Practical Grammar Engineering using HPSG
Exercises day4: Modifiers and Agreement

Goals:

1. Introduce modifiers (adjectives and PPs).
2. Introduce modifier rules.
3. Introduce agreement.

Background
In Day 2 we built a small grammar, consisting of some lexical entries and a small number of rules. This grammar enabeled us to parse a sentence. In Day 3, we entered MRS semantics into your grammar, and we could ask the LKB system for scope resolved MRSs. In this exercise we will expand the grammar in order to handle modification and agreement.

Exercises:

1. Introduce adjectives and prepositions.
   (a) Introduce *adj* and *prep* as subtypes of *head*. Constrain the subtypes of *head* that can not be modifiers to have empty MOD lists. Introduce a type *adj-relation* as a subtype of *arg1-ev-relation*.
   (b) Introduce a type *mod-lex-item* as a subtype of *norm-lex-item*. Let it have an element on the MOD list and constrain the index of this element to share structure with the ARG1 value of the main relation. Make the LBL of the element on the MOD list share structure with the LBL of *mod-lex-item*. Let the SUBJ, SPR and SPEC lists be empty. Let the COMPS list of the element on the MOD list be empty. Constrain the HCONS list of *mod-lex-item* to be empty.
   (c) Introduce a type *adj-lxm* as a subtype of *mod-lex-item* and *lexeme*. Give it the correct head value and relation, and constrain its COMPS list. Constrain the modified element to be of category noun.
   (d) Introduce *adj-word* under *word* and give it some subtypes, reflecting for example case, number, gender or definiteness in the type names (if these factors doesn't show in the inflection, you only need one (positive) subtype). Don't worry about the agreement information yet. We will come back to that later. Make the subtypes of *adj-word* also inherit from either *infl-ltow-rule* or *const-ltow-rule*. Enter some inflectional rules for adjectives into "irules.tdl" and some lexical entries into "lexicon.tdl".
   (e) Introduce a type *arg2-comps-lex-item* (if you havent't done it already) under *lex-item* and link its ARG2 to the index of its first complement.
   (f) Let *prep-word* inherit from *mod-lex-item*, *word* and *arg2-comps-lex-item*. Give *prep-word* the correct head value and relation (*prep-mod-relation*), and constrain its COMPS list.
   (g) The lexical entries of *adj-lxm* and *prep-word* should look like figure 3 and figure 4.

2. Make three rules for adjunct attachment: One for combining an adjective with a noun, one for combining a PP with a noun and one for combining a PP with a VP.
   (a) Introduce a type *a-mod-phrase*. Let it inherit from either *head-adj-int-phrase* or *adj-head-int-phrase* dependent on whether an adjective attaches from the right or from the left in your language. Constrain its non-head daughter to be an adjective and the head-daughter to be a noun.
   (b) Introduce a type *pp-mod-phrase* for PP's modifying nouns or verbs. Also here you should use either *head-adj-int-phrase* or *adj-head-int-phrase* as supertype.
   (c) Introduce two subtypes of *pp-mod-phrase*. One for PPs modifying a noun *pp-mod-n-phrase* and one for PPs modifying a VP *pp-mod-vp-phrase* (remember that a VP is a sign with head value *verb*, a *synsem* on the SUBJ list and an empty COMPS list.

Figure 1: Agreement in the verb “bark”

$$\begin{bmatrix} \text{STEM} \left\langle \text{“bark”} \right\rangle \\ \text{SYNSEM} \begin{bmatrix} \text{LOCAL} \begin{bmatrix} \text{CAT} \mid \text{HEAD} \mid \text{verb} \\ \text{CONT} \mid \text{HOOK} \mid \text{INDEX } \boxed{2} \text{ event} \\ \text{AGR} \mid \text{PNG} \mid \text{PN 3sg} \end{bmatrix} \\ \text{LKEYS} \mid \text{KEYREL} \begin{bmatrix} \text{PRED} & \text{“\_bark\_rel”} \\ \text{ARG0} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(d) Enter the rules *a-mod-rule*, *pp-mod-n-rule* and *pp-mod-vp-rule* into the ‘rules.tdl’ file. The rule for adjectives modifying nouns should look like figure 5.

(e) Check that your grammar can parse sentences with modifiers.

3. We will now introduce person, number and gender agreement into our grammar. This can be done by representing agreement information on the indices. The ‘matrix.tdl’ has the following type:

```
ref-ind &
  [ PNG png ].
```

ref-ind is a referential index, and the PNG feature it carries is intended to contain information about person, numer and gender of the referent. How this information should be structured will vary from language to language. In languages where determiners, adjectives, nouns and verbs agree in person, number and maybe also gender, one needs to make this information very explicit, while in other languages this need is not the same. In the English Resource Grammar (ERG), *png* looks like this:

```
png := *avm* &
  [ PN pernum,
    GEN gender ].
```

The information about person and number is merged into one type *pn*, while gender information is put into a separate type *gen*. This makes it possible to have agreement either with just *pn* or with the whole *png*. If you feel this is a distinction that also makes sense for your language, you may adopt it. If you don’t feel it is a good distinction, you should make one by yourself.

(a) Decide how the person, number and gender information should be represented in your grammar, and introduce the types you need. You may want to add some features to the *png* type in ‘matrix.tdl’.

(b) Introduce subtypes of *noun-lxm* and specify which gender information they carry.

(c) Introduce some subtypes of the noun, determiner and adjective *word* types. Specify them with regard to number(/person) (nouns) and number/gender (adjectives and determiners). In nouns and determiners you may do the specifications on the index, while in adjectives you may use the value of AGR. Also make the new inflectional rules you need, and constrain the old rules to apply only on lexemes that are appropriate.

(d) Make sure that nouns and determiners link their indices to their AGR feature, and that adjectives link their ARG1 to the AGR feature. If verb words in your language carry *png* agreement information, state this on the index which is the value of their AGR feature. The word “barks” will contain the constraints in figure 1, and lexical entry of “big” should have the constraints in figure 2.

(e) Check that person, number and gender agreement works. Activate the ‘(index-for-generator)’ function in the end of the script file, and use the generator to confirm that your grammar is in a good condition.

Figure 2: Agreement linking in the adjective “big”

$$\begin{bmatrix} \text{STEM} \left\langle \text{“big”} \right\rangle \\ \text{SYNSEM} \begin{bmatrix} \text{LOCAL} \begin{bmatrix} \text{CAT}\,|\,\text{HEAD} \begin{bmatrix} \text{adj} \\ \text{MOD} \left\langle \begin{bmatrix} \text{LOCAL}\,|\,\text{CONT}\,|\,\text{HOOK}\,|\,\text{INDEX}\ \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{CONT}\,|\,\text{HOOK}\,|\,\text{INDEX}\ \boxed{2}\ \text{event} \\ \text{AGR}\ \boxed{1} \end{bmatrix} \\ \text{LKEYS}\,|\,\text{KEYREL} \begin{bmatrix} \text{PRED} & \text{“\_big\_rel”} \\ \text{ARG0} & \boxed{2} \\ \text{ARG1} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(f) If you still have problems in your grammar with the definite - indefinite distinction, you may consider introducing *definiteness* as a semantic or syntactic factor.

Figure 3: Lexical entry of the adjective “big”

$$\left[\begin{array}{ll}
\text{STEM} & \left\langle \text{“big”} \right\rangle \\
\text{SYNSEM} & \left[\begin{array}{ll}
\text{LOCAL} & \left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{ll}
\text{HEAD} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\text{adj}} \\
\text{MOD} & \left\langle \left[\begin{array}{ll}
\text{LOCAL} & \left[\begin{array}{l}
\text{CAT}\,|\,\text{VAL}\,|\,\text{COMPS}\,\langle\rangle \\
\text{CONT}\,|\,\text{HOOK} \left[\begin{array}{ll}
\text{LTOP} & \boxed{0} \\
\text{INDEX} & \boxed{1}
\end{array}\right]
\end{array}\right]
\end{array}\right] \right\rangle
\end{array}\right] \\
\text{VAL} & \left[\begin{array}{ll}
\text{SUBJ} & \langle\rangle \\
\text{SPR} & \langle\rangle \\
\text{SPEC} & \langle\rangle \\
\text{COMPS} & \langle\rangle
\end{array}\right]
\end{array}\right] \\
\text{CONT} & \left[\begin{array}{ll}
\text{HOOK} & \left[\begin{array}{ll}
\text{LTOP} & \boxed{0} \\
\text{INDEX} & \boxed{2}\ \text{event}
\end{array}\right] \\
\text{RELS} & \left\langle !\ \boxed{3} \left[\begin{array}{ll}
\multicolumn{2}{l}{\text{adj-relation}} \\
\text{LBL} & \boxed{0} \\
\text{PRED} & \text{“_big_rel”} \\
\text{ARG0} & \boxed{2} \\
\text{ARG1} & \boxed{1}
\end{array}\right] ! \right\rangle \\
\text{HCONS} & \left\langle !\ ! \right\rangle
\end{array}\right]
\end{array}\right] \\
\multicolumn{2}{l}{\text{LKEYS}\,|\,\text{KEYREL}\ \boxed{3}}
\end{array}\right]
\end{array}\right]$$

Figure 4: Lexical entry of the preposition "near"

$$
\left[\begin{array}{ll}
\text{STEM} & \left\langle \text{“near”} \right\rangle \\
\text{SYNSEM} & \left[\begin{array}{ll}
\text{LOCAL} & \left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{ll}
\text{HEAD} & \left[\begin{array}{ll}
\textit{prep} & \\
\text{MOD} & \left\langle \left[\begin{array}{ll} \text{LOCAL} & \left[\begin{array}{l} \text{CAT}\,|\,\text{VAL}\,|\,\text{COMPS}\,\langle\rangle \\ \text{CONT}\,|\,\text{HOOK} \left[\begin{array}{ll} \text{LTOP} & \boxed{0} \\ \text{INDEX} & \boxed{1} \end{array}\right] \end{array}\right] \end{array}\right] \right\rangle
\end{array}\right] \\
\text{VAL} & \left[\begin{array}{ll}
\text{SUBJ} & \langle\rangle \\
\text{SPR} & \langle\rangle \\
\text{SPEC} & \langle\rangle \\
\text{COMPS} & \left\langle \left[\text{LOCAL}\,|\,\text{CONT}\,|\,\text{HOOK}\,|\,\text{INDEX} \quad \boxed{2} \right] \right\rangle
\end{array}\right]
\end{array}\right] \\
\text{CONT} & \left[\begin{array}{ll}
\text{HOOK} & \left[\begin{array}{ll} \text{LTOP} & \boxed{0} \\ \text{INDEX} & \boxed{3}\ \textit{event} \end{array}\right] \\
\text{RELS} & \left\langle !\ \boxed{4} \left[\begin{array}{l} \textit{prep-mod-relation} \\ \text{LBL}\ \boxed{0} \\ \text{PRED}\ \text{“\_near\_rel”} \\ \text{ARG0}\ \boxed{3} \\ \text{ARG1}\ \boxed{1} \\ \text{ARG2}\ \boxed{2} \end{array}\right] ! \right\rangle \\
\text{HCONS} & \left\langle !\ ! \right\rangle
\end{array}\right]
\end{array}\right] \\
\text{LKEYS}\,|\,\text{KEYREL}\ \boxed{4} &
\end{array}\right]
\end{array}\right]
$$

Figure 5: Rule for adjectives modifying nouns

$$
\left[\begin{array}{l}
\textbf{a-mod-rule} \\
\text{SYNSEM}\,|\,\text{LOCAL} \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD } \boxed{0} \text{ noun} \\ \text{VAL} \left[\begin{array}{l} \text{SUBJ } \boxed{1} \\ \text{SPR } \boxed{2} \\ \text{COMPS } \langle\rangle \end{array}\right] \end{array}\right] \\ \text{CONT}\,|\,\text{HOOK } \boxed{3} \\ \text{AGR } \boxed{4} \end{array}\right] \\
\text{HEAD-DTR}\,|\,\text{SYNSEM}\,|\,\text{LOCAL} \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD } \boxed{0} \\ \text{VAL } \boxed{5} \left[\begin{array}{l} \text{SUBJ } \boxed{1} \\ \text{SPR } \boxed{2} \\ \text{COMPS olist} \end{array}\right] \end{array}\right] \\ \text{CONT}\,|\,\text{HOOK } \boxed{3} \\ \text{AGR } \boxed{4} \end{array}\right] \\
\text{NON-HEAD-DTR}\,|\,\text{SYNSEM}\,|\,\text{LOCAL}\,|\,\text{CAT} \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \text{adj} \\ \text{MOD} \left\langle \left[\begin{array}{l} \text{LOCAL} \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD } \boxed{0} \\ \text{VAL } \boxed{5} \end{array}\right] \\ \text{CONT}\,|\,\text{HOOK } \boxed{3} \\ \text{AGR } \boxed{4} \end{array}\right] \end{array}\right] \right\rangle \end{array}\right] \\ \text{VAL} \left[\begin{array}{l} \text{SPR olist} \\ \text{COMPS onull} \end{array}\right] \end{array}\right] \\
\text{C-CONT} \left[\begin{array}{l} \text{HOOK } \boxed{3} \\ \text{RELS } \left\langle !\ ! \right\rangle \\ \text{HCONS } \left\langle !\ ! \right\rangle \end{array}\right]
\end{array}\right]
$$