CNLI-TR Project - Handout 1

Büşra Marşan

Abstract

This project aims to explore De Re/De Dicto phenomenon from various aspects including its signficance in computational linguistics and Natural Language Processing. Our primary goal is to create a De Re/De Dicto ambiguity data set to explore a language model's capabilities of "understanding" this ambiguity. Within the framework of this project, we define "understanding" a sentence as knowing what can and cannot be entailed from that. In this handout, we will get familiar with the De Re/De Dicto phenomenon, explore the NLP task we aim to do, and discuss your first task as the contributors.

1 An Introduction to De Re/De Dicto Phenomenon

An ambiguous sentence lies at the core of De Re/De Dicto phenomenon. The interaction between noun phrases and intensional operators trigger the ambiguity (see Keshet and Schwarz (2019) for a detailed discussion). As a result, it is possible to get at least two readings from the ambiguous sentence: De Re and De Dicto.

Let's say that we have a common friend, Ayşe. I learned something about her and I am sharing it with you:

(1) Ayşe bir tesisatçı ile evlenmek istiyor.
"Ayşe wants to marry a plumber."

(1) contains both elements that are required for a De Re/De Dicto ambiguous sentence: A noun phrase (bir tesisatçı *"a plumber"*) and an intensional operator (istemek *"to want"*). Hence, it is possible to get the following interpretations:

Specific De Re: There is an individual who is a plumber, let's call him Ali. Ayşe wants to marry this person. She may or may not know that he is a plumber. In this interpretation, "a plumber" refers to an actual person[1]. Hence, we have a De Re reading.

Non-Specific De Re: There is another individual who is a plumber, let's call him Veli. Ayşe wants to marry either Ali or Veli. She hasn't decided which one yet but it is certain that she wants to marry one of them. In this interpretation, "a plumber" refers to actual people. Hence, we have another De Re reading.

De Dicto: Ayşe has very specific desires: She wants to get married and she wants to marry a person who is a plumber. She hasn't met anyone special yet, there is not a specific person in her mind. She just has this affection towards plumbers and she can marry any plumber.

Another interesting thing to note about De Re/De Dicto ambiguity is that strong quantifiers and operators that shift the time of evaluation can cause De Re/De Dicto Ambiguity:

(2) 2050 senesinde Dilbilim bölümündeki profesörlerin çoğu emekli olacak.
"By 2050, most professors in Linguistics department will retire."

Here our noun phrase is "Dilbilim bölümündeki profesörlerin çoğu" and our operator is "emekli olacak." Hence, it is possible to get the following interpretations:

De Re: The majority of the professors who are working at the linguistics department **NOW** will be retired by 2050. In this interpretation, the noun phrase ("Dilbilim bölümündeki profesörlerin çoğu") is

---

[1] The terms De Re and De Dicto can be confusing but learning their literal meaning can help you distinguish them: "De Re" means "of thing", while "De Dicto" means "of meaning/concept." Note that "dicto" is "saying, expression" in Latin and "dictionary" is derived from this root.
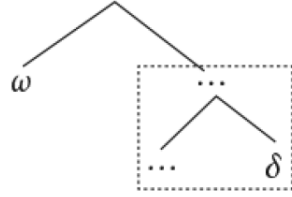
anchored in the utterance context with respect to the time and world of evaluation.

De Dicto: The majority of the professors who **WILL BE** working at the linguistics department **in 2050** will retire in 2050. Here, the noun phrase ("Dilbilim bölümündeki profesörlerin çoğu") is evaluated relative to the time and world at which the main proposition expressed by the sentence is evaluated (year 2050).
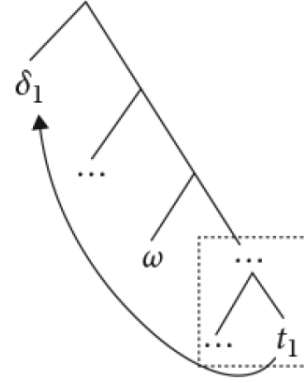
1.1 Semantics of De Re/De Dicto

I don't think we need to go into the intricate semantic details of the issue since there is a lot to cover and quite frankly, they are very peripheral to our task. I just want to point out that the widely accepted consensus is that De Re is interpreted above the intensional operator while De Dicto is interpreted in situ. Below you can find a simple schema where w is the intensional operator and d is the noun phrase.

2 The Goal(s) of The Magnificent De Re/De Dicto Project

What we want to achieve through CNLI-TR Project is to see if language models like BERT are able to detect De Re/De Dicto ambiguities. In order to do that, we aim to design an inference task where the language model is presented with the ambiguous sentence and one of the possible readings (De Re or De Dicto). If the model is able to "see" the ambiguity, it must be able to say that both De Re and De Dicto readings can be inferred from the ambiguous sentence.

In order to train the model, we need large quantities of data. This is where you as contributors come into play!

For this purpose, 500+ novel De Re/De Dicto ambiguous sentences were written by me (Büşra) and we need your support to expand the data set. After your contributions, we will feed the seed data (manually generated sentences) into a language augmentation pipeline to create similar sentences.

3 Our First Task: Sentence Generation

You can find the sentences written by me on this link: https://docs.google.com/spreadsheets/d/1tEPO-4B9Z9QOulKEBB-Y6CwwZxGOvWF1m_DJTxDTRFY/edit?usp=sharing

First of all we would like you to review them, get familiar with them. Then we would like you to write 50 novel sentences each. In order to make sure that we are on the same page, we should start with 10 sentences each and review them. After the review session, you can write 40 more.

You can get as creative as you want with the sentences. All we ask is that they should contain an intensional operator, a proper name, and a noun phrase. You can try writing conditional sentences, questions etc. or play with the word order.

Following this task, we will write De Re and De Dicto readings for these ambiguous sentences. Then the augmentation task will be done and we will ask for your help for your final task: Reviewing the augmented data. Theoretically, the augmentation pipeline must provide us with meaningful sentences but in practice, we probably will come across some nonsensical and/or freak of nature sentences. We need to eliminate such sentences to ensure that our data is not noisy and the language model can properly learn from it.

References

Keshet, E., & Schwarz, F. (2019). De re/de dicto. *The Oxford handbook of reference*, 167–202.