Answers to post-lecture exercises for NLP course, 2004

Ann Copestake (aac@cl.cam.ac.uk)

1 Lecture 2: Post-lecture exercises

1. Q: For each of the following surface forms, give a list of the states that the FST given in the lecture notes for e-insertion passes through, and the corresponding underlying forms:

   (a) c a t s
   c:c 1, a:a 1, t:t 1, s:s 4 (final): underlying cats
   c:c 1, a:a 1, t:t 1, ε:ˆ2, s:s 3 (final): underlying catˆs
   Also 'dead-ends' like: c 1, a 1, ˆ2 — fail, but I will omit these.

   (b) c o r p u s
   c:c 1 etc, s:s 4 (final): underlying corpus
   c:c 1 etc, ε:ˆ2, s:s 3 (final): underlying corpuˆs

   (c) a s s e s
   a:a 1, s:s 4, s:s 4, e:e 1, s:s 4 (final): underlying asses
   a:a 1, s:s 4, s:s 4, e:e 1, ε:ˆ2, s:s 3: underlying asseˆs
   a:a 1, s:s 4, s:s 4, e:ˆ2, s:s 4 (final): underlying assˆs

   (d) a s s e s s
   a:a 1, s:s 4, s:s 4, e:e 1, s:s 4, s:s 4 (final): underlying assess

   (e) a x e s
   a:a 1, x:x 4, e:ˆ2, s:s 4 (final): underlying axˆs
   a:a 1, x:x 4, e:e 1, ε:ˆ2, s:s 3: underlying axeˆs
   a:a 1, x:x 4, e:e 1, s:s 4 (final): underlying axes
   Notice that in this case, all three of the segmentations are plausible, since *axe*, *ax* (US spelling) and *axes* (irregular plural of *axis*) might well all be in a lexicon.

2. Q: Modify the FSA for dates so that it only accepts valid months.
   A: Replace 0,1 on the 4-to-5 transition with 1. Replace 'digit' on the 4-to-6 transition with 1, . . . 9.
   (Note: you might like to consider what would be needed to ensure that the FST only accepted valid day/month pairs).
   Q: Turn your revised FSA into a FST which maps between the numerical representation of months and their abbreviations (Jan . . . Dec).
   A: If we assume that 'Jan' etc are allowed as symbols, this just amounts to associating 4-to-5 with the pair 1:empty and then associating each digit on 4-to-6 and 5-to-6 with the appropriate month. i.e., on 4-to-6, 1:Jan, and so on, and on 5-to-6, 0:Oct, 1:Nov, 2:Dec. If we restrict ourselves to one character per transition (as is required for morphology), then we need a 1:J transition to a new state, followed by empty:A and empty:N leading to a final state. And so on for the other months.

2 Lecture 4: Post-lecture

Using the CFG given in the lecture notes (section 4.3):

1. Q: show the edges generated when parsing *they fish in rivers in December* with the simple chart parser in 4.7

id	left	right	mother	daughters
1	0	1	NP	(they)
2	1	2	V	(fish)
3	1	2	VP	(2)
4	0	2	S	(1 3)
5	1	2	NP	(fish)
6	2	3	P	(in)
7	3	4	NP	(rivers)
8	2	4	PP	(6,7)
9	1	4	VP	(3,8)
10	0	4	S	(1,9)
11	1	4	NP	(5,8)
12	4	5	P	(in)
13	5	6	NP	(December)
14	4	6	PP	(12,13)
15	1	6	VP	(9,14)
16	0	6	S	(1,15)
17	1	6	NP	(5,14)
18	3	6	NP	(7,14)
19	2	6	PP	(6,18)
20	1	6	VP	(3,19)
21	0	6	S	(1,20)

2. Q: show the edges generated for this sentence if packing is used (as described in 4.9)
   Edges 1 to 19 are initially generated in the same way as above except that the daughters are a singleton set. But instead of constructing a new edge 20, (3,19) is added to the daughters set for 15. Edge 21 is not generated. When the covering edge 16 is expanded, the expansion of 15 will give both alternative bracketings.

3. Q: show the edges generated for *they fish in rivers* if an active chart parser is used (as in 4.10).

   The lectures notes don't give enough detail for the answer to be completely determined, but the following is one reasonable possibility:

id	left	right	mother	expected	daughters
1	0	1	NP		(they)
2	0	1	S	VP	(1,?)
3	0	1	NP	PP	(1,?)
4	1	2	V		(fish)
5	1	2	VP	PP	(4,?)
6	1	2	VP		(4)
7	1	2	VP	NP	(4,?)
8	1	2	VP	VP	(4,?)
9	0	2	S		(2,6)
10	1	2	NP		(fish)
11	1	2	S	VP	(10,?)
12	1	2	NP	PP	(10,?)
13	2	3	P		in
14	2	3	PP	NP	(13,?)
15	3	4	NP		(rivers)
16	2	4	PP		(13,15)
17	1	4	VP		(5,16)
18	0	4	S		(2,17)
19	1	4	NP		(12,16)

Note that edge 9, 16, 17, 18 and 19 are completions of an active edge to form a passive edge. We assume no active edges are postulated from 15, 16, 17 or 19 because there's nothing following vertex 4. In principle, we could also use *top-down* information to prune the search space to avoid postulating unnecessary active edges although practical results from doing this seem to vary considerably with different styles of grammar.

3 Lecture 5: Post-lecture exercise answers

1. Q: Give the unification of the following feature structures:

   (a) $\begin{bmatrix} \text{CAT} & [\;] \\ \text{AGR} & \textbf{pl} \end{bmatrix}$ unified with $\begin{bmatrix} \text{CAT} & \textbf{VP} \\ \text{AGR} & [\;] \end{bmatrix} = \begin{bmatrix} \text{CAT} & \textbf{VP} \\ \text{AGR} & \textbf{pl} \end{bmatrix}$

   (b) $\begin{bmatrix} \text{MOTHER} & \begin{bmatrix} \text{CAT} & \textbf{VP} \\ \text{AGR} & \boxed{1} \end{bmatrix} \\ \text{DTR1} & \begin{bmatrix} \text{CAT} & \textbf{V} \\ \text{AGR} & \boxed{1} \end{bmatrix} \\ \text{DTR2} & \begin{bmatrix} \text{CAT} & \textbf{NP} \\ \text{AGR} & [\;] \end{bmatrix} \end{bmatrix}$ unified with $\begin{bmatrix} \text{DTR1} & \begin{bmatrix} \text{CAT} & \textbf{V} \\ \text{AGR} & \textbf{sg} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \text{MOTHER} & \begin{bmatrix} \text{CAT} & \textbf{VP} \\ \text{AGR} & \boxed{1}\ \textbf{sg} \end{bmatrix} \\ \text{DTR1} & \begin{bmatrix} \text{CAT} & \textbf{V} \\ \text{AGR} & \boxed{1} \end{bmatrix} \\ \text{DTR2} & \begin{bmatrix} \text{CAT} & \textbf{NP} \\ \text{AGR} & [\;] \end{bmatrix} \end{bmatrix}$

   Notice that, by convention, the value is shown in the first place it can be in the AVM. However, it's not actually wrong to show it in the part of the AVM following the DTR1 feature since the paths are describing the same node.

   (c) $\begin{bmatrix} \text{F} & \boxed{1} \\ \text{G} & \boxed{1} \end{bmatrix}$ unified with $\begin{bmatrix} \text{F} & \begin{bmatrix} \text{J} & \textbf{a} \end{bmatrix} \\ \text{G} & \begin{bmatrix} \text{J} & [\;] \\ \text{K} & \textbf{b} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \text{F} & \boxed{1} \begin{bmatrix} \text{J} & \textbf{a} \\ \text{K} & \textbf{b} \end{bmatrix} \\ \text{G} & \boxed{1} \end{bmatrix}$

   (d) $\begin{bmatrix} \text{F} & \boxed{1}\ \textbf{a} \\ \text{G} & \boxed{1} \end{bmatrix}$ unified with $\begin{bmatrix} \text{G} & \textbf{b} \end{bmatrix} = \bot$

   i.e. unification fails (because **a** and **b** clash)

   (e) $\begin{bmatrix} \text{F} & \boxed{1} \\ \text{G} & \boxed{1} \end{bmatrix}$ unified with $\begin{bmatrix} \text{F} & \begin{bmatrix} \text{J} & \textbf{a} \end{bmatrix} \\ \text{G} & \begin{bmatrix} \text{J} & \textbf{b} \\ \text{K} & \textbf{b} \end{bmatrix} \end{bmatrix} = \bot$

   unification fails (because F.J and G.J would have to lead to the same node and **a** and **b** clash).

   (f) $\begin{bmatrix} \text{F} & \begin{bmatrix} \text{G} & \boxed{1} \end{bmatrix} \\ \text{H} & \boxed{1} \end{bmatrix}$ unified with $\begin{bmatrix} \text{F} & \boxed{1} \\ \text{H} & \boxed{1} \end{bmatrix} = \bot$

   unification fails (because the result would be a cyclic feature structure)

   (g) $\begin{bmatrix} \text{F} & \boxed{1} \\ \text{G} & \boxed{1} \\ \text{H} & \boxed{2} \\ \text{J} & \boxed{2} \end{bmatrix}$ unified with $\begin{bmatrix} \text{F} & \boxed{1} \\ \text{J} & \boxed{1} \end{bmatrix} = \begin{bmatrix} \text{F} & \boxed{1} \\ \text{G} & \boxed{1} \\ \text{H} & \boxed{1} \\ \text{J} & \boxed{1} \end{bmatrix}$

   Note that the number convention is arbitrary: I could have used $\boxed{2}$ here, or indeed $\boxed{3}$ or $\boxed{33}$ — they all mean the same.

   (h) $\begin{bmatrix} \text{F} & \begin{bmatrix} \text{G} & \boxed{1} \end{bmatrix} \\ \text{H} & \boxed{1} \end{bmatrix}$ unified with $\begin{bmatrix} \text{F} & \boxed{2} \\ \text{H} & \begin{bmatrix} \text{J} & \boxed{2} \end{bmatrix} \end{bmatrix} = \bot$

   unification fails (because the result would be a cyclic feature structure)

2. Q: Add case to the initial FS grammar in order to prevent sentences such as *they can they* from parsing.

   The following is a simple way of doing this:

   Grammar rules

   Rule1 $\begin{bmatrix} \text{CAT} & \textbf{S} \\ \text{AGR} & \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & \textbf{NP} \\ \text{CASE} & \textbf{nom} \\ \text{AGR} & \boxed{1} \end{bmatrix}, \begin{bmatrix} \text{CAT} & \textbf{VP} \\ \text{AGR} & \boxed{1} \end{bmatrix}$

   Rule2 $\begin{bmatrix} \text{CAT} & \textbf{VP} \\ \text{AGR} & \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & \textbf{V} \\ \text{AGR} & \boxed{1} \end{bmatrix}, \begin{bmatrix} \text{CAT} & \textbf{NP} \\ \text{CASE} & \textbf{acc} \\ \text{AGR} & [\;] \end{bmatrix}$

Lexicon:

;;; noun phrases

they $\begin{bmatrix} \text{CAT} & \textbf{noun} \\ \text{CASE} & \textbf{nom} \\ \text{AGR} & \textbf{pl} \end{bmatrix}$

them $\begin{bmatrix} \text{CAT} & \textbf{noun} \\ \text{CASE} & \textbf{acc} \\ \text{AGR} & \textbf{pl} \end{bmatrix}$

fish $\begin{bmatrix} \text{CAT} & \textbf{noun} \\ \text{AGR} & [\;] \end{bmatrix}$

it $\begin{bmatrix} \text{CAT} & \textbf{noun} \\ \text{AGR} & \textbf{sg} \end{bmatrix}$

;;; verbs

like $\begin{bmatrix} \text{CAT} & \textbf{verb} \\ \text{AGR} & \textbf{pl} \end{bmatrix}$

likes $\begin{bmatrix} \text{CAT} & \textbf{verb} \\ \text{AGR} & \textbf{sg} \end{bmatrix}$

Root structure:

$\begin{bmatrix} \text{CAT} & \textbf{s} \end{bmatrix}$

3. Q: Work though parses of the following strings for the second FS grammar, deciding whether they parse or not:

   (a) fish fish
   This parses, by application of Rule 2 to the nominal *fish* and the verbal *fish*.

   (b) they can fish
   can fish has two possible analyses, both involving application of Rule 1, either auxiliary *can* with verbal *fish* or transitive *can* with the nominal sense of *fish*. (The first one of these is ‘wrong’ in the sense that the verbal use of *fish* given in the grammar is inflected and the auxiliary *can* should take the infinitival form rather than the inflected form, but this is one of many distinctions outside the scope of this tiny fragment.) The auxiliary *can* won’t combine with the nominal use of *fish* because its COMP value specifies that the HEAD.CAT is **verb** — similarly the transitive use won’t combine with verbal *fish*, because its HEAD.CAT is **noun**.

   Both of the phrasal feature structures formed for *can fish* can take *they* as a specifier by application of Rule 2, so there are two parses.

   (c) it fish
   This fails to parse because the agreement values clash.

   (d) they can
   This fails to parse because for both structures for *can* the COMP value is a complex feature structure, which fails to unify with the value **filled** on Rule 2.

   (e) they fish it
   This fails to parse because the verb *fish* is specified as having COMP value filled (it’s assumed to be intransitive in this grammar), hence Rule 1 won’t combine it with *it*.

4. Q: Modify the second FS grammar to allow for verbs which take two complements. Also add a lexical entry for *give* (just do the variant which takes two noun phrases).

   A: We need another slot for the second complement, which we’ll call COMP2, and either a ternary rule for COMP filling which fills both slots at once, or a second binary rule which fills in the COMP2 slot.

   Rule 1a ;;; Option 1 Filling two complements at once

$$\begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{COMP} & \textbf{filled} \\ \text{COMP2} & \textbf{filled} \\ \text{SPR} & \boxed{3} \end{bmatrix} \rightarrow \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{COMP} & \boxed{2} \\ \text{COMP2} & \boxed{4} \\ \text{SPR} & \boxed{3} \end{bmatrix}, \boxed{2} \begin{bmatrix} \text{COMP} & \textbf{filled} \\ \text{COMP2} & \textbf{filled} \end{bmatrix}, \boxed{4} \begin{bmatrix} \text{COMP} & \textbf{filled} \\ \text{COMP2} & \textbf{filled} \end{bmatrix}$$

Rule 1a ;;; Option 2 Filling COMP2 only

$$\begin{bmatrix} \text{HEAD } \boxed{1} \\ \text{COMP } \textbf{filled} \\ \text{COMP2 } \textbf{filled} \\ \text{SPR } \boxed{3} \end{bmatrix} \rightarrow \begin{bmatrix} \text{HEAD } \boxed{1} \\ \text{COMP } \textbf{filled} \\ \text{COMP2 } \boxed{2} \\ \text{SPR } \boxed{3} \end{bmatrix}, \boxed{2} \begin{bmatrix} \text{COMP } \textbf{filled} \\ \text{COMP2 } \textbf{filled} \end{bmatrix}$$

Notice that in this second case we assume this rule applies to a structure whose COMP has already been filled. We want the COMP2 of the first daughter in the existing Rule 1 to be unspecified, so it applies both to ditransitives and ordinary transitives.

In both alternatives, structures other than lexical entries for two complement verbs have to be changed so that COMP2 is set to filled.

The lexical entry for *give* would be:

$$\text{give} \quad \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CAT } \textbf{verb} \\ \text{AGR } \textbf{pl} \end{bmatrix} \\ \text{COMP} & \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \\ \text{COMP } \textbf{filled} \end{bmatrix} \\ \text{COMP2} & \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \\ \text{COMP } \textbf{filled} \end{bmatrix} \\ \text{SPR} & \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \end{bmatrix} \end{bmatrix} \qquad \text{;;; ditransitive verb}$$

4 Lecture 6: Post-lecture

Q: If you did the exercise associated with the previous lecture to add ditransitive verbs to the grammar, amend your modified grammar so that it produces semantic representations.

Lexical entry for *give*:

$$\text{give} \quad \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CAT } \textbf{verb} \\ \text{AGR } \textbf{pl} \end{bmatrix} \\ \text{COMP} & \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \\ \text{COMP } \textbf{filled} \\ \text{COMP2 } \textbf{filled} \\ \text{SEM} \begin{bmatrix} \text{INDEX } \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{COMP2} & \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \\ \text{COMP } \textbf{filled} \\ \text{COMP2 } \textbf{filled} \\ \text{SEM} \begin{bmatrix} \text{INDEX } \boxed{3} \end{bmatrix} \end{bmatrix} \\ \text{SPR} & \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \\ \text{SEM} \begin{bmatrix} \text{INDEX } \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{PRED } \textbf{like\_v} \\ \text{ARG1 } \boxed{1} \\ \text{ARG2 } \boxed{3} \\ \text{ARG3 } \boxed{2} \end{bmatrix} \end{bmatrix} \qquad \text{;;; ditransitive verb}$$

Notice that I've specified the entry so there's a mismatch between the linear order of complements and the semantic order (i.e., COMP is the ARG3 and COMP2 is the ARG2). Although in some ways ordering is arbitrary, we want the semantics for *give Sandy the book* and *give the book to Sandy* to be the same. In the latter case, the first complement (*the book*) is the ARG2 and the second complement (*Sandy*) corresponds to the ARG3 (the *to* is conventionally assumed to be semantically empty). It's more consistent with the rest of the grammar to have this ordering rather than the alternative, because for ordinary transitive verbs, the direct object is the ARG2.

Rule 1a ;;; Option 1 Filling two complements at once

$$\begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{COMP} & \textbf{filled} \\ \text{COMP2} & \textbf{filled} \\ \text{SPR} & \boxed{3} \\ \text{SEM} & \begin{bmatrix} \text{PRED} & \textbf{and} \\ \text{ARG1} & \boxed{5} \\ \text{ARG2} & \begin{bmatrix} \text{PRED} & \textbf{and} \\ \text{ARG1} & \boxed{6} \\ \text{ARG2} & \boxed{7} \end{bmatrix} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{COMP} & \boxed{2} \\ \text{COMP2} & \boxed{4} \\ \text{SPR} & \boxed{3} \\ \text{SEM} & \boxed{5} \end{bmatrix}, \boxed{2} \begin{bmatrix} \text{COMP} & \textbf{filled} \\ \text{SEM} & \boxed{6} \end{bmatrix}, \boxed{4} \begin{bmatrix} \text{COMP} & \textbf{filled} \\ \text{COMP2} & \textbf{filled} \\ \text{SEM} & \boxed{7} \end{bmatrix}$$

Rule 1a ;;; Option 2 Filling COMP2 only

$$\begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{COMP} & \textbf{filled} \\ \text{COMP2} & \textbf{filled} \\ \text{SPR} & \boxed{3} \\ \text{SEM} & \begin{bmatrix} \text{PRED} & \textbf{and} \\ \text{ARG1} & \boxed{4} \\ \text{ARG2} & \boxed{5} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{COMP} & \textbf{filled} \\ \text{COMP2} & \boxed{2} \\ \text{SPR} & \boxed{3} \\ \text{SEM} & \boxed{4} \end{bmatrix}, \boxed{2} \begin{bmatrix} \text{COMP} & \textbf{filled} \\ \text{COMP2} & \textbf{filled} \\ \text{SEM} & \boxed{5} \end{bmatrix}$$