Annotation Guidelines

Targeted Aspect Based Sentiment Analysis Youtube Comments in Nepali language

# Introduction

Goal of this project is to identify abusive sentiment in the text collected from Nepali YouTube videos under “News & Politics” category. In this project, the main is to annotate the aspect terms (which represent abusive terms), the targets (towards whom such terms are targeted) and to identify the polarity of each annotated aspect term.

Overall, we need to annotate the following types of information:

1. **Aspect Terms** - Identify any terms that can be categorized into General, Profanity, and Violence.
2. **Polarity** of Aspect Term - Each aspect term has its associated Polarity [0/1]. In the case of General/Feedback, it is positive or negative. In the case of Profanity and Violence, it is less profane/violence or high.
3. **Targets** - Identify any terms that are targeted by Aspect Terms. Targets can be categorized into Person, Organization, Location, and Misc.

# Aspect Terms

All the noun phrases that express any sentiment carrying subjectivity that lies under the following categories should be tagged. Hence, each *aspect term* has its *polarity* and *aspect terms* might be targeted towards *targets* or not.

Note: Examples in each category are taken directly from comments of social media users, so there might be cases of spelling mistakes.

## General

Annotators should tag all the terms which express the general positive or negative opinions. Polarity:

1. 0: Aspect terms with positive sentiment not limited to commendations and applauds. (Example - रा ो कु रा, उ कृ ट, मन पय , माया, fठकछ, बधाई छ, य नेपाल, वा वा, देश ेमी, गब छ)
2. 1: Aspect terms with negative sentiment without any profane or violent terms, which are meant to disparage, defame or hurt others. (Example - गाेबर कु रा, fछ fछ fछ, ग़लत, बेकार, काम छै न, ताइन तुइ, धाक, काम गदे'न, कु नेताकाे झाेला बाेRने, पागल, बैजfत, भतुवा)

Examples:

1. देश को  हV को सुची मा  पfन पछ ।

*(Translation:* ***Rabindra*** *also lies in the list of country’s* ***worst journalist****)*

* 1. **(*worst journalist*)**-> General [1]
  2. **(*Rabindra*)** -> Person
  3.  **(*worst journalist*)** is targeted towards  **(*Rabindra*)**
  4. In this statement, someone is expressing his/her negative sentiment towards a journalist ***Rabindra*** using negative terms like ***worst journalist***.

2.  जी को अ ोस मा के हk  ।

*(Translation: There is* ***nothing wrong*** *in anger of* ***Dwarika*** *jee)*

1.  (***nothing wrong***)-> General [0]
2. (***Dwarika***) -> Person
3. (*nothing wrong*) is targeted towards  (***Dwarika***)
4. In this statement, someone is expressing his/her positive sentiment towards a

***Dwarika*** using positive terms like ***nothing wrong***.

3.  काे  ले  गरे का छन ।

*(Translation:* ***UML politician*** *has done* ***shameless talking****)*

1. (***shameless talking***)-> General [1]
2. (***politician***) -> Miscellaneous
3. (**UML**) -> Organization
4.  (***shameless talking***) is targeted towards  (***politician***)
5. In this statement, someone is expressing his/her negative sentiment towards a

***politician*** using negative terms like ***shameless talking.***

## Profanity

Annotators should tag all the terms which express insults, cursing, or swearing, swear words, bad words, or expletive words1.

Polarity:

1. 0: Aspect terms including rude, bad or slander which are not very harsh and but offensive words used on day-to-day lives in Nepal. (Example - मुला, ज5या, साला, ते र मा टोRने, खाते, कु कु र, चोर, चो र, डाका, भु याहा कु कु र, जंगलk fहचडा, खाfतनी, धfत क बोझ, छRका,

कं लक, चुfतया, जाठ , ड ना, गददार, हुf हारा, नकचरा, अलfछना, घ टा, च डािǐन, ह यारा, आतंकवादk, आतंककारk)

1 "Profanity - Wikipedia." <https://en.wikipedia.org/wiki/Profanity>. Accessed 7 Apr. 2020.

1. 1: Aspect terms include swear or curse words which are very harsh (Example - र डी, रनfड को चेलो, मुजी, याfचRनी, रि ड, बे यालय को माfलक, बǐ कार, fछRने, राणी, भालुनी, य या)

Examples:

1.  खालk fबदेfस को गुनगान र बढाइचढाइ मा ै गछ ।

*(Translation:* ***Stupid d runkard middleman Punya Gautam****, you are only admiring foreigners too much)*

* 1. (***Stupid drunkard middleman***)-> Profanity [0]
  2. (***Punya Gautam***) -> Person
  3. (***Stupid drunkard middleman***) is targeted towards  (***Punya Gautam***)
  4. In this statement, someone is expressing his/her profane sentiment towards

***Punya Gautam*** using profane terms like ***Stupid drunkard middleman***.

2. यो  मलाई भन तेरो पाट को के योगदान छ ।

*(Translation: You* ***cunt Nagendra Sahi****, can you tell me what is the contribution of your*

*party)*

1. (***cunt***)-> Profanity [1]
2. (***Nagendra Sahi***) -> Person
3. In this statement, someone is expressing his/her profane expression towards

***Nagendra Sahi*** using profane terms like ***cunt***.

## Violence

Annotators should tag all the terms which express, incite or encourage any kinds of violence, physical, mental or sexual. It also includes destroying or burning the property.

Polarity:

* 1. 0: Comments with physical assault like beating or physical harm (Example - ना गो परेर, मुख मा हानु, पा ा फका उनु पछ , हा न मन ् ला यो, हा न सRनु भएन, नुन खुसा fन दǐलनु पछ )
  2. 1: Comments with fatal assault like killing or rape including pyromaniac act. (Example -

करे ट लगाउनु पछ , खुकु र ले घाfट रेटनु पनˇ, फासी गनुपछ)

Examples:

1. यो  लाई  को जाम मा उfभन लगाउन पछ ।

*(Translation:* ***Guy wearing a blue coat*** *should be made to stand* ***bare naked*** *in the traffic of* ***Koteshwor****)*

* + 1. (***bare naked***)-> Violence [0]
    2. (***Guy wearing a blue coat***) -> Miscellaneous [1]
    3.  **(*Koteshwor) -*** *Location*
    4. In this statement, someone is expressing his/her violent sentiment towards a ***Guy wearing a blue coat*** talking about making him **bare naked**. ***Koteshwor*** is simply tagged as it is part of Location.

2. यो  लाई कसै ले  ज ता हेन चाहा छन ।

*(Translation: Someone should* ***beat to break a nose*** *of this* ***bastard****, people wants to see that)*

1. (***beat to break a nose***)-> Violence [0]
2. (***bastard***) -> Profanity [0]
3. In this statement, someone is expressing his/her violent sentiment towards a **bastard** using physical assault terms. Since **bastard** is not a proper noun and also abusive term, we tag it as Profanity [0] as it carries negative sentiment.

## Feedback

Annotators should tag all the terms that express genuine advice, suggestion or complaint to the program, anchor (program host) or the guests in the program.

Polarity:

1. 0: Comments with positive feedback or suggestions.
2. 1: Comments with negative feedback or complaints. Examples:

1.  बोǐन देउ न बौदfधक डा. लाई Tवाक Tवाक  ।

*(Translation: Dhamala let intellectual Dr. Acharya speak, do not speak in middle)*

* 1.  **(*do not speak in middle*)** -> Feedback [0]
  2. **(*Dhamala*)** -> Person
  3. **(*Acharya*)** -> Person
  4. In this statement, someone is giving his/her feedback to ***Dhamala*** to **not speak in the middle**. **Acharya** is simply tagged as a **Person**.

2. घाँस का न  सोǐटा ।

*(Translation: You should go for some menial job Bamey dude)*

1.  **(*should go*)** -> Feedback [1]
2. बामे **(*Bamey*)** -> Person
3. In this statement, someone is giving negative feedback to ***Bamey*** that he better

**do** some menial job.

## Sarcasm

Annotators should tag the whole sentence, unlike specific phrases in previous categories. Comments that express sarcasm based on the video should be tagged.

Polarity:

1. 0: Sarcastic comment without insulting or harming anyone.
2. 1: Sarcastic comment with negative sense or insulting others.

Examples:

3. प च भ ने नेपालk श दै छै न भ ने क तो आलु fप एच fड गरे को हो ?

*(Translation: What is the worth of that PhD degree, when he says there is no such word like treachery in Nepali. )*

* 1. Comment - Sarcasm [1]
  2. In this statement, someone is expressing his/her sarcastic comment in a negative way by insulting or questioning someone else’s Ph.D. degree.

1. अब fतमेर को बोलk सु दा रस होइन हास उ छ भ या ।

*(Translation: I don’t feel angry but funny when I listen to you guys.)*

* 1. Comment - Sarcasm [0]
  2. In this statement, someone is expressing his/her sarcastic comment, since there is nothing negative, we tag **0** under **Sarcasm** for such comment

## Out Of Scope

Comments that are beyond the above categories.

1. रातो कोट लगाउने ज तो ।

*(Translation: Like the one wearing a red coat)*

* 1. This statement does not express any sentiment that lies under the given categories.

1. आfखर जे सोचे को, यहk पु गो ।

*(Translation: Finally, it happened as I expected)*

* 1. This statement does not express any sentiment that lies under the given categories.

1. कु रा मा ठु लो?

*(Translation: Only talking big ?)*

* 1. This statement does not express any sentiment that lies under any given categories.

# Targets

All the aspect terms are either targeted or not. If it is targeted, then it is mostly targeted to people, organization or locations. Therefore, we tag all the proper nouns for Person, Organization, and Location, whereas common nouns for Miscellaneous.

## Person

All the names of people are tagged whether they are targeted or untargeted by an aspect term in a sentence. Example - पु य गौतम (*Punya Gautam*), fदल fनशानी मगर (*Dil Nishani Magar*), रfब लाfमछाने (*Ravi Lamichhane*), एfलशा (*Alisha*)

## Organization

All the names of an organization are tagged whether they are targeted or untargeted by an aspect term in a sentence. Example - काठमाड महानगरपाfलका, fस धुपाǐचोक गाfबस, नेकपा एमाले, प कार महासंघ, गृह म ालय

## Location

All the names of an organization are tagged whether they are targeted or untargeted by an aspect term in a sentence. Example - नेपाल, काठमाड , fस धुपाǐचोक, कोतेeवोर, fतनकु ने, रंगशाला

## Miscellaneous

All the names besides the proper nouns which are targets of an aspect term are tagged as Miscellaneous. Example - नेता, प कार, राजा, कि न ट, सरकार, police, mantri, गणत , राजत , मेयर

# References

1. SemEval [2014](http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools), [2015](http://alt.qcri.org/semeval2015/task12/) and [2016](http://alt.qcri.org/semeval2016/task5/) Aspect Based Sentiment Analysis
2. [Annotating Targets of Opinions in Arabic using Crowdsourcing](https://www.aclweb.org/anthology/W15-3210.pdf)