Chapter 1

Introduction

1.1 Studying syntax

- **What is grammar?**:

- **Generative syntax**: developing formal theories of natural language grammars to:
  i. make precise, testable generalizations and predictions about sentence structure
  ii. understand what are possible and impossible natural language grammars
  iii. better understand how humans develop, process, and reason about grammar.

- **Why study syntax?** (*for linguists*):

- **Why study syntax?** (*for everyone*):

1.2 Grammaticality

- The goals and terminology of modern syntactic theories have been largely influenced by Chomsky 1957 (*Syntactic Structures*).[1]
- Chomsky takes a language, like English or Japanese, to be a set of possible sentences.

(1.1) English = {
My name is James.
Sweden may export synthetic wolf urine — sprayed along roads to keep elk away — to Kuwait for use against camels.
Myserious condition dubbed 'scromiting' hits weed smokers across the US and causes them to vomit AND scream.
...
}

[1] Itself a condensed version of Chomsky 1956.

- The set excludes impossible English sentences like **James my is name*.
- **A central methodological question**: How do we *know* whether any given sentence *S* should be in English or not? i.e., is *S grammatical* or *ungrammatical*?
  - The idea from Chomsky and those influenced by his work is that speakers can *intuitively recognize* sentences as grammatical or ungrammatical.
  - We can measure the grammaticality of a sentence by asking questions like "Can you say _?", "Can you imagine someone saying _?", "Does _ sound natural?", etc.
  - Can you think of methodological issues with this approach?

- Can't we just look in a corpus? Could we take a set of sentences (e.g., all sentences ever published in *The New York Times*) and take that to be English? Why or why not?

- Chomsky's central argument against this corpus-based approach:

(1.2) a. Colorless green ideas sleep furiously
b. Furiously sleep green ideas colorless.

- What do these sentences have in common?

- What's the difference between these two sentences?

- Chomsky 1957: the difference between (a) and (b) can only be observed through intuitive judgement, not through meaningfulness or frequency in a corpus.
- *However*:
  - Given a corpus, more sophisticated statistical models *do* observe a difference between (a) and (b) (e.g., Perreira 2000, Lappin and Shieber 2007).
  - Without corpus data, we can't answer certain syntactic questions, such as what determines preferences between different structures (see Bresnan et al. 2005).
  - The debate goes on (see Phillips 2011, Gibson and Fedorenko 2013, Schütze 2016).
- In this class we rely heavily on the intuitive judgements of native speakers, but let's keep talking critically about linguistic methodology!

Exercise: Mark each of the following sentences as grammatical or ungrammatical (following Chomsky's distinction between "grammatical" and "non-occurring" (1.2):

(1.3)
1. Wash your own monkey's ears.
2. Wash your monkey's own ears.
3. Harvey wants me to wash myself.
4. I want Harvey to wash myself.
5. The rumor that I was dead surprised myself.
6. A pig that the sky is blue is a happy pig.
7. A pig that can fly is a happy pig.

8. It was me.
9. There were assumed to be errors in the proof.
10. The errors were assumed to be in the first two steps.
11. There weren't any errors, were there?
12. There goes the argument against the proof, doesn't there?
13. The children aren't going to be hard to persuade to paint themselves blue.
14. The children aren't going to be hard to promise to paint themselves blue.
15. Bob seems to want to like abstract art.
16. Bob wants to seem to like abstract art.
17. Bob wants to seem to want to like abstract art.
18. Bob wants to seem to persuade to like abstract art.
19. Andy tends to be easy to persuade to like any kind of food.
20. Andy tends to be likely to persuade to like any kind of food.
21. Andy tends to be likely to pretend to like whatever is put before him.
22. I have an old tin can to keep my money in.

1.3 Formal approaches to grammaticality

- Syntactic theories aim to model data like the grammaticality judgements in (1.3).
- We want a precise, well-defined theory which predicts which sentences are judged as grammatical and which are not.
- The central insight of Chomsky 1957: we can use **Context Free Grammars** in order to precisely define a notion of 'grammaticality' – let's see how below.

1.3.1 Context Free Grammars

- A Context Free Grammar (CFG) is a formal tool for determining whether or not a list of symbols (a "string") is accepted or not.
- Sentences can be understood as strings of symbols (i.e., words).

(1.4) **Grammar**: Defining a CFG:

a. Term: a list of terminal symbols, e.g., *pig*, *pitchfork*, *the*, *lied*, ...
b. NTerm: a list of non-terminal symbols, e.g., NP, VP, S, D, ...
c. Start: a "starting symbol", a member of NTerm, e.g., S
d. Rules: a list of "replacement rules" of the form $a \rightarrow b$, where a is a non-terminal and b is a sequence of any symbols.

- Which of these are possible replacement rules by this definition? Why or why not?:

a. *pig* $\rightarrow$ *pitchfork* S
b. $S \rightarrow$ *pig pitchfork*
c. $NP \rightarrow VP$ *pig*
d. *Replace every word starting with 'p' with 'frog'*

- We can make "derivations" by taking a sequence of symbols, and making alterations to it according to the replacement rules. NB: derivations have to start with the starting symbol.

(1.5) **Derivation**: Given a grammar G, a derivation is a list of sequences of symbols, $x_1, x_2, ..., x_n$, such that:

- $x_1 =$ Start
- x_{i+1} is obtained from x_i by an application of a replacement rule in Rules

- Now we are in a position to give a formal definition of "grammaticality".

(1.6) **Grammaticality**: a string σ, consisting of symbols in Term, is grammatical just in case there is a derivation which has σ as the last step ($\sigma = x_n$).

(1.7) G:
Term = {*pig*, *lied*, *the*}
NTerm = {S, NP, VP, N, D, V}
Start = S

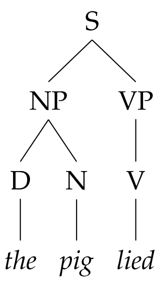
$$\text{Rules} = \left\{ \begin{array}{cc} \text{S} \rightarrow \text{NP VP} & \text{N} \rightarrow \textit{pig} \\ \text{NP} \rightarrow \text{D N} & \text{D} \rightarrow \textit{the} \\ \text{NP} \rightarrow \text{N D} & \text{V} \rightarrow \textit{lied} \\ \multicolumn{2}{c}{\text{VP} \rightarrow \text{V}} \end{array} \right\}$$

(1.8) *The pig lied* is a grammatical string with respect to G. What is its derivation?

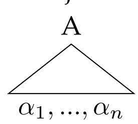
(1.9) a. Is the string *The lied pig* grammatical with respect to G?
b. Is the string *Pig the lied* grammatical with respect to G?

- A derivation can be represented in a tree format like in (1.10).
- Usually this notation is much more perspicuous, so it's standard in theoretical syntax.

(1.10)



(1.11) **Generation**: A grammar generates a tree just in case each subtree of the form:



corresponds to a rule $A \rightarrow \alpha_1, ..., \alpha_n$

- Talking about trees:

  i. **Parent/Sibling/Child**: immediate relations between nodes are supplied by the phrase structure rules in Rule. Given a rule "$A \to a\ b$":

     a. A is the parent of a and b. Branches in trees represent parenthood.

     b. a and b are children of A.

     c. a is b's left sibling, b is a's right sibling.

  ii. **Dominance**: x *dominates* y if we can draw a line from a node x downward along the branches to a node y. NB: every node dominates itself.

  iii. **Precedence**: x *precedes* y if x is dominated by a node a, y is dominated by a node b, and a is the left sister of b.

- Properties of trees:

  i. **Single root** (*just one root node*):
  There is exactly one node x that dominates all nodes. Which is the root in (1.10)?

  ii. **Exclusivity** (*every node has at most one parent*):
  For any pair of nodes x and y, either they are in a dominance relationship or a precedence relationship. *Why does this rule out a node having multiple parents?*

  iii. **No-tangling** (*no crossing branches*): If x precedes y, then all nodes dominated by x precede all nodes dominated by y. *Why does this rule out crossing branches?*

1.3.2 Using CFGs to model natural language

(1.12) G_1:
Term = {*pig*, *answer*, *pitchfork*, *knew*, *lied*, *came*, *became*, *a*, *the*}
NTerm = {S, NP, VP, N, D, V}
Start = S

$$\text{Rules} = \left\{ \begin{array}{cc} \text{S} \to \text{NP VP} & \text{N} \to \textit{pig} \\ \text{VP} \to \text{V NP} & \text{N} \to \textit{answer} \\ \text{VP} \to \text{V} & \text{N} \to \textit{pitchfork} \\ \text{NP} \to \text{D N} & \text{V} \to \textit{knew} \\ \text{NP} \to \text{N} & \text{V} \to \textit{lied} \\ \text{V} \to \textit{became} & \text{V} \to \textit{came} \\ \text{D} \to \textit{a} & \text{D} \to \textit{the} \end{array} \right\}$$

adapted from Hankamer 2017

- For each sentence below, determine

  i. Is it a grammatical sentence of English?

  ii. Is it grammatical relative to G_1?

  iii. If the answers to (i) and (ii) are different, what modifications can we make to the grammar G_1? You can add, delete, or replace rules in Rules.

(1.13) The pig knew the answer.

(1.14) The answer knew the pig.

(1.15) Knew the answer the pig.

(1.16) Knew the pig the answer.

(1.17) The pig lied.

(1.18) The pitchfork lied the pig.

(1.19) The pig lied the pitchfork.

(1.20) Pig the pitchfork the lied the.

(1.21) The pig came a pitchfork.

(1.22) The pig became a pitchfork.

(1.23) The pig became.

(1.24) An answer came to the pig.

(1.25) The pitchfork lied to the pig.

(1.26) My pig lied to my pitchfork.

(1.27) Draw a tree for (1.24)

1.4 Expanding beyond CFGs

1.4.1 Transformations

- Chomsky 1957 proposes CFGs as a way of understanding NL syntax. But he includes an additional mechanism "transformations" over and above the underlying CFG.
- Transformations are used for a variety of phenomena, but prototypically for "movement" phenomena, like (1.28). More on these later.

(1.28)
  a. An answer came to the pig.
  b. To the pig, an answer came.

- Chomsky proposes that strings generated by CFGs are numbered like in (1.29).
- The numbers (and thus the words) can be re-ordered or deleted, or new numbers can be inserted, yielding new sentences.

(1.29)

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

- But this negates the original promise to generate a constrained set of structures corresponding to what we observe in natural language.
- In fact Peters and Ritchie 1973 prove that any arbitrary set of strings can be generated by a grammar using transformations like (1.29).
- Thus transformations like (1.29) aren't useful in delimiting possible human languages.

1.4.2 Theories of syntax and CFGs

- CFGs maintain a central role in practically all modern formal theories of syntax.
- Here are some theories of syntax. In each case, CFGs are an important component.
  - **Government and Binding Theory** (GB): uses a constrained CFG (called X′-syntax), combined with some non-CFG mechanisms like overarching constraints on trees, and a constrained set of transformations.
  - **Generalized Phrase Structure Grammar** (GPSG) and its descendant **Head Driven Phrase Structure Grammar** (HPSG): also uses a constrained CFG, as well as a system of "indexing" nodes to record (sometimes long-distance) relationships between them.
  - **Lexical Functional Grammar** (LFG): a CFG-style tree is only one part of the representation of a sentence, alongside completely different types of structures which determine things like semantics, intonation, etc.
  - **Categorial Grammar** (CG) and its many descendants: developed independently of CFGs but the simplest CGs can be rewritten as CFGs. Instead of phrase structure rules, lexical items can specify the category of their sibling and their parent (e.g., "*arrived* needs an NP sibling and an S parent").
  - **Minimalism** (MP): simultaneously a set of assumptions about language/cognition, and a family of representational systems (Collins and Stabler 2011 provide a precise version). Usually combines simplified X′-syntax/transformations from GB with categorial grammar-type lexical items.

1.5 Further readings

- The first two chapters of Sag et al. 2003 provide an accessible overview of the issues discussed in this handout.
- Chomsky 1957 is a must-read, a highly influential approach to describing natural languages with CFGs and transformations. Pullum 2010 provides a challenging critique.
- Chomsky 1965 elaborates on the earlier work, outlining a perspective on goals for linguistic theory, all of which have been both influential and frequently contested.
- Gazdar et al. 1985 is a sophisticated application of (an enriched class of) CFGs to natural language, full of rich insights. Pollard and Sag 1994 and Sag et al. 2003 are similarly insightful descendants.
- Partee et al. 1990 (esp. chapters 16 and 18) is essential reading for those interested in the properties of trees, CFGs, and other aspects of formal language theory.
- There are many incisive critiques of the empirical foundations and applicability of generative syntax. Ferreira 2005 is an accessible entry point.

1.6 Possible paper topics

- Variation (especially sociolinguistically meaningful variation) is central topic in linguistic theory but not often incorporated into formal syntax. How could this be done? (NB: see Antilla 1997, 2002 for an approach to variation in formal phonology).
- The formalism in this handout is *categorical* ("a structure is generated or it isn't"). Some recent work denies this assumptions, incorporating a probabilistic component into grammar (see Manning 2003, Levy 2015). Does this approach provide new insight into some phenomena?
- Some work in syntax (e.g., Frampton 2004, Chen-Main 2006, Citko 2011) suspends the *exclusivity* constraint on trees (thus nodes can have multiple parents). Is this relaxation motivated? Does it offer new insight?

References

Anttila, A. 1997. Deriving variation from grammar. *Variation, Change and Phonological Theory*, ed. by F. Hinskens et al., 35–68. Amsterdam, John Benjamins.

Anttila, A. 2002. Variation and phonological theory. *Handbook of Language Variation and Change*, ed. by J. Chambers et al., 206–243. Oxford: Blackwell.

Bresnan, J., A. Cueni, T. Nikitina, and R. H. Baayen. 2007. Predicting the dative alternation. *Cognitive Foundations and Interpretation*, ed. by G. Boume. Royal Netherlands Academy of Science.

Chen-Main, J. 2006. On the generation and linearization of multi-dominance structures. Ph.D. thesis, Johns Hopkins University.

Chomsky, N. 1956. *The Logical Structure of Linguistic Theory*. Cambridge, MA: MIT Library.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press. 3–37.

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

Ferreira, F. 2005. Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review* 22: 365–380.

Frampton, J. 2004. Copies, traces, occurrences, and all that. Ms., Northeastern University.

Gazdar, G., E. Klein, G. K. Pullum, and I. A. Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.

Gibson, E. and E. Fedorenko. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28: 1–2, 88–124.

Hankamer, J. 2017. Lecture notes for Syntax I. Ms., UC Santa Cruz.

Katz, J. and P. Postal. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: MIT Press.

Lappin, S. and S. M. Shieber. 2007. Machine learning theory and practice as a source of insight into Universal Grammar. *Journal of Linguistics* 43: 393–427.

Levy, R. 2015. Grammatical knowledge is fundamentally probabilistic. Paper presented at the 2015 Annual Meeting of the Linguistics Society of America, Portland OR.

Manning, C. D. 2003. Probabilistic syntax. *Probabilistic Linguistics*, ed. by R. Bod et al., 289–341. Cambridge MA: The MIT Press.

Partee, B. H., A. G. ter Meulen, and R. Wall. 1990. *Mathematical Methods in Linguistics*. Dordrecht: Springer.

Pereira, F. 2000. Formal grammar and information theory: Together again? *Philosophical Transactions of the Royal Society* 1239–1253. London: Royal Society.

Peters, P. S. and R. W. Ritchie. 1973. On the generative power of transformational grammars. *Information Sciences* 6: 49–83.

Phillips, C. 2013. Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28: 1–2, 156–187.

Pollard, C., and I. A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.

Pullum, G. 2010. Creation myths of generative grammar and the mathematics of Syntactic Structures. *The Mathematics of Language*, ed. by C. Ebert et al., 238–254. Springer Verlag.

Sag, I. A., T. Wasow, and E. Bender. 2003. *Syntactic Theory: A Formal Introduction*, 2nd edition. Stanford, CA: CSLI Publications.

Schütze, C. T. 2016. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Berlin: Language Science Press.