Perceptyx Data Tagging Instructions

Tagging the dataset:

Firstly, the Data Annotator should revise the sample attached with the instructions. The Data should be tagged in accordance with the Unicode Format and correct Encoding/Decoding errors.

1. **Sentiment Tags** (Some are Given): The dataset usually have Sentiment Associated with it given. However, the Data Annotator will revise if the Sentiment Tags are in accordance with the overall sentiment of the Paragraph. If it is very apparent that the Sentiment is contradictory than the given Sentiment then correct them. In absence of Sentiment Tags, read the paragraph and try and pick off what is the sentiment conveyed in the Text. There are three tags Positive, Negative and Neutral.

2. **Intent Tags:** The intent tags follow similar principle of Sentiment Tags. There are :

- Expression (Opinion) – These are the responses where usually personal beliefs or ideas are dominant. Another way to look at them is that if none of the below classes are relevant then its this class.

- Demands(Ask) – What is needed? What is to be done? Responses which asks for actions with directions, has requirements.

- Complaint – What is wrong? People discussing what they dislike, suggestion or opinion are very different from this.

- Suggestion – What should/could be done? This is straight–forward but might be confusing to tag Demands and Suggestion. So please check the intonation of the document by revising them.

- Appreciation – Praise or Approval or Likeness

- Information – Information or Facts or News. Different from Expression.

- Query – Questions, the responses where the Employee or User ask questions or show confusion due to lack of information.

3. **Phrase Tags:** To tag the phrases each word has following attribute

I) Subject–Verb–Object Tags: S for subject, if the word lies in Verb Phrase then tag them even if they might be adjectives or Adverbs as V. (See Examples). And Finally, O for object. One important thing to note is that in our case Subject is usually "the product, Company Policy, or the main entity the response

is talking about". Similarly, object is often the medium the above described subject affects. Verb phrases are picked off based on actions regarding these Subjects and Objects other "Filler" S—V—O are ignored. See Examples.

II) Intent – if certain verb phrase exhibits certain intent then they are tagged themselves

III) Sentiment Tag – p for positive, n negative intonation in the phrase, neu for Neutral

IV) Parts of Speech – rechecking the parts of Speech is infact valid (intuitive) || PENN Tree Bank to be followed

– POStags are given alongside the word. The Data Annotator should check if the POS tags are consistent. Since the responses do not necessarily follow conventional syntactic grammar rules and might be treated as Informal Dialogue Text. Current POS tagger might not parse them correctly.

V) Phrase chuck to be saved or not pr or npr

Examples:

"V_a_p_VBZ_pr" means the word belongs to a Verb Phrase(V) chunk which conveys appreciative words(a) from employee, has positive sentiment(p) and is a Verb word(VBZ). "pr" conveys that the phrase should be saved.

"S_n_neu_NN_pr" means the word is a Subject(S) chunk which conveys no apparent intent from the person, has no apparent sentiment/ is neutral(neu) and is a Noun word(NN). "pr" conveys that the phrase should be saved.

Note: Samples attached is to provide how the dataset is to be tagged, and structural similarity is pivotal.