More on phrase structure

LING 200B · Ethan Poole · 13 October 2021

1 The beginning: Phrase-structure rules

- The classical device for describing phrase structure (PS) is PHRASE-STRUCTURE RULES (Chomsky 1956, 1957, 1965). PS rules are context-free string-rewrite rules.[1]

  (1) **PS-rule schema**[2]
  $\alpha \rightarrow \beta\ \gamma$

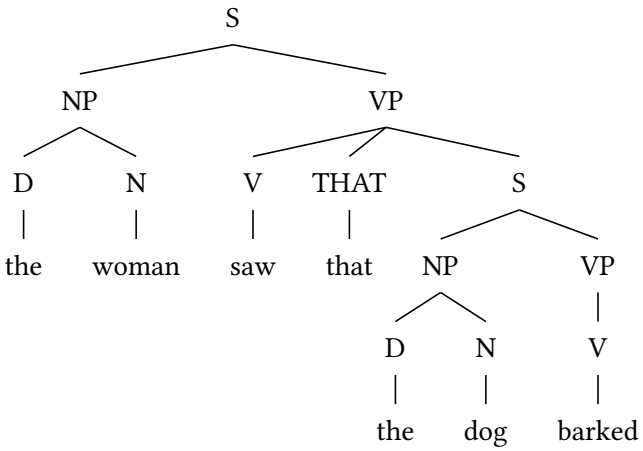
  (2) **Sample phrase-structure grammar**

S → NP VP	D → the | a
NP → (D) N	N → man | woman | dog
VP → V (NP)	V → barked | saw
VP → V THAT S	THAT → that

⇒ Because PS rules can be recursive, they are able to capture the infinity of natural-language syntax.

- PS rules give rise to hierarchical 'tree' structures and thereby capture the fact that sentences are not just strings of words, but are organized into constituents:

(3)



- ***Linear order***
  - PS rules not only encode structural relations, but also the linear order of elements:

    (4) **PRECEDENCE**
    A node X PRECEDES a node Y iff X is to the left of Y and X does not dominate Y and Y does not dominate X.

    (5) **NONTANGLING CONDITION**
    In any wellformed tree, for any nodes X and Y, if X precedes Y, then all nodes dominated by X precede all nodes dominated by Y.

  - The view that PS trees encode linear order held sway up to GB, but was eventually abandoned in the 90s.

[1] We now know that natural language is *not* context free (Culy 1985; Shieber 1985; Miller 1991).

[2] 'Rewrite α as β and γ' or 'β and γ can combine to make α.'

(6) a. **English**
John saw the dog. *VP → V NP*

b. **Hindi-Urdu**
John-ne kuttaa dekhaa *VP → NP V*
John-ERG dog saw
'John saw a/the dog.'

- ***Problem: Unbounded dependencies***
  - Despite their versatility, PS rules have limitations. Most importantly, PS rules cannot capture unbounded dependencies, where the presence of one element depends on the absence (or presence) of another element unboundedly far away:

    (7) **Who** does Rose think that Dorothy believes that … that Blanche met with ___ in the park?

  - This problem was solved by having a transformational component of the grammar:

    (8) a. **Base component**
    phrase-structure rules → D-STRUCTURE

    b. **Transformational component**
    structural changes to PS trees (movement, deletion, copying, …) → S-STRUCTURE

2 From phrase-structure rules to $\overline{\text{X}}$-Theory

2.1 Problem 1: Endocentricity

- Because every PS rule is in fact a separate stipulation, patterns that hold across all PS rules cannot be captured on a systematic basis.
- Not every logically possible rule is in fact attested: every XP contains an X head, even if other elements within XP are optional.
- For example, an NP will always contain a noun, a VP will always contain a verb, a PP will always contain a pre- or postposition, etc.

  (9) a. Alex [likes cats]

  b. Alex [danced]

  c. *Alex [fond of Susan]

  d. *Alex [the next President]

- Using PS rules, this is a profound mystery. The rules in (10) are all perfectly possible PS rules, but are unattested:

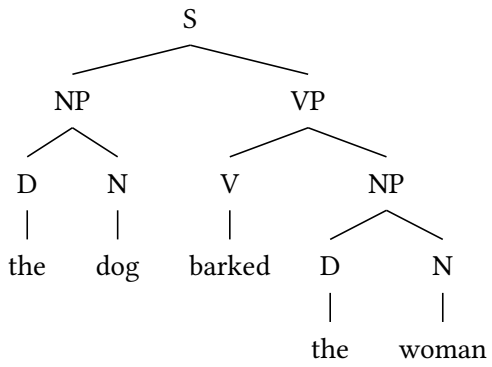
  (10) **Unattested rules**
  VP → D N
  VP → AP
  NP → Adj V
  ⋮

⇒ All attested rules are ENDOCENTRIC. That is, a phrase always contains an element of the same category.

2.2 Problem 2: Redundancy between rules and the lexicon

- All else equal, our toy set of PS rules overgenerates:

(11)



- ***Verb classes***
  To ensure that verbs appear with the appropriate number of arguments, additional symbols referring to classes of verbs need to be introduced.

⇒ Using PS rules, this information has to be stated twice: First in the PS rules for the entire class of verbs, and second for each individual verb in that class. This redundancy is undesirable.[3]

[3] In our Minimalist system, this kind of information is only stated once: on the verb.

(12) a. $VP_{intr} \rightarrow V$
$VP_{trans} \rightarrow V\ NP$

b. $V_{intr} \rightarrow$ barked
$V_{trans} \rightarrow$ saw

2.3 Problem 3: Crosscategorial patterns

- The general shape of phrases seems to be constant across different categories:
  - For both VPs and NPs, the agent is the 'subject' in the active and is in a *by*-phrase in the passive.[4]
  - For both VPs and NPs, the theme is the 'object' in the active and is the 'subject' in the passive.
  - Both VPs and NPs allow PP adjuncts.

[4] The terms 'subject' and 'object' should be understood descriptively, as these notions have no formal status in GB or Minimalist syntax.

(13) **Active voice**

a. The Romans destroyed Carthage in the third Punic war. *VP*

b. the Romans' destruction of Carthage in the third Punic war *NP*

(14) **Passive voice**

a. The city was destroyed (by the enemy). *VP*

b. the city's destruction (by the enemy) *NP*

- These parallels hold for non-action nominalizations as well:[5]

[5] Examples from Stowell (1981:18).

(15) **Agent nominals**

a. i. Someone killed his brother. *VP*

ii. the killer of his brother *NP*

b. i. His brother was killed. *VP*

ii. his brother's killer *NP*

(16) **Adjective-based nominals**

a. i. It is likely that Jim will come. *VP*

ii. the likelihood that Jim will come *NP*

b. i. Joe got angry with me. *VP*

ii. Joe's anger with me *NP*

⇒ PS rules have no means of expressing such commonalities. Such pairs were standardly related via a TRANSFORMATION that turns a clause into a nominal.

- However, Chomsky (1970) points out that these processes are normally unproductive and that their phonological and semantic effects are entirely unpredictable. This is a problem for a transformational account.[6]

[6] Examples from Chomsky (1970).

(17) a. John is easy to please.

b. *John's easiness to please

(18) a. John is likely to win the prize.

b. *John's likelihood to win the prize

(19) a. John { amused / interested } the children with his stories

b. *John's { amusement / interest } of the children with his stories

(20) a. Tomatoes grow ⇝ the growth of tomatoes

b. John grows tomatoes ↛ *John's growth of tomatoes

(21) laughter, marriage, construction, actions, revolution, belief, trial, residence, qualifications, …

- Furthermore, there are also nominals without a verbal counterpart:[7]

[7] Chomsky cites *author* as not having a verbal counterpart, but at least for me, it does: *She authored many books*. The relevant point still stands, though, because there are nominals without verbal counterparts.

(22) a. the weather in England

b. his habit of interrupting

c. the author of the book

d. the algebra of revolution

e. the bottom of the barrel

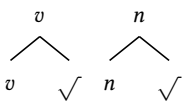
2.4 $\overline{\text{X}}$-Theory

- ***The birth of $\overline{X}$-Theory***
  - Chomsky (1970) proposed that VPs and NPs are both base-generated. That is, there are no transformations turning VPs into NPs (or vice versa).
  - Lexical elements can be unspecified for their syntactic category, thus appearing, e.g., in both VPs and NPs.[8]
  - Idiosyncratic morphological rules then determine the shape of the root if it is specified for a category.

  ⇒ It follows then that the crosscategorial patterns must be a property of the categorial phrase structure itself, and thus $\overline{\text{X}}$-Theory was born.

[8] Nowadays, in Distributed Morphology, there is the idea that all roots are uncategorized and categorization is the result of a categorizing functional head:

(23)



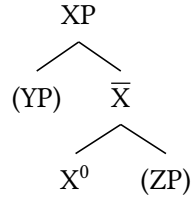
- ***Further developments of $\overline{X}$-Theory***
  - Chomsky (1970) treated $\overline{\text{X}}$-schema as a wellformedness condition on PS rules. PS rules were still needed because a node could still have more than two daughters and they needed to be ordered with respect to each other on a rule-by-rule basis.
  - Stowell (1981) proposed that PS rules should be dispensed with altogether in favor of having just the $\overline{\text{X}}$-schema. Ordering restrictions then have to be attributed to other aspects of the system, like Case.[9]
  - Finally, Kayne (1984) argued that all branching is maximally binary.

[9] Case = abstract Case

∗ ***$\overline{X}$-schema***
Together, these developments yielded the general $\overline{\text{X}}$-schema of GB syntax:

(24) a. $\text{XP} \rightarrow (\text{YP})\ \overline{\text{X}}$
$\overline{\text{X}} \rightarrow \text{X}^0\ (\text{ZP})$

b.



∗ ***Projection***
  - Under $\overline{\text{X}}$-Theory, the properties of a phrase are determined by its head.
  - The head PROJECTS the structure by inheriting (some of) its information, like its category, up through the $\overline{\text{X}}$-schema.
  - Concomitantly, the analytical emphasis in syntactic theory began to shift to features encoded on heads.

- ***Ever-present structure***
One distinctive feature of $\overline{\text{X}}$-Theory (and different from BPS) is that intermediate levels are present even in nonbranching projections:

(25) $[_{\text{NP}}\,[_{\overline{\text{N}}}\,[_{\text{N}}\,\text{cat}\,]\,]\,]$

3 From $\overline{\text{X}}$-Theory to Bare Phrase Structure

- $\overline{\text{X}}$-Theory shifted the focus towards features of lexical items and viewed properties of the phrases that they project as derived from these lexical items.

- ***Two problems***
  - $\overline{\text{X}}$-Theory maintained a division between the heads (like N^0) and the lexical item (like *Alex*). Because the properties of the two have to match, it has to be stated twice, again creating a redundancy.
  - At the same time, $\overline{\text{X}}$-Theory maintained properties of phrase structure that are primitive in the sense that they are not derivative of properties of the lexical items, namely the levels.

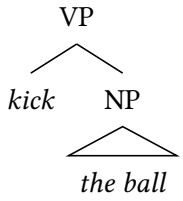
✳ ***Bare Phrase Structure (BPS)***

- The motivation for BPS:[10]

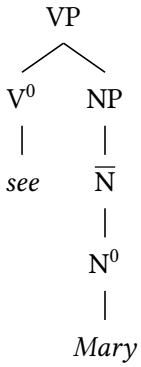
  (26) "In a 'perfect language,' any structure Σ formed by the computation – hence π and λ – is constituted of elements already present in the lexical elements selected for N [the numeration]; no new objects are added in the course of the computation (in particular, no indices, bar-levels in the sense of X-bar theory, etc.)." (Chomsky 1995:393–394)

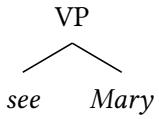
  (27) **Inclusiveness Condition**
  No new objects are added in the course of the computation.

- What if lexical items *directly* combined with other expressions, without projecting their category as a separate node?[11]

  (28)
  

- Vacuous intermediate positions are now eliminated as well:

  (29)
  

  (30)
  

- Moreover, there is no fixed template anymore, either for instances of the same projection (e.g. VPs) or across different projections.

[10] The NUMERATION is the set of lexical items that the computation may access at any given time. In other words, the lexical items that are needed to construct the current phrase.

[11] Having a label is still useful for clarity and legibility, but in order to signify that it is not a separate projection, we place the label just above the lexical item without a branch.

- ***Generalized transformations***
  - In BPS, MERGE is intertwined with transformations (movement). This differs from previous models (Chomsky 1965–Chomsky 1986) in that transformations no longer apply after all structure building has occurred.

⇒ This marks a return to GENERALIZED TRANSFORMATIONS from Chomsky (1957).

4 Other notions of selection

⇒ ***TL;DR***
selection = c-selection = subcategorization

- ***S-selection vs. c-selection***

(31) a. **S(EMANTIC)-SELECTION**
Predicates semantically require certain types of arguments.

b. **C(ATEGORY)-SELECTION**
Predicates impose syntactic requirements on their arguments that do not follow from their semantics.

(32) **Example of s-selection**[12]

a. Björk { **wondered** / ***thought** } [$_{CP}$ who she saw]. *question*

b. Björk { ***wondered** / **thought** } [$_{CP}$ that she saw someone]. *declarative*

[12] Nowadays, this particular contrast might be encoded in the predicates' denotations, since a declarative clause denotes a proposition and a question denotes a set of propositions.

- ***Subcategorization***
  - SUBCATEGORIZATION is used interchangeably with c-selection.[13]
  - The term comes from Chomsky (1965), where grammatical categories are further *subcategorized* based on what they c-select for.
  - For example, V further divides into intransitive, transitive, ditransitive, etc.

[13] I believe that this use of 'subcategorization' to mean 'c-selection' originates in Grimshaw (1977, 1979).

- ***Theta-roles***
  - Traditionally, the number of arguments that a predicate takes is encoded in the predicate's Θ-ROLES.

(33) **Θ-CRITERION**
Every θ-role must be assigned to an argument and every argument must receive one and only one θ-role.

  - THEMATIC ROLES like agent, patient, and goal are *not* the same thing as θ-roles. θ-roles are intended to be more abstract syntactic objects.
  - If we are being charitable, θ-roles are a syntactic way of talking about the semantic properties of a predicate:

(34) *love*:
$[\theta_1, \theta_2] \Rightarrow \lambda x\, \lambda y\, .\, y$ loves x

  - However, θ-roles are entirely superfluous under very basic semantic assumptions (e.g. Heim and Kratzer 1998).[14] Nevertheless, the concept is unfortunately still used in some circles, so it is worth being aware of.

[14] For discussion, see Heim and Kratzer (1998:49–58).

5 Adjunction

- Some constituents are not selected by the lexical items with which they combine:

  (35) a. Björk sang the song **beautifully**.

  b. I ate the nattoo **with a fork**.

  c. The **orange** cat likes lasagna.

  d. The doctor had never seen **quite so rapid** a recovery.

✳ These phrases are called ADJUNCTS, and the phenomenon is called ADJUNCTION.

- ***Some properties of adjuncts***
  - Adjuncts are optional:[15]

    (36) Björk sang the song **(beautifully)**.

  - Adjuncts can be stacked or iterated on a given phrase:

    (37) I ate the nattoo [**quickly**] [**with a fork**] [**every Tuesday**] [**for a year**].

  - A useful diagnostic: *do so* and *one*-anaphora can strand adjuncts, but not arguments:

    (38) I ate the nattoo with a fork.

    a. I did so **with a fork**.

    b. *I did so **nattoo**.

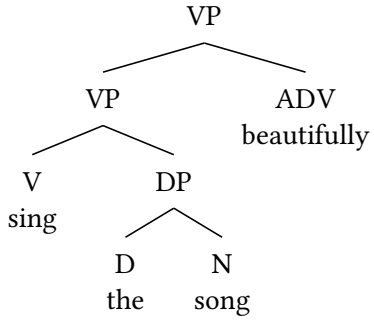
    (39) a. the student **of Finnish** → *the one **of Finnish**

    b. the student **from Finland** → the one **from Finland**

- ***Adjunction targets XP***
  By assumption, adjunction targets maximal projections. The motivation for this assumption is that adjuncts cannot generally intervene between a verb and its object:

  (40) Björk sang **(*beautifully)** the song **(beautifully)**.

  (41)



- ***Adjunction in Minimalist syntax***
  - Unlike $\overline{\text{X}}$-Theory, Bare Phrase Structure does not assign a distinct representation to adjuncts.
  - In Minimalist syntax, adjunction is generally taken to involve an operation different from MERGE.

[15] This implication crucially only goes one way. In other words, optionality does not entail adjuncthood.

– For example, Chomsky (2000, 2001, 2004) argues that adjunction structures are produced with Pair-MERGE, an operation that takes two elements and yields an *ordered pair*.

(42) Pair-MERGE$(\alpha, \beta) = \langle \alpha, \beta \rangle$

– Other approaches decompose MERGE into two operations, e.g. CONCATENATE and LABEL, where arguments involve both operations, and adjuncts involve only the combinatory operation.[16]

[16] e.g. Hornstein and Pietroski (2009); Hunter (2011, 2015)

What to read if you want to learn more?

- Stowell (1981): Foundational text on phrase structure
- Grimshaw (1979): Influential paper on selection

References

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2:113–124.

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

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English Transformational Grammar*, eds. Roderick A. Jacobs and Peter S. Rosenbaum, 184–221. Waltham, MA: Ginn.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. Bare phrase structure. In *Government and Binding Theory and the Minimalist Program*, ed. Gert Webelhuth, 383–439. Cambridge, MA: Blackwell.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, eds. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and beyond*, ed. Adriana Belletti, 104–131. Oxford: Oxford University Press.

Culy, Christopher. 1985. The complexity of the vocabulary of Bambara. *Linguistics and Philosophy* 8:345–351.

Grimshaw, Jane. 1977. English *wh*-constructions and the theory of grammar. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10:279–326.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Hornstein, Norbert, and Paul Pietroski. 2009. Basic operations: Minimal syntax-semantics. *Catalan Journal of Linguistics* 8:113–139.

Hunter, Tim. 2011. *Syntactic Effects of Conjunctivist Semantics: Unifying movement and adjunction*. Amsterdam: John Benjamins.

Hunter, Tim. 2015. Deconstructing merge and move to make room for adjunction. *Syntax* 18:266–319.

Kayne, Richard. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris.

Miller, Philip. 1991. Scandinavian extraction phenomena revisited: Weak and strong generative capacity. *Linguistics and Philosophy* 14:101–113.

Shieber, Stuart. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8:333–343.
Stowell, Timothy. 1981. Origins of phrase structure. Ph.D. dissertation, MIT, Cambridge, MA.