Grammar Formalisms

Yoav Goldberg

(the nice slides are by Julia Hockenmaier)

The dependencies so far:

- **Arguments:**
  - verbs take arguments: subject, object, complements, ...
  - **Heads subcategorize for their arguments**

- **Adjuncts/Modifiers:**
  - adjectives modify nouns,
  - adverbs modify VPs or adjectives,
  - PPs modify NPs or VPs
  - **Heads do not subcategorize for their modifiers (modifiers subcategorize for the head)**

These are all "local" dependencies that can typically be expressed in a CFG.
Each word is the dependant of *one other word*.
Hence, dependency *trees.*

Non-local dependencies

Long-range dependencies

- **Bounded long-range dependencies:**
  - limited distance between the head and argument

- **Unbounded long-range dependencies:**
  - arbitrary distance (within the same sentence) between the head and argument

Long-range dependencies are represented with null elements and coindexation.

Now, a word can be dependent on several other words. Hence, dependency *graphs*.

Bounded dependencies

- **Raising:**
  *He **seems** to sleep in class.*
  (you can't say "what does he seem?")

- **(Subject/Object) Control:**
  *He **wants** to sleep in class.*
  (you can say "what does he want?")
  *He **promises** her not to sleep in class.*
  *She **persuades** him not to sleep in class.*

Bounded dependencies: Raising/Control in the Treebank

```
(S (NP-SBJ-1 Every Japanese under 40)
   (VP (VBZ seems)
       (S (NP-SBJ (-NONE- *-1))
          (VP to be fluent in Beatles lyrics))))
```

Annotation: a *pro* null element (*), co-indexed (-1) with its antecedent

Infinitival/Participial VPs, Gerunds in the Treebank

```
(PP (IN over)
    (S-NOM (NP-SBJ (-NONE- *))
           (VP (VBG cutting)
               (NP capital-gains taxes))))
```

Annotation: a *pro* null element (*), without co-indexation

Passive in the Treebank

```
(S (NP-SBJ-1 accountants)
   (VP (VBP are)
       (RB n't)
       (VP (VBN noted)
           (NP-2 (-NONE- *-1))
           (PP-CLR as being deeply emotional))))
```

Annotation: a *pro* null element (*), co-indexed (-1) with its antecedent

Unbounded long-range dependencies I: Extraction

- ***wh*-movement (relative clauses, wh-questions)**
  the guy that *[I believe Peter told me you thought]* ***you*** *like.*
  who do *[you believe Peter told you I thought]* I *like?*

- **Topicalization:**
  That guy, [I believe Peter told me you thought] you *like.*

- **Clefts:**
  It's that guy that *I believe Peter told me you thought]* you *like*

Unbounded long-range dependencies II: Coordination

- **Right-node raising:**
  [[she would have bought] and [he might sell]] shares.

- **Argument-cluster coordination:**
  I give [[you an apple] and [him a pear]].

- **Gapping:**
  [She likes sushi], and [he sashimi].

Extraction in the Treebank: *T* (Relative clauses)

```
(NP-SBJ (NP Brooks Brothers))
        (, ,)
        (SBAR (WHNP-1 (WDT which))
              (S (NP-SBJ NNP Marks))
                 (VP (VBD bought)
                     (NP (-NONE- *T*-1))
                     (NP-TMP last year))))))
```

Treebank Annotation: a trace (*T*), co-indexed (-1) with its antecedent (the relative pronoun)

All extraction uses *T* traces.

Long-range dependencies in coordination

Standard (constituent) coordination does not create long-range dependencies:

X → ....Y....
Y → Y conj Y

The dependencies in non-standard coordination cannot be expressed with CFG rules.

Right-Node Raising in the Treebank: *RNR*

```
(SBARQ (SBARQ (WHNP-5 (WP Who))
              (SQ (NP-SBJ (-NONE- *T*-5))
                  (VP (VBZ is)
                      (VP (-NONE- *RNR*-4)))))
       (CC and)
       (SBARQ (WHNP-6 (WP who))
              (SQ (NP-SBJ (-NONE- *T*-6))
                  (VP (MD should)
                      (VP (VB be)
                          (VP (-NONE- *RNR*-4))))))
       (VP-4 (VBG making)
             (NP the criminal law)
             (ADVP-LOC (RB here)))
       (. ?))
```

Treebank Annotation:
Two RNR traces (*RNR*), co-indexed (-4) with the shared element

Argument-cluster coordination in the Treebank: =-co-indexation

```
(VP (VP (VB pay)
        (NP HealthVest)
        (NP-2 $ 5 million)
        (ADVP-TMP-3 right away))
    (CC and)
    (VP (NP=2 additional amounts)
        (PP-TMP=3 in the future))
```

Treebank Annotation:
Instead of null element for the verb,
a new co-indexation (=2) of second conjunct with corresponding element (-2) in first conjunct

Gapping in the Treebank: =-coindexation

```
(S (S (NP-SBJ-1 Only the assistant manager)
      (VP (MD can)
          (VP (VB talk)
              (PP-CLR-2 to the manager))))
   (CC and)
   (S (NP-SBJ=1 the manager)
      (PP-CLR=2 to the general manager)))
```

Treebank Annotation:
Just like argument cluster coordination, but here the conjuncts are S, not VP.