

Syntax in the Minimalist tradition

Michael Yoshitaka ERLEWINE (mitcho)

National University of Singapore

mitcho@nus.edu.sg <https://mitcho.com>

Me: I study the syntax and semantics of understudied languages of Southeast Asia (Austronesian, Tibeto-Burman) through fieldwork, as well as of Mandarin, Japanese, English. I am particularly interested in (a) the mapping between syntax and semantics, and their relationship to discourse and (b) understanding and explaining the range and shape of cross-linguistic variation. The theory of grammar that I assume/build on in my work is Minimalism. Hi!

You?

1 The program

Government & Binding (GB; Chomsky 1981) / Principles & Parameters (P&P) → The Minimalist Program (MP; Chomsky 1995...). From Sag, Wasow, and Bender (2003: p. 531), annotated:

Many linguists since the early 1980s have framed their grammatical studies in terms of this framework, yielding a large literature that represents analyses of a much wider range of languages and phenomena than any of the other theories listed here. But the analyses developed within GB are often inconsistent with one another. In addition, these analyses are seldom formulated with a precision comparable to that assumed in this text. For these reasons (and the further absence of any theory of what could count as a possible 'parameter'), particular GB analyses and the general claims about crosslinguistic parametric variation are often quite difficult to evaluate. Nonetheless, it is clear that GB analyses tend to share certain noteworthy characteristics, including the following:

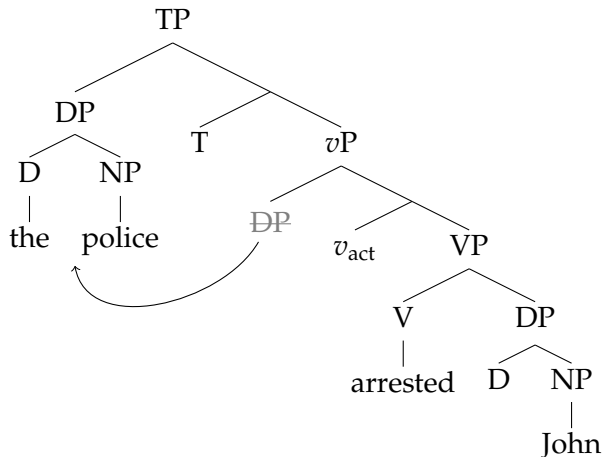
- Highly articulated phrase structures (linguistically significant distinctions and relations are encoded; into tree configurations);
- Use of movement (that is, the transformation 'Move α ');
- Extensive use of empty categories [= unpronounced nodes in the syntax];
- A rich set of universal principles, some of which are parameterized;
- Avoidance of language-particular rules (properties specific to a language are to be expressed in terms of values of universally available parameters);
- Deductive structure (small changes in a grammar should have far-reaching consequences for the language, so that stipulation is minimized).

This strikes me as a pretty accurate characterization. We'll see the underlined today. I will comment on the wavy parts below.

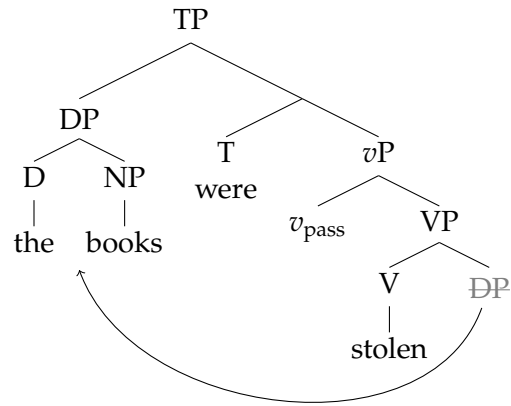
2 A first look

- (1) The police arrested John.
- (2) The books were stolen.

Active transitive:



Passive:



Many things to observe here:

- Nodes have labels, determined by their category.
 - Category is just one type of feature which can be on a node.
 - T = Tense; D = Determiner; *v* = “little *v*”¹
 - Here I follow a pretty common modern approach to node-labeling, where XP is used for maximal projections of category X. (Instead of adopting the X'-bar schema.)
 - Parts are abbreviated. For example, the mother of T and *v*P is a node of category T, but since it's clear that T is the head, I didn't give it a label (as a kind of T, e.g. T').
- Specifiers are on the left; complements are on the right. These are parameterized.²
- Some nodes are unpronounced, but their existence is motivated by...
 - their overt appearance in other sentences in the language, cf *The students will steal the books.*, together with consideration of pros/cons for explaining other phenomena in the language (e.g. a characterization of *do*-support);
 - their overt appearance in other languages:³
 - (3) O Giorgos ephuge.
the George left
'George left.'

(ex Adger, 2003)

¹The quality of a theory and the quality of naming conventions are logically independent.

²Even for languages with mixed headedness, it is claimed that there are interesting universals on what patterns are possible, such that simply specifying headedness head-by-head (or category-by-category) would overgenerate. See Biberauer, Holmberg, and Roberts 2014.

³But there has recently been much discussion on whether phenomena observed in some languages (e.g. object agreement) therefore really occurs in *all* languages. This line of reasoning assumes UG.

– their semantics, which can be logically factored out of the rest of the sentence.

- Adoption of UTAH:


(4) **Uniformity of Thematic Alignment Hypothesis (UTAH; from Baker 1988, statement from Adger 2003: p. 138):**

Identical thematic relationships between predicates and their arguments are represented syntactically by identical structural relationships when items are Merged.

In particular: agent \leftrightarrow Specifier of vP ; theme \leftrightarrow Complement of V.

- The subject is then derived by *movement* of a DP to the Specifier position of TP.

Q: If we assume UTAH, can a subject ever be generated directly (“base-generated”) as the Specifier of TP?

- Movement leaves a *trace*, which is unpronounced; traditionally t . Since the mid-nineties, this has also been thought of as “copying with deletion,” in the *Copy Theory of movement*:
[the police] ... [~~the police~~] arrested John.

- Why did the subject start in the UTAH-following position and then “move to” its pronounced position? Minimalism is regularly conceived of as a *bottom-up, derivational* theory of grammar.

2.1 Let’s build them!

One elementary operation:

(5) **Merge**(α, β):⁴ (read: ‘merge β to α ’)

For any syntactic objects α, β , where α bears an unchecked selectional feature [uF], and β bears a matching categorial feature [F]:

- a. check the feature F on α : \bar{F} ;
- b. let the label γ be the unchecked features of α ; and
- c. return $\begin{array}{c} \gamma \\ \wedge \\ \alpha \quad \beta \end{array}$ if α is non-branching and $\begin{array}{c} \gamma \\ \wedge \\ \beta \quad \alpha \end{array}$ otherwise.

⁴Formulation of Merge and Agree from handouts of Jason Merchant’s. Notice the parameterization of headedness for English in (c).

Lexicon:

- [D; uN] the / \emptyset
- [N] police / books / John
- [T, PAST; uV, uD] \emptyset / were
- $v_{act} = [v; uV, uD]$
- $v_{pass} = [v; uV]$
- [V; uD] arrest(ed) / steal(en)

Hints: (1) takes 7 steps of Merge; (2) takes 5 steps of Merge.

Note: Something tricky happened at the end. We Merged part of the tree back together with itself, which we used to give us the “movement” representation.

- Internal Merge (=Move): Merge(α, β) when α dominates β , resulting in copying. (Generally,) the higher position is pronounced and the lower position is unpronounced.
- External Merge: Merge(α, β) when α does not dominate β .

Many remaining questions at this stage:

Q1: Why does *the* combine with *police* but \emptyset takes *John*?

A1: Broadly, either (a) their selectional requirements ([uN]) can be refined to refer to subclasses of nouns or (b) *the* and \emptyset are both pronunciations of the same D head, but whose form is determined by the type of its complement (a form of contextual allomorphy).

Q2: Why does the copula have to appear in T when we choose v_{pass} ? Why does the verb (V) appear as a passive participle when we choose v_{pass} ?

A2: Similarly, this could be the result of refined selectional requirements on T and v_{pass} , each of which selects for the correct next head down.⁵

Q3: How did inflectional features (tense, ϕ -agreement) end up realized on the verb in (1)?

Q4: What ensures subject-verb agreement?

Q5: Why can't we say **It_(expletive) was stolen the books*?

Other questions you might have: adjunction, the order in which selectional features should be satisfied (the “extension condition”)...

⁵Both parts of this answer are radical simplifications of the Truth.

3 Where things came from

- ▶ The “movement” theory relates the pronounced position of “displaced” material to its “gap” by positing that *it used to be there* or *it has a copy there*.

3.1 Passivization

Consider the following Japanese data from Miyagawa (1989):

(6) **Quantifier adjacent to its noun phrase:**

Taroo-ga **hon-o san-satu** ka-tta.
TARO-NOM book-ACC 3-CL buy-PAST

‘Taro bought three books.’

(7) **Quantifier separated from subject: Actives**

- a. * [Tomodati-no kuruma]-ga **futa-ri** nusum-are-ta.
friend-GEN car-NOM 2-CL steal-PASS-PAST
- b. * **Gakusei-ga** hon-o **futa-ri** ka-tta.
student-NOM book-ACC 2-CL buy-PAST
- c. ?* **Kodomo-ga** [kono kagi]-de **futa-ri** doa-o ake-ta.
child-NOM this key-by 2-CL door-ACC opened
- d. ?* **Gakusei-ga** [zibun-no kane]-de **futa-ri** denwa si-ta
student-NOM self-GEN money-with 2-CL telephone do-PAST

(8) **Quantifier separated from subject: Passives**

- a. **Kuruma-ga** doroboo-ni **san-dai** nusum-are-ta.
car-NOM thief-by 3-CL steal-PASS-PAST
‘Three cars were stolen by the thief.’
- b. Kinoo, **gakusei-ga** [ano otoko]-ni **futa-ri** koros-are-ta.
yesterday student-NOM that man-by 2-CL kill-PASS-PAST
‘Yesterday two students were killed by that man.’

Cf passivization as a lexical rule, which manipulates ARG-ST so that an argument which originally is a complement of the verb becomes a specifier (subject). No corresponding “trace” of the original complement position is predicted.

3.2 Wh-questions

- (9) *What* did he say that he read ___?

There is a range of evidence showing that (a) material between the gap and the pronounced position (here: *what*) can make reference to this item, but that (b) this only occurs in certain points in the derivation. These points (b) roughly correspond to clause edges.⁶

Here are two such arguments from different Englishes:

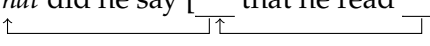
1. Quantifier float in West Ulster English (McCloskey, 2000):

- (10) a. *What all* do you think (that) he'll say (that) we should buy ___?
 b. *What* do you think **all** (that) he'll say (that) we should buy ___?
 c. *What* do you think (that) he'll say **all** (that) we should buy ___?
 d. *What* do you think (that) he'll say (that) we should buy **all**?

2. Intermediate copies in child English (exx Crain and Lillo-Martin, 1999: p. 238):

- (11) a. *What* do you think *what* Cookie Monster eats? (age 5;0)
 b. *Who* do you think *who* Grover wants to hug? (age 4;9)
 c. *What* do you think *what's* in that box? (3;3)

Within the Minimalist tradition, such effects are explained by principles which require that (even unbounded) movement actually never travels that far. This idea is called *successive cyclic movement*:⁷

- (12) *What* did he say [___ that he read ___]?


Cf propagated GAP feature values in HPSG, which suggest possible reference to the “moved” constituent every step along the way.

4 Variation, parameters, features

4.1 Tense and head movement

When a tense involves an auxiliary in English and French, agreement appears on the auxiliary (T), rather than the verb. Note that adverbs (and negation) appear in between the auxiliary and verb:

- (13) John has often eaten apples.
 (14) Jean a souvent mang-é des pommes.
 Jean have.3SG often eat-PAST some apples

⁶CP = Complementizer Phrases. There is growing evidence that *vP* may also be such a region; see Van Urk and Richards (2015) and references there.

⁷It has traditionally been enforced by principles such as the “Subjacency” of yore and, more recently, “Phase Impenetrability” theory. See also footnote 6 above.

Assume such adverbs are adjoined to *vP*. If the auxiliary is pronounced at T and the verb in *vP*, this word order is explained in both languages.

Q: How does the tense and the main verb get pronounced together as one word, for example in the English past or present or French future or present? Two options:

(Somehow) pronounce tense low on *v/V*: (Somehow) pronounce the verb high with T:



The answer in English and French seem to be different!

- (15) John (often) ate/eats (*often) apples.
- (16) Jean (*souvent) manger-a/mang-e (souvent) des pommes.
 Jean often eat-FUT.3sg/eat-PRESENT.3sg often some apples

We say that French has *head movement* of V to T, but English main verbs do not.⁸

- (17) The logic of head-movement in French:
 V moves to T (the main verb is pronounced in that higher position, together with tense/agreement information from T) *if and only if* there is no separate auxiliary in T.

This distinction between main verbs and auxiliaries in English but not French (in being in T or not) in part explains the necessity of *do*-support in English NICE contexts⁹ but not in French.

- (18) * John not eats/ate a sandwich.
- (19) John does/did not eat a sandwich.

4.2 EPP

We moved the subject to Spec,TP in English above because T has a [uD] feature. This property is called the *EPP*¹⁰ and is shared with many languages, but certainly not all.

Now consider Irish, commonly described as a VSO language (but see also (21)):

- (20) Phóg Máire an lucharachán
 kissed Mary the leprechaun
 ‘Mary kissed the leprechaun.’

⁸Head movement may have to be an operation distinct from Merge, or perhaps not... see e.g. Matushansky (2006).

⁹Sag et al. (2003: chapter 13): Negation, Inversion, Contraction, and Ellipsis.

¹⁰which might stand for “Extra Peripheral Position”

- (21) Tá Máire ag-pógail an lucharachán
is Mary PROG-kiss the leprechaun
'Mary is kissing the leprechaun.'

Q: How can we account for Irish word order given the points of variation discussed above?¹¹

4.3 A note on variation

We've seen a lot of thinking about underlying universals and cross-linguistic parameters of syntactic variation. Sag et al. (2003) say that there is an "absence of any theory of what could count as a possible 'parameter.'" But many serious people do think seriously about the shape and locus of "parameters." One now widely-adopted approach is in fact to put "parameters" of variation in the lexicon:

- (22) **The Borer-Chomsky Conjecture (so named by Baker 2008; see Borer 1984)**

All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon.

Summary

Let's look back at the characterization of P&P/MP syntactic theory from Sag et al. 2003. We have indeed seen:

- Highly articulated phrase structures (linguistically significant distinctions and relations are encoded; into tree configurations);
- Use of movement;
- Extensive use of empty categories;
- A rich set of universal principles, some of which are parameterized;
- Avoidance of language-particular rules (properties specific to a language are to be expressed in terms of values of universally available parameters);
- Deductive structure (small changes in a grammar should have far-reaching consequences for the language, so that stipulation is minimized).

References

- Adger, David. 2003. *Core syntax: A minimalist approach*. Oxford University Press.
- Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Baker, Mark C. 2008. The macroparameter in a microparametric world. In *The limits of syntactic variation*, ed. Theresa Biberauer, 351–373. John Benjamins.

¹¹See McCloskey (1997) and related work for further discussion of Irish clause structure.

- Biberauer, Theresa, Anders Holmberg, and Ian Roberts. 2014. A syntactic universal and its consequences. *Linguistic Inquiry* 45:169–225.
- Borer, Hagit. 1984. Restrictive relatives in Modern Hebrew. *Natural Language & Linguistic Theory* 2:219–260.
- Burzio, Luigi. 1986. *Italian syntax: A Government-Binding approach*. Springer.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Foris.
- Chomsky, Noam. 1995. *The minimalist program*. MIT Press.
- Crain, Stephen, and Diane Lillo-Martin. 1999. *An introduction to linguistic theory and language acquisition*. Blackwell.
- Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37:69–109.
- McCloskey, James. 1997. Subjecthood and subject positions. In *Elements of grammar*, ed. Liliane Haegeman, 197–235. Kluwer Academic Publishers.
- McCloskey, James. 2000. Quantifier float and *wh*-movement in an Irish English. *Linguistic Inquiry* 31:57–84.
- Miyagawa, Shigeru. 1989. *Structure and case marking in Japanese*. Academic Press.
- Sag, Ivan A., Thomas Wasow, and Emily M. Bender. 2003. *Syntactic theory: A formal introduction*. CSLI Publications.
- van Urk, Coppe, and Norvin Richards. 2015. Two components of long-distance extraction: Successive cyclicity in Dinka. *Linguistic Inquiry* 46:113–155.
- Wurmbrand, Susi. 2001. *Infinitives: Restructuring and clause structure*. Mouton de Gruyter.

Appendix: Case and agreement with Agree

Continued from section 2.1... A slightly improved version of the above system will (go further toward) actually account(ing) for case and agreement facts, helping with Q4 and Q5. We add another operation, which facilitates the transfer of information between different nodes in the tree (*cf unification*).

(23) **Agree**(α , β ; F) (read: ' α and β agree in F')

For any syntactic objects α and β with matching feature F where α 's sister dominates β ¹²:

- a. let the value of F on α and the value of F on β be equal;
- b. if F is uninterpretable ([u-]) on α or β , check the feature (let [uF:val] = [uF:val]).

Lexicon:

- [D; uN; uCase: ___] the / \emptyset

¹²" α 's sister dominates β " is often stated as " α c-commands β ."

- [N; ϕ :3pl] books
- [N; ϕ :3sg] police / John
- [T, PAST; uv, uD; u ϕ : __, Case:NOM] \emptyset / were / was ...
- $v_{act} = [v; uV, uD; Case:ACC]$
- $v_{pass} = [v; uV]$
- [V; uD] arrest(ed) / steal(en)

What Agree steps are involved in the derivation of (1) and (2) above?

(24) **Grammaticality/convergence:**

A structure is grammatical (a derivation *converges*) if it can be built from items in the *lexicon*, using the operations available (Merge,...), and has no unchecked features.

- The requirement that all nouns (DPs) need Case — even in languages or environments without overt morphological case — has been described as a universal (*the Case Filter*; formerly a “principle”) but it is a lexical property here: the [uCase: __] feature on all D.
- Agreement on T is required by [u ϕ : __], which affects the choice of realization of T. (We still haven’t explained why this and tense information are sometimes pronounced on T and sometimes on the verb.)
- ▶ Notice that there are just two types of v : one with [uD] and [Case:ACC] and one with neither. This is a way of capturing Burzio’s generalization:

(25) **Burzio’s generalization (Burzio, 1986):**

If a verb licenses accusative case, it has an agent.

This accounts for the unavailability of licensing accusative in the passive, forcing the theme to receive nominative from T (and move to Spec,TP together with that). (Q5 above)

- Furthermore, German data suggests that ACC is not linked to the local verb (V), but is instead linked to the choice of v . Data and argument from Wurmbrand (2001):

(26) **German “restructuring”:**

...weil Hans den Traktor zu reparieren versuchte
 since John the tractor.ACC to repair tried
 ‘...since John tried to repair the tractor’
 ‘Tractor’ is the object of ‘repair,’ not ‘try.’

(27) **The so-called “long passive”:**

...dass der Traktor zu reparieren versucht wurde
 that the tractor.NOM to repair tried was

'...that they tried to repair the tractor'

Passivization of 'try' leads to the lack of ACC on 'tractor'; note that 'repair' is unchanged and not in a passive form.

Further questions you may have:

Q1: Why do we move the agent in (1), not the theme?

Q2: Why do we give nominative to the agent in (1), not the theme?

Q3: Why can't we use the same Case feature value twice, for two different DPs?

You're right that our current system is unconstrained in these regards, and will continue to overgenerate ungrammatical structures. Some "principles" must be added, perhaps by refining the definition of Agree. Some candidates:

- A particular feature-value can only Agree once?
- Feature valuation by Agree must be "downward"? "Upward"?
- Agree cannot cross (and similarly for internal Merge)?

Such refinements to the theory are choices that people continue to debate, based on empirical evidence from a broad range of languages.