SICK Re-Annotation Task

Aikaterini-Lida Kalouli Annebeth Buis Livy Real
Martha Palmer Valeria de Paiva

1 Annotation Guidelines

This task concerns the annotation of the textual inference relation between pairs of sentences.

1.1 Entailment between sentences

Textual entailment is a relation that holds between two sentences (or propositions) when from the one proposition we can infer the other, i.e. one proposition entails another. This means that the relation is directed, in the sense that we have to define which proposition implies which. If A entails B, it does not necessarily mean that B entails A as well. Entailment is a directed relation between two sentences. Textual entailment recognition is the task of deciding whether the meaning of one text is entailed (can be inferred) from another text.

You will receive two sentences, *A* and *B*. You should imagine *A* as a caption of a picture, describing whatever is on that picture. You should consider only the inference relation from *A* to *B*. You should also assume that sentence *A* represents everything you know about the world of the picture; *A* represents the truth based on which you have to judge sentence *B*. This means that if sentence *A* is talking about *a man in red pants walking* and sentence *B* is also talking about *a man in red pants running*, you should assume that both sentences are talking about the same man and the same event.

1.2 SICK Annotation Labels

The task of textual inference attempts to find what is the inferential relation between two sentences, *A* and *B*. The three labels to be used are:

- *E* for the entailment relation
- *C* for contradiction
- *N* for the neutral relation

In exceptional cases where you cannot decide at all for the correct inference relation you can also use a fourth label: *D* (don't know). Independently from the label you choose, you should write a justification for why you chose it, i.e.

which aspects of the sentences drove your decision. The justification doesn't have to be long, as long as it is clear why you made your choice. You will also be given extra room for comments, in case you wish to comment on something else about the sentences: you can cross the corresponding field if you think that one of the sentences is ungrammatical or non-sensical or just write some other comment in the extra field provided.

In the following, you will get some detailed examples of each inference relation and how to deal with them.

1.2.1 ENTAILMENT Label (E)

There is entailment from A to B when the validity of A implies the validity of B. This is to say, if A is true, B is necessarily true or, in other words, A implies B. Example:

(1) **A** = A skilled person is riding a bicycle on one wheel. *(remember, this is all that you know about the world of the picture)*
B = A person is riding the bicycle on one wheel.
Label = E
Justification = a *skilled* person is a person

1.2.2 CONTRADICTION Label (C)

There is a contradiction between A and B when the validity of A implies that B is not valid. This is to say, if A is true, B is false or, in other words, A and B cannot be true at the same time. Example:

(2) **A** = A child is hitting a baseball. *(remember, this is all that you know about the world of the picture)*
B = A child is missing a baseball.
Label = C
Justification = the child of the picture cannot be hitting and missing the same ball at the same time

1.2.3 NEUTRAL Label (N)

When the validity of A does not imply that B is either false of true, we label it NEUTRAL. This is to say, if A is true, we cannot say if B is also true or not; it could be true or it could be false. Consider that some NEUTRAL pairs may still refer to the same event or referent, but the truth value of B cannot be predicted from A. This is to say that there is no inference relation between the expressed content. Example:

(3) **A** = A man is crying. *(remember, this is all that you know about the world of the picture)*
B = A man is screaming.
Label = N
Justification = when somebody cries, it doesn't necessarily mean that

they are also screaming. They might be crying very very quietly, thus neutral.

1.3 Some important remarks

There are cases in which the referents of the sentences are not straightforward, i.e. the determiners/quantifiers *(a, the, one, some, few, two, etc.)* tend to confuse us whether we talk about the same entities and referents (people, events, etc). In these cases, you should assume that both sentences are indeed talking about the same referents and the determiners/quantifiers used *(a, the, one, some, few, two, etc.)* should NOT be sufficient to imply a non-contradiction relation. You should always remember that we are trying to describe this one specific picture and NOT what we might know about the world in general. Example:

(4) **A** = An Asian woman is holding a bag. *(remember, this is all that you know about the world of the picture)*
B = An Asian woman is not holding a bag.
Label = C
Justification = the woman of the picture cannot be holding a bag and not holding the same bag at a given moment. (don't get confused by the indefinite determiners here: we could NOT be talking about two different women, one holding a bag and one not holding a bag, because we are describing this one specific picture where the same woman is doing one specific thing)

Consider also that there are borderline cases in which it's not obvious if A contradicts B. Example:

(5) **A** = A dog is carrying a huge stick in its mouth. (remember, this is all that you know about the world of the picture)
B = A dog is carrying a small stick in its mouth.
Label = C
Justification = using common sense, something that is described as a huge stick cannot be a small stick at the same time, thus *huge* contradicts *small* (this matter is quite subjective but you should try and use common sense to judge pairs that might seem controversial).

1.4 Data Format

The format of the data you will be given is the following:

1. **column A:** id of the pair
2. **column B:** the sentence A of the pair
3. **column C:** the sentence B of the pair

4. **column D:** mandatory field: this will be your label for the pair (*E*, *C*, *N* or *D*)

5. **column E:** mandatory field: justification for the label you gave

6. **column F:** mandatory field: a "computational feasibility" score, i.e. a score from 1-5 of how hard you think this pair would be for an automatic system to get right (1 means that this pair would be very easily solved by an automatic system and 5 means that this pair is very hard for a system to get right)

7. **column G:** optional field: in this cell write A or B if you think that A or B, respectively, is ungrammatical

8. **column H:** optional field: in this cell write A or B if you think that A or B, respectively, is non-sensical

9. **column I:** optional field: a comment field where you can write any other comments you might have about the pair