Chapter 2

Constituency

2.1 Recapping CFGs

- (2.1) is a new CFG. It contains some helpful abbreviations:
  - [X → (Y) Z] collapses these two rules: [X → Z] and [X → Y Z]
  - $(X)^*$ abbreviates (((((...) X) X) X) X)
  - $X \rightarrow \left\{ \begin{array}{c} Y \\ Z \end{array} \right\}$ collapses these two rules: [X → Z] and [X → Y]
- Also, instead of writing a rewrite rule for every word (e.g., [N → *farmers*], [N → *ducks*], [N → *monkeys*]), we can abbreviate things using the Lexicon below.

(2.1) *G*:
NTerm = {S, NP, VP, AP, PP, D, A, N, Pro, P, V, Deg}
Start = S

$$\text{Rules} = \left\{ \begin{array}{c} S \rightarrow NP\ VP \\ NP \rightarrow \left\{ \begin{array}{c} (D)\ (AP)^*\ N \\ Pro \end{array} \right\} \\ PP \rightarrow P\ NP \\ VP \rightarrow V\ (NP)\ (PP) \\ AP \rightarrow (Deg)\ A \end{array} \right\}$$

Lexicon:

N:	farmers, ducks, monkeys, windows, vacuums
V:	stab, appear, run, die, assemble, triangulate
D:	the, my
A:	good, bad, ugly, mean, bold, spotted, ruthless
P:	in, on, at, beside, under, through
Deg:	very, somewhat, rather, really
Pro:	I, me, you, he, she, it, we, us, they, them

(2.2) Write three sentences that are generated by this grammar.
Draw a tree for one of them.

(2.3) Write three grammatical sentences that are not generated by this grammar.

(2.4) For at least one of these, revise the grammar so that it does generate that sentence. Then draw its tree, according to the new grammar.

(2.5) Write three ungrammatical sentences that are generated by the original grammar.

(2.6) For at least one of these, propose a revision to the grammar so that the revised grammar will not generate the ungrammatical sentence.

2.2 Constituency

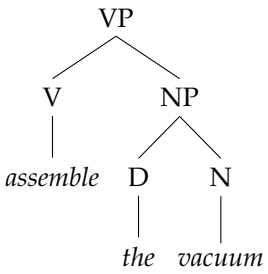
- Basically all theories of syntax make the observation that certain words cluster together in sentence structure. These clusters are referred to as "constituents".
- If we observe a sentence like *We assemble the vacuum*, we get the sense that *the vacuum* is a closer knit string of words than *assemble the*, despite both being strings of adjacent words.
- We also observe that syntactic rules like coordination, deletion, movement, and so on seem to only make reference to constituents – we'll look at examples in detail below.

- Constituents are an empirical phenomenon (something we observe from natural language data), but we can use our CFG-based theory to *predict which strings are constituents and which aren't*. Let's see how:

(2.7) **The yield:** For any non-terminal node in a tree, its yield is largest string consisting of terminal symbols which it dominates.

- What is the yield of the following VP node? Of the NP node? Of the D node?

(2.8)



(2.9) **Constituency Hypothesis:** A string is a constituent just in case it's the *yield* of some node.

- Take the tree drawn in (2.2) and list the constituents according to (2.9).

- We see a fair bit of looseness in how syntacticians talk about constituents:
  - The term constituent often refers to both a string like *the vacuum*, or the non-terminal node yielding it, the NP.
  - We also often refer to the string, like *the vacuum*, by the name of the non-terminal node yielding it, the NP.
  - Constituency is such an important notion in syntax, sometimes CFG structures like (2.8) are referred to as "constituency structures".

2.3 Nominal anaphora

- Anaphora refers to the phenomenon of replacing strings of words with a shorter expression, like a pronoun.
- It usually has quite strict discourse-pragmatic conditions.

2.3.1 Pronouns

- Pronouns are potentially the most pervasive anaphoric phenomenon, widely observed cross-linguistically.
- What do we observe about English pronouns, based on the following data?:

(2.10) a. *My ostentatious spider has long legs*

b. **My ostentatious it has long legs*
c. **My it has long legs*
d. *It has long legs*
e. **My ostentatious spider has long them*
f. *My ostentatious spider has them*
g. **My ostentatious spider them*

- **Generalization**:
- We can state this generalization more precisely using our CFG-based theory.
- **Pronouns hypothesis**:

- Luckily, our CFG in (2.1) already derives this.
- Question: can we have a transformational rule, which takes a sentence with a full NP and replace the NP with a pronoun? E.g.,

(2.11)

An	answer	came	to	the	pig
1	2	3	4	5	6
		⇓			
An	answer	came	to	it	
1	2	3	4	5	

- This was a standard analysis until the 1970s when Bach 1970, Karttunen 1971 observed the following kinds of sentences, referred to as *Bach-Peters sentences*.

(2.12) a. [The pilot who shot at [it]$_i$]$_j$ hit [the MIG that shot at [him]$_j$]$_i$.
b. [The man who shows [he]$_j$ deserves [it]$_i$]$_j$ will get [the prize [he]$_j$ desires]$_i$.

- What are the identifying characteristics of Bach-Peters sentences?

- Something to think about: why are Bach-Peters sentences problematic for analyses like (2.11)? Hint: what happens when we try to 'undo' the transformation?

2.3.2 *one*-anaphora

- What structure does our CFG give to *My very ostentation spider with long legs arrived.*

- *one*-anaphora is another type of English anaphora.
- Question: based on the following data, can *one*-anaphora reduce to a constituent-replacement rule, based on our underlying CFG?

(2.13)
a. (Her very ostentatious spider with long legs left, and)
my very ostentatious spider with long legs arrived

b. (Her very ostentatious spider with long legs left, and)
my one arrived

c. (Her very ostentatious spider with big eyes left, and)
my one with long legs arrived

d. (Her rather humble spider with long legs left, and)
my very ostentatious one arrived

e. (Her rather humble spider with big eyes left, and)
my very ostentatious one with long legs arrived

f. (Her very ostentatious spider with long legs left, and)
**one rather humble cockroach arrived* (where *one* replaces *her*)

g. (Her very humble cockroach with long legs left, and)
**my very one with long legs arrived*

h. (Her very humble cockroach with long legs left, and)
**my rather ostentatious one long legs left arrived*

- **Hint:** List out the strings which *one* seems to replace, and which strings it doesn't replace:

- **Answer**:

- We need to alter our grammar in order to propose a replacement rule for *one*-anaphora. The following solution comes from Baker 1978.
- The grammar now has a constituent which is bigger than an N, but smaller than an NP, called N′. It excludes the determiner, but can include APs and PPs.

(2.14) $G_{N'}$:
NTerm = {S, NP, N', VP, AP, PP, D, A, N, Pro, P, V, Deg}
Start = S

$$\text{Rules} = \left\{ \begin{array}{c} \text{S} \rightarrow \text{NP VP} \\ \text{NP} \rightarrow \left\{ \begin{array}{c} \text{(D) N}' \\ \text{Pro} \end{array} \right\} \\ \text{N}' \rightarrow \left\{ \begin{array}{c} \text{AP N}' \\ \text{N}' \text{ PP} \\ \text{N} \end{array} \right\} \\ \text{PP} \rightarrow \text{P NP} \\ \text{VP} \rightarrow \text{V (NP) (PP)} \\ \text{AP} \rightarrow \text{(Deg) A} \end{array} \right\}$$

- Now we have the tools to make a replacement rule for *one*-anaphora
- **Generalization**:

- Draw a tree for *My very ostentatious spider with long legs arrived*, using $G_{N'}$. Note that there are two solutions. Check that it makes the right predictions for *one*-anaphora.

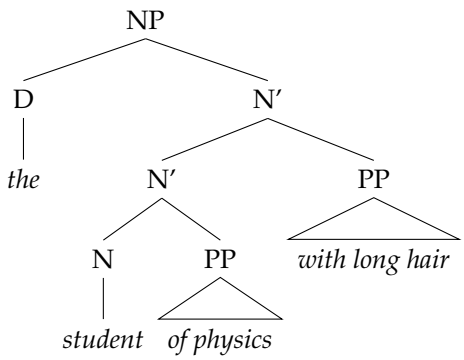
(2.15)

- *one*-anaphora is a nice case study of how a theory (our previous CFG) made the wrong prediction with respect to a observed phenomenon, which led us to amend the theory so that it made the right prediction (adding the N′-constituent).
- How do we integrate *one*-anaphora into the grammar?
  a. Option 1: add a new lexical item s.t. [N → *one*]
  b. Option 2: (Baker's (1978) original proposal) *one* is itself an N′ s.t. [N′ → *one*]
- Baker's Option 2 is based on his following grammaticality judgements.

(2.16) a. The student of physics with short hair is smarter than the one with long hair.
b. *The student of physics with short hair is smarter than the one of mathematics with long hair.

- Baker's analysis has *of*-PPs as direct siblings of the N, as in (2.18).
- This predicts (b) is ungrammatical according to Option 2.

(2.17)



- Baker's observation that (b) is ungrammatical is usually adopted in intro syntax textbooks (see Carnie 2013:§6 for example, though see his fn5). But subsequent work just denies that (b) is ungrammatical. Especially Payne et al. 2013. What do you think?

2.4 Verbal anaphora

- Anaphoric phenomena aren't only restricted to the NP domain.

- English demonstrates at least two anaphoric phenomena in the VP-domain: *do so*-anaphora and VP ellipsis. We'll talk about VP ellipsis in later weeks.
- Here's our most recent grammar, except I altered the VP rule so we can have as many PPs as we like.

(2.18) $G_{N'2}$:
NTerm = {S, NP, N', VP, AP, PP, D, A, N, Pro, P, V, Deg}
Start = S

$$\text{Rules} = \left\{ \begin{array}{c} \text{S} \rightarrow \text{NP VP} \\ \text{NP} \rightarrow \left\{ \begin{array}{c} \text{(D) N}' \\ \text{Pro} \end{array} \right\} \\ \text{N}' \rightarrow \left\{ \begin{array}{c} \text{AP N}' \\ \text{N}' \text{ PP} \\ \text{N} \end{array} \right\} \\ \text{PP} \rightarrow \text{P NP} \\ \text{VP} \rightarrow \text{V (NP) (PP)*} \\ \text{AP} \rightarrow \text{(Deg) A} \end{array} \right\}$$

- What structure does $G_{N'2}$ assign to *assemble the barbecue on the patio with the screw driver*.

(2.19)

- The following data give some judgements about *do so*-anaphora.
- Question: can we give a replacement rule for *do so*-anaphora given our grammar?

(2.20) a. (She assembled the barbecue on the patio with the screwdriver, and)
and I disassembled the deck chair in the garden with the sledgehammer

b. (She assembled the barbecue on the patio with the screwdriver, and)
and I did so with the Allen key

c. (She assembled the barbecue on the patio with the screwdriver, and)
and I did so in the garden with the Allen key

d. (She assembled the barbecue on the patio with the screwdriver, and)
**and I assembled did so the patio with the Allen key*

e. (She assembled the barbecue on the patio with the screwdriver, and)
**and I assembled the barbecue did so with the Allen key*

f. (She assembled the barbecue on the patio with the screwdriver, and)
**and I assembled did so with the Allen key*

g. (She assembled the barbecue on the patio with the screwdriver, and)
**and I did so the barbecue in the garden with the Allen key*

- **Hint:** List the strings which *do so* can replace, and which strings it doesn't replace:

- **Answer:**
- Again we need to alter our grammar if we want *do so* to replace a constituent. This solution comes from Lakoff and Ross 1976.

(2.21) G_{VPs}:
NTerm = {S, NP, N′, VP, AP, PP, D, A, N, Pro, P, V, Deg}
Start = S

$$\text{Rules} = \left\{ \begin{array}{c} \text{S} \rightarrow \text{NP VP} \\ \text{NP} \rightarrow \left\{ \begin{array}{c} \text{(D) N}' \\ \text{Pro} \end{array} \right\} \\ \text{N}' \rightarrow \left\{ \begin{array}{c} \text{AP N}' \\ \text{N}' \text{ PP} \\ \text{N} \end{array} \right\} \\ \text{PP} \rightarrow \text{P NP} \\ \text{VP} \rightarrow \left\{ \begin{array}{c} \text{VP PP} \\ \text{V (NP)} \end{array} \right\} \\ \text{AP} \rightarrow \text{(Deg) A} \end{array} \right\}$$

- Using G_{VPs}, what is the structure for *assembled the barbecue on the patio with the screwdriver*?

(2.22)

- Now we have all the tools to give a rule for *do so*-anaphora.
- **Generalization**:
- Additionally we can explain why *I did so with the screwdriver* is grammatical, but **I did so the barbecue with the screwdriver* is ungrammatical. How?

- We have a structural distinction between "objects" like *the barbecue*, and "verbal modifiers" like *with the screwdriver*.
  - **Objects:**

- **VP modifiers:**

- The following is a Groucho Marx joke (from the 1930 movie *Animal Crackers*).

(2.23) *One morning I shot an elephant in my pajamas...*
...how he got into my pajamas I'll never know.

- The joke comes from the ambiguity of the first sentence. We can model this ambiguity structurally.
- Sketch two trees, representing the two readings of the first sentence. Use our most recent grammar. Verify the structures with pronouns, *one*-anaphora, *do so*-anaphora. Ignore "*One morning*".

(2.24)

(2.25)

- Embedded in this discussion is a hypothesis about the phenomenon of "ambiguity".

(2.26) **Ambiguity hypothesis**: A string s is n-ways ambiguous relative to a grammar G if G provides n possible derivations for s.

- Therefore our grammar (correctly) predicts the Groucho Marx has two readings.
- Our Groucho Marx example is a case of structural ambiguity. But there are other types:

(2.27) a. *I went to the bank.* (Lexical ambiguity)
b. *Every student learned two languages.* (Scopal ambiguity)

- Describe how (a) and (b) are ambiguous.
- Capturing the ambiguity in (a) structurally is easy: add the rules [N $\rightarrow$ $bank_1$] and [N $\rightarrow$ $bank_2$] to the grammar.
- Capturing the ambiguity in (b) structurally is way harder. In fact, its probably *the* biggest question in the study of the syntax-semantics interface. See Barker 2015 for a taste of what modern approaches to this problem look like.

2.4.1 VP ellipsis

- "VP Ellipsis" is a term applied to examples like the following

(2.28) a. I never put a snake in my pocket before, but I might.
b. A: Did Harvey go to the store on his motorcycle yesterday?
B: Yes, he did. / B: Well, he might have. / B: No, but he will tomorrow.
c. They think I'm afraid of them, but I'm not.

- Often (see, e.g., Sag 1976) ellipsis is analyzed as a "deletion" transformation — generate a tree and then delete a part of it.

(2.29) Some questions about VP ellipsis?
a. For each example, what strings seem to have been "deleted".

b. What are the conditions on deletion?
  - What constituent is deleted?
  - What gets left behind?
- Ellipsis is a standard test for constituency.
  - If a string can undergo ellipsis, then it is the yield of a constituent.

2.5 Coordination

- So far our grammar doesn't generate these sentences involving coordination (adapted from Carnie 2013:p87).
- What generalization can we make about coordination based on the data below?

(2.30) a. the [brilliant blue and pale red] station wagon
b. I saw [these dancers and those musicians] smoking something suspicious.
c. I am [drinking lemonade and eating a brownie]
d. [I've lost my wallet or I've lost my mind]
e. We went [through the woods and over the bridge]
f. My [talented daughter and precocious son] performed last night
g. She [fixed the fuse box and packed up her tools] in the garden.
h. ??I [eat a and cook my brownie]
i. ??We [went through and crossed over the bridge]

- We can add a generalized rule for coordination into our CFG. NB: XP stands for any category. Conj is a new category, assigned to *and*, *but*, *or*, etc.

(2.31) **Coordination**: [XP → XP Conj XP]

- What structure should we assign to *through the woods and over the bridge*?

(2.32)

- What do we do about the following? (adapted from Sag et al. 1985)

(2.33)
a. Pat is stupid or a liar
b. Pat is a Republican and proud of it.
c. Pat is healthy and of sound mind.
d. Pat is asleep or at the office.
e. The was a rude remark and in very bad taste.
f. *The stupid or a liar person arrived.
g. *a Republican and proud of it arrived.
h. *The comment a rude remark and in very bad taste was met with silence.

- Under what conditions does English seem to allow coordination of unlike constituents?

- Stipulating a rule which accounts for (2.34) is tricky given our current tools. We may come back to it.
- More problems in (2.35)... These kinds of examples are referred to as "non-constituent coordination" (some examples adapted from Steedman 2017).

(2.34)
a. I saw Ike on Monday and Adlai on Wednesday.
b. Anna married and Sue divorced the same fellow.
c. The red car and blue bus in the driveway.
d. I journeyed through and returned from the woods.

- The literature on how to account for data like (2.35) while preserving (2.32) is large, and there are varieties of proposals. Most involve movement and/or deletion.

2.6 Further readings

- Cardinaletti and Starke 1999 argue based on data from many languages that pronouns and coordination interact in interesting ways, motivating a more complex syntax than the one proposed in 2.3.1.
- Baker's original 1978 proposal about *one*-anaphora has been influential in the acquisition literature arguing that certain components of syntactic information are innate (see e.g., Lidz et al. 2003). Using a corpus study, Payne et al. 2013 refute both Baker's analysis and the subsequent acquisition claims.

- The most comprehensive modern account of *do so*-anaphora is probably Houser 2010. Houser et al. 2007 provide an analysis of a related phenomenon in Danish.
- There's a lot to be said about coordination. Zhang 2009 is an interesting starting point, especially for coordination structures involving more than two conjuncts, and coordination modifiers like *both* and *either*.
- Various accounts of apparent non-constituent coordination exist, starting with Ross 1970, McCawley 1982, and Dowty 1989. Some modern accounts which use movement and/or deletion include Nunes 2006, Johnson 2009, and Citko 2011. Steedman 2017:§7 is an account using neither movement nor deletion.

2.7 Possible paper topics

- Suggestion by Rory Turnbull: the notion of ambiguous parses of strings may be overstated once intonational information is taken into consideration. Do the cases of ambiguity from PP-adjunction discussed above truly demonstrate string ambiguity. What implications does intonation have for the *ambiguity hypothesis* in (2.27)?
- Here we dissected the notions of NP and VP anaphora and proposed the N'-constituent. Is there evidence for PP or AP anaphora? Is a P' or A' constituent useful in describing some natural language phenomenon?
- Carnie 2013:p181 suggests that the acceptability of *one*-anaphora like *the one of physics with long hair* is subject to dialectal variation. Is there any evidence for this claim? Does this feature vary along any sociolinguistic category (region, class, race, gender, etc.)?
- The syntax of coordination is of perpetual interest, especially coordination of unlike categories (Sag et al. 1985, Breuning and Khalaf 2017), and apparent coordination of non-constituents (Dowty 1988, Steedman 1990). But there needs to be more cross-linguistic work. Though see Farudi 2013 on Farsi, and Wyngaerd 2009 on German and Dutch.

References

Bach, E. 1970. Problominalization. *Linguistic Inquiry* 1: 121–122.

Baker, C. L. 1978. *Introduction to Generative Transformational Syntax*. Englewood Cliffs, NJ: Prentice-Hall.

Barker, C. 2015. Scope. *The Handbook of Contemporary Semantic Theory*, ed. by S. Lappin and C. Fox, 47–87. Malden, MA: Wiley-Blackwell.

Breuning, B. and E. Al Khalaf. 2017. Category mismatches in coordination revisited. Ms., University of Delaware and University of Jordan.

Cardinaletti, A. and M. Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. *Clitics in the Languages of Europe*, ed. by H. van Riemsdijk, 145-233. Berlin: Mouton de Gruyter.

Carnie, A. 2013. *Syntax: A Generative Introduction*. 3rd edition. Malden, MA: Wiley-Blackwell.

Citko, B, 2011. Multidominance. *The Oxford Handbook of Linguistic Minimalism*, ed. by C. Boeckx, 119–142. Oxford: Oxford University Press.

Dowty, D. 1988. Type raising, functional composition, and non-constituent conjunction. *Categorial Grammars and Natural Language Structures*, ed. by R. T. Oehrle et al., 153–197. Dordrecht: Reidel.

Farudi, A. 2013. Gapping in Farsi: A Crosslinguistic Investigation. Ph.D. thesis, MIT.

Houser, M. 2010. The syntax and semantics of *do so* anaphora. Ph.D. thesis, UC Berkeley.

Houser, M. L. Mikkelsen, and M. Toosarvandani. 2007. Verb phrase pronominalization in Danish: Deep or surface anaphora?. *Proceedings of WCCFL 34*, ed. by E. Brainbridge and B. Agbayani, 183–195. Somerville, CA: Cascadilla Press.

Johnson, K. 2009. Gapping is not (VP-) ellipsis. *Linguistic Inquiry* 40: 289–328.

Karttunen, L. 1971. Definite descriptions with crossing coreference: A study of the Bach-Peters paradox. *Linguistic Inquiry* 7: 157–182.

Lakoff, G. and J. R. Ross. 1976. Why you can't do so into the kitchen sink. *Syntax and Semantics 7: Notes from the Linguistic Underground*, ed. by J. D. McCawley, 101–111. New York, NY: Academic Press.

Lidz, J., S. Waxman, and J. Freedman. 2003. What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition* 89: B65–B73.

McCawley, J. 1982. Parentheticals and discontinuous constituent structure. *Linguistic Inquiry* 13: 91–106.

Nunes, J. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, MA: MIT Press.

Payne, J., G. K. Pullum, B. C. Scholz, and E. Berlage. 2013. Anaphoric *one* and its implications. *Language* 89: 794–829.

Ross, J. R. 1970. Gapping and the order of constituents. *Progress in Linguistics*, ed. by M. Bierwisch and K. Heidolph, 249–259. The Hague: Mouton.

Sag, I. A. 1976. Deletion and logical form. Ph.D. thesis, Massachusetts Institute of Technology.

Sag, I. A., G. Gazdar, T. Wasow, and S. Weisler. 1985. Coordination and how to distinguish categories. *Natural Language and Linguistic Theory* 3: 117–171.

Steedman, M. J. 1990. Gapping as constituent coordination. *Linguistics and Philosophy* 13: 207–263.

Steedman, M. J. 2017. Combinatory Categorial Grammar: An Introduction. Lecture notes for the 2017 LSA Summer Institute, Lexington KY.

Wyngaerd, G. V. 2007. Gapping constituents. Ms., FWO/K.U. Brussel.

Zhang, N. N. 2009. *Coordination in Syntax*. Cambridge: Cambridge University Press.