Features and Augmented Grammars

Allen’s Chapter 4
J&M’s Chapter 11
Augmenting CFGs with Features

• Certain linguistic constraints are not naturally described via CFGs
• Example: *Number Agreement* between constituents - “a boys”
• Possible to describe using refined CF rules:
  - NP-Sing --> ART-Sing N-Sing
  - NP-Plural --> ART-Plural N-Plural
• Much more natural to describe via a *single* feature Number in an augmented CF rule:
  - NP --> ART N (only when $\text{Number}_1 = \text{Number}_2$)
Feature Structures

- *Constituents* can be viewed as feature *structures* that map *values to features*
- Features can be *shared between constituents*
- Some basic features for English:
  - Number, Gender and Person agreement
  - Verb form features and sub-categorizations
- Complex Feature Structures: Feature values can themselves be feature structures
An Example

ART1: (CAT ART ROOT a NUMBER s)

Represents a particular use of word a

Often summarized as:

ART1: (ART ROOT a NUMBER s)
Larger constituents:

Representing an NP constituent for the phrase a fish

NP1: (NP NUMBER s

1 (ART ROOT a

   NUMBER s))

2 (N ROOT fish

   NUMBER s))
Feature Structure

Figure 4.1 Viewing a feature structure as an extended parse tree
Augmented Rules

(NP NUMBER ?n) →

(A RT NUMBER ?n) (N NUMBER ?n)

*(NP 1 (ART NUMBER s)
  2 (N NUMBER s))

*(NP NUMBER s
  1 (ART NUMBER s)
  2 (N NUMBER p))
Using Variables to Express ambiguities

(N ROOT fish NUMBER ?n)

(N ROOT fish NUMBER ?b {s, p})

(N ROOT fish NUMBER {s, p})
Basic Features in English: Number Agreement

Singular and plural markers must agree.

- article and noun
  "a man"
  "a men" (*)

- determiner and noun
  "that book"
  "that books" (*)

- subject and verb
  "Fred sings"
  "Fred sing" (*)
Person and number agreement are inherently linked, so they are often combined in a single feature.

1s: first person singular
2s: second person singular
3s: third person singular
1p: first person plural
2p: second person plural
3p: third person plural

Ex: "am" → {1s}
    "are" → {2s, 1p, 2p, 3p}
    "is" → {3s}
**PForm Feature**

<table>
<thead>
<tr>
<th>Value</th>
<th>Example Prepositions</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>TO</td>
<td>to</td>
<td>I gave it to the bank.</td>
</tr>
<tr>
<td>LOC</td>
<td>in, on, by, inside, on top of</td>
<td>I put it on the desk.</td>
</tr>
<tr>
<td>MOT</td>
<td>to, from, along, ...</td>
<td>I walked to the store.</td>
</tr>
</tbody>
</table>

*Figure 4.3* Some values of the PFORM feature for prepositional phrases
Verb Sub categorizations: SUBCAT Feature

<table>
<thead>
<tr>
<th>Value</th>
<th>Example Verb</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>_none</td>
<td>laugh</td>
<td>Jack laughed.</td>
</tr>
<tr>
<td>_np</td>
<td>find</td>
<td>Jack found a key.</td>
</tr>
<tr>
<td>_np_np</td>
<td>give</td>
<td>Jack gave Sue the paper.</td>
</tr>
<tr>
<td>_vp:inf</td>
<td>want</td>
<td>Jack wants to fly.</td>
</tr>
<tr>
<td>_np_vp:inf</td>
<td>tell</td>
<td>Jack told the man to go.</td>
</tr>
<tr>
<td>_vp:ing</td>
<td>keep</td>
<td>Jack keeps hoping for the best.</td>
</tr>
<tr>
<td>_np_vp:ing</td>
<td>catch</td>
<td>Jack caught Sam looking at his desk.</td>
</tr>
<tr>
<td>_np_vp:base</td>
<td>watch</td>
<td>Jack watched Sam look at his desk.</td>
</tr>
</tbody>
</table>

Figure 4.2 The SUBCAT values for NP/VP combinations
SUBCAT Examples:

(VP) $\rightarrow$ (V SUBCAT _np_vp:inf)
   (NP)
   (VP VFORM inf)

(VP) $\rightarrow$ (V SUBCAT _np_pp:loc)
   (NP)
   (PP PFORM LOC)
Binary Features

Whether the constituent does or does not have a feature
INV feature has two values + and –

Jack Laughed
has an INV feature with value – = (-INV)

Did Jack Laugh
has an INV feature with value + = (+INV)
Morphological Analysis

Present Tense
1. (V ROOT ?r SUBCAT ?s VFORM pres AGR 3s) →
   (V ROOT ?r SUBCAT ?s VFORM base IRREG-PRES −) +S
2. (V ROOT ?r SUBCAT ?s VFORM pres AGR {1s 2s 1p 2p 3p}) →
   (V ROOT ?r SUBCAT ?s VFORM base IRREG-PRES −)

Past Tense
3. (V ROOT ?r SUBCAT ?s VFORM past AGR {1s 2s 3s 1p 2p 3p}) →
   (V ROOT ?r SUBCAT ?s VFORM base IRREG-PAST −) +ED

Past Participle
4. (V ROOT ?r SUBCAT ?s VFORM pastprr) →
   (V ROOT ?r SUBCAT ?s VFORM base EN-PASTPRT −) +ED
5. (V ROOT ?r SUBCAT ?s VFORM pastprr) →
   (V ROOT ?r SUBCAT ?s VFORM base EN-PASTPRT +) +EN

Present Participle
6. (V ROOT ?r SUBCAT ?s VFORM ing) →
   (V ROOT ?r SUBCAT ?s VFORM base) +ING

Plural Nouns
7. (N ROOT ?r AGR 3p) →
   (N ROOT ?r AGR 3s IRREG-PL −) +S

Grammar 4.5 Some lexical rules for common suffixes on verbs and nouns
| a: | (CAT ART  
ROOT A1  
AGR 3s) | saw: | (CAT N  
ROOT SAW1  
AGR 3s) |
| --- | --- | --- | --- |
| be: | (CAT V  
ROOT BE1  
VFORM base  
IRREG-PRES +  
IRREG-PAST +  
SUBCAT { _adjp_np} ) | saw: | (CAT V  
ROOT SAW2  
VFORM base  
SUBCAT _np) |
| cry: | (CAT V  
ROOT CRY1  
VFORM base  
SUBCAT _none) | see: | (CAT V  
ROOT SEE1  
VFORM past  
SUBCAT _np) |
| dog: | (CAT N  
ROOT DOG1  
AGR 3s) | seed: | (CAT N  
ROOT SEED1  
AGR 3s) |
| fish: | (CAT N  
ROOT FISH1  
AGR {3s 3p}  
IRREG-PL +) | the: | (CAT ART  
ROOT THE1  
AGR {3s 3p}) |
| happy: | (CAT ADJ  
SUBCAT _vp:inf) | to: | (CAT TO) |
| he: | (CAT PRO  
ROOT HE1  
AGR 3s) | want: | (CAT V  
ROOT WANT1  
VFORM base  
SUBCAT { _np_vp:inf _np_vp:inf}) |
| is: | (CAT V  
ROOT BE1  
VFORM pres  
SUBCAT { _adjp_np}  
AGR 3s) | was: | (CAT V  
ROOT BE1  
VFORM past  
AGR {1s 3s}) |
| Jack: | (CAT NAME  
AGR 3s) | were: | (CAT V  
ROOT BE  
VFORM past  
AGR {2s 1p 2p 3p}  
SUBCAT { _adjp_np}) |
| man: | (CAT N1  
ROOT MAN1  
AGR 3s) |  |  |
| men: | (CAT N  
ROOT MAN1  
AGR 3p) |  |  |
Sample Grammar (Abbreviated Form)

1. S[-inv] → (NP AGR ?a) (VP[\{pres past\}] AGR ?a)
2. NP → (ART AGR ?a) (N AGR ?a)
3. NP → PRO
4. VP → V[_none]
5. VP → V[_np] NP
7. VP → V[_np_vp:inf] NP VP[inf]
8. VP → V[_adjp] ADJP
9. VP[inf] → TO VP[base]
10. ADJP → ADJ
11. ADJP → ADJ[_vp:inf] VP[inf]

Head features for S, VP: VFORM, AGR
Head features for NP: AGR

Grammar 4.7 A simple grammar in abbreviated form
Sample Grammar (Detailed Form)

1. \[(S INV - VFORM \ ?v\{pres past\} AGR \ ?a) \rightarrow \]
   \[(NP AGR \ ?a) (VP VFORM \ ?v\{pres past\} AGR \ ?a)\]
2. \[(NP AGR \ ?a) \rightarrow (ART AGR \ ?a) (N AGR \ ?a)\]
3. \[(NP AGR \ ?a) \rightarrow (PRO AGR \ ?a)\]
4. \[(VP AGR \ ?a VFORM \ ?v) \rightarrow (V SUBCAT _none AGR \ ?a VFORM \ ?v)\]
5. \[(VP AGR \ ?a VFORM \ ?v) \rightarrow (V SUBCAT _np AGR \ ?a VFORM \ ?v) NP\]
6. \[(VP AGR \ ?a VFORM \ ?v) \rightarrow \]
   \[(V SUBCAT _vp:inf AGR \ ?a VFORM \ ?v) (VP VFORM inf)\]
7. \[(VP AGR \ ?a VFORM \ ?v) \rightarrow \]
   \[(V SUBCAT _np_v:inf AGR \ ?a VFORM \ ?v) NP (VP VFORM inf)\]
8. \[(VP AGR \ ?a VFORM \ ?v) \rightarrow \]
   \[(V SUBCAT _adjp AGR \ ?a VFORM \ ?v) ADJP\]
9. \[(VP SUBCAT inf AGR \ ?a VFORM inf) \rightarrow \]
   \[(TO AGR \ ?a VFORM inf) (VP VFORM base)\]
10. \[ADJP \rightarrow ADJ\]
11. \[ADJP \rightarrow ADJ(SUBCAT _inf) (VP VFORM inf)\]

Grammar 4.8 The expanded grammar showing all features
Parsing with Features

1. \((\text{NP AGR } ?a) \rightarrow \bullet (\text{ART AGR } ?a) (\text{N AGR } ?a)\)
2. \((\text{ART ROOT A AGR } 3s)\)
3. \((\text{NP AGR } 3s) \rightarrow \bullet (\text{ART AGR } 3s) (\text{N AGR } 3s)\)
4. \((\text{NP AGR } 3s) \rightarrow (\text{ART AGR } 3s) \bullet (\text{N AGR } 3s)\)
5. \((\text{N ROOT DOG1 AGR } 3s)\)
6. \((\text{NP AGR } 3s) \rightarrow (\text{ART AGR } 3s) (\text{N AGR } 3s) \bullet\)
Given an arc $A$, and $\text{NEXT}$ as the constituent following $\bullet$, and a new constituent $X$, which is used to extend $A$

1. Find an instantiation of variables such that all features specified in $\text{NEXT}$ are found in $X$.
2. Create a new arc $A'$, which is a copy of $A$ except for the instantiations of the variables determined in step 1
3. Update $A'$ as usual in a Chart parser
Parse Trees with Features

**Figure 4.9** Two sample parse trees with feature values
Chart with Features

Figure 4.10  The chart for *He wants to cry.*