distance (such as from `A1` to `C3`, `A4`).

7 Check List

Before you hand in your work, please check if you finish all these steps:

- 3 pilot dialogues completed

- 10 individual dialogues completed

- For each dialogue, no isolated `EDUs`

- For each dialogue, every pair of linked `EDUs` has one relation type.

- in total 13 `json` files compressed in zip which is named as "packagenumber-yourfamilyname-yourfirstname.zip".

---

[1]For reference, in STAC corpus annotation, the Kappa scores for structure attachment and labelling are 0.72 and 0.58 respectively.

References

[1] Nicholas Asher, Julie Hunter, Mathieu Morey, Farah Benamara, and Stergos Afantenos. Discourse structure and dialogue acts in multiparty dialogue: the stac corpus. In *10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2721–2727, 2016.

[2] Nicholas Asher, Julie Hunter, and Kate Thompson. Modelling structures for situated discourse. *Dialogue & Discourse*, 11(1):89–121, 2020.

[3] Alex Lascarides and Nicholas Asher. Segmented discourse representation theory: Dynamic semantics with discourse structure. In *Computing meaning*, pages 87–124. Springer, 2008.

[4] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995, 2017.