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.

   Solution

   (a) Mary noticed her own sister at the financial institute
   (b) Mary noticed her own sister at the river bank
   (c) Mary noticed someone's sister at the financial institute
   (d) Mary noticed someone's sister at the river bank

   Note that there are two ambiguities: what does "bank" mean? (this is a word sense ambiguity) and who does "her" refer to? (reference ambiguity), and we can combine them freely, so there are $2 \cdot 2 = 4$ possible readings in total.

2. Every student wants to win the first prize in a programming competition with a robot.

   Solution There are at least the following ambiguities:

   (a) A "programming competition" could be a competition that involves programming (plausible) or a competition 'that programs' (implausible). This is a semantic ambiguity.
   (b) "with a robot" could describe the competition, i.e. a competition involving a robot, or it could describe the winning, i.e. winning the competition by using a robot or with a robot as a teammate. This is a structural ambiguity.
   (c) It could be the case that speaker was trying to say that there is (at least) one competition in which every student wants to win the first prize or there could be multiple competitions and every student wants to win first prize in (at least) one of them. This is a semantic ambiguity, more specifically, a scope ambiguity (because the scope of the quantifiers are ambiguous).
   (d) It could be the case that speaker was trying to say that the students want to win the first prize because it has the property of being the first prize (they want to come first), no matter what the prize is or the speaker meant that they are keen on the first prize because it is something specific (e.g. a laptop) that they all would like to win. This is a semantic ambiguity, more specifically it's a de re/de dicto ambiguity[1].

   Note that the ambiguity described in (d) does *not* arise from the problem that we cannot infer the students' intentions in the competition but rather that we cannot unambiguously infer what the *speaker* of (2) meant.

[1] You can find more examples at https://en.wikipedia.org/wiki/De_dicto_and_de_re

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?

   Solution There is often no single optimal solution for annotation; whether something is a good annotation depends on what it is used for. For example, should "the Hilton" be a named entity or just "Hilton"? Can named entities overlap?

   The following should be a reasonable annotation though for many applications:

   (a) [Paris Hilton]$_{\text{PER}}$ stayed at the [Hilton]$_{\text{ORG}}$ in [Paris]$_{\text{LOC}}$.

   (b) [Alan]$_{\text{PER}}$ Bleeding [Turing]$_{\text{PER}}$.

   (c) [Tom]$_{\text{PER}}$ works for the [[[Dumfries]$_{\text{LOC}}$ & [Galloway]$_{\text{LOC}}$]$_{\text{LOC}}$ Standard]$_{\text{ORG}}$.

   1b can be considered a *single* discontinuous named entity, rather than two named entities. In 1c, there is an ambiguity if *Dumfries & Galloway* is a location or an organisation (the council) that is difficult to resolve.

   An automated system might struggle with cases that require context: the token "Paris" can refer to a location or a person, "Hilton" can refer to a person or to an organisation (the hotel chain). A named entity recogniser that does not use a list of council areas of Scotland might also struggle with detecting that the span *Dumfries & Galloway* is one named entity.

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?

Solution

1. There are 3 true positives, 1 false positive and 3 false negatives. Therefore, the precision is 0.75, the recall is 0.5, and the F_1-score is $2 \cdot \frac{0.75 \cdot 0.5}{0.75 + 0.5} = 0.6$.
2. We could, for example, use a simple annotation with perfect precision, such as $\{\langle 0, 1\rangle\}$ that gives a recall of $\frac{1}{6}$, and an F_1-score of approximately 0.286. This F_1-score lies far below the arithmetic mean of the two, and lies, relatively speaking, closer to the recall compared to (1) above. This illustrates why the harmonic mean is used: it penalises extreme values for either precision or recall. This is desirable since a trivial prediction – such as predicting all candidates, or no candidates at all – can yield a high recall or a high precision, but not both.
3. The span $\langle 9, 10\rangle$ is incorrect, but "James Clerk" is a part of a named entity. Nonetheless, this partial match between the prediction and the target is not captured by the current metrics.