

# Event Paraphrase Annotation Guidelines

State: 0.4; 25.03.17

Author: Darina Benikova, darina.benikova@uni-due.de

## Table of Contents

Introduction .....	2
Definitions .....	3
Paraphrase .....	3
Event .....	3
<i>Event Elements</i> .....	3
Annotation Process.....	4
Paraphrases on sentence level.....	4
Annotation of event elements .....	5
Paraphrases on the event level .....	6
Annotation of Element Paraphrases .....	7
Special cases .....	9

## Introduction

**Events in Social Media are textual descriptions of actions and facts in the real world.** The same real world events can be expressed in different ways. For instance, “Trump lives in Washington.” and “Donald Trump resides in the White House.” are paraphrases of the same event.

Methodologies for detecting event paraphrases would be helpful in many tasks of natural language processing, such as summarization, information extraction, plagiarism detection, machine translation, and question answering. **Paraphrases describe differently worded pieces of text with the same content or in other words they can be described as bidirectional entailment.**

We believe that by detecting paraphrases on a more granular level than on the sentence level as it is currently performed, we detect more precise and accurate paraphrases.

However, previous works have mostly focused on paraphrases on the sentence level.

To measure and quantify the difference between paraphrases on different levels, namely the sentence and the event level, the annotation on both levels will be conducted as herein described.

You will be provided with tuples from existing sentence paraphrase corpora.

The three levels that you will be annotating on are depicted in Figure 1.

The two lower levels will be explained in more detail further on.

**Shortly, it shall be stated that we consider events as so-called verb-argument structures: one main verb and all its arguments (subject, objects, ...).**

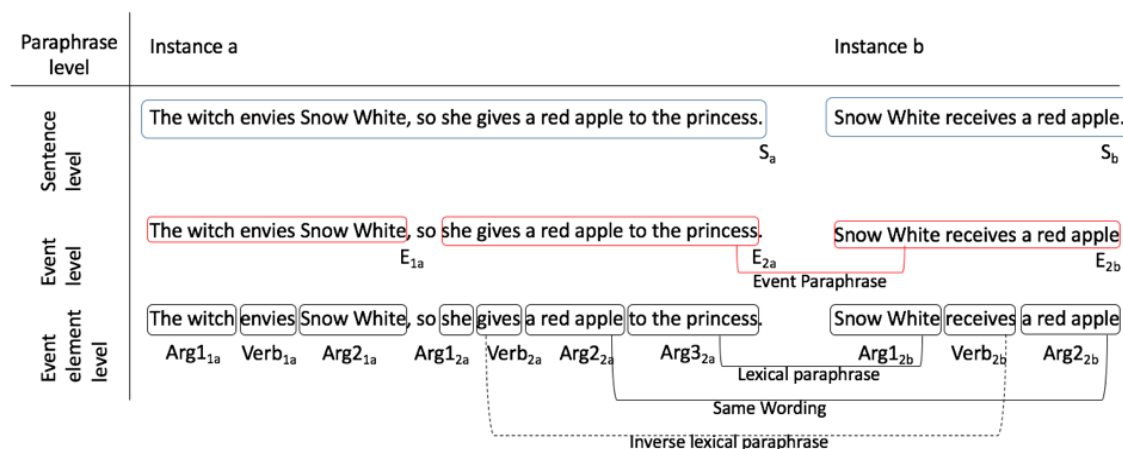


Figure 1 Example of paraphrases on sentence and event level

### Definitions

#### Paraphrase

Here, a paraphrase will be defined as a piece of text with clear boundaries (Instance a) that has the same meaning as another piece of text (Instance b) with clear boundaries.

Also, we define paraphrase a bi-directional entailment, meaning that one text (Instance a) must result in the other (Instance 2) and the other way around. According to this definition, in Figure 1, the two instances

- a) She gives a red apple to the princess
- b) Snow White receives an apple

have the same meaning and also, a) results in b) and b) results in a).

However, the two instances

- a) She gives a red apple to the princess
- b) Snow White holds an apple

are not paraphrases. Although a) results in b), b) does not result in a).

#### Event

Here, an event consists of the main verb and its arguments. A main verb is sometimes also called *full verb* and is the opposite of an auxiliary verb. Arguments are those pieces of text that are grammatically dependent on the main verb, e.g. the subject and the objects.

In Figure 1, the first sentence, “The witch envies Snow White, so she gives a red apple to the princess.”, has two events, as it has two main verbs. The second event consists of the main verb, “gives” and the arguments “she”, “a red apple”, and “to the princess”.

#### Event Elements

Event Elements are the main verb and the arguments.

In case the main verb is a phrasal verb (consisting of more than one token), the whole span of the main verb is regarded as such, e.g. in the event

She asked him out

the verb would consist of both “asked” and “out”.

The arguments consist of the whole span of the argument, meaning of the head of the argument and everything that is dependent on it, e.g. in Figure 1, the argument “a red apple” consists of the head of the argument, namely “apple” and all its dependents, in this case the article “a” and the adjective “red”. Articles may consist of full clauses, e.g. “a red apple, that was given to her by her evil stepmother”.

### Annotation Process

This section gives a short overview of the three sequential annotation steps. Each step will be discussed in a dedicated subsection.

First, you will be provided with two sentences. If you see any annotations (but your own), you are requested to turn them off.

In the first step of the annotation, you are requested to annotate whether the two presented sentences are paraphrases.

In the second step, you are requested to annotate the events, so switch the event structure annotations on and decide whether the span it is encompassing is a paraphrase of an event in the second sentence.

In the third step, you decide whether the individual event elements are paraphrases of event elements in the second sentence.

In this annotation study, we will be using the web-based annotation tool WebAnno. In case of questions concerning the tool, first consult

<https://webanno.github.io/webanno/releases/2.3.0/docs/user-guide.html> .

In each of the levels you will have to make an alignment label decision between three different kinds of paraphrase link:

- Same wording, meaning that the lemmas of the word(s) in the span have the same lemma(s)<sup>1</sup>
- Sure Paraphrasing, meaning that the wording differs, but you are sure that this is a paraphrase
- Unsure paraphrase, meaning that you are not entirely sure whether the words are paraphrases<sup>2</sup>

The last does not have to be explicitly annotated. The second one should only be chosen in cases in which you are unable to decide. As it is easiest to show this on the event element level, the examples for each of the paraphrase link kinds are presented in the corresponding section.

### Paraphrases on sentence level

In the first step, you have to judge whether the two presented sentences are paraphrases. In order to be as unbiased as possible, please switch off the event annotation<sup>3</sup> and just make your decision based on the definition provided in the previous section.

---

<sup>1</sup> A lemma is the canonical form, dictionary form, or citation form of a set of words, e.g. the lemma of *giving*, *gave*, *gives*, and *given* is *give*.

<sup>2</sup> This label should only be used in case you are really unsure concerning the paraphrase, so please use it only in this occasion.

<sup>3</sup> The description of how to switch layers on and off can be found here:

[https://webanno.github.io/webanno/releases/2.3.0/docs/user-guide.html#\\_layers](https://webanno.github.io/webanno/releases/2.3.0/docs/user-guide.html#_layers)

## Event Paraphrase Annotation Guidelines

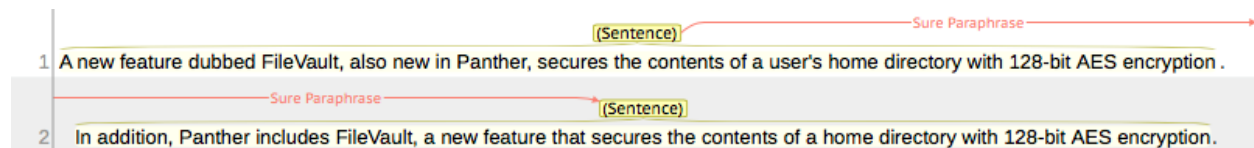


Figure 2 Example for sentence paraphrase annotation

### Annotation of event elements

The definition of event elements is described in the previous section. You will be presented with automatic pre-processing, the full verbs as well as their argument spans will be annotated in the first sentence.

First check the event elements marked in the first sentence.

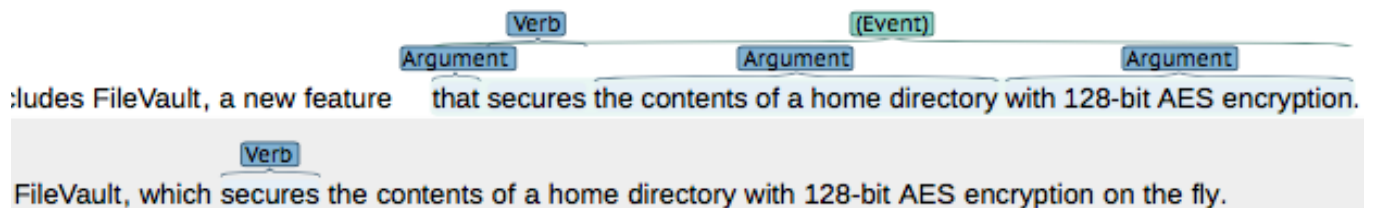
Then perform the following steps:

- 1) Find the event in the second sentence that corresponds best to the event in the first.

If it is difficult to find a matching event in the second sentence, first try to find an event that has similar arguments or a high lexical overlap. If this is also impossible, choose the first event in the sentence.

- 2) Annotate the verb of the event

When annotating the verb, please remember that we only consider main verbs, meaning that auxiliary and modal verbs, participles, and gerunds are not considered as verbs.



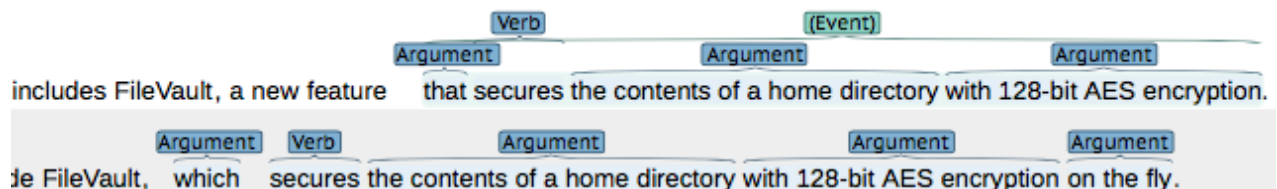
- 3) Find the heads of the arguments of the verb

The arguments of the verb are the subject and the objects on the one hand, but also temporal or local information, e.g. expressed through short phrases such as “last year” or “in Paris”, but also through clauses, such as “after seeing what had happened” or “where the statue was standing”.

Connectives such as e.g. “furthermore”, “in addition”, “but”, or “whether” are not considered as arguments.

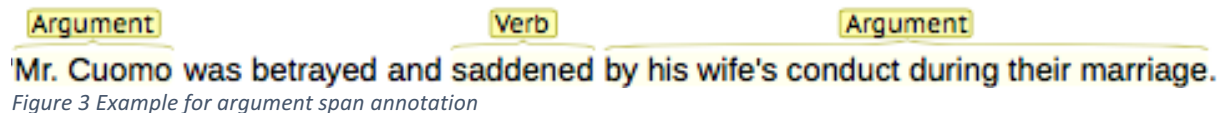
- 4) Find all the elements that are dependent on the heads

Elements depending on the head might be adjectives, or clauses describing the head. Mark all the elements that semantically depend on the head of the argument as part of the argument.



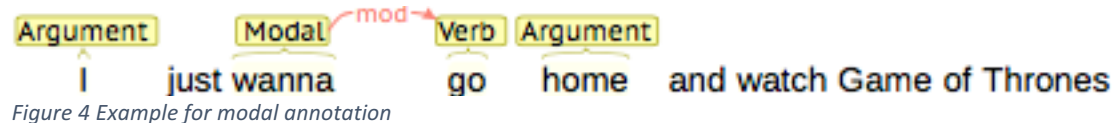
In the example in Figure 5, the second argument is “by his wife’s conduct during their marriage”, as all the phrases refer to the argument head, “wife”. It could be assumed that “during their marriage” could be an argument of its own, but I does not refer to “saddened”.

## Event Paraphrase Annotation Guidelines



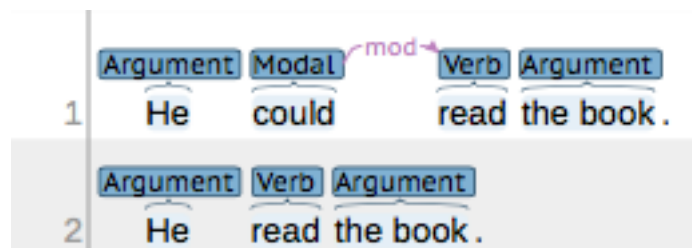
The example also shows that the span of the argument should also encompass the preposition referring to the verb.

If the main verb is used in connection with modals or negations, this should be annotated as shown in the following:

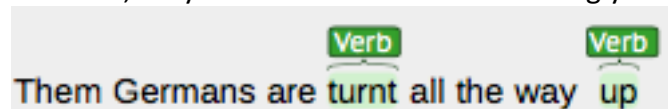


The link between the modal and the verb is only for your understanding, you do not need to annotate this in your own annotation.

If the verb is linked to a modal, it should be annotated considering this information, e.g. in Figure 5, the verbs are not paraphrases due to the use of “could” in the first event.

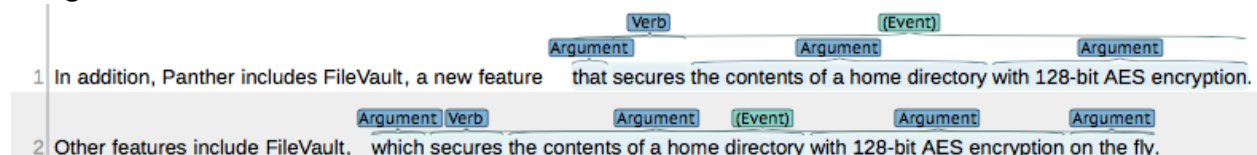


Phrasal verbs cannot be annotated as one span if there would be also other words involved. In this case, they should be annotated followingly:



### Paraphrases on the event level

Having annotated the event elements of the second sentence, mark the event span, as shown in Figure 7.



Please consider that the event consists of the marked elements, so in case there is additional information in the span that is not marked through elements, it is not part of the event. For

## Event Paraphrase Annotation Guidelines

instance, in Figure 8 the event is “That fire charred 469,000 acres”, not including the information that “[it] devastated timber on the reservation”.



Figure 8 Span with information that is not part of the event

After having annotated the whole span of the event and align it with an event in the second sentence, if applicable.

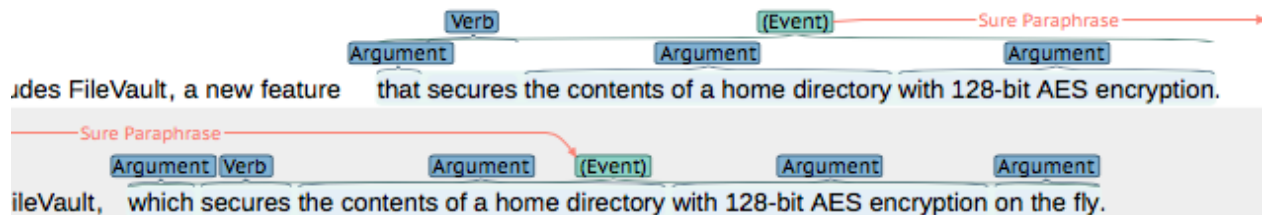


Figure 9 Example of event paraphrase alignment

## Annotation of Element Paraphrases

Given the event element annotation in the two sentences, you should link those that are paraphrases of each other according to the classes mentioned previously. Figure 10 shows a full exemplary annotation of the event elements.

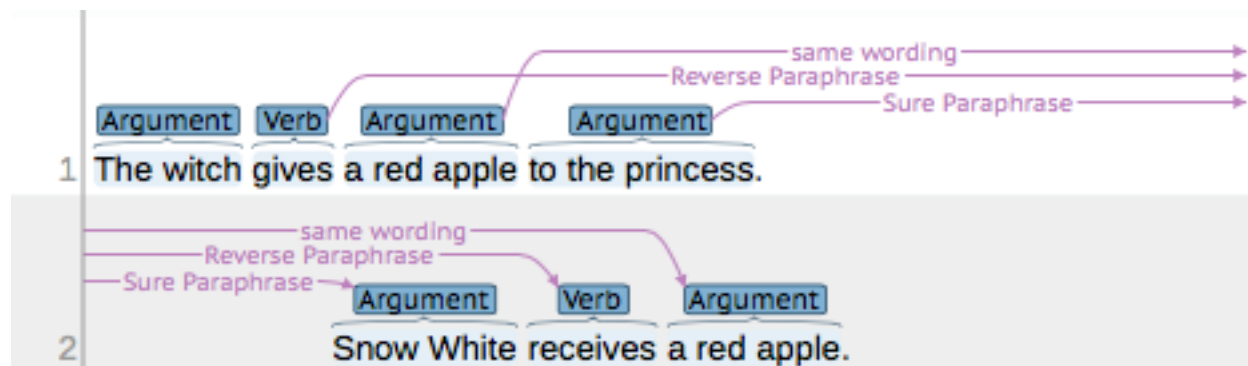


Figure 10 Example with all paraphrase link classes on the event element level

In the following, examples for the paraphrase classes “Same Wording” and “Sure Paraphrase” are shown:

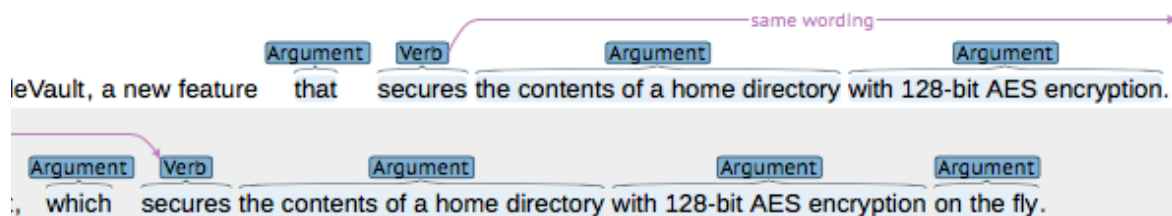


Figure 11 Example for same wording paraphrase annotation

## Event Paraphrase Annotation Guidelines

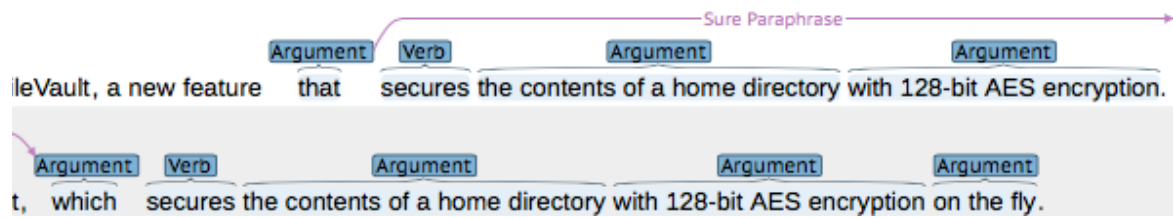


Figure 12 Example for sure paraphrase annotation

As the “Unsure paraphrase” class should only be used in cases you are not sure, we cannot give an example for that.

On the event paraphrase level, there is an additional paraphrase class for the verb element, the “Reverse paraphrase” and it should be used in cases where the same event is described with antonymic or opposite words, e.g.

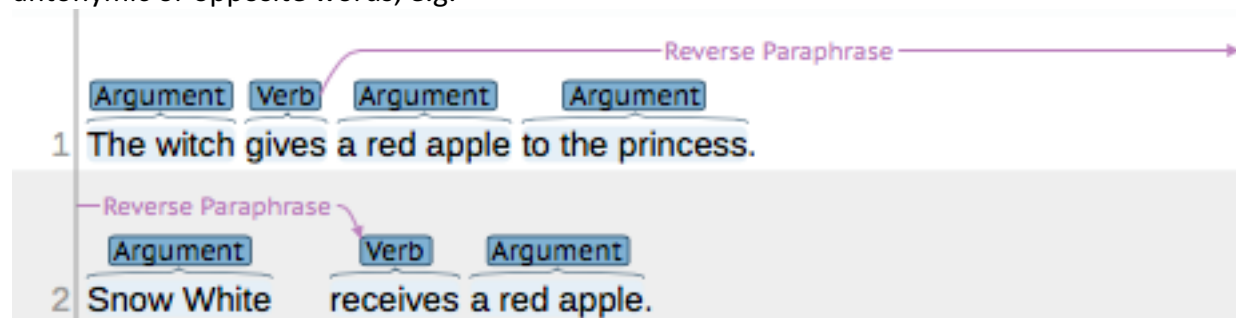


Figure 13 Example for reverse paraphrase

If there is no paraphrase, nothing is aligned or annotated.



## Event Paraphrase Annotation Guidelines

### Special cases

If one instance contains more or less information than the other, it may still be a paraphrase.

e.g.

They left the house.

They left the house on Monday.

If one instance contains information contradicting the other instance, they cannot be paraphrases.

They left the house on Wednesday.

They left the house on Monday.

If you have relative pronouns that are used instead of a full phrase, mark the pronoun as argument. In the first sentence there are co-references marked for your convenience, but you are not asked to do so.

The argument of interest is marked in bold in the following example.

e.g.

The pastry cook, who makes wonderful cakes, made a statement in the morning.

There are 2 events:

The pastry cook, [**who** makes wonderful cakes], made a statement in the morning.

[**The pastry cook, who makes wonderful cakes**, made a statement in the morning].

If you find orthographic or syntactic mistakes, please annotate as if they weren't there.

e.g.

Theyve done a great job.

"Theyve" would be an argument of done.