

Topic segmentation & labeling

In the first annotation assignment, we asked you to segment a telephone conversation into subtopics and provide a label for each topic. In this report, we're interested in your reflections regarding these annotation tasks.

The questions are intended for you to discuss your insights in group, so you should all engage in the discussion of each question. However, you don't need to reach agreement on every answer. When more than one point of view/idea come up, feel free to include them all. Please include **a contribution statement** at the end of the assignment.

Please make sure to include all team members' **full names + UBIT Name**.

Please submit only **one report** per group in **pdf format** on UBLearn.

I) Segmenting and labeling topics (5p)

1. In your experience segmenting the text into topics, how clear were topic boundaries? What information did you use to help you identify boundaries? What was hard about identifying boundaries?
2. How much overlap in topic segmentation did you find there was across annotators?
3. To what extent do you think topics can be found in adjacent segments? Do you think this could vary across different text types?
4. What criteria did you use to come up with labels for topics?
5. Did you all come up with similar labels? For which topics did you find the most/least agreement?

II) Annotation and Machine Learning models (4p)

1. If we were to set up a human-in-the-loop model, where the model queries an annotator about specific annotation instances, how could a statistical model help annotators make decisions?

2. What do you think are the biggest challenges for annotating language datasets for ML models?
3. What kind of model do you think would benefit from the data you annotated?
4. How would you (partially) automate these tasks?

III) Annotation interface (2p)

1. Discuss advantages and disadvantages about the way the tasks were set up. You may consider:
 - a. An alternate setup for segmentation in which longer sections of the conversation would be shown per screen. Would this be beneficial? How would you decide how many turns to show? What would be possible effects on annotators?
 - b. An alternate setup for segmentation in which the task is other than a binary decision task (same vs new/different topic). What could that task be?
2. How would you improve the annotation interface?