Annotation scheme for grammatical roles (AS4)

Katja Markert and Malvina Nissim
School of Computing
University of Leeds
Leeds, UK
and University of Bologna
Bologna, Italy
{markert@comp.leeds.ac.uk} {malvina.nissim@unibo.it}

December 31, 2006

1 Kinds of grammatical annotation

There are two kinds of grammatical annotation per possibly metonymic word (PMW). The first expresses the grammatical role(s) of the PMW. The range of grammatical roles follows partially the one in the GNOME scheme for grammatical functions of NPs (`http://www.hcrc.ed.ac.uk/~gnome/anno_manual.html`). At the moment we annotate the following grammatical roles that express that the PMW is the modifier of another word: active subjects (`subj`), passive subjects (`subjpassive`), direct objects (`obj`), genitives (`gen`), proper name premodifiers of common nouns (`premod`), predicates (`pred`), indirect objects (`iobj`), pps (`pp`), appositions (`app`) and dependencies on adjectives (`adj`). The PMW can also have modifiers itself, of course, and we consider it having an adjective modifier (`hasadj`), a proper name premodifier (`haspremod`), having an apposition (`hasapp`), a genitive modifier (`hasgen`), or a prepositional attachment (`haspp`).

Should a PMW have several grammatical relations, all are annotated (each on a separate line in the `grammannot` file). Should a PMW have none of the mentioned grammatical relations, it has a single annotation (`nohead`) (with lemma annotation NONE).

The second kind of grammatical annotation concerns the lemma to which the PMW is linked, both heads and modifiers. The notion of head and modifier corresponds approximately to the notion of head/modifier in dependency grammar but we use **lexical heads** in case of modals and auxiliary verbs and also in the case of some other explicitly named constructions.

2 Grammatical Role

For most of the information on grammatical roles see the GNOME manual. Most are also intuitively plausible but here are some additional pointers. An expression can only have the value `subj` or `obj` if it is in a unit that contains a verb (see GNOME for unit definitions). The only exception are cases of ellipsis (see below).

- **Coordination.** Noun phrases in a coordination inherit the grammatical role of the coordination. Thus *"Hungary"* in the following example is to be annotated as `subj`: *"**Hungary** and Germany signed the treaty."*.

- **Non-restrictive Appositions/Parentheticals.** Non-restrictive appositions (i.e., appositions between hyphens or in parenthesis) are in generally only marked as appositions `app` and do not inherit the grammatical roles of their head. Thus in the following example the nouns in bold face are not subjects: *"two countries — **Hungary** and **Germany** — signed the treaty."*

- **Restrictive Appositions.** Restrictive appositions inherit the grammatical roles of their heads. Thus, in *"The two countries **Hungary** and **Poland** signed the treaty"*, "Hungary" is to be marked as `subj` as well as as `app`.

- **Non-finite verbs.** The verbed unit can be non-finite. Thus, in *"He travelled to Europe, visiting **Hungary**"* "Hungary" still has the role `obj`.

- **Verbal gaps.** In sentences such as *John likes dogs, and **Mary** cats."*, "Mary" is to be annotated as subject of "like".

2.1 Subjects

By subject with the short notation `subj` we mean the structural subject of the sentence/unit. Structural subjects of passive sentences are marked with `subjpassive`.

- In **passive sentences**, where the structural subject is not the agent, we note the role `subjpassiv`. Example *"**Hungary** was proclaimed a republic."*

- **Inversions.** Take care to still note the subject role in inversions. Example: *"Here come **the buses**."*

- **Questions.** The same holds for questions.

- **Direct speech, tags.** This can lead to inversions as well. The subject is still to be marked ordinarily. For example: *"So am **I**"*; *"She has plenty of sense, has **Joan**."* and *"Oh — said **Claire**."*.

2.2 Direct objects and pred

Direct objects are marked with `obj`.

- **Non-finite verbs.** See above.

- **Phrasal verbs** The value `obj` should also be used with NPs as an object argument of a transitive phrasal verb. Following GNOME we regard only verbs as phrasal verbs where the particle can either precede or follow the argument (except when the argument is a pronoun). Thus "turn on" is a phrasal verb as both *"They turned on **the light**"* and *"They turned **the light** on."* are possible. Thus , "(the) light" is to be annotated as `obj` in both cases.

- **Verb "to be" or "to become"**: Use the value `pred` instead of `obj`. Example sentence: *"One of the countries is **Hungary**"*, where "Hungary" is to be marked as `pred`.

- **Inversions.** Take care to still note the object role in inversions. Example: *"**The letters** he read to Julia."*

2.3 Indirect Objects

For ditransitive verbs, the second object (if not expressed with a *to*-prepositional phrase) is annotated as `iobj`. Thus, in *"Mary gave **John** a book"*, "John" is the `iobj` of "give".

2.4 PPs

In *"He came to **the house**"* "(the) house" to be marked as "pp". Similar in *"**the house** with the windows"* "(the) house" has the role `haspp`. PPs can be attached to nouns, verbs or adjectives/adverbs.

2.5 Genitives

The value `gen` is to be assigned to nouns with the possessive marker *'s*, functioning as a determiner. Thus, in *"in **Hungary's** case"*, "Hungary" is to be marked with the role `gen`. In *"in Hitler's **Germany**"*, "Germany" then has the role `hasgen` as one of its roles.

2.6 Premods

Proper names that premodify another noun but are not genitives, are to be marked as `premod`. Examples are *"**Vietnam** war"* or *"**UK** companies"*.

2.7 Adjectives

If the PMW has an adjectival modifier, it receives a role `hasadj` as in *"in northern Spain"*. In the rare case that the PMW is the modifier of an adjective itself, it receives the role `adj`.

2.8 Appositions

Appositions (whether restrictive or non-restrictive) are marked as `app`/`hasapp`. Thus, in *"the country **Canada** is in North America"*, "Canada" has the role `app` with head "country" whereas "country" has the role `hasapp` with modifier "Canada".

3 Heads and Modifiers

We use a notion of head/modifier corresponding approximately to dependency grammar's notion of head/modifier.

- **Form of the lexical Head/Modifier.** We lemmatise the head/modifier and lower case all lemmata but proper names. Regarding phrasal verbs: if the phrasal verb is in Wordnet we use the full phrasal verb (e.g., "account for"), otherwise we leave out the particle. If the head/modifier is a compound noun (e.g., for the role `gen` as in *"**Hungary's** prime minister"*) then if the compound is in Wordnet we use the full compound, otherwise we use the head in its lemmatised form. Disregard all adjective or similar modifiers of the head/modifier.

- **Coordination.** If the head/modifier is a coordinated unit, then annotate all members of the coordination as heads/modifiers For example in *"**Hungary** signed and ratified the treaty."* write down "sign" AND "ratify" as lexical heads.

- **Ellipsis.** If the head/modifier is elliptic, then replace it. Thus in *"I like Scotland's coast and **Norway's**, too."*, the head of "Norway" is "coast".

- **Heads of subjects on modal and similar constructions.** When the role is `subj` or `subjpassive` the lexical head can vary from the dependency head. We use the lexical verb as head if only the following modals, auxiliaries or special constructions separate the PMW from the lexical head: the auxiliaries have ("America had tried"), do ("America did try..."), be ("America is trying") or modals can/could ("America can try"), may/might ("America may try"), shall/should ("America should try"), ought (to) ("America ought to try"), will/would ("America would try"), must ("America must try"), need ("America needs to try"), dare ("America dares to try"), used to ("America used to try") or special constructions have (to) ("America has to try"), be to ("America is to try") be able to ("America is able to try") and going to ("America is going to try"). In all these cases the lexical head of "America" to be annotated is "try". This also holds for combinations like "America would have been able to try".

- **Heads of pps.** The preposition is the head of the pp. Thus in *"He came to the house"*, "to" is head of "(the)house". We do not annotate pp-attachment.