

# STM-Corpus Coreference-Resolution Annotation Guidelines

Supplementary material for the paper: Brack A., Müller D., Hoppe A., Ewerth R.: Coreference Resolution in Research Papers from Multiple Domains, ECIR 2021.

20th December 2020

## Noun phrases

A noun phrase is a phrase, where the head of the phrase is a noun modified by zero or more components in the phrase. These components can range from determiners (“*this*”, “*the*”, “*our*”), to noun premodifiers (“*turbine*” in “*turbine performance*”), and attributive adjectives (“*fast*”, “*passive*”), which appear before the head of a phrase. Components can also be prepositional phrases (“*in the guidelines*”) or relative clauses (“*which was used*”). Furthermore, a noun phrase can contain multiple smaller noun phrases. A noun phrase can be identified by checking if replacing it with a pronoun would change the meaning of the sentence. For example “*The wind speed*” in “*The wind speed changes rapidly.*” can be replaced by “*it*” without changing the meaning of the sentence, thus “*The wind speed*” is a noun phrase.

(1) A noun phrase can be marked as a coreference mention, if the phrase contains at least one entity mention and if marking the noun phrase doesn’t violate any of the other rules.

## Concept Mentions

The STM-Corpus already contains concept mentions, which in most cases are nouns or noun phrases, in some cases verbs and in very few cases numbers.

(2) Phrases contained entirely inside concept mentions cannot be marked as coreference mentions.

(3) Exception to (2): pronouns.

(4) Concept mentions can be marked as coreference mentions if marking the concept mentions doesn’t violate any of the other rules.

Examples:

- Neither “*runtime*” nor “*monitoring*” in the concept mention “*runtime monitoring*” can be marked as a coreference mention (2).
- The pronoun “*our*” in “*our model*” can be marked as a coreference mention (3).

## Pronouns

There are different kinds of pronouns: personal pronouns (“*it*”, “*they*”), possessive pronouns (“*our*”, “*theirs*”), relative pronouns (“*which*”, “*who*”), reflexive pronouns (“*themselves*”, “*itself*”) and

demonstrative pronouns (*“this”, “that”*). Demonstrative pronouns are not to be confused with demonstrative determiners (*“this protein”, “that model”*). Here possessive determiners (*“its results”, “our model”, “their isotopes”*) are regarded as possessive pronouns.

(5) Pronouns can be marked as coreference mentions.

(6) Exception to (5): pronouns which don't refer to a specific noun or noun phrase.

Examples:

- In *“We release the model [...] It was first developed [...] Our model [...]”* “We” and “Our” can be marked as being coreferent, as well as *“the model”, “It”* and *“Our model”* (5).
- *“it”* in *“it is also observed that”, “it is clear that”* cannot be marked (6).

### Generic Expressions

In the annotation guidelines of the OntoNotes-Corpus<sup>[1]</sup> generic expressions are defined as bare plurals - nouns or noun phrases without a determiner (*“proteins”*) - or indefinite noun phrases (*“a protein”*).

(7) There are no restrictions on marking generic expressions as coreference mentions.

### Apposition

An apposition consists of a head and an attribute, which describes the head in more detail, where the head is only separated from the attribute by a colon, comma, dash or parentheses. Both the head and the attribute are usually noun phrases.

(8) Neither the head nor the attribute of an apposition can be marked as a coreference mention.

(9) The whole apposition can be marked as a coreference mention.

Examples:

- In *“SpiNNaker, the massively-parallel computer”* neither *“SpiNNaker”* (head) nor *“the massively-parallel computer”* (attribute) can be marked as a coreference mentions (8).
- In *“SpiNNaker, the massively-parallel computer”* the entire phrase *“SpiNNaker, the massively-parallel computer”* can be marked as a coreference mention (9).

### Phrases connected by copular verbs

Copular verbs are verbs such as *“be”, “seem”, “become”, “remain”, “look”* which occur between a noun phrase and an attribute that describes or renames the noun phrase.

(10) Phrases which are connected by a copular verb cannot be marked as being coreferent with each other.

(11) Only the noun phrase in such a copular structure can be marked as a coreference mention.

Examples:

- In “*SpiNNaker is a massively-parallel computer*” “*SpiNNaker*” and “*a massively-parallel computer*” cannot be marked as being coreferent with each other (10).
- in “*SpiNNaker is a massively-parallel computer*” “*SpiNNaker*” can be marked as a coreference mention (11).

## Verbs

(12) A verb can be marked as a coreference mention, but only if its cluster would contain at least one noun phrase.

(13) Gerunds (verbs ending on “*ing*”) should be treated as noun phrases.

Examples:

- in “*Wind turbines are optimized. The optimization [...] “optimized” and “The optimization” can be marked as being coreferent, since the cluster would contain at least one noun phrase (“The optimization”)*” (12).
- “*decreases*” cannot be marked as a coreference mention if its cluster would only contain “*decreased*”, but can be marked if its cluster would contain “*the decrease in*” or “*decreasing*” (12).

## Overlapping noun phrases

(14) When multiple noun phrases overlap and share the same head, only the longest phrase can be marked as a coreference mention. Thus it is possible that multiple coreference mentions overlap (as long as they don’t share the same head).

Examples:

- The phrase “*Experimental varieties selected from previous cycles also revealed [...] contains three noun phrases: a) “Experimental varieties”, b) “previous cycles” and c) “Experimental varieties selected from previous cycles”. According to (14) b) and c) can be marked as coreference mentions, whereas a) cannot be, since a) and c) overlap, share the same head-word “varieties” and c) is longer than a).*

## Conjoined noun phrases

(15) When multiple noun phrases are conjoined by “*and*”, “*or*”, “*as well as*”, or comma then these noun phrases, as well as the noun phrase containing the conjoined noun phrases can be marked as a coreference mention.

Examples:

- In “*normal TCP flows and LDDoS flows*” the following phrases can be marked as coreference mentions: “*normal TCP flows*”, “*LDDoS flows*” and “*normal TCP flows and LDDoS flows*”.

[1] <https://catalog ldc.upenn.edu/docs/LDC2007T21/coreference/english-coref.pdf>