The Minimalist toolkit

LING 200B · Ethan Poole · 29 September 2021

1 Basic facts about natural-language syntax

❶ ***Recursion***
Syntactic structure can be recursive:

(1) a. Rose ate a cheesecake.
b. Blanche said that [Rose ate a cheesecake].
c. Dorothy thought that [Blanche said that [Rose ate a cheesecake]].

❷ ***Hierarchical grouping***
Syntactic elements are grouped hierarchically:[1]

(2) Can [eagles that fly] swim?
a. ✓Is it the case that eagles that fly can swim?
b. *Is it the case that eagles that can fly swim?

[1] The * here indicates that the sentence cannot have this interpretation. That is, it is ungrammatical on this interpretation.

❸ ***Endocentricity (a.k.a. headedness)***
Phrases have the same distribution as one of their subparts:

(3) a. the { **kumquats** / lamps / *eat / *grow } — *Test for noun (N)*
b. I can { **eat** / grow / *kumquats / *lamps } — *Test for verb (V)*
c. *the **eat kumquats** — *Cannot go where Ns go*
d. I can **eat kumquats** — *Can go where Vs go*

❹ ***Combinatoric potential (e.g. selection, subcategorization, theta-roles)***
Elements can specify what they can or must combine with:

(4) a. I { **asked** / **wondered** } [$_{\text{CP}}$ what the time was].
b. I { **asked** / ***wondered** } [$_{\text{DP}}$ the time].

❺ ***Displacement (i.e. movement)***
An element can be pronounced in a position different from where it is interpreted:[2]

[2] Also vice versa, i.e. QR.

(5) **What** did Rose eat { ___ / *cheesecake }?
⇝ What is the x such that Rose ate x?

❻ ***Dependency of form (e.g. agreement, case)***
An element's morphosyntactic form can depend on other elements in the structure and its configuration with respect to them:

(6) a. There { **seems** / *seem } to be [**a unicorn**] in the garden.
b. There { ?seems / **seem** } to be [**two unicorns**] in the garden.

2 Building structure

⇒ ***Task***
Develop a system for combining atomic linguistic elements[3] (❶, ❷), and account for basic distributional facts (❸).

[3] For now, let us consider the atomic elements to be WORDS.

- This notion is called PHRASE STRUCTURE. The most basic instantiation of such a system is a grammar made up of PHRASE-STRUCTURE RULES (PSRs):

(7)
S → NP VP V → chases
NP → D N D → the
VP → V NP N → cat, dog
VP → sleeps

- We will be adopting the Minimalist conception of phrase structure, which is known as BARE PHRASE STRUCTURE (BPS).[4] We will discuss the progression from phrase-structure rules to $\overline{\text{X}}$-theory to BPS later in the quarter.

[4] Speas (1986, 1990); Chomsky (1995a)

2.1 MERGE

⁎ ***The operation MERGE***
Structure is built using MERGE, which takes two arguments and combines them:

(8) MERGE$(\alpha, \beta) = \{\alpha, \beta\}$

- Chomsky often refers to MERGE as a "virtual conceptual necessity".[5]
  - Understanding the term: If it is essentially impossible to imagine a theory of syntax without X, then X is a VIRTUAL CONCEPTUAL NECESSITY.[6]
  - As it pertains to MERGE: Any theory of syntax needs to have a way to put two elements together, and so needs MERGE or something equivalent.

[5] More recently, Chomsky (2005) has argued that only Internal MERGE (responsible for movement) is a virtual conceptual necessity, and anything beyond that carries the burden of proof.

[6] 'Virtual' here meaning 'in effect', not in the sense of computers.

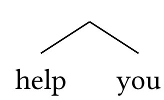
- ***Representing structure***
The three representations below are equivalent. Trees are arguably the most clear. Brackets are useful for preserving space and for highlighting only parts of the structure.[7]

[7] We will turn to category labels below!

(9) a. **With nested sets**
MERGE(help, you) = {help, you}

b. **With brackets**
MERGE(help, you) = [help you]

c. **With a tree**
MERGE(help, you) =



⇒ ***Recursive*** (❶)

MERGE can apply to its own output, thereby generating recursion:

(10) a. MERGE(help, you) = $[_{\text{VP}}$ help you $]$

b. MERGE(to, $[_{\text{VP}}$ help you $]$) = $[_{\text{TP}}$ to $[_{\text{VP}}$ help you $]]$

c. MERGE(trying, $[_{\text{TP}}$ to $[_{\text{VP}}$ help you $]]$) = $[_{\underline{\text{VP}}}$ trying $[_{\text{TP}}$ to $[_{\underline{\text{VP}}}$ help you $]]]$

⇒ ***Hierarchical*** (❷)

– Elements are grouped according to the order of MERGE:[8]

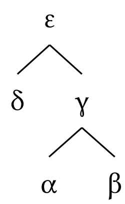
[8] Note that α, β, and δ in (11) could themselves be complex and thus have further hierarchical groups inside them.

(11) a. MERGE$(\alpha, \beta) = [_\gamma\ \alpha\ \beta\ ]$ groups: $\gamma = \{\alpha, \beta\}$

b. MERGE$(\delta, \gamma) = [_\varepsilon\ \delta\ [_\gamma\ \alpha\ \beta\ ]]$ groups: $\gamma = \{\alpha, \beta\}$, $\varepsilon = \{\alpha, \beta, \delta\}$

– In a tree, each NODE represents a hierarchical grouping (i.e. a CONSTITUENT):

(12)

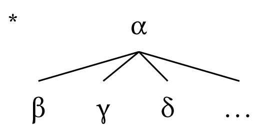


⁕ ***Binarity***

– Since Kayne (1981, 1984), it has generally been assumed that all structure *can* be binary-branching, thus ruling out the need for *n*-ary branching, where $n > 2$:[9]

[9] See also Larson (1988).

(13) **No ternary-, quaternary-, etc. branching**



– BPS takes this hypothesis one step further. It claims that all structure *must* be binary-branching, thus also ruling out unary branching:[10]

[10] There is also a simple conceptual argument against unary branching: it is unnecessary.

(14) **BINARITY**

Every syntactic structure is binary-branching.

(15) **No unary branching**

– MERGE offers a principled explanation for Binarity: MERGE only takes two arguments and thus all structure is binary-branching.

– Binarity is obviously a very strong hypothesis, but if we can maintain it, we have a more *constrained* theory of syntax.

✎ ***Question***

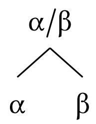
What about structures that at least appear to involve ternary branching, e.g. coordination and ditransitives?

2.2 Projection

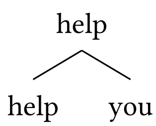
- When *help* merges with *you*, it forms the phrase *help you*. This phrase distributes like *help* and not like *you*:

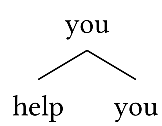
(16) a. We are trying to **help (you)**. [verb]

b. **(*Help) you** are very difficult. [(pro)noun]

> ⁕ ***Projection***
> When α and β merge, one of them PROJECTS:
>
> (17) 
>
>
> - The one that projects—called the HEAD—determines the grammatical properties of the phrase. Thus, a phrase is a projection of the head.

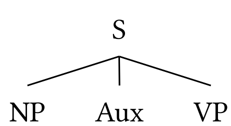
- We can describe the data in (16) in terms of *help* projecting and thus heading the resulting phrase:

(18) 

(19) * 

> ⁕ ***Headedness*** (❸)
> Endocentricity generalizes, from which we can posit a more general principle:
>
> (20) **HEADEDNESS**
> Every syntactic structure is a projection of a head.

✎ ***Question***
What about a classic 'S → NP (Aux) VP' structure (setting aside Binarity)?

(21) 

- ***What projects and why?***
  - For any given instance of MERGE(α, β), we can always (in principle) tell whether α or β projects, based on how the resulting phrase distributes.
  - However, why it is α that projects and not β, or vice versa, is a nontrivial question.
  - For this class: Because we know the "answer" of what projects, we can and will set aside this deeper question (though see Chomsky 2013).[11]

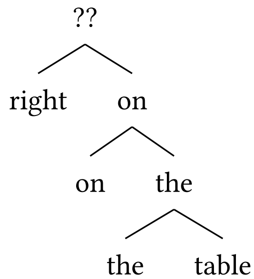
[11] Phrase-structure rules, interestingly, do not face this issue, because they hard-code what projects. As a consequence though, they provide no principled explanation for Headedness, since they can also hard-code exocentric phrases like 'DP → V NP'.

- ***Projecting more than once***
  - Consider the following pair. When *right* merges with *on the table*, what projects?[12]

    (22) a. on the table

    b. right on the table

    (23)



  - Evidence that *on* projects again in (23) comes from cleft sentences:

    (24) **Baseline sentences**

    a. It was [$_{\text{PP}}$ **with great happiness**] that Alex peeled the pomelo.

    b. *It was [$_{\text{AdvP}}$ **very happily**] that Alex peeled the pomelo.

    (25) **Target sentences**

    a. It was [$_{\text{PP}}$ **on the table**] that Alex peeled the pomelo.

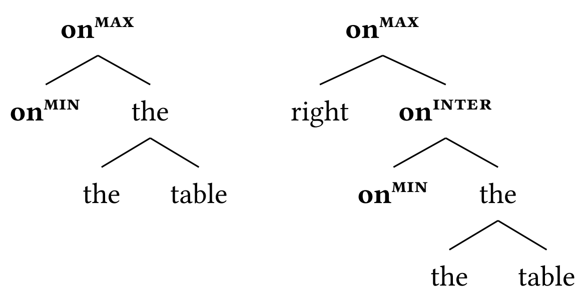
    b. It was [**right on the table**] that Alex peeled the pomelo.

  - The logic of this evidence is as follows:[13]
    * The pivot of a cleft can be a PP (24a), but not an AdvP (24b).
    * Because *right on the table* can be the cleft pivot (25b), it must not be an AdvP.
    * The only other option of what it could be is a PP. Therefore, it must be a PP—i.e. a projection of *on*.

- ***Levels of projection***
  - MINIMAL PROJECTION: the head itself
  - MAXIMAL PROJECTION: the largest projection of a head
  - INTERMEDIATE PROJECTIONS: all projections of a head that are neither minimal nor maximal
  - If a head does not project, then it is both a minimal and maximal projection.

⇒ Note that the maximal projection of a head can change during the course of a derivation!

(26)

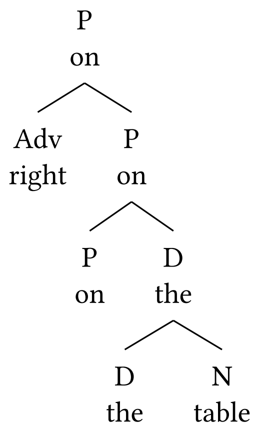


[12] (23) assumes that nominals are projections of determiners, which we will discuss later. For the present discussion, this choice is inconsequential.

[13] PP = projection of a preposition
AdvP = projection of an adverb

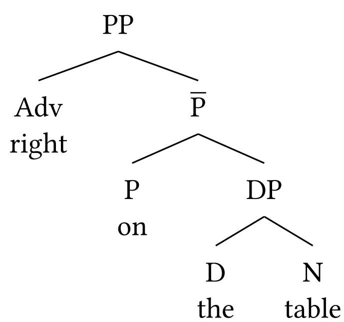
- ***Adding some helpful notational conventions***
  - Under BPS, **what projects is the head**. This differs from pre-BPS theories, where what projects is a *category label*.
  - BPS treats categories as part of the linguistic atom, and thus a head's category projects along with it:[14]

(27)



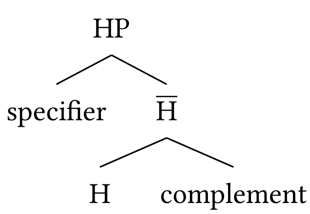
⇒ Even though it is the head that projects in BPS, it is standard to adopt a version of $\overline{\text{X}}$-theory's labels for notational convenience (i.e. it is easier to read):
  - ⋆ minimal projection: X (or X^0)
  - ⋆ maximal projection: XP
  - ⋆ intermediate projection: $\overline{\text{X}}$ (or X′)

(28)



2.3 Anatomy of a phrase

(29)



- ***Some terminology***
  - CONSTITUENT: a structural unit; an element that is one of the components out of which a phrase is built up
  - HEAD: the constituent of a phrase that determines the properties of the phrase
  - COMPLEMENT: the constituent of a phrase that is merged directly with the head ⇝ abbreviated as [Comp, HP] or comp-HP

[14] In Distributed Morphology, roots have no category. Instead, the category comes from categorizing functional heads.

  - specifier: the constituent of a phrase that is merged with a non-minimal projection of the head
    ⤳ abbreviated as [Spec, HP] or spec-HP

2.4 Practical differences with PSRs and $\overline{\text{X}}$-theory

⇒ BPS is derivational, rather than representational. It is concerned not only with what the structure is, but also with how that structure is built.

  - representational: defines wellformed structures
  - derivational: defines wellformed derivations

- Generally, representational approaches can be recast in terms of derivations, and vice versa. They offer different perspectives.

- ***No empty projections***
  - In $\overline{\text{X}}$-theory, every phrase has at least the following amount of structure:

(30)
```
XP
|
X̄
|
X
|
α
```

  - BPS does away with all empty projections. The bar-level in particular is projected only on an 'as needed' basis.

- ***Linearization***
  - Like $\overline{\text{X}}$-theory and unlike PSRs, in BPS, syntactic structure only contains information about *hierarchical* structure and nothing specific about linear order.
  - This is because linear information is redundant, in the sense that it can be predicted from hierarchical structure and simple word-order rules:

(31) **Head-initial linearization**
The complement of HP is linearized to the right of H, and the specifier to the left of H.

(32) **Head-final linearization**
The complement and the specifier of HP are linearized to the left of H.

3 Selection

- merge builds syntactic structure, but it is itself unconstrained. Our system, though, needs to be able to distinguish between outputs of merge that are wellformed and outputs of merge that are not wellformed.

- In particular, with respect to phrase structure, our system needs to handle combinatoric potential, i.e. which elements can combine with which elements.

⇒ The most important dimension of combinatoric potential is SELECTION: a head 'selects' for the elements with which it can syntactically combine:[15]

(33) a. They { **ought** / ***should** } [$_{TP}$ to help you].

b. They { ***ought** / **should** } [$_{VP}$ help you].

(34) a. I { **asked** / **wondered** } [$_{CP}$ what the time was].

b. I { **asked** / ***wondered** } [$_{DP}$ the time].

(35) a. Björk **expects** { to win / that she will win }.

b. Björk **believes** { *to win / that she will win }.

c. Björk **tries** { to win / *that she will win }.

- The selectional properties of lexical items are language-particular and cannot be reduced to meaning:

(36) a. I want (to have) a sandwich.

b. Jag vill *(ha) en smörgås [Swedish]
I want have a sandwich
'I want a sandwich'

(37) a. discriminate *(against) cats

b. ask (*after) the time[16]

c. (*gegen) Katzen diskriminieren [German]
against cats discriminate
'discriminate against cats'

d. *(nach) der Zeit fragen [German]
after the time ask
'ask the time'

⇒ Thus, while the capacity for selection should be part of UG, the actual selectional properties of lexical items must be language-specific.

⁎ ***Modelling selection*** (❹)
Alongside phonological and semantic information, selection information must be encoded in lexical items.

– Let us model selection in terms of FEATURES, where a feature is a diacritic on lexical items used to encode a particular grammatical property.

– The particular features responsible for selection are:[17, 18]

(38) BULLET FEATURES
Where H is a head bearing [•x•], [•x•] is satisfied by merging H (or a projection of H) with an element of category X.

[15] We will discuss other notions of combinatoric potential later.

[16] With the intended interpretation: ask what the current time is.

[17] This notation comes from Heck and Müller (2007).

[18] An 'element of category X' will in effect be a maximal projection of X.

(39) **Examples**

a. [•D•]: Satisfied by merging with an element of category D

b. [•V•]: Satisfied by merging with an element of category V

c. [•A•]: Satisfied by merging with an element of category A

d. [•N•]: Satisfied by merging with an element of category N

– Features must be satisfied in the course of a derivation:

(40) **Principle of Full Interpretation**
Every element of PF and LF must receive an appropriate "interpretation"; they must be licensed. *[scare quotes added]* [Chomsky 1986]

– The formulation of bullet features allows for heads to select their specifiers; that is, features project along with the head.[19] We can make this more explicit:

(41) If [•X•]′ is a projected occurrence of [•X•], then satisfaction of [•X•]′ entails satisfaction of [•X•].

[19] This follows for free under Bare Phrase Structure (Rezac 2003; Béjar and Rezac 2009).

- ***Terminological note***
What we are calling selection or 'bullet' features are standardly called UNINTERPRETABLE FEATURES (Chomsky 2000, 2001). 'Uninterpretable' is intended to imply that they must be "deleted" before reaching LF.

- ***Example of lexical items***

(42) a. **Should**

category:	T
features:	[•V•]
phonology:	. . .
denotation:	. . .

b. **Ought**

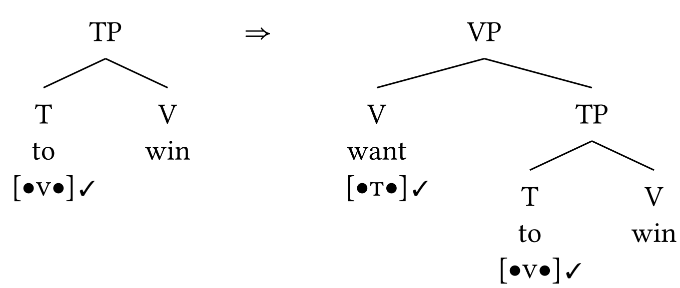
category:	T
features:	[•T•]
phonology:	. . .
denotation:	. . .

c. **Help**

category:	V
features:	[•D•]
phonology:	. . .
denotation:	. . .

- ***Illustration of a derivation***

(43)



⇒ ***Consequences***

– α cannot merge with β and project if β still bears unsatisfied bullet features, because then there would be no way to satisfy those features.[20]

– A transitive verb plus its object has the same distribution as an intransitive verb.

[20] Put differently: only maximal projections can be sisters to c-selecting heads.

⇒ Instances of MERGE are essentially invoked to satisfy bullet features. In this sense, they are what *trigger* MERGE.

4 In brief

- We will revisit these topics throughout the quarter. These are just previews.

4.1 Movement

- Recall that an element can be pronounced in a position different from where it is interpreted:

(44) **What** did Rose eat { ___ / *cheesecake }?
⇝ What is the x such that Rose ate x?

- Displacement is standardly analyzed as MOVEMENT:

(45) What did Rose eat ___ ?

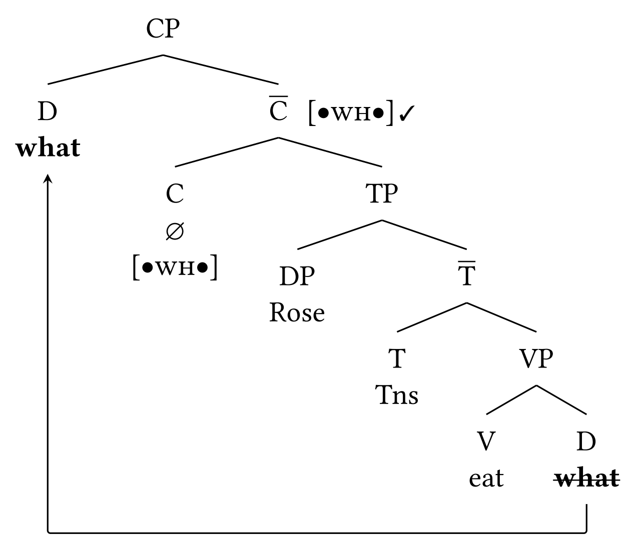
∗ ***Copy Theory of Movement*** (❺)

– In Minimalist syntax, movement of α is the result of merging a copy of α (or α itself) back into the structure:[21,22]

[21] Chomsky (1993, 1995b)

[22] The tree in (46) ignores T-to-C movement.

(46)



– At PF, the lower copies are not pronounced.

- ***Internal and external MERGE***[23]

– Merging a new element into the structure is EXTERNAL MERGE.

– Merging an element contained within an existing structure to a new position in that structure is INTERNAL MERGE.

[23] Chomsky (2001)

4.2 Feature valuation

- Recall that an element's morphosyntactic form can depend on other elements in the structure and its configuration with respect to them:

(47) a. There { **seems** / *seem } to be [**a unicorn**] in the garden.

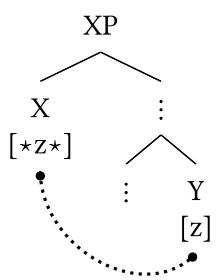
b. There { ?seems / **seem** } to be [**two unicorns**] in the garden.

⁎ ***The operation AGREE*** (❻)
AGREE allows for information (namely, features) to be passed around the structure:

(48) **Agree**
AGREE(f): Given an unvalued feature f on a head H, look for an XP bearing a valued instance of f and assign that value to H. [Preminger 2011, 2014]

(49) **Star features**
[⋆X⋆] is satisfied by triggering AGREE(X). [Heck and Müller 2007]

- ***Probes and goals***
The head bearing [⋆X⋆] is the PROBE; it initiates the search. The element bearing [X] is the GOAL; it is the finding of which terminates the search of the probe.

(50)


- ***Examples of star features***[24]

(51) a. [⋆π⋆]: Triggers agreement for person

b. [⋆#⋆]: Triggers agreement for number

c. [⋆GEN⋆]: Triggers agreement for gender

[24] The more standard way of representing probes is as [*u*X], where *u* means 'unvalued' (and uninterpretable) (Chomsky 2000, 2001).

What to read if you want to learn more?

- Chomsky (1995a): the original BPS paper
- Chomsky (2005): discusses foundational issues concerning MERGE
- Chomsky (2013): labelling algorithm
- Heck and Müller (2007): bullet/star notation (also an interesting theory of phasehood)

References

Béjar, Susana, and Milan Rezac. 2009. Cyclic agree. *Linguistic Inquiry* 40:35–73.

Chomsky, Noam. 1986. *Knowledge of Language.* New York: Praeger.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, eds. Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, MA: MIT Press.

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

Chomsky, Noam. 1995b. *The Minimalist Program.* Cambridge, MA: MIT Press.

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. 2005. Three factors in language design. *Linguistic Inquiry* 36:1–22.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130:33–49.

Heck, Fabian, and Gereon Müller. 2007. Extremely local optimization. In *Proceedings of the 26th Western Conference on Linguistics (WECOL 26)*, eds. Erin Brainbridge and Brian Agbayani, 170–183. Fresno, CA: California State University.

Kayne, Richard. 1981. Unambiguous paths. In *Levels of Syntactic Representation*, eds. Robert May and Jan Koster, 143–183. Dordrecht: Reidel.

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

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19:335–391.

Preminger, Omer. 2011. Agreement as a fallible operation. Ph.D. dissertation, MIT, Cambridge, MA.

Preminger, Omer. 2014. *Agreement and its failures.* Cambridge, MA: MIT Press.

Rezac, Milan. 2003. The fine structure of Cyclic Agree. *Syntax* 6:156–182.

Speas, Margaret. 1986. Adjunction and projections in syntax. Ph.D. dissertation, MIT, Cambridge, MA.

Speas, Margaret. 1990. *Phrase Structure in Natural Language.* Dordrecht: Kluwer.