Annotation Guidelines for Monolingual Alignment

Victoria Xiao

February 6, 2017

1 Introduction

For this task, you are given two sentences which are paraphrases of each other. Your task will be to show which parts of the sentence are in correspondence by aligning them on a word-to-word basis. This kind of correspondence is called an "alignment".

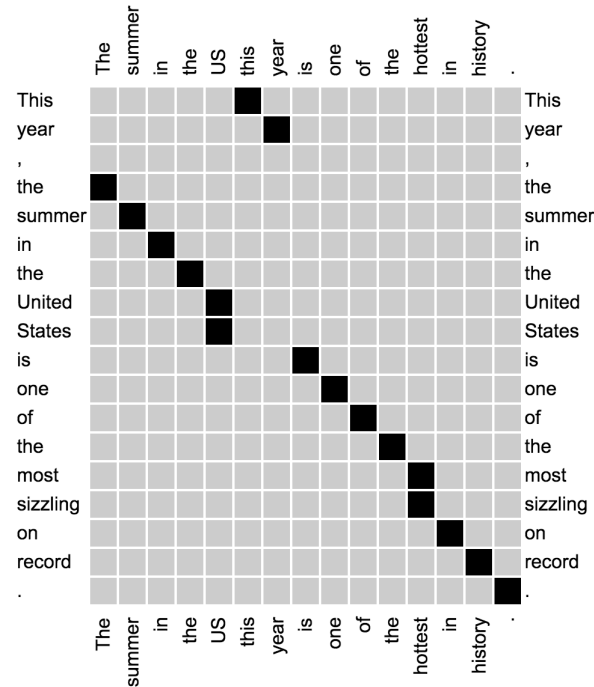Here is an example of the kind of alignment we would like you to produce.
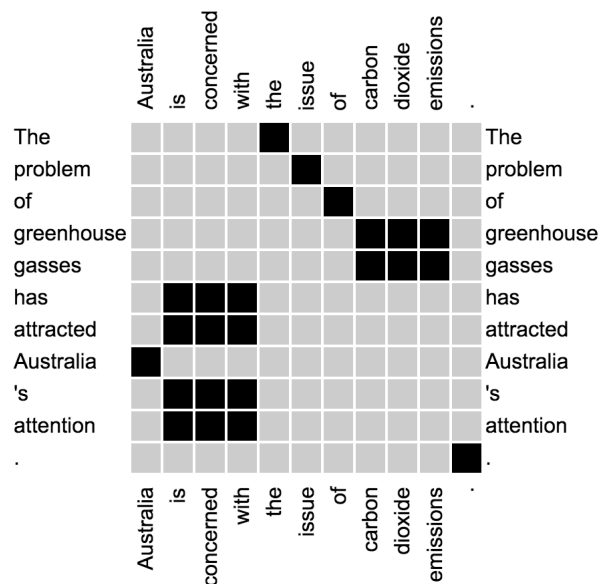
2 Aligning Sentence Pairs

The majority of paraphrases in this project will express similar concepts using different words and phrases. In this case, your task is to mark alignments between words that correspond to each other. Words correspond to each other when they express the same idea or when they can be substituted for each other in their respective sentences.

Click on a square in the grid to turn the square black. This indicates a correspondence between two words.

If you encounter two sentence pairs that are completely unrelated to each other, check the box for Not a paraphrase. If the sentence pairs have approximate but not identical meanings, you should still attempt to mark correspondences.

	The	summer	in	the	US	this	year	is	one	of	the	hottest	in	history	.	
This						■										This
year							■									year
,																,
the	■															the
summer		■														summer
in			■													in
the				■												the
United					■											United
States					■											States
is								■								is
one									■							one
of										■						of
the											■					the
most												■				most
sizzling												■				sizzling
on													■			on
record														■		record
.															■	.

Often, you will encounter sentences that are constructed very differently, making it difficult to make word-for-word alignments. In this case, you may use a many-to-many block alignment to mark phrases that correspond to each other. You should always aim to make alignment blocks as small as possible. Only create alignments between the smallest units of corresponding meaning.

	Australia	is	concerned	with	the	issue	of	carbon	dioxide	emissions	.	
The					■							The
problem						■						problem
of							■					of
greenhouse								■	■	■		greenhouse
gasses								■	■	■		gasses
has			■	■	■							has
attracted			■	■	■							attracted
Australia	■											Australia
's			■	■	■							's
attention			■	■	■							attention
.											■	.

Words do not have to be be exact synonyms in order to be in correspondence with each other. They only have to convey the same idea. For example, in the following example, "lived in" should be aligned with "were inhabited by", since both of these phrases indicate the same kind of relationship.
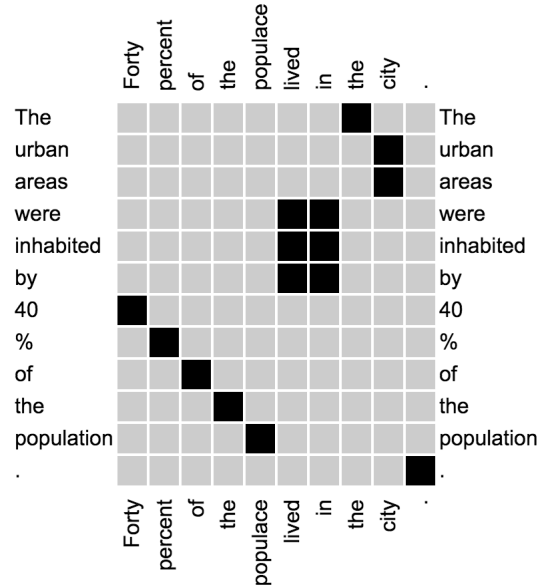

3 New information

Sometimes, one sentence will contain information that is not present in the other sentence. Indicate that words convey new information by clicking on them. This will turn its corresponding row/column pink. Only mark words that convey actual content as new information. Do not mark function words (eg. and, that, for, to), punctuation, connectives (eg. however, therefore) or temporal markers (eg. then) as new information unless they are part of a phrase that conveys content.
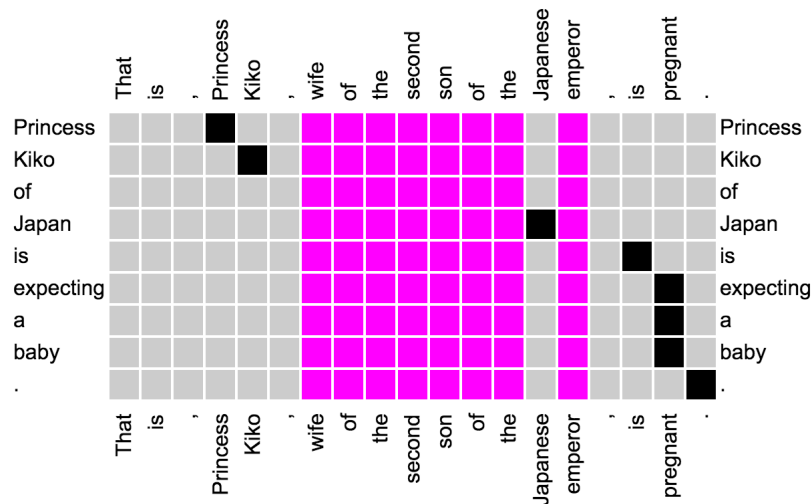
4 Possible alignments

There are some phrases across sentences which have very loose correspondence. These may be phrases that have similar, but not identical meanings. They can also be phrases that fulfill the same function in the sentence but technically mean different things. Mark the words in question as possible alignments by clicking their corresponding square twice. This will turn the square grey.
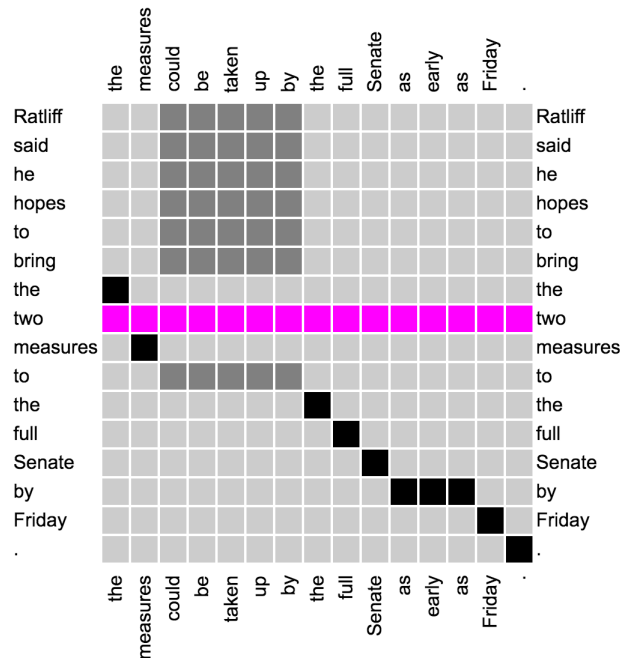
5 Special Cases

Typos and Approximate Correspondences
Some of the sentences you encounter will contain typos (eg. help meet instead of helpmate). You may also encounter sentence pairs with non-identical numbers (eg. 0.7 litres vs. 1.5 litres) Both of these instances should be marked with possible alignments.
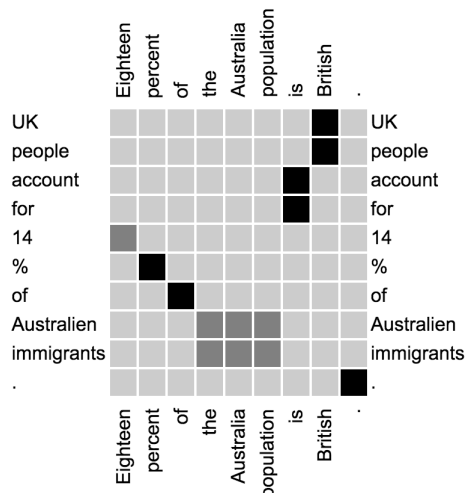
Named entities

Some of the sentences will refer to specific people, places, organizations or other named entities.
Differing titles that refer to the same individual, location or organization should be block-aligned.
For example, Elizabeth II should be block-aligned with the Queen of England.

If a pronoun (eg. she, him, herself) corresponds with a name or title, mark the correspondence as
a possible alignment.

Repetition

Sometimes, you will encounter an idea or named entity that is repeated multiple times in a sentence. In this case, mark one instance of the idea or named entity as a sure alignment, then mark any other instances as possible alignments.
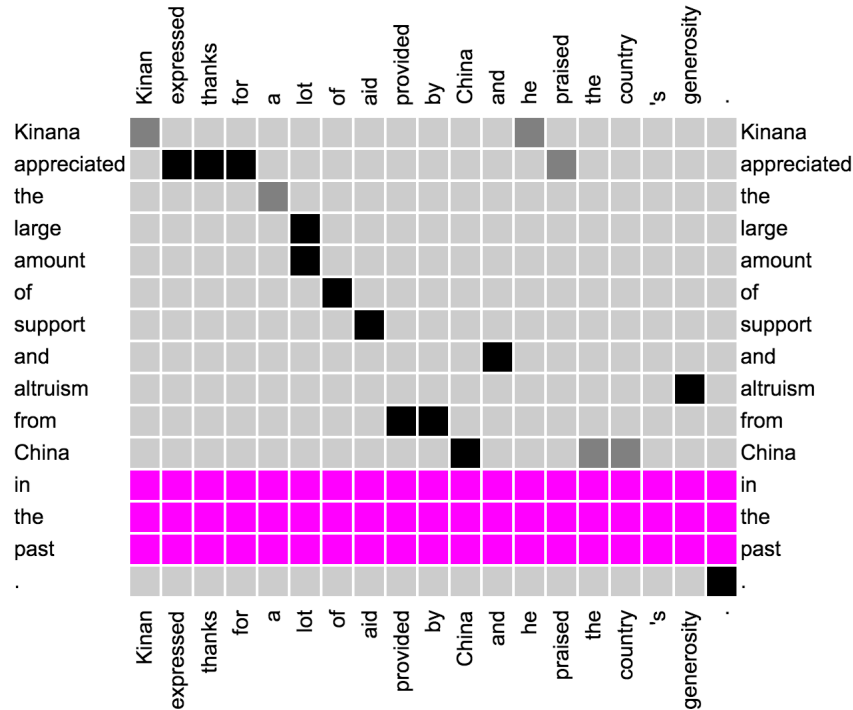
Tenses

Some of the sentence pairs will contain verbs with different tenses (eg. walked vs walk or walked vs. will walk). These verbs should be marked as a possible alignment.
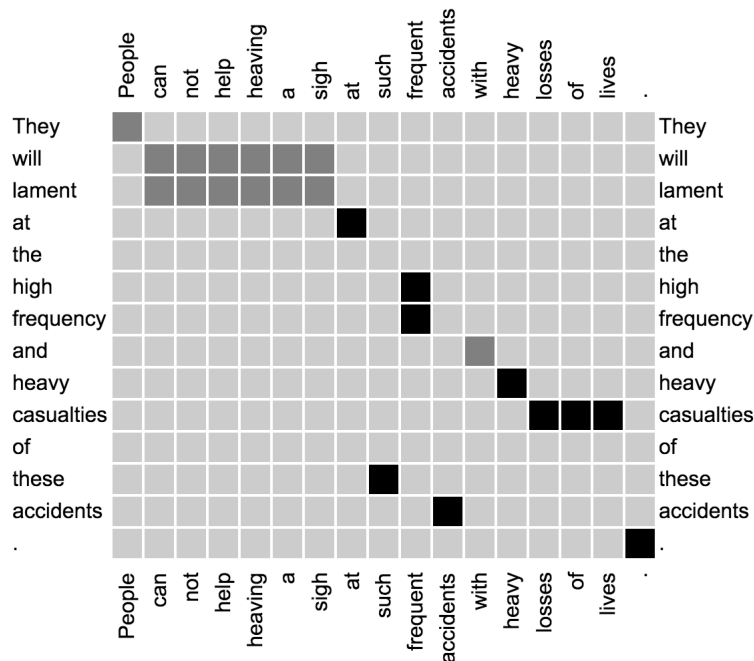
Genitives

If it is ever encountered, 's should be aligned with of, as well as any determiner (eg. the, my, a) that is introduced.

Determiners

Determiners are words that come before a noun and give us context about the noun (ie. the, a, my, this). Sometimes, the same noun will be introduced with a different determiner (ie. a man vs the man. In this case, mark the two determiners as a possibly alignment.

You may also encounter cases where one word is missing a determiner altogether (ie. frost vs. the frost . In this case, do not align the determiner.