1 Goal

The goal of tagging is to find sentences and parts of sentences which convey identical, or near-identical, information, articulated in different forms.

2 Rules for Tagging

Given two sentences $S_1, S_2$, the annotator should decide between the following options:

2.1 Paraphrase

the pair would be considered a paraphrase in the following case:

1. A human who fully accepts $S_1$, must therefore accept $S_2$ as a whole.

2. The same holds for the second direction as well. Namely, accepting $S_2$, must yield acceptance of $S_1$.

2.1.1 Restrictions

• The acceptance can derive also by former knowledge, yet it must not be based solely on it. (examples 1,2)

• The pair will be considered as a paraphrase, when there is sufficient confidence that they indeed convey the same information. (example 3)

• Given a pair for which their meaning is ambiguous (for example - there is not enough information to be certain regarding the identities which appear in them). One must assume that the ambiguity is solved in such a way that the pair relate to same incident. (example 4)
• The pair should be judged from a timeless point of view. Namely, if the pair relate the same information, yet treat it from a different point in time - it should be tagged as paraphrase.

• The pair should be judged from a location-less point of view. Namely, if the same incident is reported, yet it is regarded from a different relative point of view, it should be considered paraphrase.

• If one of the sentences, or both of them, contain additional information, or contradictory information, the pair should not be tagged as a paraphrase.

2.2 Partial Paraphrase

If $S_1, S_2$ are not paraphrases, by the definition given above, yet certain clauses can be removed to form a paraphrase, then $S_1, S_2$ should be tagged as partial paraphrases. The additional clauses should be annotated as such. (example 6)

2.3 Negative

A catch-all rule. If $S_1, S_2$ do not fall under any of the previous tags, then they should be tagged as Negative, meaning they do not convey any shared information.

3 Examples

1. $S_1$ - .

   ступност Глфтес Тетас нус эм Хасти.

   China has reached an oil trade contract with Russia

   $S_2$ - .

   The world most populated nation has reached an oil trade contract with Russia

   Tag: Paraphrase.

   (It is of common knowledge that China is the world most populated nation)

2. $S_1$ - .

   ступност Глфтес Тетас нус эм Хасти.

   China has reached an oil trade contract with Russia

   $S_2$ - .

   China is the world most populated nation.

   Tag: Negative.

   (Although the second sentence is true, it does not follow from the first, nor does the second derive from the first).
3. 
USA has closed its embassy in Syria.

Tag: Paraphrase.
(It is very probable to assume that closing the embassy yields calling back the ambassador, and vice versa)

4. 
The network has covered latest events in Syria.

Tag: Paraphrase.
(It should be assumed that the network mentioned in the second sentence is the same network mentioned in the first sentence)

5. 
Trains will not work on the line to Beer Sheva this Sunday, due to strike.

Tag: Paraphrase.
(Although the pair convey the same information from different points in time, it should be regarded as paraphrase)

6. 
Benyamin Netanyahu has met with his Russian counterpart, an later returned to Israel.

Tag: Partial Paraphrase.
(Without the marked parts, the pair consists a paraphrase)
7. \( S_1 \) - Ehud Barak spoke in front of new army recruits.

\( S_2 \) - Ehud Barak decided to run for the next elections.

**Tag**: Negative.

(Both sentences do not convey the same information, nor can any clauses be removed from them so that they do)