Constructional Natural Language Inference Annotation Guidelines

We have developed a dataset of sentences featuring different linguistic "constructions"—pairing of form and meaning. The constructions exemplified in this dataset range from purely *substantive* (the words filling the constructional slots are fixed), such as the Much-less construction, e.g., "He won't eat shrimp, **much less** squid;" to purely *schematic* (the words filling the constructional slots can vary, but fulfill some general semantic and syntactic requirements), such as the Caused-Motion construction, e.g., "She blinked the snow off her eyelashes." It's ok if you aren't familiar with this terminology or the idea of these constructions!

Your job is to read the sentences from this dataset, which are presented as the Premise in a set of triples for the Natural Language Inference (NLI) task. Also known as Recognizing Textual Entailment (RTE), NLI is the task of determining the inference relation between two (short, ordered) texts: *entailment*, *contradiction*, or *neutral* (MacCartney and Manning 2008). For example, consider the following premise, taken from the Stanford NLI corpus:

Premise: A man inspects the uniform of a figure in some East Asian country.

Given this premise, you will also be supplied with a Hypothesis, for the above example:

Hypothesis: The man is sleeping

Then, you will fill in the Relation between the Premise and the Hypothesis; which indicates the kind of entailment between the two sentences. We are using numerical coding, which is also listed in your annotation spreadsheet as a reminder:

0 – entailment – The hypothesis *must be true* given the premise
1 – neutral – The hypothesis *may or may not be true* given the premise
2 – contradiction – The hypothesis *must not be true* given the premise

So for the above premise/hypothesis pair, the answer would be

2 – contradiction – If the man is inspecting a uniform, as the premise states, then it *must not be true* that the man is sleeping.

Note that the two sentences are distinct descriptions of the same scenario. Any entities mentioned are the same entities; for example, if a/the "man" is mentioned in the premise and hypothesis, these refer to the same man. They do not describe the same scenario at two different times, but rather the same time snapshot; for example, the "man" in the above premise may sleep at a later time, but it is not possible that he is sleeping at the same time as the scenario in which he is inspecting the uniform.

If you encounter any words that you are not familiar with, you may look up the word in a dictionary. However, it is not expected or encouraged that you would have to "do research" into a topic in order to determine a relation between a premise and hypothesis. Instead, you should rely on common sense and your understanding of the words.

You will be completing these annotations in a spreadsheet, where there is a relation space available below a given premise/hypothesis pair, like this:

CxN Type	**Number**	**P/H/R**	**Annotation Target**				
let-alone	2	premise	A ceasefire, let alone lasting peace, will take long negotiation.				
	2	hypothesis	Negotiating lasting peace will take less time than negotiating a ceasefire.				
	2	relation					

In the cell to the right of "relation", you will provide the appropriate relation number; such that your final annotation will look like this (see highlighted number inserted):

CxN Type	**Number**	**P/H/R**	**Annotation Target**				
let-alone	2	premise	A ceasefire, let alone lasting peace, will take long negotiation.				
	2	hypothesis	Negotiating lasting peace will take less time than negotiating a ceasefire.				
	2	relation	2				

This space should include nothing except for the numbers 0, 1, or 2.

If you would like to note instances that are problematic, please add a Notes column to the right of Annotation Target and make your note as relevant to the right of the premise, hypothesis or relation.

There is also a space in column L where you can note the start and end times of each annotation session, or "sitting." Please kindly track about how many judgments you are able to do in each sitting, so we can get a sense of how long the annotation task takes.

There are about 100 distinct NLI judgments or triples per annotation spreadsheet. Once you have completed all the relation annotations, please save and send the spreadsheet back.

Get your inference hat on. Happy annotating!