MSDS Annotation Guidelines

Hello, dear annotator, and thank you for your time!
Here is a brief background for the task and the specific annotation guidelines. This task is a part of the DharmaBench paper and evaluation benchmark for Sanskrit and Tibetan.

Introduction

Task: Metaphors and Similes Detection Sanskrit (MSDS)

Goal: Identifying similes (upamā) in and metaphors (rūpaka) in the text.

Background: The boundary between simile and metaphor is not always rigid, and the two often appear to overlap in poetic usage. However, within the tradition of classical Sanskrit, a distinction is maintained: simile (*upamā*) expresses a comparison, whereas metaphor (*rūpaka*) indicates an assimilation or complete identification between the subject or referent and the image.

Real-world relevancy: Identifying metaphors and similes across a large corpus of texts represents the foundational step toward understanding how similes develop and function in Tibetan literature. This analysis could facilitate future research endeavors, such as tracing the evolution of specific ideas expressed through various literary devices, including similes. When combined with other analytical and computational capabilities, this approach can also help trace the historical development of particular concepts, including the range of similes used to express them and their transformation over time. Since this task is applied to both Sanskrit and Tibetan texts, the ultimate objective is to identify similes both monolingually within each corpus and cross-lingually across the two corpora.

Examples

The similes are ==highlighted==, and the metaphors are ==<u>highlighted and underlined</u>==.

1. vahati ca calitavilocanajaladhāram==<u>ānanakamalam</u>==udāram |
   ==vidhum iva== vikaṭavidhuntudadantadalanagalitāmṛtadhāram || Gītagovinda ||

   ==<u>ānana-kamalam</u>== "lotus-face" > *Rūpaka* (**MET**)
   ==vidhum iva== "like a moon" > *upamā*(**SIM**)

2. uddharatyandhatamasād viśvamānandavarṣiṇī |
   paripūrṇā jayatyekā devī ==<u>ciccandracandrikā</u>== || Śivastotrāvalī ||

   In this verse, the compound **cic-candra-candrikā** ("moonlight of the moon of consciousness" or "[Goddess] whose consciousness is the moonlight of moon") is a good example. It metaphorically modifies the Goddess, identifying her consciousness

with the cool, luminous essence of moonlight. It evokes both the nature of consciousness (cit) and the poetic image of the moonlight (candrikā), fusing them inseparably.

Data collection guidelines

1. Go over texts from the defined sources list.
2. For negative examples, cut samples of texts into segments no longer than two verses consisting of four lines each, which do not contain a simile.
3. For positive samples, detect similes and metaphors that appear in the texts.
4. Collect all in a CSV file with a column for source text name and the cut text chunk.

Annotation guidelines

Annotation platform

We will be using the Label Studio app. We will upload the collected CSV files and set up the annotation project.

Preparation

1. Log in to our Label Studio organization here.
2. Let us know, and we will approve your user and assign you to the respective project.

Annotation guidelines

1. Read the sentence/paragraph thoroughly.
2. Identify the simile(s) in the sentence.
3. Mark with "SIM" label the minimal span which contains the simile, if found.
    a. Potentially, use comparison markers: iva, yathā, -vat, sama
4. Mark with "MET" label the minimal span which contains the metaphor, if found.
    a. A *rūpaka* compound is a metaphorical compound that fuses two nouns or a noun and a qualifier.
5. Re-read and re-evaluate your annotation/s, focusing on boundaries.
6. Annotation Speed vs. Accuracy Tradeoff: Emphasize **accuracy over speed** if necessary.

Edge cases

1. In the case of ambiguity, please consult the team and the supervisors.
    a. If you can't reach an agreement, discard the sample.

2. **Semantic fusion:** Metaphors often merge entities, collapsing the source and target into one expression (e.g., ciccandracandrikā).

Something is off?

Reach out to us.

Thank you very much!