

f) Google was in talks to buy YouTube.
Google bought YouTube.

g) Mary and Wendy went to the cinema .
Mary and Wendy like each other .

4. Identical tag: Once all other phenomena are annotated, snippets which are identical in both sentences may remain. We should annotate as IDENTICAL these **snippet (not linguistic unit)** residues (h). In this case, we do not follow the linguistic unit criteria (Section A.2.3.1).

Only one (discontinuous) identical tag will be used in each pair of sentences.

Punctuation marks will also be annotated as IDENTICAL if they effectively are.

h) The two argued that only a new board would have had the credibility to restore El Paso to health.
The two believed that only a new board would have had the credibility to restore El Paso to health.

Finally, it should be noted that tags overlap on many occasions. In (i), a SAMEPOLARITY tag overlaps an ORDER one.

i) shaking his head wisely .
sagely shaking his head.

A.2.3.3 Should Punctuation Marks Be Included?

When a whole phrase/cause/sentence is annotated, **the closing (and opening) punctuation mark (if any) is(are) included**. Some examples are (a) and (b), which are cases of DIATHESIS and ADDITION/ELETION, respectively. In contrast, in (c) and (d), the commas are not included as they are not the opening and closing punctuation marks of the paraphrase phenomenon tagged (SAMEPOLARITY), but of a bigger unit.

a) This song (John sang it last year in the festival) will be a great success.
This song (it was sung by John last year in the festival) will be a great success.

b) His judgment have kept equal pace in that conclusion.
His judgment and interest may , however , have kept equal pace in that conclusion.

- c) Before leaving and before saying goodbye , I looked around.
Before leaving and before the bye bye moment , I looked around.
- d) My sisters, lovely girls, live in Melbourne.
My sisters, nice girls, live in Melbourne.

A.3 Tagset Definition

In the following, the annotation specifics are presented. For each tag, we provide (1) the definition and (2) examples both for “positive sense preserving” and “negative sense preserving” instances, where applicable.

A.3.1 Morphology based changes

Morphology based changes stand for those paraphrases that take place at the morphology level of language. Some changes in this class arise at the morphology level, but entail significant structural implications in the sentence. Only the linguistic unit affected by the trigger morphology change is annotated.

A.3.1.1 Inflectional changes

Definition: Inflectional changes consist in changing inflectional affixes of words. In the case of verbs, this type includes all changes within the verbal paradigm.

Negative sense preserving inflectional changes lead to significant changes in the meaning of the whole text, thus giving raise to non-paraphrases.

- Positive sense preserving:
It was with difficulty that the course of **streets** could be followed.
You couldn’t even follow the path of the **street**.
- Negative sense preserving:
You can’t travel from Barcelona to Mallorca with the **boat**.
Boats can’t travel from Barcelona to Mallorca.

A.3.1.2 Modal verb changes

Definition: The MODAL VERB tag stands for changes of modality using modal verbs.

- Positive sense preserving:
I was still lost in conjectures who they **might** be.
I was pondering who they **could** be.

A.3.1.3 Derivational changes

Definition: The DERIVATIONAL tag stands for changes of category by adding derivational affixes to words. These changes comprise a syntactic reorganization in the sentence where they occur.

- Positive sense preserving:
I have heard many accounts of him all **differing** from each other.
I have heard many **different** things about him.

Although drivers and driving are linked by a derivational process, in the following example this type is classified as SAME-POLARITY, and not as DERIVATIONAL, because there is not an actual change of category, both are acting as nouns.

- I dislike rash drivers.
I dislike rash driving.

A.3.2 Lexicon based changes

Lexicon based change tags stand for those paraphrases that arise at the lexical level.

Always the **smallest** possible lexical unit has to be annotated. In (a), we should not consider one single paraphrase phenomenon because it can be divided into two lexical units pairs: often-debated/much-disputed (1) and issue/question (2). These SAME-POLARITY substitutions are independent paraphrase phenomena, as we could substitute often-debated by much-disputed, leaving issue unchanged (much-disputed issue). Thus, two different SAME-POLARITY tags should be used. In contrast, in (b), lies and is revealed should not be tagged on their own as SAME-POLARITY substitutions, as they are semantically embedded in the wider lexical units lies its appeal and its appeal is revealed, respectively. The tag used in this case is, again, SAME-POLARITY.

- a) often-debated₁ issue₂
much-disputed₁ question₂
- b) Here by virtue of humanity's vestures, lies its appeal .
Here by virtue of humanity's vestures, its appeal is revealed .

Auxiliaries and infinitive marks are not tagged within the lexical unit in question. Only the verb to be, when it is part of a passive voice, should be included in the scope (c).

- c) The viewpoint of these lands had been altered .
The whole aspect of the land had changed.