

- a) how the headmaster failed / the failure of the headmaster
- b) We were able to hear the report of a gun on shore intermittently / the report of a gun on shore was still heard at intervals
- c) I'm guessing we won't be done for some time / I've got a hunch that we 're not through with that game yet

Miscellaneous changes comprise types not directly related to one single class. Finally, in paraphrase extremes, two special cases of paraphrase phenomena should be considered: they consist of the extremes of the paraphrase continuum, which goes from the highest level of paraphrasability (identity) to the lowest limits of the paraphrase phenomenon (entailment). Non-paraphrase fragments within paraphrase pairs are also part of the class paraphrase extremes.

As some of the names of our types explicitly reflect (e.g. ADDITION / DELETION), they are **bidirectional**: in a paraphrase pair, they can be applied from the first member of the pair to the second and vice versa.

ETP contains both "sense preserving" atomic phenomena (atomic paraphrases) and "non sense preserving" atomic phenomena (atomic non-paraphrases). While some phenomena are considered to (almost) always preserve the meaning (ex.: abbreviation, habitual same polarity substitution), other phenomena are not innately preserving the meaning and can lead both to paraphrasing and to non-paraphrasing at the textual level (ex.: In (d) and (e) the involved phenomena is the same - "inflectional change", however in (d) the two texts are paraphrases, while in (e) they are not). The "sense preserving" feature is required for the annotation of the "non-paraphrases".

- d It was with difficulty that the course of streets could be followed.
You couldn't even follow the path of the street.
- e You can't travel from Barcelona to Mallorca with the boat.
underlineBoats can't travel from Barcelona to Mallorca.

A.2.3 The Scope

The scope refers to the selection of the tokens to be annotated within each tag. In what follows, we first define the type of units we are willing to annotate (Section A.2.3.1), the criteria followed in the scope selection (Section A.2.3.2) and when the punctuation marks should be included (Section A.2.3.3).

A.2.3.1 Kind of Units to Be Annotated

We annotate **linguistic units**, not strings that do not correspond to a full linguistic unit. These linguistic units can go from the word to the (multiple-)sentence level.

In the paraphrase pair in (a), although a change takes place between the snippets *here by* and *it is there in*, two paraphrase mappings have to be established between *here* and *there* (1), and *by virtue of* and *in virtue of* (2), two different pairs of linguistic units.

- a) Here ₁ by virtue of ₂ humanity's vestures.
It is there ₁ in virtue of ₂ the vesture of humanity in which it is clothed.

However, selecting full linguistic units is not always possible or adequate from the paraphrase annotation point of view. In the following, we set out some exceptions to the above rule:

1. Cases in which only one member of the paraphrase pair corresponds to a linguistic unit. In (b), a SEMANTICS BASED CHANGE occurs between the underlined fragments. In the first sentence, it consists in a full linguistic unit, namely a causal clause; in the second sentence, the semantic content in the first appears divided into a nominal phrase and part of a verbal phrase, i.e., the verb *has impressed*. This nominal phrase plus the verb, although they do not constitute a full linguistic unit, are the scope of the phenomenon in the second sentence

- b) There is a pattern of regularity and order in the entire cosmos, due to some hints that science provides us.
A presiding mind has impressed the stamp of order and regularity upon the whole cosmos.

2. Cases in which none of the members of the paraphrase pair correspond to a linguistic unit. The prototypical example of this situation are contractions, within the SPELLING tag. In (c), *I* constitutes a nominal phrase and *will* is part of a verbal phrase. As the contraction is produced between these two pieces, they and only they constitute the scope of the phenomenon.

- c) I will go to the cinema.
I'll go to the cinema.

3. Cases of identical (see Section A.2.3.2)

A.2.3.2 Scope Annotation Criteria

The way the scope should be annotated depends on the class of the tag. Three criteria should be followed:

1. Morpho-lexicon based changes, semantics based changes and miscellaneous changes: only the linguistic units affected by the trigger change are tagged.

- a) I dislike rash motorists .
I dislike rash drivers .
- b) He rarely makes us smile .
He has little to do with making us smile .

2. Structure based changes: the whole linguistic unit suffering the syntactic or discourse reorganization is tagged (light green rectangle in Figure 2). If the reorganization takes place within a phrase, the phrase is tagged. If the reorganization takes place within a clause, the clause is tagged. If the reorganization takes place within a sentence, the sentence is tagged. If the reorganization takes place between different phrases/clauses/sentences (mainly coordination and subordination phenomena), all and only the phrases/clauses/sentences affected are tagged. In the case of clause changes, if the reorganizations takes place within the subordinate clause, only this one is annotated (not the main clause) and vice versa.

Moreover, all structure based changes (except from diathesis alternations) have a **key element** that gives rise to the change and/or distinguishes it from others. This key element is also annotated. First, the whole linguistic unit (including the key element) is tagged, and then the key element is annotated independently.

In (d), an active/passive alternation takes place (DIATHESIS tag). As the change takes place within the subordinate clause, only this clause is tagged. In (e), a change in the subordination form takes place (SUBORDINATION & NESTIG tag). As the change affects the way the two clauses (the main and the subordinate) are connected, the whole sentence is tagged. The connective mechanisms (the conjunction and the gerund clause) are annotated as key elements.

- d) When she sings that song, everything seems possible.
When that song is sang, everything seems possible.
- e) **When** we hear that song, everything seems possible.
Hearing that song, everything seems possible.

3. Entailment and non-paraphrase tags: the affected linguistic unit is tagged. The example in (f) is a case of ENTAILMENT; the example in (g) is a NON-PARAPHRASE.