

# 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:

- a. 0: Aspect terms with positive sentiment not limited to commendations and applauds.  
(Example - राम्रो कुरा, उत्कृष्ट, मन पर्यो, माया, ठिकछ, बधाई छ, ज्य नेपाल, वा वा, देशप्रेमी, गर्ब छ)

- b. 1: Aspect terms with negative sentiment without any profane or violent terms, which are meant to disparage, defame or hurt others. (Example - गोबर कुरा, छि छि छि, गलत, बेकार, काम छैन, ताइन तुइ, धाक, काम गर्दैन, कुनेताको झोला बोक्ने, पागल, बैजति, भतुवा)

Examples:

1. देश को **भाइरस पत्रकार** हरु को सुची मा **रबिन्द्र** पनि पर्छ ।  
(Translation: **Rabindra** also lies in the list of country's **worst journalist**)
  - a. **भाइरस पत्रकार (worst journalist)** -> General [1]
  - b. **रबिन्द्र (Rabindra)** -> Person
  - c. **भाइरस पत्रकार (worst journalist)** is targeted towards **रबिन्द्र (Rabindra)**
  - d. In this statement, someone is expressing his/her negative sentiment towards a journalist **Rabindra** using negative terms like **worst journalist**.
2. **द्वारिका** जी को अक्रोस मा केही **गलत देखिँदैन** ।  
(Translation: There is **nothing wrong** in anger of **Dwarika jee**)
  - a. **गलत देखिँदैन (nothing wrong)** -> General [0]
  - b. **द्वारिका (Dwarika)** -> Person
  - c. **गलत देखिँदैन (nothing wrong)** is targeted towards **द्वारिका (Dwarika)**
  - d. In this statement, someone is expressing his/her positive sentiment towards a **Dwarika** using positive terms like **nothing wrong**.
3. **नेकपा** को नेता ले **लाजमर्दो कुरा** गरे का छन ।  
(Translation: **UML politician** has done **shameless talking**)
  - a. **लाजमर्दो कुरा (shameless talking)** -> General [1]
  - b. **नेता (politician)** -> Miscellaneous
  - c. **नेकपा (UML)** -> Organization
  - d. **लाजमर्दो कुरा (shameless talking)** is targeted towards **नेता (politician)**
  - e. 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 words<sup>1</sup>.

Polarity:

- a. 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, जड्या, साला, तेरि मा टोक्ने, खाते, कुकुर, चोर, चोरि, डाका, भुस्याहा कुकुर, जंगली हिचडा, खातिनी, धर्ति की बोझ, छक्का, कलक, चुतिया, जाठी, डम्ना, गद्दार, हुतिहारा, नकचरा, अलछिना, घन्टा, चन्डाल्लि, हत्यारा, आतंकवादी, आतंककारी)

---

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

- b. 1: Aspect terms include swear or curse words which are very harsh (Example - रण्डी, रनडि को चेलो, मुजी, म्याचिक्नी, रन्डि, बेस्यालय को मालिक, बल्त्कार, छिक्ने, राणी, भालुनी, ब्यस्या)

Examples:

1. मुला जइया दलाली पुण्य गौतम खाली बिदेसि को गुनगान र बढाइचढाइ मात्रै गर्छ ।  
(Translation: **Stupid d drunkard middleman Punya Gautam**, you are only admiring foreigners too much)
  - a. मुला जइया दलाली (**Stupid drunkard middleman**)-> Profanity [0]
  - b. पुण्य गौतम (**Punya Gautam**) -> Person
  - c. मुला जइया दलाली (**Stupid drunkard middleman**) is targeted towards पुण्य गौतम (**Punya Gautam**)
  - d. In this statement, someone is expressing his/her profane sentiment towards **Punya Gautam** using profane terms like **Stupid drunkard middleman**.
2. यो मुजि नगेन्द्र साही मलाई भन तेरो पार्टी को के योगदान छ ।  
(Translation: You **cunt Nagendra Sahi**, can you tell me what is the contribution of your party)
  - a. मुजि (**cunt**)-> Profanity [1]
  - b. नगेन्द्र साही (**Nagendra Sahi**) -> Person
  - c. 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:

- a. 0: Comments with physical assault like beating or physical harm (Example - नाङ्गो परेर, मुख मा हानु, पात्ता फर्काउनु पर्छ, हान्न मन् लाग्यो, हान्न सक्नु भएन, नुन खुर्सानि दल्लनु पर्छ)
- b. 1: Comments with fatal assault like killing or rape including pyromaniac act. (Example - करेन्ट लगाउनु पर्छ, खुकुरि ले घाटि रेटनु पर्ने, फासी गनुपर्छ)

Examples:

1. यो निलो कोटे लाई नाङ्गो परेर कोटेश्वर को जाम मा उभिन लगाउन पर्छ ।  
(Translation: **Guy wearing a blue coat** should be made to stand **bare naked** in the traffic of **Koteshwor**)
  - a. नाङ्गो परेर (**bare naked**)-> Violence [0]
  - b. निलो कोटे (**Guy wearing a blue coat**) -> Miscellaneous [1]
  - c. कोटेश्वर (**Koteshwor**) - Location
  - d. 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 खाते लाई कसै ले नाक भाच्नेगरि हान्न जन्ता हेर्न चाहान्छन् ।

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

- a. नाक भाच्नेगरि हान्न (**beat to break a nose**)-> Violence [0]
- b. खाते (**bastard**) -> Profanity [0]
- c. 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:

- a. 0: Comments with positive feedback or suggestions.
- b. 1: Comments with negative feedback or complaints.

Examples:

1. धमला बोल्न देउ न बौद्धिक डा.आचार्य लाई प्वाक प्वाक नबोल ।

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

- a. प्वाक प्वाक नबोल (**do not speak in middle**) -> Feedback [0]
- b. धमला (**Dhamala**) -> Person
- c. आचार्य (**Acharya**) -> Person
- d. 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 काट्न गए हुन्छ बामे सोल्टा ।

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

- a. गए हुन्छ (**should go**) -> Feedback [1]
- b. बामे (**Bamey**) -> Person
- c. 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:

- a. 0: Sarcastic comment without insulting or harming anyone.
- b. 1: Sarcastic comment with negative sense or insulting others.

Examples:

3. प्रपञ्च भन्ने नेपाली शब्दै छैन भन्ने कस्तो आलु पि एच डि गरे को हो ?  
(Translation: What is the worth of that PhD degree, when he says there is no such word like treachery in Nepali. )
- a. Comment - Sarcasm [1]
  - b. In this statement, someone is expressing his/her sarcastic comment in a negative way by insulting or questioning someone else's Ph.D. degree.
4. अब तिमेर को बोली सुन्दा रिस होइन हास उठ्छ भन्या ।  
(Translation: I don't feel angry but funny when I listen to you guys.)
- a. Comment - Sarcasm [0]
  - b. 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.

5. रातो कोट लगाउने जस्तो ।  
(Translation: Like the one wearing a red coat)
- a. This statement does not express any sentiment that lies under the given categories.
6. आखिर जे सोचे को, त्यही पुग्यो ।  
(Translation: Finally, it happened as I expected)
- a. This statement does not express any sentiment that lies under the given categories.
7. कुरा मात्र ठुलो?  
(Translation: Only talking big ?)
- a. 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 - पुण्य गौतम (Punya Gautam), दिल निशानी मगर (Dil Nishani Magar), रवि लामिछाने (Ravi Lamichhane), एलिशा (Alisha)

## Organization

All the names of an organization are tagged whether they are targeted or untargeted by an aspect term in a sentence. Example - काठमाडौं महानगरपालिका, सिन्धुपाल्चोक गाबिस, नेकपा एमाले, पत्रकार महासंघ, गृह मन्त्रालय

## Location

All the names of an organization are tagged whether they are targeted or untargeted by an aspect term in a sentence. Example - नेपाल, काठमाडौं, सिन्धुपाल्चोक, कोतेश्वोर, तिनकुने, रंगशाला

## Miscellaneous

All the names besides the proper nouns which are targets of an aspect term are tagged as Miscellaneous. Example - नेता, पत्रकार, राजा, कम्निस्ट, सरकार, police, mantri, गणतन्त्र, राजतन्त्र, मेयर

## References

1. SemEval [2014](#), [2015](#) and [2016](#) Aspect Based Sentiment Analysis
2. [Annotating Targets of Opinions in Arabic using Crowdsourcing](#)