FNLP Tutorial 1

1 Ambiguities

Ambiguities are pervasive in natural language, but often go unnoticed when we use language because humans are so good at resolving them. In this exercise, we want you to find ambiguities in the example sentences and attempt to articulate a paraphrase that as much as possible removes the ambiguity (similar to section 1.2 of J&M, 2nd edition). Categorise the different ambiguities you observe: e.g., word sense ambiguity, structural ambiguity, phonetic ambiguity and so on.

1. At the bank, Mary noticed her sister.
2. Every student wants to win the first prize in a programming competition with a robot.

2 Corpora and annotation

In this exercise, we want you to get some insights into the challenges that humans and machines face when it comes to annotation. Consider the following corpus:

1. Paris Hilton stayed at the Hilton in Paris.
2. Alan Bleeding Turing.
3. Tom works for the Dumfries & Galloway Standard.

1. Annotate the above utterances with named entities. For our purposes, a named entity is a single word or multiple words that refer to a person (`PER`), location (`LOC`) or organisation (`ORG`). Are there cases that you found difficult? Which cases do you think are difficult for an automated system? And why?

Question 3. Evaluating data annotations

Imagine that this is your small corpus of named entities in a simple task, where we ignore a named entity's type and annotate the real named entities with square brackets:

> [Paris Hilton] stayed at the [Hilton] in [Paris] and
> [James Clerk Maxwell] was educated at [Edinburgh] and [Cambridge]

We formalise the annotation of a single sentence as a set A. Each element $a \in A$ represents a named entity as a span through an ordered pair of zero-based indices (a named entity $\langle a, b \rangle$ starts at position a and ends at b, including b). Assume a computational model predicts the following named entities: $\{\langle 0, 1 \rangle, \langle 5, 5 \rangle, \langle 9, 10 \rangle, \langle 15, 15 \rangle\}$

1. Compute the precision, recall, and F_1-score of this annotation.
2. Provide an annotation that would give a precision of more than 0.8 and a recall of less than 0.2, and use your answer to explain why the F_1-score uses the *harmonic* mean.
3. While the F_1-score is a better metric than precision and recall in isolation, there are other flaws all three metrics suffer from. What, specifically in the context of span identification, does it fail to capture about the model's predictions provided above?