Guidelines for the genericity annotation task

Aurelie Herbelot

May 16, 2010

1 Material

You will be given a list of 'triples' of the type 'A does B' and the sentence from which each triple was extracted. For instance:

TRIPLE: state win case

ORIGINAL: Few roads even entered the area before 1929, when the state won another court case and built what is now known as the Pacific Coast Highway.

Your task is to annotate the noun phrase (NP) that contains the first element of the triple, that is, the **subject** of the triple. In the example above, you should annotate 'the state'. Your annotation should follow the following format:

- Triple Number – subject NP – annotation for subject NP

So, for the above example:

- 1 – the state – one

2 The task

The annotation that you give for each sentence will depend on instructions given later in these guidelines. Those instructions are given in the form of decision trees and ask you to provide three types of annotations: you are requested to **quantify** the noun phrase under consideration, to tell whether it is a **kind** or not, and to decide whether the verb phrase in the sentence applies to the subject noun phrase **distributively** or **collectively**. To help you, definitions of the terms in bold are provided below.

2.1 Some definitions

2.1.1 Quantification

By 'quantifying' we mean paraphrasing the NP using a term expressing some quantity, such as *some* or *most*:

- *Europeans* discovered the Tuggerah Lakes in 1796 = *Some Europeans* discovered the Tuggerah Lakes in 1796.

  Note that the meaning of the whole sentence *mustn't* change.

2.1.2 Collective vs Distributive

A distributive statement is one where every entity referred to by the subject is individually involved in the verb's action:

- *The students* took an exam = Each student, individually, took the exam.

A collective statement is one where the group referred to by the subject, as opposed to individuals, performs the action:

- *The residents* founded a self-help group = The residents, together, founded a self-help group (and not: each residents founded their separate self-help group).

2.1.3 Kinds

You will sometimes be asked to annotate a noun phrase as a kind. No decision is needed there, but for your guidance, we define a 'kind' as the entire group of entities described by the noun phrase under consideration:

- *Ducks* lay eggs = The group of all ducks, collectively, lays eggs.

3 More on quantification

You will be asked to explicitly quantify each noun phrase. The quantifier should be one of several suggested by the scheme (see Section 4) – try each one and add the most fitting one to your annotation (the quantifier must be added to the noun phrase without changing the meaning of the sentence.) If there is already an article in front of the noun phrase, try using *some of, etc* instead of *some, etc*.

1. [Some] Europeans discovered the Tuggerah Lakes in 1796.

2. [Most of] The organizers of the exhibition were appaled.

3. [All] Such drifts hurt portfolios that are built with diversification as a high priority.

3.1 The annotation labels

You will be using five labels in the course of the annotation: *quant, one, some, most* and *all*.

quant: used when a noun phrase is already quantified:

- some people/6 million inhabitants/most of the workers

one: used when the noun phrase refers to an individual, distinguishable entity – or when the entity is unique in the world:

- My cat = a given individual, distinguishable from other cats, possibly also from other cats of mine. I can point at it.
- Bohr's model of the atom = a unique entity (there are not several such models).
- The Eiffel Tower = a unique entity.

some: self-explanatory. Used when part of a group is involved in the action:

- *Europeans* discovered the Tuggerah Lakes in 1796 = a few individuals taken out of the group of all Europeans.

most: self-explanatory. Used when the majority of a group is involved in the action:

- *Cats* have four legs = most of them – some of them are injured or have a birth defect.

all: self-explanatory. Used when all members of a group are involved:

- *Cats* are mammals = every cat is necessarily a mammal.

Note that cases where the quantifier *both* would normally be used map onto *all*:

- *The brothers* [Romulus and Remus] were good warriors = all of (the two of) them.

3.2 What to do when you hesitate?

In general, try to imagine the referent set of the noun phrase – that is, what the NP refers to in the whole sentence, as opposed to what it refers to at the 'point of hearing':

- Recent research papers have described modern dolomite formations under anaerobic conditions.

  [When I hear of *recent research papers*, I think of all recent research papers. The sentence is obviously only referring to some of those, so the annotation is *some*.]

- The filmmakers achieved the reinstatement of the President, and they founded the Film Directors' Society that same year.

  [When I hear of *the filmmakers*, I think of a certain group of filmmakers. The sentence is referring to the whole group, so I annotate as *all*.]

- The supermarket round the corner has an offer: pineapples are on three for two.

  [By the time I hear of *pineapples*, I think of all pineapples at the local supermarket. The sentence is referring to them as a whole, so I annotate as *all*.]

  If you are still hesitating: the label *most* has priority over *all*, and *some* has priority over the other two.

4 The annotation process

A different annotation process should be followed depending on the type of noun phrase that you are annotating. There are five basic types: already quantified noun phrases, proper nouns, (non-bare) singulars, plurals and bare singulars. Instructions for each of those types are given below. Simply follow the instructions until you get to the keyword 'finish'. It is helpful to refer to Section 2.1.2 to decide between collective and distributive annotations, and to Section 3 for a description of the annotation labels.

4.1 Quantified NPs

If the noun phrase is already quantified: annotate as *quant*.

- *Some governments* have labelled the church as a cult.

- *Many players* today use plastic plectra.

4.2   Proper nouns

For the purpose of this annotation, we define a proper noun as a noun phrase that contains capitalised words and that refers to a concept that doesn't have instances:

1. Proper: *John Smith, Easter Island, World War II...* (Nothing 'is a' John Smith or 'an instance of' World War II)

2. Non proper: *The Romans, The Chicago Bulls* (Caesar is a Roman, Michael Jordan is a Chicago Bull)

3. A case to consider:

   (a) *The First Circle* is an important book = one of Soljenitsyn's novels, no instances possible. Proper.

   (b) He still hasn't returned my *First Circle* = a copy of the book The First Circle. All such copies are instances of the concept 'physical copy of the First Circle'. Non proper.

All proper nouns should be annotated as *one.*

4.3   (Non-bare) singulars

Those are singular noun phrases introduced by an article, e.g. *a car, the fish and chips shop in the town centre...*

1. Try to pluralise the noun phrase. If it is possible to pluralise the noun phrase and keep the original sentence meaning, go to 3, otherwise go to 2.

   NB: if the noun phrase is an indefinite singular, just bare pluralise it. If the noun phrase is a singular introduced by a demonstrative or possessive article, try to pluralise keeping the article. If the noun phrase is a definite singular, use a bare plural or keep the article, whichever feels most natural.

2. Annotate as *one.* Finish.

   - *The film* featured two songs by Radiohead , Fake Plastic Trees and My Iron Lung taken from the album The Bends. [one]

3. Annotate as *distributive* or *collective.* Go to 4.

5

4. Annotate as *kind-*. Go to 5.

5. Annotate as *some*, *most* or *all*, with reference to the pluralised sentence. Finish.

   NB: Don't be fooled by sentences with modals:

   - *Individual provinces* may accord their primate with more or less authority. [all] (Don't be biased by the fact that only some or most will indeed do it. All of them **may** do it.)

4.4   Plurals

Those are all plural noun phrases, bare or introduced by an article, e.g. *conscripts*, *the founders of the club...*

1. Annotate as either *distributive* or *collective*.

2. Try to singularise the sentence with either *a* or *the*, keeping its meaning. If it is possible, go to 3. Otherwise, annotate as *some*, *most*, or *all* and finish.

   - *Europeans* discovered the Tuggerah Lakes in 1796 $\neq$ *A European* discovered the Tuggerah Lakes in 1796. [collective/some]

3. Annotate as *kind*.

   - *Voluntary female conscripts* receive a small additional benefit = *A voluntary female conscript* receives a small additional benefit. [distributive/kind/]

4. Annotate as *some*, *most*, or *all*. Finish.

   - *New York residents* founded the theatre in 1928. [collective/some]
   - *Community members* debate all issues once a week. [collective/all]

4.5   Bare singulars

Those are singular noun phrases without article, e.g. *psychology*, *water*, *modern Finnish popular music...*

   N.B.: if the bare singular is the result of an ellipsis, add the appropriate article and annotate with reference to the section on non-bare singulars:

- Once deployed, EPIRBs can be activated, depending on the circumstances, either manually (*crewman* flicks a switch) or automatically (as soon as water comes into contact with the unit's see-switch) = a crewman...

Bare singulars are notoriously difficult to annotate. So instead of annotating the actual noun phrase, try first to paraphrase the sentence with a plural noun phrase of your choice (the paraphrase does not have to be perfect, and if you cannot find a synonym of the noun phrase under consideration, use something that is an instance of it):

- *Free software* allows users to co-operate in enhancing and refining the programs they use = *Open source programs* allow users to co-operate...

- *Damage* showed seismic resistance deficiencies in modern apartment construction = *Cracks* showed seismic resistance deficiencies...

If you are able to do such a conversion, simply annotate the noun phrase using the section on plurals and write which word(s) you chose as a paraphrase:

- 32 – free software – distributive/kind/all (open source programs)

If the paraphrase is impossible, annotate as *one*:

- Although *modern analytic celestial mechanics* started 400 years ago with Isaac Newton, prior studies addressing the problem of planetary positions are known going back perhaps 3000 or more years. [one]

5   Decision trees

This section is here to help you make a decision for each type of noun phrase encountered. Simply follow up the relevant decision tree, referring to the detailed instructions when needed. Decision points enclosed in square boxes indicate that an annotation is needed.
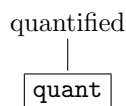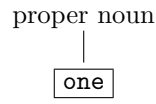
quantified
|
`quant`

Figure 1: Quantified case.

```
                    proper noun
                         |
                       ┌─────┐
                       │ one │
                       └─────┘
```
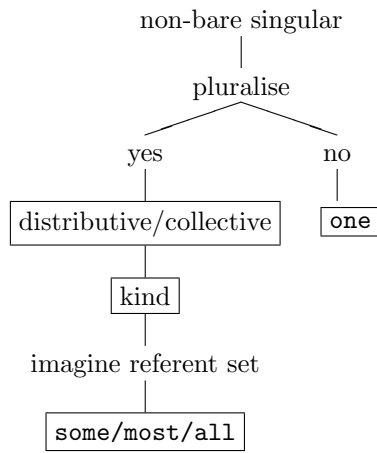
Figure 2: Proper noun case.

```
                 non-bare singular
                         |
                     pluralise
                    ╱         ╲
                 yes            no
        ┌────────────────────┐   ┌─────┐
        │ distributive/collective │   │ one │
        └────────────────────┘   └─────┘
                    |
                 ┌──────┐
                 │ kind │
                 └──────┘
                    |
            imagine referent set
                    |
            ┌───────────────┐
            │ some/most/all │
            └───────────────┘
```

Figure 3: Non-bare singular case.

```
                    plural
                      |
          ┌────────────────────────┐
          │ distributive/collective │
          └────────────────────────┘
                      |
                 singularise
                ╱           ╲
             yes              no
          ┌──────┐      imagine referent set
          │ kind │              |
          └──────┘      ┌───────────────┐
             |          │ some/most/all │
   imagine referent set └───────────────┘
             |
     ┌───────────────┐
     │ some/most/all │
     └───────────────┘
```
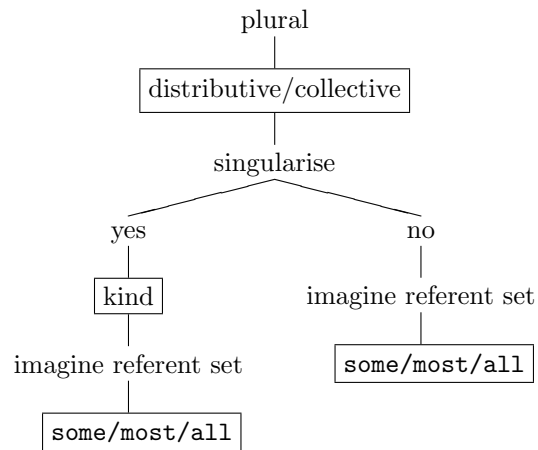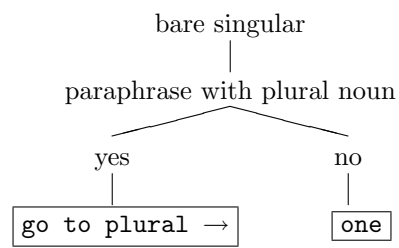
Figure 4: Plural case.

Figure 5: Bare singular case.