

# Generating Natural Language Entailments and Contradictions to assist Natural Language Inference

Ryan Sie

## 1 Introduction

## 2 Previous Work

## 3 Linguistic Motivation

In the linguistic subfield of semantics, the relationships of **entailments** and **contradictions** between sentences in natural language are well studied phenomena. The models discussed in this paper attempt to make use of various parts of semantic theory in their design. This section will provide a brief introduction and overview of the core ideas that motivate the models' architectures.

A common way to formalize the representation of natural language sentences are as sets of possible worlds in which they are true, in the mathematical sense of *set*. A possible world can simply be thought of any particular arrangement of circumstances in the real world that can be imagined. For our purposes it will suffice to describe a possible world  $w$  in words. For instance, in one world you might own a goldfish, while in another, you might not. Then a sentence  $S = \{w_1, \dots, w_k, \dots\}$  is a simple set containing possible worlds  $w_i$  in which the sentence is true. Note that this set may be infinite (in fact, many are), and even empty. To help illustrate this formulation, below are some example sentences with descriptions of some possible worlds they contain.

- (1) I ate breakfast today
  - a. I ate a bagel for breakfast today
  - b. I ate a bowl of cereal for breakfast today
- (2) The sky is not blue
  - a. The sky is purple
  - b. The sky is white (imagine a different world than ours)
- (3) This sentence is false =
  - a. asdf

**Definition 3.1.** Let  $A, B$  be two sentences. We say that  $A$  **entails**  $B$  if when  $A$  is true, it follows that  $B$  is true.

**4 Model**

**5 Conclusion**