



A Novel Annotation Schema for Conversational Humor: Capturing the cultural nuances in Kanyasulkam

ABSTRACT

Humor research is a multifaceted field that has led to a better understanding of humor’s psychological effects and the development of different theories of humor. This paper’s main objective is to develop a hierarchical schema for a fine-grained annotation of Conversational Humor. A prominent 19th century play from Telugu, Kanyasulkam, is annotated to substantiate the work across cultures at multiple levels. Based on the Benign Violation Theory, the benignity or non-benignity of the interlocutor’s intentions is included within the framework. Under the categories mentioned above, in addition to different types of humor, the techniques utilized by these types are identified. Furthermore, the inter-annotator agreement is calculated to assess the accuracy and validity of the dataset. An in-depth analysis of the disagreement is performed to understand the subjectivity of humor better.

DATA AND ANNOTATION

- The full text of the Telugu play, Kanyasulkam, is annotated by two people, A1 and A2.
- For the pre-processing of the data, the whole text was split first by each character’s dialogue, and each utterance by the character was further split into single sentences/segments.
- In the presence of poems, lists, etc. the utterance remains as is, and this final output is used for annotation giving a total of 6645 segments to be annotated.
- After developing the gold standard corpus, 2710 utterances were classified as humorous, 1782 were given the tag dialogue, 1881 conversational, and 892 benign.
- The annotation was done with appropriate checkpoints after every 2000 segments to identify any new techniques or revise the schema.

SCHEMA AND ANALYSIS

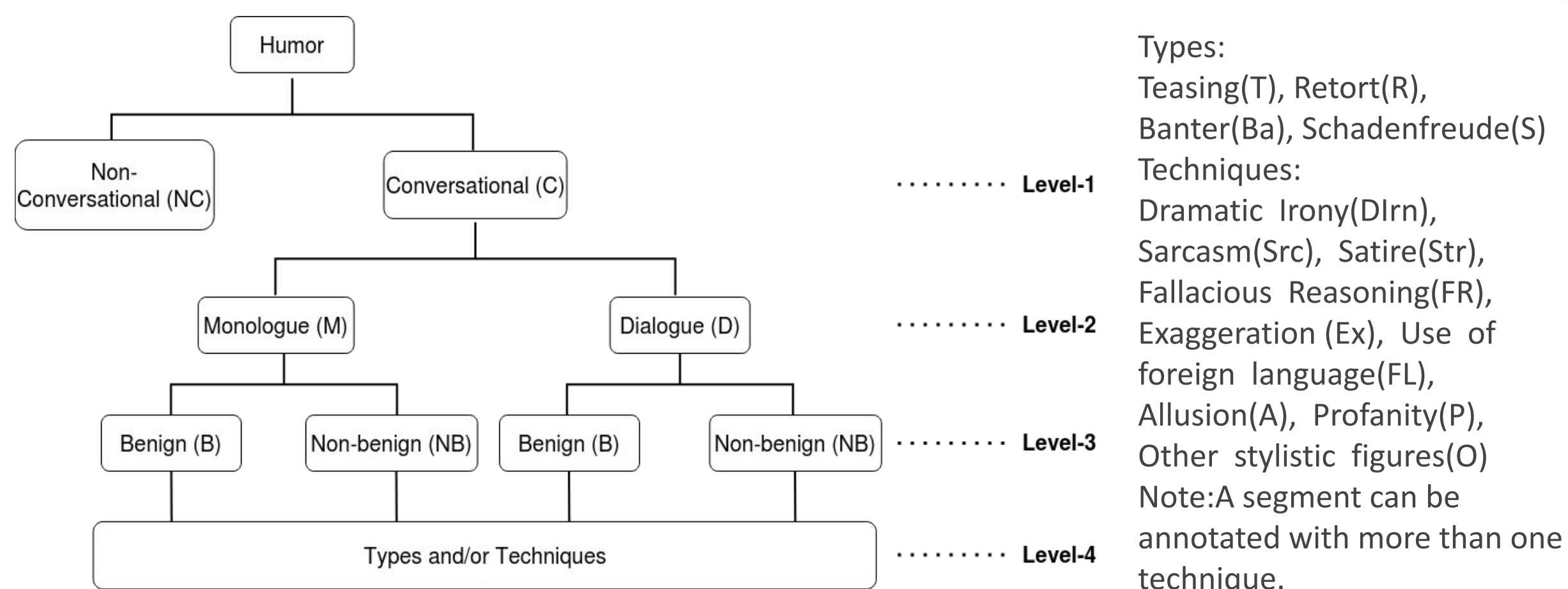


Figure 1. Hierarchical schema for Conversational humor.

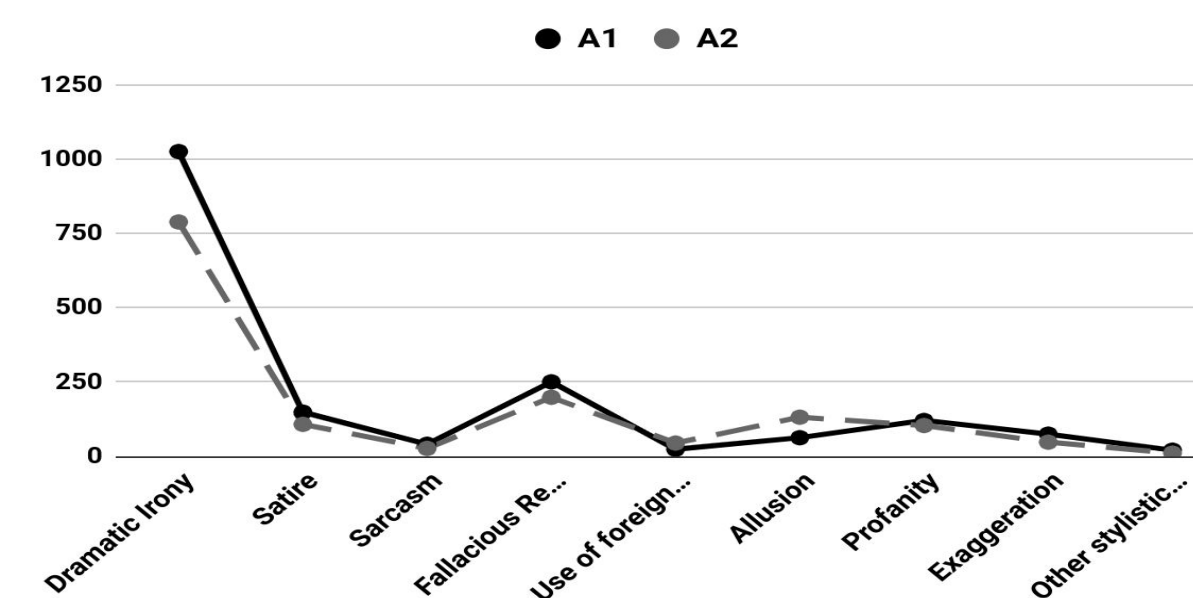


Figure 2. Distribution of Techniques Annotated by A1 and A2

Level	Level Name	Agreement	Interpretation
1	Non-Conversational / Conversational	0.48	Moderate Agreement
3	Benign / Non-Benign	0.42	Moderate Agreement
4	Types	0.49	Moderate Agreement

Figure 3. Disagreement Analysis on Different Levels