

# Genericity Annotation Guidelines

Aurelie Herbelot

28th December 2007

## 1 The Annotation Scheme

The annotation scheme is designed as a list of 14 steps. Each step corresponds to a test, the answer to which decides of the next step to take. Annotators should start at step 1 and follow instructions in each subsequent step. When one of the five labels SPEC, NON-SPEC, GEN, GROUP or AMB appears, the noun phrase should be marked appropriately and the annotator should stop unless a further step is specified.

In what follows, the letter P refers to the noun phrase being annotated.

1. Does P appear in an existential construct of the type ‘there [be] (a) P(s)’? The copula *be* can appear conjugated in any tense and there may be other phrases separating it from its logical subject and object. Yes → SPEC  
No → 2.
2. Is P a proper noun? Yes → SPEC No → 3.

Examples: *John Smith, War and Peace, Easter Island, World War II...*

3. Does P start with an indefinite quantifier **other than ‘all’**? This includes *some, a few, few, many, most, one of, a couple of*, etc. Yes → NON-SPEC  
No, quantifier is ‘all’ → read 4 and jump to 13 No → 4.
4. It will often be the case that P refers to something more specific than it seems. We will call that more specific reading P2. Consider the following cases:

*Make sure you have a hammer at hand. The tool will be needed in step 4.*  
The tool = the hammer.

*The cathedral is magnificent. The stained-glass windows date from the 13th century.* The stained-glass windows = the stained-glass windows of the cathedral X.

(In a book about 19th century Europe) *Women could not vote.* Women = women in 19th century Europe.

It helps to ask whether the entity could be ‘any P in existence’. If not, there is probably a more specific reading involved (P2). Sometimes, it is

difficult to answer the question (for instance if the entity is not a familiar concept). In this case, try to specify as much as possible using explicit location/time details from the context. If the word is an anaphoric reference to a previous word, let P2 be the previous word. If the identity of P can be precised, through context or general knowledge, record the precised entity in P2. Possessives should also be resolved:

*his car* = Paul's car.

If P cannot be precised further, then P2 = P. Go to step 5.

5. Does P appear in a construct of the type 'A [be] P(s)'? The copula *be* can appear conjugated in any tense and there may be other phrases separating it from its logical subject and object. Implicit copulas also call for an affirmative answer: for instance, 'X classified as Y' or 'X named Y' will, in certain cases, mean 'X is a Y.' Yes -> 6 No -> 7.
6. In the identified construct, 'A [be] P (s)', are A and P two appellations for the same thing? Consider for example: *Elizabeth II is the Queen of England* or *The morning Star is the evening Star*. Yes -> 7 No -> 13.
7. Does the lexical realisation of P2 refer to an entity which is unique in the world? (Plurals are necessarily non-unique.) Yes -> 8 No -> 9. (If unsure, go to 9.)
  - *The Daily Mail was looking for a new chief editor. Paul went for **the job**.* There are not several jobs of chief editor for the Daily Mail, so this is unique.
  - ***The lion** bit my toe.* There is more than one lion in the world, possibly even more than one lion who bit my toe, so this is not unique.
8. Is P2 a common noun that could have taxonomical children? (When considering complex entities, e.g. 'X of Y', the taxonomical child must belong to the head of the phrase - X in the example.) Yes -> 13 No -> SPEC.
  - *Psychology*: yes, because experimental psychology and behavioural psychology are forms of psychology.
  - *Mozart's death*: no, nothing is a form of Mozart's death.
9. Does P have a determiner? Yes -> 10 No -> 11.
10. Does P2 refer to a particular object, or group of objects, in the world? Ie, out of all possible P2s, is the text only talking about one/some of them? (The determiner is sometimes a very good clue.) Yes -> 12 No -> 13.
  - *Can you see **the lion**?* (Assuming P2=the lion at the London zoo.) Particular object = the lion being pointed at, as opposed to all possible lions at the London zoo.

- *The lion is a mammal.* Non-particular = this is talking about lions in general.
11. Is there a **reading of** the sentence where the P2(s) in the text can be distinguished from other P2s (or group of P2s)? Ie, out of all possible P2s, is the text only talking about some of them? Yes -> 12 No -> 13. (See examples in step 10.)
  12. Is/Are the P2(s) in the text identifiable or is the text talking about, potentially, any of them? Identifiable -> SPEC Not identifiable and last step = 10 -> NON-SPEC Not identifiable and last step = 11 -> NON-SPEC + 14.
    - *Mary has a new bike.* Identifiable = one bike, Mary's bike.
    - *I would like a new bike.* Not identifiable = one bike, but any will do.
  13. Is P2 = P? Yes -> GEN No -> GROUP.
  14. Is there a reading of the sentence where P2 means 'all' P2(s)? Yes -> AMB No -> STOP.