A Minimalistic Introduction to Minimalism

Avery D Andrews
ANU, July 2023

The purpose of this little paper is to provide just enough of a background in Chomsky's 'Minimalism' to understand what the motivations and use might be for the mathematical constructions in Marcolli, Chomsky & Berwick (2023) (henceforth, at least occasionally, MCB) are, and very little more. But I think a super-limited bit of historical background might be useful.

In particular, there was a substantial split in generative grammar in the early 1970s, between those who were enthusiastic about Chomksy's (1973) 'Conditions on Transformations', and those who were not.[1] The 'Mainstream Generative Grammarians' followed Chomsky in the direction of Condtions, leading to Government-Binding Theory in the 1980s, and on to Minimalism in the early 1990s. The dissenters started devising and working on a variety of 'alternative generative theories', such as Categorial Grammar, Generalized Phrase Structure Grammar, Lexical-Functional Grammar, Head-driven Phrase Structure Grammar, and more, along with what might be described as 'semi-generative theories' such as Relational Grammar and Role and Reference Grammar, which did not sneer at trying to formalize grammatical theory, but put that activity at a lower level of priority than the AGTs.

A defining feature of the mainstream has been to follow Chomsky's pursuit (evident since Conditions) of the idea that human language has a very simple 'core', which can be reasonably located in what people tend to call 'syntax', which is some kind of facility for putting concepts/ideas/thoughts together. This pursuit involves a tendency to put a variety of apparently discrepant phenomena, typically complex and with exceptions, into a 'periphery' which was not always paid much attention to, a practice which made (and still makes) AGT followers uncomfortable. Chomsky's current account of this core is called 'Merge', and is proposed to today be possessed only by Homo sapiens, although I don't think it can be ruled out for Neanderthals or Heidelbergensis (what he says would seem to definitely exclude Homo erectus).

The development of Merge seems to be extensively fueled by rhetoric about Simplicity which does not do much for me, personally, but the prospect that it might fall under some 80yo mathematics originally developed for completely different purposes makes it a lot more interesting. Especially because the math might make it clearer what things belong in the core = Merge, vs what can plausibly be left out of it and put elsewhere.

[1] I recall Arlene Berman recounting to me, not long after the draft of this paper started circulating in summer 1971 (iirc), Susumu Kuno saying to her "What we need is Remarks on Nominalizations [Chomsky (1970)] *without* Conditions on Transformations".

In order to produce anything recognizable as a generative grammar, the core, often called the 'narrow faculty of language' (NFL) has to be supplemented with additional material, producing the 'Broad Faculty of Language'. The aspects of BFL that are closest to NFL are the 'interfaces', of which there are generally supposed to be two:

(1) a. the Conceptual-Intentional Interface (c-i), to semantics ('Nobody knows what semantics is, but most people think it has something to do with meaning' — Robert Meyers (logician) at an ANU seminar, early in this century).

b. the Sensory-Motor Interface (s-m), to overt the overt performance of utterances.

Harley (2014), recently referenced by Chomsky as providing the version he will be adopting for various things, provides a substantial discussion of what is supposed to happen at both interfaces. It seems evident to me that most of what would have been traditionally regarded as 'real' linguistics happens at these interfaces, although the use of the term 'linguistics' has become broader in recent years. NFL is also in particular limited to sentence-internal structure; anything involving discourse is BFL.

Marcolli, Chomsky & Berwick (2023: 3) characterized syntax as the generation, interpretation and performance of 'syntactic objects', that is, grammatical structures, which are generated by using Merege to combine into larger structures an initial set $\mathcal{SO}_0$ of what might be called basic syntactic objects, listed in a lexicon. They are divided into 'lexical items' and 'syntactic features', terms that are not explained in the text, but I think they can be taken to coincide at least roughly with the traditional distinction between major/open class items, and minor/closed class ones. I will identify the 'lexical items' with the 'roots' of Harley (2014), to be discussed shortly below, and the 'syntactic featres' as everything else.

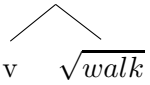
Following Borer (2009), Harley treats roots as essentially anonymous blobs in the syntax itself, which however have various kinds of instructions connecting to the interfaces; prononunciation and morpholgy for the s-m interface, and semantic and contextual information for the c-i interface, which is where most of the traditional material of descriptive syntax lies (on this account). She represents as numbers next to a square root sign, e.g. $\sqrt{42}$, to emphasize their lack of inherent properties. This notation makes a conceptual point, but I don't think it is very suitable for normal use, so will simply designate roots by putting their conventional citation forms under the square-root symbol, e.g. $\sqrt{walk}$ (they would appear to be essentially the same things as the PRED-features values of Andrews (2008) and Andrews (2019)).

To get actual words with a part of speech, we combine these with 'categorizers', such as 'n', 'v' and more. The general approach to building com-

binations is to first put all the stuff you are going to use into a collection called a 'Workspace',[2] and then start combining its initially disconnected structures into bigger and bigger trees (from a crassly empirical viewpoint, I have no idea what, if anything the Workspace actually does for you, but it has a conceptually interesting resemblance to the premise set of a deduction in Linear Logic, the basis for LFG's current method for semantic interpretation, 'Glue Semantics'. Be that as it may, MCB characterizes Merge as an operation on Workspaces.

Threfore, if we have a workspace containing 'v' and $\sqrt{walk}$ into a workspace, we can apply Merge, produce (2), and add it to the Workspace, removing the two ingredients used:

(2) ,s sep=0, l=0



This is a binary, unordered, tree. Unordered means that there are no ordering relations between the nodes, so that it would just be a different picture of the same thing if we flipped the order of the daughter nodes. Another difference between (2) and a normal PS tree is that there is no label on the mother; this is a feature of recent Minimalism, which is justified on the basis that removing conventions for labelling mother nodes reduces complexity of the theory. Of course there is usually some kind of cost for such a move, and the cost is that we need to have some way to find the essential properties of the composite, which are traditionally determined by its 'head', which would be assumed to be the v. The generalization would be that if we combine a root with a categorizer, the categorizer is treated as the 'head'.[3]

It is proposed that this is accomplished by a principle of 'minimal search', namely, do as little work as possible to find the daughter determining the properties of the head. I wonder, however, how much of the work of minimal search can actually be done by the interface constructions associated with the various kinds of items, about which it is time to say something.

The members of $\mathcal{SO}_0$ are associated with two lists of instructions, one for each interface, in Harley's paper.[4] List 1 is the list of roots, which I don't think serves any actual function, List 2 specifies the s-m interface for each item in List 1, while List 3 specifies the c-i interface for these items. List 2 for our root $\sqrt{walk}$ is very simple, because it is spelled only as /walk/

[2]It is tempting to think of this as a multiset, but it MCB it is probably best conceived of as a graph composed of trees, as expressed more clearly in Marcolli, Berwick & Chomsky (2023) than in MCB itself.

[3]I this may amount to starting to deconstruct the concept of 'head', which has always been problematic in various ways, along the lines that what appears to be the head is just the place where you find the properties that are relevant for what you are doing.

[4]She considers only roots, I will be extending this to features.

(diffent details in different varieties of English). But List 3 is a bit more complicated, because it needs to cover its uses and noun and verb. For $\sqrt{walk}$, there would be components like this (notation adapted in various ways from Harley, to fit in better with our mission and my 'alternative generative theory' preferences for more explicitness in formalism:

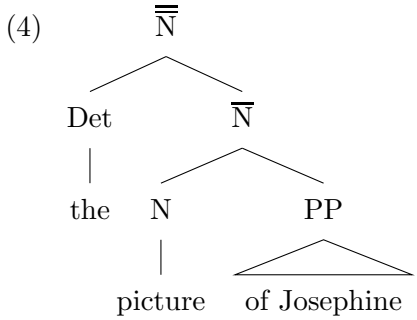
(3) List 3 components for $\sqrt{walk}$:

$$\begin{array}{lcll} \sqrt{walk} & \longleftrightarrow & \lambda x.Walk(x) & / \quad [\text{v} -] \\ & \longleftrightarrow & \text{“episode of walking”} & / \quad [\text{n} -] \\ & \longleftrightarrow & \ldots & \end{array}$$

The first line is the verbal entry; keep in mind that because the trees are unordered, the order of the 'v' and the environment dash is irrelevant. It presents the verb meaning as an 'activity' which will be applied to an argument later. This will be an 'external argument', basically the same thing as the 'I' of Relational Grammar, or the 'Actor' of Role and Reference Grammar', about which more will be said later (and is said by Harley). The second is one of the nominal entries, with a grossly oversimplified meaning; note that a 'walk' is not any old episode of walking, such as from one bus stop to another to get somewhere, but an episode specifically for purposes of health or enjoyment.[5] An additional use is "manner of walking", appearing only when there is an adjective modifying the noun.

'Walk' is an 'unergative' intransitive verb taking an external argument, but there are also transitive verbs, and Harley (drawing on earler work by Doron and others) takes these as having a complement introduced under the categorizer node. So for its semantics we will need a notation to deal with that. The tradition of X-bar theory[6] provides the notions 'complement' (C) and 'specifier' (S) that can be used here. The 'complement' of a terminal category is its phrasal sister, its 'specifier', the sister of its mother, provided that these all have the same category feature. Traditionally, for example:

(4)



[5]Except in sardonic uses such as 'Let's go for a walk', meaning that I need to take you to some other place in order to do something like 'discuss something important".

[6]Originally due to **Harris1948**, adapted and imported into Generative Grammar in Chomsky (1970).

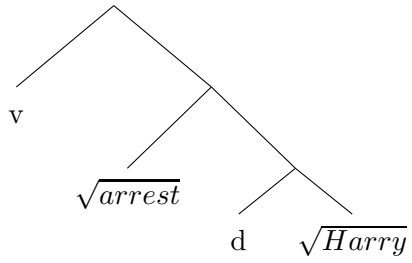
In this picture, *the* is in the specifier position, and the PP in complement position.

This configuration has retained its importance for a long time, and we will designate these positions as 'S' and 'C' respectively, for grammatical elements in the tree. So now we can propose this List 3 entry for the transitive verb *arrest*L as the List 3 entry for $\sqrt{arrest}$:

(5) $\sqrt{arrest} \quad \Leftrightarrow \quad \lambda x.Arrest(x, C) \quad / \quad$ [v –]

This is supposed to produce an interpretation of the form $\lambda x.Arrest(x, Harry)$ from a structure like this, formed by three applications of (External) Merge:

(6)



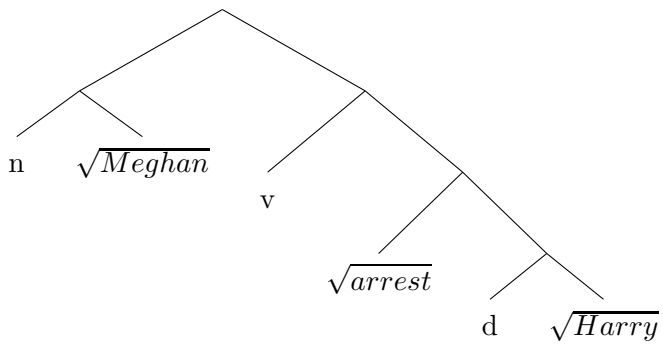
'd' is here the categorizer for proper names, and also determiners.

With this together with (5), we can plausibly get $\lambda x.Arrest(x, Harry)$ as the interpretation of (6), but what about 'subjects' as external arguments? What I propose here is a substantial simplification of the more complex Mainstream treatment (which is motivated by the behavior of causatives and various other things). There is quite a lot of literature on the properties of external arguments, but the only one I will mention here is their ability to be freely omitted in the passive, resulting in the 'agentless passive' construction:

(7) Harry was arrested/left alone/criticized

My (oversimplified) analysis here says that external arguments are attached to 'v' as a 'Specifier', giving structures like this (in the actual current analysis, the external argument goes higher):

(8)



We can get an interpretation for this by including in the List 3 instructions for 'v' some that apply the meaning of the item in the complement position (the verb and object) to the meaning of that in the specifier postion:

(9) v $\Leftrightarrow$ $C(S)$

This rule doesn't need an environment, because if C or S don't find anything in the appropriate position, it won't apply (or what it comes up with will make no sense, and the structure will be blocked for that reason).

But 'unergatives' are part of an opposition, the other, older, member beging 'unaccusative', so called by David Perlmutter in some foundational papers for Relational Grammar in the 1970s. Unaccusative verbs, such as *fall*, at least in its nonagentive meanings, can be described as taking a complement, which fits the semantic profile of an Undergoer, but doesn't provide any semantic role for an Actor/Agent. So its List 3 instruction will be (a) below, an additional instruction for 'v' (b):

(10) a. $\sqrt{fall}$ $\Leftrightarrow$ *Fall*(U)

b. v $\Leftrightarrow$ C

(b) merely passes the meaning of the complement up to become the meaning of the whole.

This account has been organized to support the possibility of using Lexical-Functional Grammar's 'glue semantics' to build the meanings, and thereby allow semantic composition to function in a rigorously specified way to constrain the syntax. But I will not pursue the details of this here (and they might fall apart when I try to work it out).

So far, all of our instances of Merge have been 'Extenal Merge', which together with some of the provisions of the c-i interface takes over the role of phrase-structure rules applying in underlying structure (recalling that linear order needs to be supplied later), but there is another instance of Merge, 'Internal Merge', which takes over the role of transformational movement. Internal Merge is what is responsible for the 'Displacement Property' of natural language that Chomsky has directed quite a lot of attention to, illustrated in examples such as:

(11) a. Which major public figure are they reporting that Empress Meghan has arrested?

b. Harry is reported to have been arrested.

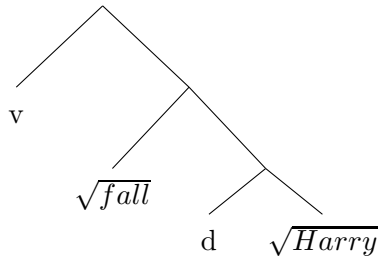
'Displacement' refers to the fact that the first NP in these examples appears a considerable distince from the position that determines its semantic role. In classic TG, transformational rules such as *Wh*-movement, Passivization and Subject-Raising were responsible for such phenomena, but in current Minimalism, they are all produced primarily by the application of Merge to

an item *inside* a tree that the item was attached to by external merge, with more things attached after that.

But the simplest case of Internal Merge would be the appearance of the subjects of unaccusative verbs in the regular subject position of English, producing *Meghan fell* rather than *Fell Meghan*.[7] In classic TG, the question about displacment would be how do the displaced items get to their new positions, but, in the approach using Internal Merge, this is not a problem, because, in a sense, IM can happen whenever it wants to. The problem however is what forces it to apply in some situations, and perhaps blocks it from applying in others. The generic answer would be 'interface conditions', in particular, Chomsky now argues, the information-structural aspects of the c-i interface. The effect of displacement seems in particular to make the displaced constituent either some kind of 'topic' or some kind of 'focus' of a clause, the quote marks indicating that our understanding of what these terms are trying to refer to is not as good as it could be. Operator scope is also sometimes involved, as with the wh-marked NP moving to the front of the clause that constitutes the question.

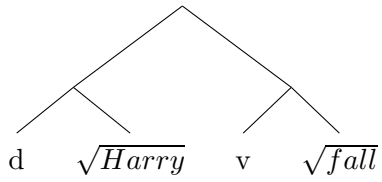
In the past the Internal Merge has often been assumed to leave some kind of 'trace', or 'copy' in the vacated position, but in the MCB version it does not. Rather, IM removes the original mother of the 'moved' node and its sister, putting the sister in the mother's position. So an input to IM could be this:

(12)



IM can then shift the [d $\sqrt{Harry}$] subtree to the top, and produce:

(13)



The usual idea for the obligatory subject position is that it has something to do with the 'tense' element T. Accepting this, T will have some requirement

[7]Importantly, there are many languages, including Italian and Icelandic, where this movement doesn't happen when the subject is indefinite.

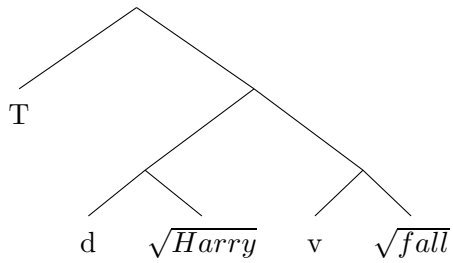
that its Specifier position be filled. This is plausibly to satisfy some kind of information-structure requirement.

There has been a long controversy as to whether items can get semantic roles from more than one predicate (the movement theory of control), but the current verdict is that they can't. So things moved by IM cannot get additional semantic roles, only ones related to information structure. For T, in addition to its relative time specification, I provisionally suggest a statement to the following effect:

(14) C says something about S

The effect is that if we apply T to (14), we get:

(15)



In order to (14) to be satisfied, IM will have to attach something to the current top node of (15), and due to least effort principles, this will be the highest item of the tree that satisfies the requirements.

The actual Minimalist analyses are rather more complicated than these, but this hopefully gives an idea how they might work. In particular, items select for various other items that must appear nearby, and for performance of the utteerance 'planarization' rules impose linear oder on the item, yielding ordering effects such SVO, VSO, variability word order of various kinds, etc. Whether this approach yields better results than any other is however beyond the scope of this sketch

References

Andrews, Avery D. 2008. The role of PRED in LFG+glue. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings of LFG08*, 46–76. Stanford CA: CSLI Publications.

Andrews, Avery D. 2019. A one-level analysis for Icelandic Quirky Case. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings LFG19*, 24–27. Stanford CA: CSLI Publications.

Borer, Hagit. 2009. Roots and categories. `http://www-rcf.usc.edu/~borer/rootscategories.pdf`.

Chomsky, Noam. 1970. Remarks on nominalizations. In R. Jacobs & P. Rosenbaum (eds.), *Readings in english transformational grammar*, 184–221. Waltham MA: Ginn.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson & Paul Kiparsky (eds.), *A festschrift for morris halle*, 232–296. Prentice-Hall.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40. 225–276. http://heidiharley.com/heidiharley/wp-content/uploads/2016/09/Harley2014IdentityOfRootsPublished.pdf.

Marcolli, Matilde, Robert Berwick & Noam Chomsky. 2023. *Old and new minimalism: a Hopf Algebra comparison*. https://arxiv.org/abs/2306.10270.

Marcolli, Matilde, Noam Chomsky & Robert Berwick. 2023. Mathematical structure of syntactic merge. https://arxiv.org/abs/2305.18278.