

# Corpus Creation

## Emotion Analysis – Assignment 1

### Data Selection

For the task of annotating emotions, we considered different amazon review corpora and several reddit threads. The amazon reviews were discarded as they mostly contained sentiments rather than emotions. A possible solution of filtering out 'neutral' reviews with a medium number of stars would distort the observable emotions. Thus, the distribution in the corpus would not contain all possible emotions, which makes the corpus unsuitable for an exemplary emotion annotation. This also lead to the elimination of most reddit threads, as their domain often primed the users for a specific subset of emotions. We concluded to elect the reddit thread *Redditors who live with their SOs, what changed?* for our annotation. The topic of this thread prompts users to describe emotional experiences and the question for *change* also includes negative experiences. The data was scraped from [https://www.reddit.com/r/AskReddit/comments/5nrh7b/redditors\\_who\\_live\\_with\\_their\\_sos\\_what\\_changed/](https://www.reddit.com/r/AskReddit/comments/5nrh7b/redditors_who_live_with_their_sos_what_changed/) and processed further using a python script. We removed deleted comments and users, subsequent conversations and bot answers, thus leaving the corpus with only comments directly answering to the question and containing around 500 documents. From those, the first 100 instances where used for the annotation task.

### Annotation

The annotation categories are modelled after Plutchik's wheel for its eight base categories and intensity rating were suitable for our specific data set. For the topic of *change*, the emotions of surprise and anticipation were needed and the feature of forming 'cross-emotions' like love from neighboring categories opened up possibilities to further study the corpus. After annotation the first 10 instances, e chose to disregard Plutchik's intensity categories in favor of unnamed intensity scores. We also encountered the problem of latent emotions implied by the corpus domain, which is why we had to specify annotation rules for those specific categories. The resulting annotation guidelines contained the following information:

*Example text is  
here*

ID000

### Annotation guidelines

The annotation categories are Plutchik’s eight base emotions of joy, sadness, trust, disgust, fear, anger, surprise and anticipation. The annotated value between 0 and 3 measures the intensity of the emotion label. The annotation environment has a document per line and eight columns for the emotion categories, thus every cell should contain a value. The categories are sorted as above with opposing emotions next to each other. Contrary to Plutchik’s wheel is the annotation of opposing emotions in one document. This is possible due to the multi-label annotation and the perspective of the annotation.

As the data is related to the event of ‘moving in together’, most documents will likely contain event descriptions or event appraisals. Therefore, the annotator should assume the emotion felt in the described situation rather than at time of writing. This also includes comical exaggerations building to a punchline, which are annotated without regards to their intended readings. As the documents address living together with a romantic partner, joy and trust are likely to be implicit in many texts. Therefore, trust is in this context defined as an explicit referral to a positive relation to another person. This definition includes the negation, i.e. negative emotions due to the absence of another person but not statements that do not refer to a second person in any way.

## 1 First iteration of the annotation guidelines

	joy	sadness	trust	disgust	fear	anger	surprise	anticipation
anticipation	1	0	1	0	0	1	0	<b>2</b>
surprise	7	1	2	1	0	3	<b>18</b>	
anger	1	2	1	1	0	<b>18</b>		
fear	0	0	1	1	<b>2</b>			
disgust	0	0	0	<b>5</b>				
trust	6	1	<b>16</b>					
sadness	1	<b>5</b>						
joy	<b>24</b>							

## 2 Number of annotated documents using the average annotation and co-annotations of emotions

	<b>Felix</b>	<b>Carlotta</b>	<b>Max</b>	<b><i>average</i></b>
<b>joy</b>	35	28	22	<i>28.33</i>
<b>sadness</b>	11	6	4	<i>7.00</i>
<b>trust</b>	21	23	16	<i>20.00</i>
<b>disgust</b>	7	6	2	<i>5.00</i>
<b>fear</b>	3	3	2	<i>2.67</i>
<b>anger</b>	26	19	19	<i>21.33</i>
<b>surprise</b>	43	24	19	<i>28.67</i>
<b>anticipation</b>	10	15	3	<i>9.33</i>
<b><i>average</i></b>	<i>1.56</i>	<i>1.24</i>	<i>0.87</i>	

**3** Number of annotations per annotator and per category