Guidelines for Aligning Predicate-Argument Structures

1   Introduction

Annotators are provided with pairs of newswire articles describing the same news from the perspective of two distinct sources. The descriptions can vary both with regard to content and linguistic realization. As an annotator, your task will be mark corresponding pairs of predicates across documents. Note that for this task, it is important to also take the context of the predicate into consideration! The predicates have been (automatically) pre-selected and marked in boldface to ease annotation. We further provide indices for all predicates, so that you can uniquely identify them in a given text. If you notice any unmarked predicate, which you do want to align, please note them separately. We show an example text pair and annotation in the following:

A   "Peru's Luis Horna **clinched**$_1$ his second career ATP **title**$_2$ with a 7-5, 6-3 **win**$_3$ over local favourite Nicolas Massu on the **clay**$_4$ of Vina del Mar on Sunday."

A' "Luis Horna of Peru **defeated**$_1$ hometown favorite Nicolas Massu 7-5 , 6-3 for the first time in the Movistar Open final on Sunday and **claimed**$_2$ his second career ATP **title**$_3$."

Annotation (A–A'):   1–2   2–3   3–1

As you can see in the given example, the differences between corresponding predicates and their contexts can be very small: in some cases, a synonymous predicate (e.g., "win"–"defeat", "clinch"–"claim") was used, and in other cases, some extra information was introduced (e.g., "for the first time", "on Sunday"). However, not all cases are this simple. There might be a correspondency, which only becomes apparent when considering the actual meaning of the concerned predicate argument structures in context. Even if

the correspondency seems rather loose, we aim to also take these cases into account. Here is an example illustrating two such correspondencies:

B "(. . .) **Spokespeople**$_1$ at Pfizer's China **operations**$_2$ were not immediately available to **comment**$_3$ on Monday."

B' "(. . .) Phone **calls**$_1$ to Pfizer's China **headquarters**$_2$ in Beijing were not **answered**$_3$."

Annotation (B–B'):   2–2   3–3–P

Your task as an annotator will be to mark alignments as in the above examples. We provide pairs of texts in a simple text format for this task. As in the example annotations, you should write down the index pairs of corresponding predicates. We will provide you with a separate file for these annotations. Here are some additional guidelines to follow:

1. You have the option to mark alignments as "possible" using the suffix '–P' for cases, in which you feel uncertain or in which the text does not make clear, which event, state or object is being referred to. However, your main focus should lie on "sure" alignments, i.e., alignments that you are certain about.

2. You should prefer to mark alignments on a 1-to-1 basis whenever it is possible. However, you can indicate n-to-m correspondences when necessary.

3. Spend as much as time as needed to think about the meaning of marked predicates to make sure that you do not miss complex correspondencies that seem unlikely on first sight.

The next section describes the overall annotation process in a bit more detail.

2   Details

Before starting to annotate a pair of text, please make sure to read both texts carefully from beginning to end. This allows you to get an overall picture of the content and details included in each text. Depending on the length of a text, you might also want to pre-structure the document and remove paragraphs that are only contained in one of the two texts. However, please keep the predicate indices as they are for your annotation!

Once you have a good feeling for what the content of each text is, you can start the actual task. We do not provide a strict definition as to when two predicates correspond and should be aligned. As a rule of thumb, you can think of correspondency as a measure for how well one predicate argument structure can be replaced in context with another. If it is possible to exchange both predicate argument structures without changing the meaning of a text, then you should probably align the two.

As mentioned before, you should try to mark alignments on a 1-to-1 basis. However, there are cases, where this rule is not possible due to syntactic constructions and the meaning of predicates. For example, "**rear** and **spew**" are two predicates that can have the same meaning as the single predicate "**erupt**", depending on the context. In these cases, you should align all affected predicates in one text with all affected predicates in the other (for example, "1–1", "2–1").

Apart from simply marking two predicates as corresponding, you have the option to mark alignments as "possible". In particular, you should make use of this option, if you think that a correspondency between two predicates depends on one particular interpretation of one of the predicates. We have seen one such example in the previous section:

"spokespeople were not available to **comment**"

"phone calls were not **answered**"

3 Special Cases

There are a number of special cases that you should pay attention to in this task:

Exact correspondency. If two predicates are identical and their arguments overlap, they should almost certainly be annotated using a sure alignment. The only exception from this rule would be if the arguments occurred in reverse order and led to a contradictive meaning. In other words, you should not align cases such as C but you should always align an example such as D:

C "VW **bought** Porsche" – "Porsche **bought** VW" (incorrect)

D "VW **bought** Porsche" – "VW **bought** Porsche for USD 5.6bn" (correct)

Pronouns. When comparing the arguments and other contexts of two predicates, you should also check whether pronouns in one structure correspond to anything in the other. Here is another example:

E "He was **joined** by the Bassac River by his wife"

E' "Hun Sen's wife **stood** at her husband's side"

Spelling mistakes. Some newswire articles contain spelling mistakes and other errors. You can simply ignore them as long as the actual meaning of the text is still clear.

Approximate corrspondencies. Two predicates can correspond, even if they are not synonymous. In particular, this can be the case even if one predicate describes a different perspective on an event, state or object (e.g., **buy** vs. **sell**). It can also be the case that one predicate only describes a part of the concept described by the other (cf. example E). If it is clear that the event, state or object is the same though, you should also annotate these pairs using sure alignments.

E "The soldier was **killed** during a patrol in the area south of Baghdad."

E' "The soldier **died** in an attack close to the capital Baghdad."

Repetitions. If one newswire article refers to the same event, state or object multiple times, but the other article only once, then only the first correspondency should be marked as a sure alignment[1]. Further correspondencies should also be annotated but only as possible alignments ('–P'!). In general, if there are multiple references in both texts, each reference should be annotated using a sure alignment at most once. In these cases, you should mark the predicates with the highest information overlap as "sure". Here is an example:

F "Susan Boyle said she will **sing**$_1$ in front of Britain's Prince Charles (...) 'It's going to be a privilege to be **performing**$_2$ before His Royal Highness', the singer said (...) British copyright laws will allow her to **perform**$_3$ the hit in front of the prince and his wife."

---

[1]The intuition behind this guideline is that the first mention introduces the actual concept while later mentions just (co-)refer or add further information, i.e., they serve a different function with respect to the discourse.

F' "British singing sensation Susan Boyle is going to **perform**$_1$ for Prince Charles (...) The show star will **perform**$_2$ her version of Perfect Day for Charles and his wife Camilla."

Annotation (F–F'): 1–1 1–2–P 2–1–P 2–2–P 3–1–P 3–2

Note that the example annotation for F–F' only shows one possible way of aligning the occurring predicates. Depending on the interpretation of each predicate and its contexts, a different annotation might be equally good.