## Natural Language Processing 2005 – Paper 9 Question 14 (AAC)

This question relates to lectures 5 and 6. The grammar rule required in part a is in the notes but it should also be possible for students to reconstruct it given Figure 1, so the question does not require memorisation of the feature structure details. The behaviour of *rains* is mentioned in the notes, but the feature structure encoding is not given. There are several ways to achieve the correct effect: part b should be reasonably easy for anyone who has understood the feature structure formalism properly. Selectional restrictions are mentioned in the lecture notes but no encoding is given: this is quite hard. The point of the examples is to illustrate that multiple attributes (e.g. animacy, sentience) must be considered.

(*a*) Give this grammar rule as a feature structure and show the results of applying the rule to the structures in Figure 1. [7 marks]
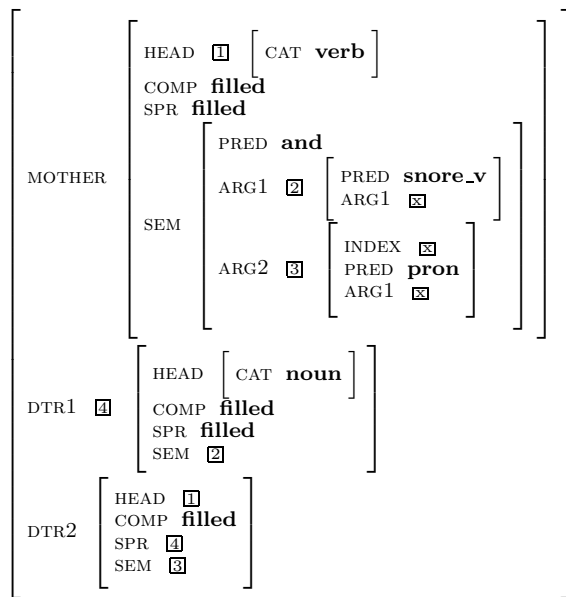
The rule is:

$$
\begin{bmatrix}
\text{HEAD} & \boxed{1} \\
\text{COMP} & \textbf{filled} \\
\text{SPR} & \textbf{filled} \\
\text{SEM} & \begin{bmatrix} \text{PRED} & \textbf{and} \\ \text{ARG1} & \boxed{2} \\ \text{ARG2} & \boxed{3} \end{bmatrix}
\end{bmatrix}
\rightarrow
\boxed{4}
\begin{bmatrix}
\text{COMP} & \textbf{filled} \\
\text{SPR} & \textbf{filled} \\
\text{SEM} & \boxed{2}
\end{bmatrix}
,
\begin{bmatrix}
\text{HEAD} & \boxed{1} \\
\text{COMP} & \textbf{filled} \\
\text{SPR} & \boxed{4} \\
\text{SEM} & \boxed{3}
\end{bmatrix}
$$

or (better, since the question asks for a feature structure):

$$
\begin{bmatrix}
\text{MOTHER} & \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{COMP} & \textbf{filled} \\ \text{SPR} & \textbf{filled} \\ \text{SEM} & \begin{bmatrix} \text{PRED} & \textbf{and} \\ \text{ARG1} & \boxed{2} \\ \text{ARG2} & \boxed{3} \end{bmatrix} \end{bmatrix} \\
\text{DTR1} \; \boxed{4} & \begin{bmatrix} \text{COMP} & \textbf{filled} \\ \text{SPR} & \textbf{filled} \\ \text{SEM} & \boxed{2} \end{bmatrix} \\
\text{DTR2} & \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{COMP} & \textbf{filled} \\ \text{SPR} & \boxed{4} \\ \text{SEM} & \boxed{3} \end{bmatrix}
\end{bmatrix}
$$

The instantiated structure is:

```
        ⎡           ⎡ HEAD  ①  [ CAT  verb ]                                    ⎤ ⎤
        ⎢           ⎢ COMP  filled                                              ⎥ ⎥
        ⎢           ⎢ SPR   filled                                              ⎥ ⎥
        ⎢           ⎢           ⎡ PRED  and                                   ⎤ ⎥ ⎥
        ⎢           ⎢           ⎢                ⎡ PRED  snore_v ⎤            ⎥ ⎥ ⎥
 MOTHER ⎢           ⎢           ⎢ ARG1  ②        ⎣ ARG1  x       ⎦            ⎥ ⎥ ⎥
        ⎢           ⎢ SEM       ⎢                                            ⎥ ⎥ ⎥
        ⎢           ⎢           ⎢                ⎡ INDEX  x   ⎤               ⎥ ⎥ ⎥
        ⎢           ⎢           ⎢ ARG2  ③        ⎢ PRED   pron ⎥               ⎥ ⎥ ⎥
        ⎢           ⎣           ⎣                ⎣ ARG1   x   ⎦               ⎦ ⎦ ⎥
        ⎢                                                                       ⎥
        ⎢           ⎡ HEAD  [ CAT  noun ]                                    ⎤  ⎥
        ⎢ DTR1  ④   ⎢ COMP  filled                                          ⎥  ⎥
        ⎢           ⎢ SPR   filled                                          ⎥  ⎥
        ⎢           ⎣ SEM   ②                                               ⎦  ⎥
        ⎢                                                                       ⎥
        ⎢           ⎡ HEAD  ①                                                ⎤  ⎥
        ⎢ DTR2      ⎢ COMP  filled                                          ⎥  ⎥
        ⎢           ⎢ SPR   ④                                               ⎥  ⎥
        ⎣           ⎣ SEM   ③                                               ⎦  ⎦
```

(Marking note: The semantics and the SPR instantiation is more important than the coindexation on the HEAD feature or the COMP value.)

(*b*) Syntactically the verb *rains* takes the pleonastic pronoun *it* as subject but semantically it has no arguments. Give possible feature structures for *rains* and pleonastic *it*. Show how ungrammatical sentences such as *he rains* are avoided, mentioning any modifications to the lexical entries in Figure 1 that might be necessary.  [6 marks]

The easiest way of doing the syntactic selection is to use the value of CAT and to specify that as 'it' for the pleonastic pronoun. The entry for *rain* says that its SPR HEAD CAT is 'it' instead of noun. This will therefore not accept 'he' as a subject since that has CAT 'noun' and unification will fail. Similarly, 'snores' will not take pleonastic 'it'. If this approach is taken, no modification is needed to the lexical entries in Figure 1.

A more conventional approach is to use another head feature FORM, which takes values 'it' and 'normal'. Ordinary verbs, such as *snore*, specify that their SPR (and COMP) are FORM 'normal'. The verb *rain* says that its SPR's form is 'it'. Everything has FORM 'normal' apart from pleonastic *it* which has value 'it'. (Marking note: although this is more standard, the CAT solution is just as good as far as the objectives of this course are concerned.)

The semantics for *rain* has a PRED value but no ARG1. The semantics for *it* is empty or PRED true, with no arguments. The solution where the semantics is empty means that the semantics for the phrase has an 'and' with only one argument, which is inelegant.

(*c*) Describe how selectional restrictions might be encoded in a feature structure grammar.  [7 marks]

The best way of doing this in a feature structure grammar is probably to have a series of binary valued attributes, such as SENTIENT and ANIMATE which have values true, false or unspecified. Then the subject of *snore* is ANIMATE true, SENTIENT unspecified;

the subject of *write* is ANIMATE true, SENTIENT true; the subject of *roll* is ANIMATE and SENTIENT unspecified etc. The noun *pebble* is ANIMATE and SENTIENT false, *dog* is ANIMATE true and SENTIENT false.[1] The reason for this rather clumsy encoding is that essentially what's being encoded is a hierarchy, but there is no direct mechanism of expressing this with feature structures. Thus to have a feature SEL-RESTR with value sentient (or whatever) doesn't work.

The best location of these features is probably on the part of the feature structure that corresponds to the variables, e.g., *snore* has SEM ARG1 ANIMATE true; *actor* has SEM ARG1 SENTIENT true etc.

---

[1]Sorry pet lovers, think of another term if you prefer, but dogs really can't write ...