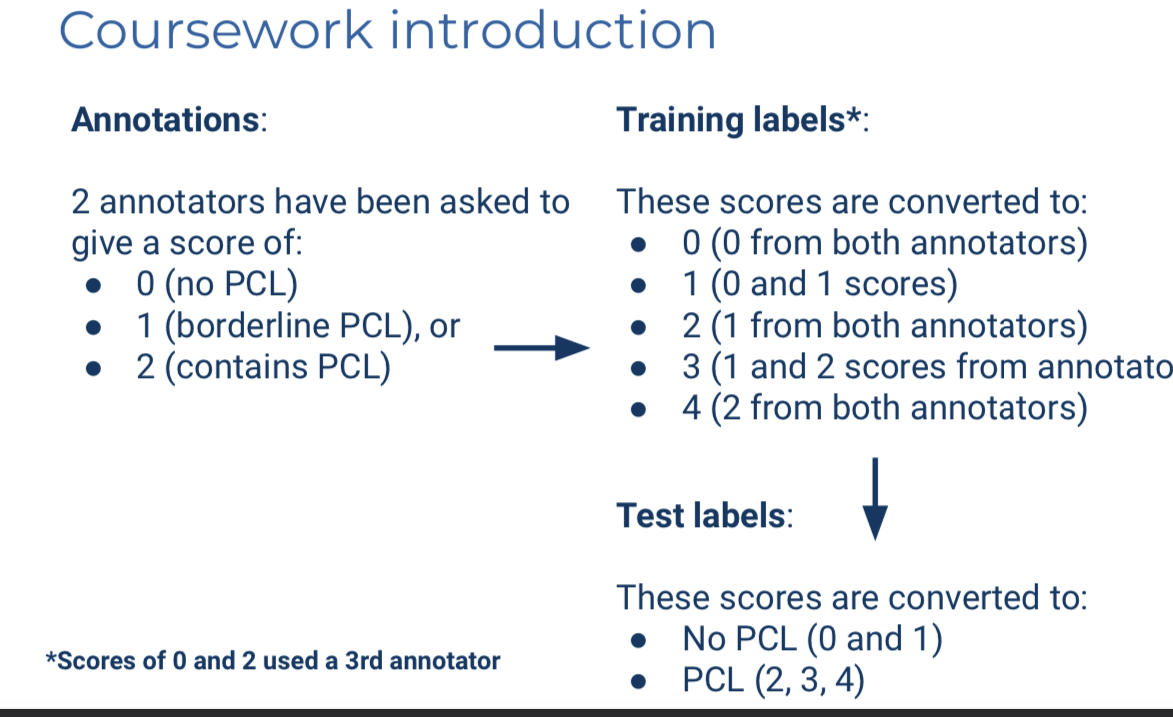
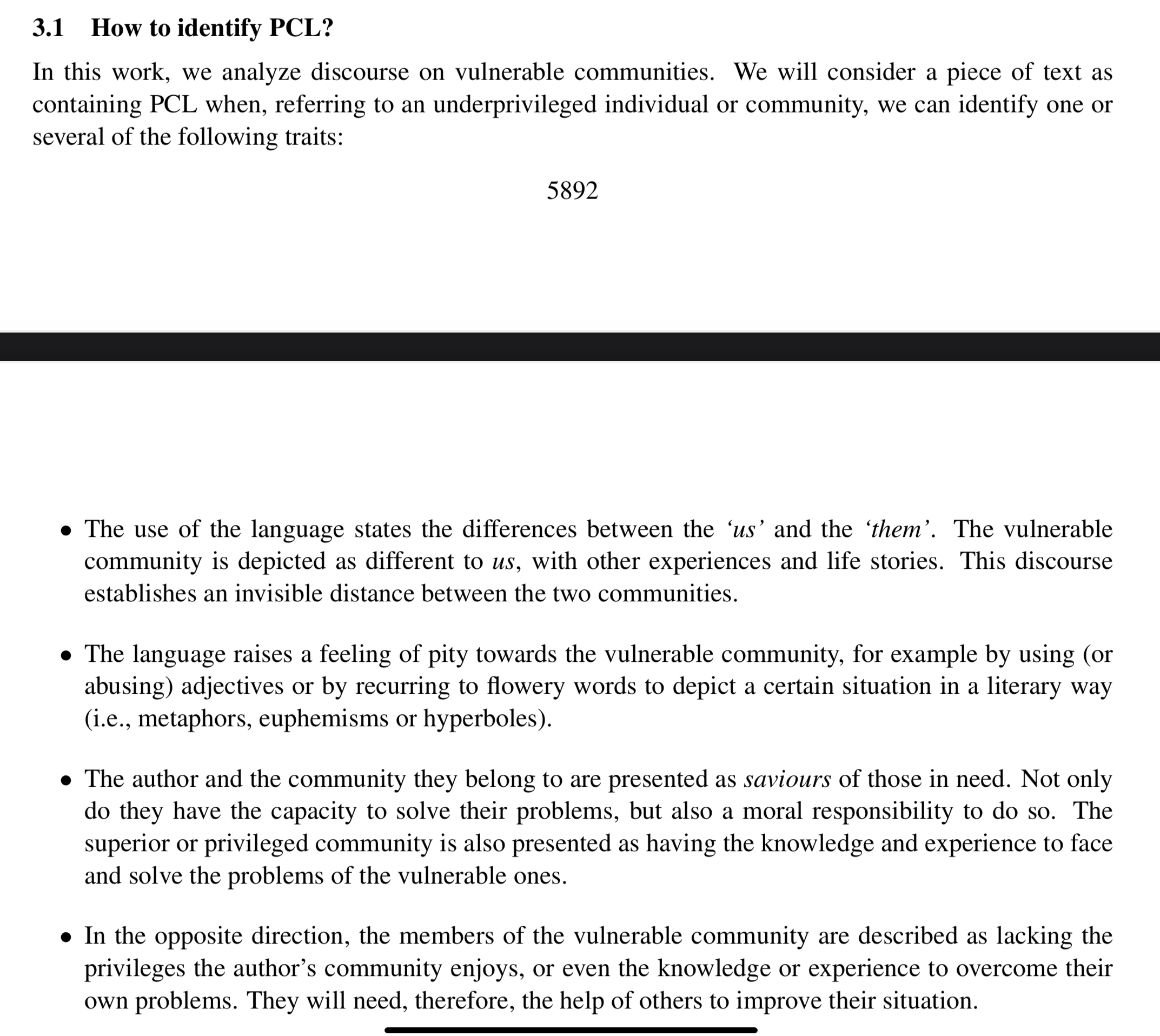
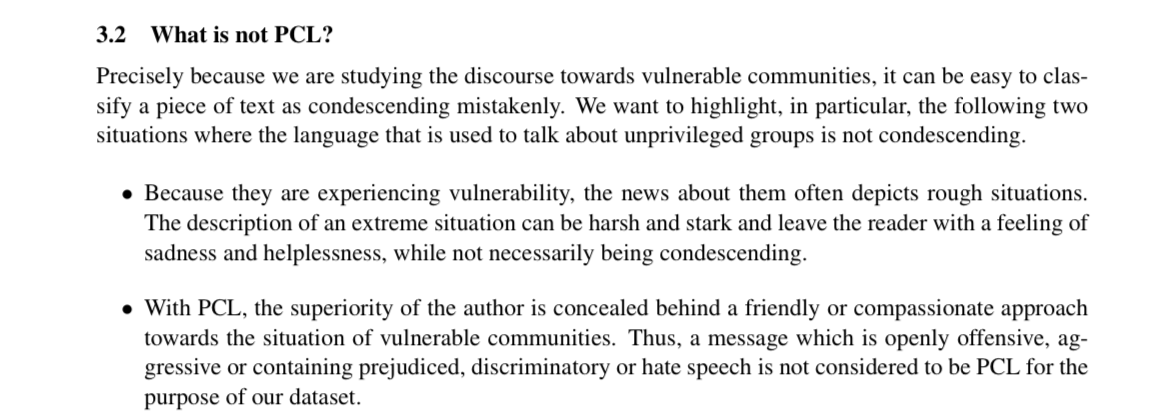
Quick ideas:

* 1. We can try debiasing at some point (e.g. with tiny BERT)
  2. 10-fold cross validation

1. There is also category and country code in the data (imo should be treated as input as well)
2. Labels define the “level” of patronizing as by 2 experts:
   1. 
3. Want to outperform the RoBERTa-baseline
4. Marked based on the report
5. On GitLab (LabTS):
   1. Code
   2. Report (maybe only on scientia)
   3. Outputs of the model
6. We get a train set which we then split into dev and internal train
7. Data analysis ideas:
   1. Plot labels vs length
   2. Labels vs country
   3. Labels vs category
   4. (3d plots of those) or Labels vs category given country
   5. Frequency of each label (we can then have weighting for performance) - how many positive vs negative ( >1 label)
8. Qualitative analysis:
   1. WHY THE TASK IS HARD/SUBJECTIVE
      1. <https://aclanthology.org/2020.coling-main.518.pdf>
      2. Some reasons for this might include that the use of PCL in the media is commonly unconscious, subtler and more subjective than the types of discourse that are typically targeted in NLP.
      3. Within a broader setting, there has been some work on PCL which is concerned with the communica- tion between two parties, where one is patronized by the other, such as in social media interactions. In particular, Wang and Potts (2019) recently published the Talkdown corpus for condescension detection in comment-reply pairs from Reddit. In this work, the authors highlight the difficulty of the task and the need for a high-quality dataset annotated by experts, which is the approach we take for studying PCL towards vulnerable communities.
      4. In particular, while some forms of PCL can be detected by identifying relatively simple linguistic patterns, many other cases seem to require a non-trivial amount of world knowledge.
   2. Identifiable features of PCL:
      1. 
   3. What is not a PCL:
      1. 
   4. Methods:
      1. To this end, we first selected ten keywords related to potentially vulnerable communities widely covered in the media and susceptible of receiving a condescending or patronizing treatment: disabled, homeless, hopeless, immigrant, in need, migrant, poor families, refugee, vulnerable and women.
      2. Categorization of the PCL (savior, expert, poet, other)