

# Intentional Ambiguity Recognition in Natural Language Texts

Lawrence J. Mazlack

## 1 Overview and Objectives

Knowledge is distilled into natural language. Much of it is computationally stored in texts. Computational natural language understanding is critical to handling the large amounts of text that already exist, as well as the geometrically growing volume that will come to exist. Knowledge cannot be readily summarized or accessed unless inherent utterance ambiguity is recognized and resolved. While there are many existing disambiguation tools, *intentional* ambiguity recognition has not received much attention. Recognizing *intentional* ambiguity is a vertical step to understanding natural language. Humor often depends on intentional ambiguity; it offers a clear, distinct research focus. Therefore, the purpose of this research is to recognize intentional ambiguity that results in humor. Humor has been studied since the ancient Greeks. However, only small steps have been made in computational humor recognition.

The long term goal of this research is to computationally recognize and handle ambiguities in text; the focus is on computational humor recognition in short, humorous texts. Considering short jokes that are dependent on lexical ambiguities and phonological similarities further focuses the proposed work. The objective of this research, which is a step toward our long-term goal, is to computationally recognize short jokes for young children. The central hypothesis of this research is that a description logic ontology that is suitable for children's text annotation is also suitable for joke recognition. The hypothesis has been formulated on the basis of preliminary analysis of short children's texts and jokes, and our prior experience in using non-ontological methods. Restricting the domain to young children's jokes is expected to decrease complexity while retaining the core issues. The domain restriction reduces needed background knowledge, which leads to a reduction of knowledge to be captured. This, in turn, leads to a smaller, more manageable ontology, and fewer inferences to be drawn from it. It allows a concentration on humor recognition methods with a reduced emphasis on the problems of capturing and inferring knowledge. The development of such methods is a step closer to an understanding of the product of human cognitive processes and "more human" intelligent systems, which is a rationale for undertaking this research. In addition to our supportive preliminary data, we are particularly well prepared to undertake the proposed research, because we have already performed and published investigations of statistical techniques; and, we have published explorations that prepare the foundation for our proposed approach. Our working environment contains investigators who are working on parallel projects involving theoretical aspects of ontologies, as well as, the ontological reasoning tools that we propose to use. This creates a rich intellectual environment in which to perform the work (See the Facilities, Equipment and Other Resources section).

The central hypothesis will be tested and the objective of this research will be accomplished by pursuing the following two specific aims:

1. ***Build a description logic-based imprecise ontology, containing minimal number of concepts and relationships between them for annotation of texts for young children.***

Based on the preliminary data, the working hypothesis is that a children's dictionary defines all nouns that are needed to create concepts for the ontology. Relationships between the concepts will be semi-automatically created with knowledge extracted from a children's dictionary and a collection of children's texts.

2. ***Determine a method for recognition of script overlap and opposition that create jokes and that are based on lexical ambiguity or phonological similarity.***

A script is a framework for typical activity. Our *working hypothesis* is that salient scripts can be determined as a composition of concepts in the jokes and inferences drawn from these concepts.

The proposed research is *creative and original* because it is ontologically based. The few humor recognizers that exist today are statistically, not ontologically based, and much more limited in the humor form. It is *expected* that the results will fundamentally advance the field of computational humor. It is expected that once an ontology-based children's joke recognizer is developed, the ontology can be expanded to include other knowledge. It is expected that knowledge contained in such ontology can be used to recognize more sophisticated jokes, using similar, or slightly modified, algorithms. The *broad applications and positive impact* can range from recognizing undesirable jokes in emails to helping children and second-language learners master the language. With increased level of sophistication, humor recognition may be used in commercial natural language translation applications. A joke recognizer can also be added to word processing software as well as enhancing search engine results. Aim One will enable the recognition of unambiguous text elements. Aim Two will enable the recognition of ambiguities. Collectively, the two aims will accomplish the overall objective of this application.