Notes: Reflections on Syntax

Joseph Galasso

California State University, Northridge~Linguistics

In prep. LINCOM Publications (2019)

Note 1. A Note on the Dual Mechanism Model: Language acquisition vs. learning and the Bell-shape curve.

In this first brief note (one of five), I'd like to reflect on how the *Dual Mechanism Model* (DMM), as compared to a *Single Mechanism Model* (SMM), might inform our more narrow discussion *of Artificial Intelligence* (AI) (discussed in Note 4), as well as inform our larger-scope discussions surrounding the 'nature of language & design' more generally. The description of our methods here will be based on the following dichotomies:

[1] **DMM vs. SMM**

> (i) Whereas an SMM is solely reliant on brute-force associations which are inherently tethered to overt **Learning**—a frequency endeavor **[+Freq]**, where frequency of item-based learning belongs on the vertical mode of processing (to be presented and discussed below). Such item-based learning could be thought of as 'structure-independent' since its focus is solely on the isolated item in question and not on the context of overall structure surrounding the item.
>
> (ii) Whereas a DMM is abstract and rule-based which is inherently tethered to tacit, covert **Acquisition**—a **[-Freq]** endeavor which doesn't rely on a one-one association of item, but rather can be both (i) item-based and (ii) **categorical** in nature, where **structure-dependency** is observant of category over item. Hence a DMM mode—a mode which is both 'item-based' when called upon (e.g. such as lexical learning, irregular formation over rule-based regulars, etc.) and 'categorical-based' when called upon to engage in the manipulation of symbols—is in a unique position to deliver the kind of 'learning curve' which is consistent with what we find of native language acquisition (to be presented and discussed below).

(T)heory.

> (T). Perhaps the sole property of what makes us uniquely human (i.e., the ability to use language) amounts to little more than the sensitivity to remove ourselves from the myopic **item,** and to place ourselves at a perspective, just a step away from the item, and to become sensitive to **structure dependency.** In this sense, T (the ultimate theory of what it means to be human) is that 'taking a step away' from the frequency of an item and seeing how the item sits in an overall structure. (This process of seeing 'item plus structure' will be what makes up T of a Dual Mechanism Model (DMM) as advanced herein these five notes, and what was considered as the core property of language discussed in the four-sentences portion of this text).

It goes without saying that items (lexical words) are quintessential *learned entities* (environmental), they are +frequency-sensitive [+Freq] and carry a classic portmanteaux of features which are typically associated with concrete and conceptual meaning (e.g., Nouns, Verbs, Adjectives). But structure is an altogether different entity. Structure is promoted not by frequency of learning since it is upheld by *categorical processes* which may strip an item away from frequency and place it into a variable standing within the structure. To see what I mean by this *stripping* of the item, let's consider an example that was presented in Sentence #4.

Consider the two bracketed *Items* (I) (say, as 'phrasal-chunks') as presented in the *Structure* (S):

(I) (i) [that is] (a two-word item*)

(S) (a) I wonder what [that is] up there.

(b) I wonder what *[that's] up there.

*(Word-item here as defined by phonological dominant stress—viz., a single word is represented by a single dominant stress pattern— if two dominant stresses, then two words, etc. Note how the word-item 'spaghetti' would have the stress pattern of 'weak-strong-weak' with the middle stress being dominant. The item [*that is*] holds two dominant stresses, (hence, a two-word item), as we hear when we clap out the two words, while [that's] holds only one stress, (hence a one-word item).

Let's restate the analysis we presented earlier in Sentence #4 below:

The base-generated structure first looks something like:

[2] Sentence # 4 (restated)

I wonder [__ [that [VP *is what*]]] up there.

In [2], the Wh-object 'what' begins as the object/complement of the verb 'is' forming a Verb Phrase (VP), and then gets displaced by moving above 'that' in the surface phonology (PF) yielding the derived structure. But if we take a closer look, we see that after such movement of 'what' out of the [VP 'is-what'] phrase, the VP survives only as a head [VP is ø] (i.e., the Head (H) 'is' survives without its complement object 'what'). Thus, the phrase is said to 'partially project'. But partial-phrase projections are indeed allowed in natural languages given that the H still remains (in situ) within the constituent phrase. Hence, we get the licit structure in (a) (as compared to the illicit vacuous/empty VP in (b):

a. I wonder [what$_j$ [that [VP is ___$_j$]]] up there? (A licit structure/ oK)

b.*I wonder [what$_j$ [that's$_k$ [VP __$_k$___ $_j$]]] up there? (An illicit structure / Not ok)

But movement, even partial movement, does have an effect: note how the H 'is' must remain phonologically intact as an H of the VP and can't become a (phonologically attached) **clitic** clinging to the adjacent 'that', as in the one-word item [that's] (whereby there is a reduction now of only one dominant stress). In other words, at least one of the two lexical items within a phrase (P) (in this case, within the VP) must be pronounced (must be phonologically projected). Hence, as we see, when both items [is] as well as [what] move out of the VP –'What' moving into a Spec of a higher P along with the item [is] moving out of its head (H) position of the P and forming itself as a clitic piggy-backing onto the item [that] of the higher P—we see the end result that the VP becomes vacuous (completely empty) and so the structure cannot survive (it becomes ungrammatical).

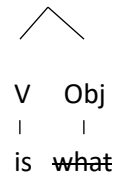
Let's restate below some points on *move* from just prior discussions.

Moved-based Hence, *[[that]'s] is an illicit structure found in (b) (asterisk* marks ungrammaticality), while **Merge-based** of the two words [that] [is] is the only licit structure. It seems simultaneous movement of both head 'is' along with its complement 'what' of the [VP is-what] renders the verb phrase vacuous (i.e., phrases can't be both without a head and without its complement at the same time). In this sense, MOVE-based *[[that]'s] is barred and only Merge-based (of the two items) [that] [is] is allowed to project—the

former (move) being affixal in nature, the latter (merge) lexical). This 'Merge vs. Move' treatment is similar to what we find with the distinction between (merge-based) **Derivational** vs. (move-based) **Inflectional** morphology, where the latter is an *affix* process, and where the former is a *word-forming* process. (For a similar treatment of 'Merge vs. Move' in child language acquisition, see Galasso 2016).

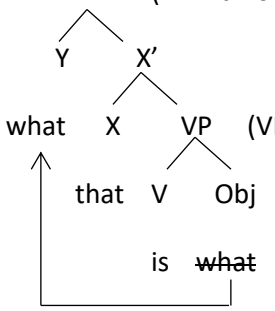
[3] **Progression of structure**

(a) 'is-what' = VP (Verb Phrase)



When object 'what' moves up, it leaves Complement/Obj of Head V still intact, still allowing a licit projection of VP.

(b)

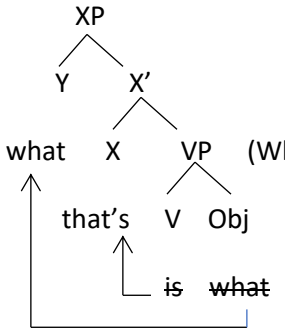


(XP marks a higher functional projection above VP)

(VP head is filled with V 'is', so VP projects)

But note how 'is' must remain as a full word and not as a clitic.

(c)



(When V 'is' is reduced to a clitic ['s], the VP becomes vacuous (i.e., both V and Comp are empty) and so the VP can't project).

But what I want to suggest here for our theory (T) having to do with the following *Five Notes* below—including, and perhaps most importantly, our discussions to come regarding artificial intelligence (AI) –is that this **sensitivity of structure over item** sits as a **core property of language**.

Let's play this out below:

Imagine asking any native speaker of English if the two items below are properly formed (Sentence #4):

(i) [that is],
(ii) [that's].

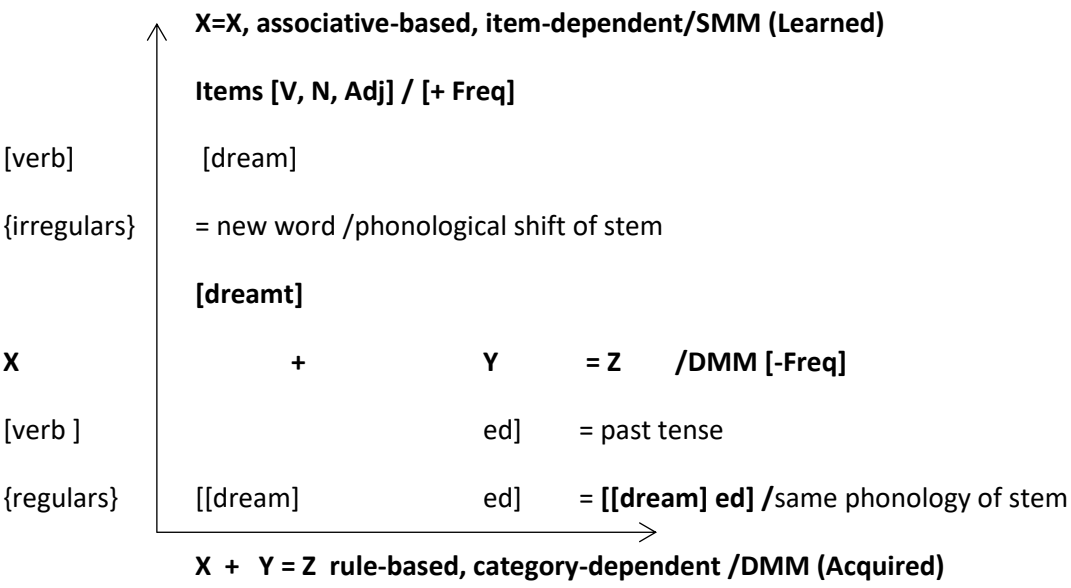
Fine, all native speakers will say both are equally proper in their form. And if there was any preference between the two, the preference would certainly go to the item which is most frequent in the input (i.e., the item most usually heard in the speech environment): that would be item (ii) [that's]. My guess would be that the frequency-count between the two versions could be as high as 'a-hundred-to-one' (if not exceedingly more) when measured in spontaneous speech. That is a perfect example of how the processing of an 'item' is [+Freq]-sensitive: clearly, we hear more numerous examples of [that's] than we do of [that is]. (I am treating the two phrases as *items* here, fragments of constituent structure we hear in the input).

But now reconsider the structure of Sentence #4 (restated in [7] below) and how the isolated item now becomes a rather peripheral feature of the overall structure. Now consider this: when those same people who were earlier asked about the two items *in isolation* are now presented with the same two, but now embedded *within a structure*, all of a sudden the aforementioned preference of [that's] not only becomes the non-preferred item, but, even more egregious, it becomes altogether ungrammatical in its usage (i.e., [that's] it can't be pronounced within Sentence #4).

Here is a perfect example of how +Frequency of **item** is trumped by [–Freq] of **structure**—'Item vs. Structure'. I say *[-Freq] of structure* because structure is not the kind of construct which carries that portmanteaux of features (semantic) which can be readily processed via a brute-force memorization scheme. (Structure is rather category-dependent, syntactic in nature (not semantic), and works on and across variables). In fact, most speakers don't know, nor can they conceptualize, what it is that allows them to tacitly know if one construct is grammatically correct over another (not unless, of course, one is a linguist who works in syntax). Rather, syntactic structure is notorious abstract and hidden, away from the mundane processes of learning a list of items. (Syntax is not simply a list of items gathered which make-up a lexicon).

In linguistic theory, much is made about *linguistic intuition* and to question from where such grammatical intuitions come. One very interesting way to talk about differences between intuition (which seem to arise in a natural way), versus a kind of learned methodology (which relies on declarative understanding of what makes a sentence grammatical) is to overlap a statistical methodology to the range of competency found for such +/-grammatical judgments. The methodology I have in mind here is the classic statistical averaging found across a given demographic range (γ), which measures a competency of a given skill <x>. The test is to see whether one finds the classic *bell-shape* curve (a universal staple behind any measurement of a learned skill), or if one finds the so-called *right-wall* (which portrays a biological endeavor of acquisition over learning). This dual outcome of **learned** (bell-shape curve) vs. **acquisition** (right-wall) might suggest how a *template scaffolding* (overlapping linguistic theory) could serve to illustrate our DMM:

[4] **Template scaffolding overlaps onto linguistic theory***



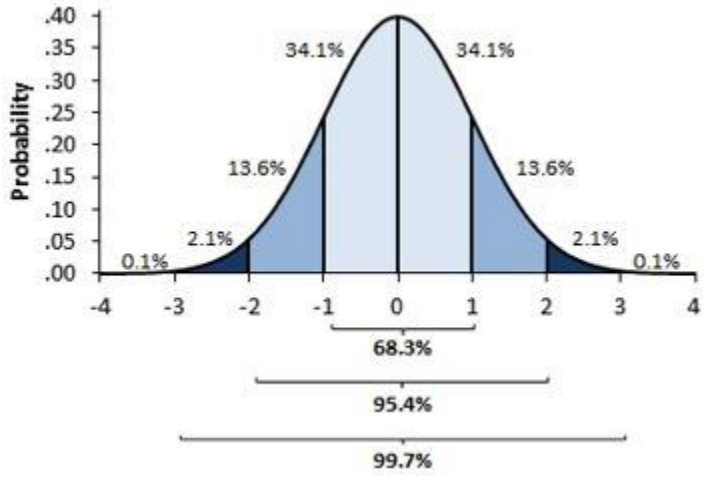
Whereas **items** extent **vertically** [x = x], **rules** spread **horizontally** [x + y =z], the former is **recurrent [ ]**, the latter **recursive [[]].** (As discussed in the Preface, this dual distinction makes-up my personal metaphor of **Items** [x-tables, y-chairs, z-nightstands] vs. **category** [α-furniture [x, y, z]]).

*(Consider such words which share semantically close stems but where the stems shift phonologically: e.g., [N glass]-[V glaze], [N grass]-[V graze] /s/>/z/, [N bath]-[V bathe] /θ/>/ð/, plus vowel shift of /æ/>/é/. Also note how irregulars such as *dream-dreamt, keep-kept, knell-knelt, dive-dove* must contain a similar phonological **sound shift** in order for the lexicon to identify the item as a new word (X=X). (Sound-shifts facilitate memorization of a new item—there is a difference between *grass* and *graze*, one is a noun-item, the other a verb-item). Also note how only a DMM could handle a certain class of words which can be both irregular and regular (both versions being accepted) at the same time: √*dive (dove or dived), √knell (knelled or knelt) √dream (dreamt or dreamed)* etc.).

So to recap, what our theory above shows (implicating a DMM as compared to an SMM) is that with such high frequency [+Freq] learning, (as with any skill which relies on brute-force memorization), what we get statistically is the bell-shape curve (below). On the other hand, when the competency level seems to reach a mastery competency across 100% of its demography, what we suggest is that such a **right-wall** is consistent with what we find of **biology**. It has long been recognized that **first language** (L1) **acquisition**, as compared to (post-critical period) **second**-**language** (L2) **learning** follows this same trajectory—with L1 biology pegged to right-wall distributions, and L2 learned skills pegged to bell-shape curves.

[5] **Bell-shape curve** (Google© 'free-to-use' image).

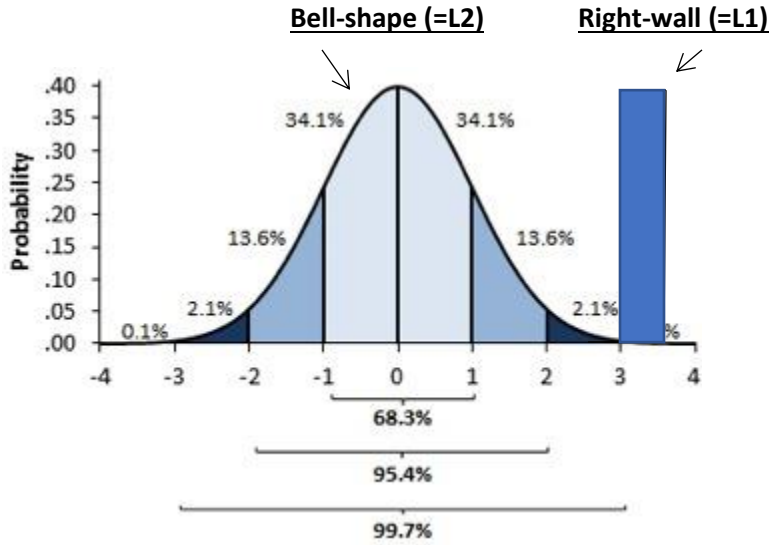
<u>Competency of a Learned Skill <X> / (L2)</u>



Whenever statistical averages of a competency of a certain Skill <x> are spread across a given demographic <γ >, what one finds is a very consistent (probabilistic) average. This average, shown above as 34.1% / 34.1% (= 68.3%) on both sloping sides of the bell, indicates the average skill set of <x>. The largest subset of 'people studies' shows an average-level of competency for skill x, and this is consistent across all skills looked at. (Note: The stableness of this bell-shape ratio comes close to what we know of the Fibonacci 'golden ratio'). The scale here -4 to +4 could be understood as *extreme incompetence* at the 'left-wall' (-4), while the **right-wall** (+4) shows a very *rare mastery*[1].

What is so intriguing about the so-called 'right-wall' <u>when it comes to a learned skill</u> is that its extreme high-level of mastery mimics what we know of any biological-bases which governs learning. In fact, it's not overt 'learning' at all, but rather a state of **biologically determined** acquisition. It is in this sense that the terms 'learned' versus 'acquisition' makes its way into the L1 vs. L2 literature— namely, L1 (*pre-critical-period* native first language) is biologically determined and so does not suffer the competency spread found of bell-shape learning, while L2 (*post-critical-period* second language learning) shows bell-shape statistics.

[6] **Bell-shape for language 'learning' vs. Right-wall for 'Biological basis' language acquisition.** (Google© 'free-to-use' image).



[1] (See Stephen Jay Gould's *Full House: The spread of Excellence from Plato to Darwin* (1996) for discussion of the 'right-wall').

What we know of *biological-based* competency distributions is that they show mastery of the acquired endeavor at mastery levels only found at the extreme right-wall—viz., of what would be considered 'the very rare extreme mastery level' of 0.1%. What right-wall mastery shows is that amongst the general population (of all biologically healthy individuals) the statistical anomaly of, say, 0.1% actually becomes the normal average. The difference here is that *learned* vs. *biologically determined* accrue very different processing costs—namely, learning a skill is a *general problem solving* skill, cognitive in manner and follows all the classical IQ-dynamic hallmarks of 'learning' (e.g., asserting oneself in such a learning environments, note-taking skills, preparation, mnemonic devices for memorization, etc., and other strategies for learning such as motivation, aptitude, as well as some physiological factors which might determine the rate and success of the attempted skill). One the other hand, *biologically-determined* acquisition accrues no cost—it comes for *free* (as part of human endowment).

So, it becomes interesting to us that the right-wall of competency only shows-up across a demographic when a biologically-based behavior is measured. This becomes important when we begin to measure linguistic intuition for first language (L1) as compared to second language (L2). Recall, that for our **four sentences**, the ability to process such recursive structure embedded in these sentence types takes on a right-wall grammatical intuition and acceptability (for L1). For instance, recall another example of the grammatical intuition that came when L1 speakers were asked 'can eagles that fly swim?' (so, what are we asking that eagles can do?). Recall, the L1 reply was 99.7% consistent across the board that what was being asked was 'if eagles can *swim'* and not *fly*. Such is a right-wall distribution on a par with any other biologically-determined processing.

It rather seems the kind of knowledge native L1 speakers bring to their L1 performance has little, if any, connection to IQ problem-solving skill capacity, motivation, or the like. In fact, it has been repeatedly shown that even low IQ children, who otherwise may suffer from general learning handicaps, seem to have their language competency completely intact and unaffected. (Even some severely mentally retarded children show little impact on their L1 language acquisition). This may be precisely because L1 is indeed innate acquisition (= biology), and not learning. Also, it becomes interesting that when the '*eagles sentence'* is presented to L1 speakers, but just visually shown to them, the innate recursive processing is not immediately made apparent to them (many students initially stumble on which is the right answer). Perhaps, this is because reading is a 'learned' processing (unlike speech) and so it doesn't necessarily map onto the internal language mode of processing. Interesting, once the L1 speaker says the sentence out-loud, and hears the construction via speech, the hidden-internal recursive mechanism becomes activated and immediately the L1 speaker instinctively knows that we are asking if 'eagles can swim' and not 'fly' (again, despite the surface-level phonology that indicates the first, closest verb as 'fly').

(Note how when sequential bell-shape curves spiral out and get spread out over a time span (evolution)—and then when something emerges along the way as some constraint or human barrier, or upper ceiling—that what we find is the horizontal spread of the bells becomes smaller and smaller (in longitude) until such a time that a right-wall develops (latitude). In a sense, the right wall is a natural outcome of a collapse of space and time, as some consequence of human capacity to statistical convergence).

L1 is biologically-determined (right wall) vs. L2 is learned (bell-shape)

Recall, that regarding the four language modes (speaking, listening, reading, writing), only the former two are natural and biologically determined which bring on the right-wall distribution of competency. The latter two are artificial skills, hence, their bell-shape curve of competency. These latter two modes, which are culture-bound and must be practiced, rely on a kind of 'frequency-effect' for its level of competency. (Such 'frequency-effect' bases of learning are altogether reliant on memorization, among other cognitive strategies).

Let's keep this dual distinction in mind when we come to discuss the intuitive grammatical judgment of Sentence 4a vs. 4b of Sentence #4 (restated here in [7]:

[7] 'I wonder [what [that is]]… up there?'

(i) [what [that is]]__
(ii) [what *[that's]__

The judgment is even more fascinating given that fact that (ii) [that's] is abundantly more prevalent in the frequency-data as compared to [that is]. Still, despite the higher frequency of [that's] as found in the Bell-shape data, the ruling against frequency and rather for structure (even when the structure goes against the frequency) suggest that a very different kind of an operating system (OS), using Artificial Intelligence (AI) terms, is being employed. (The theoretical linguist reminds us that language is indeed *structure dependent*, not frequency dependent).

The best way to test this is by simply asking a native English speaker: Which of the two utterances do you prefer, (i) that is, or (ii) that's…? The latter is overwhelmingly approved above the former, perhaps for reasons that are not at all syntactic in nature, such as simplicity, ease of speech, economy, etc. In any event, the fact that (i) is close to 100% judged as the only possible structure (for sentence #4), a fact which flies in the face of a statistical/frequency-based analysis, suggests that a larger, and rather hidden deep-state structure is active below what we find at the surface level phonology. Couple this with the notion that close to 100% of native speakers come

to the same conclusion argues against a learned, bell-curve response and rather speaks to how such a deep structure, in Chomsky terms, is indeed a biological determined a right-wall. 'Language is biology through and through'…and may not be something that can be learned, as if learning to play the piano. This demarcation of *natural* **acquisition** (pre-critical period), as found in child first language acquisition, as compared to what we know of (post-critical period) *artificial* second language **learning** allows us to see the bell-shape curve for what it really is—a probabilistic *informatique* of math and statistics which we reach when observing competency of a non-innate learned skill.

Note 2

A Note on Chomsky's (2013) *Lingua* paper 'Problems of projection'

<page 44, between examples (17) and (20)>

[1]. *The most general case of lack of label is successive-cyclic movement.* [2]. *The intermediate steps are of the form* $\{_{\alpha}$ XP, YP$\}$*, where XP can be for example a wh-phrase with YP a CP.* [3]. *The syntactic object α cannot be labeled, but it must be interpreted, if only for theta-marking.* [4]. *If XP raises, then α will be labeled Y, as required.* [5]. *Therefore XP must raise, and successive cyclic movement is forced*].

[1]. [*The most general case of lack of label is successive-cyclic movement*].

Labeling

Starting out with the theory-internal assumption that labeling comes out of a movement operation, which as a result forces **Dynamic antisymmetry** (DA) (cf. Moro 2000, Ott 2011), then, for example, two lexical items within a **set** {…} necessarily come without order. Unordered members (X, Y) make-up the set {X, Y}. So the set {X, Y} (where both are heads) retain their c-command sisterhood status and thus don't derive order. Or where {XP, YP] (where both are phrases) behave similarly without hierarchy, and don't derive order[2]. This has been sharpened by Chomsky's analysis that the theory internal and *ad hoc* configuration of Spec-Head no longer holds, and that the basic configuration is reduced to a {Head-Head} merge relation, whereby one of the two terms {H} must be complete (i.e., where its phi-features* are already drawn from the lexicon, such as with nouns or verbs) (Chomsky BEA 2001 ms. p. 12).[3]

The interest here lies in the fact that one of the two Hs must be replete with features so that it may serve as a Probe in a **probe-goal relation**. In a sense, the traditional Spec-Head has been reduced to a probe-goal relation whereby a symmetrically and unordered merged set {H, H} then

[2] Adjunction processes of {XP, YP} which seem to show asymmetry (thought which still must force an interpretation) have been discussed in the literature (e.g., see BEA p.18).
[3] BEA 'Beyond Explanatory Adequacy' ms. *MIT working papers*. May 2001. *Interpretable n, v-roots act as 'completed' Heads while uninterpretable Case and Phi features act as probes.

becomes an asymmetrical (ordered) pair <H, XP>[4] and where the modified H (Head) seeks out a goal within XP. The question is how does this asymmetry establish itself?

[1a] Turning to the recent incarnation of the minimalist program (MP) (Chomsky 1995), perhaps the most essential property which has come out is the notion that all *syntactic objects* (SO) are the result of a (re)combinatory operation called **Merge**, whereby SO = {X, Y} is the mere result of an unordered binary merge of two items: SO = the unordered members of (X, Y). In this 'first-instance' merge, SO renders two lexical items (X, Y) as an unordered set {X, Y} (where {X, Y} = {Y, X}). Call this **External merge** (EM) since X remains independent of Y (they are not part of one another), and as defined by set {…} as being unordered, there is no intrinsic order to (X, Y).

But this is not 'Language' as there is yet no hierarchy/order within the set—it still being a primitive membership composed of nothing other than 'flat' symmetric sister relations (in logic, what we call 'logical-and' Λ, as in the Λ-expression: 'I need to buy a and b and c and d' where set{a, b, c, d} are unordered). Hence, this 'first-instance' EM of the set {X, Y} doesn't necessarily give us syntax (but merely a string of items, a lexicon)—viz., there is no hierarchy of the kind required of syntax.

[1b] At the very minimum, what we need as a prerequisite for syntax is an ordered pair <X, Y> (where X precedes Y). This renders the unordered {set} now as an ordered <pair> (see fn. 40). Specifically, the problem with first-instance merge (EM) is that SO is yet to be labeled: viz., *there is a labeling problem*. 'Labeling of {H, XP} requires that the **Head** (H) not be of the form <X, Y>' (Chomsky, *Lingua* paper p. 46). In other words, SO cannot emerge as a simpleton instance of an arrangement of two heads <X, Y>. Some form of **Labeling** must ensure. In order for SO to be labeled, a 'second-instance' merge operation must take place creating **Displacement** (hence, labeling comes via displacement). While displacement yields a hierarchical expression, it does so in a unique way as specified by the theory: 'Labeling differs from other notions in that it is not virtually detectable by direct inspection of expressions' (Chomsky, *Lingua* 37). In other words, labeling via a movement operation is theory-internal.

[4] Using MP terminology, a **membership set** {X, Y} comes freely unordered, while a **pair** <X, Y> is necessarily ordered. When movement/raising is shown of an element within a set (DA), we can speak of an ordered **recursive set** as {X {X, Y}}. (See Galasso 2016 for full account of Merge as related to the development of child syntax).

So, a theory-internal operation is needed that takes us beyond a mere grouping of items of a (first) **membership set** {X, Y} (say, of two equal heads) to a (second) **recursive set** {X, {~~X~~, Y}} (or pair <X, Y>) which establishes an ordered syntax—noting that {X} of {X, Y} raises and gets displaced from its original membership set, this creating hierarchy. Call this displacement operation **Internal merge** (IM) of **H** since one of the two Heads (H) displaced must leave a **Copy** of itself behind. This forms a recursive structure [SO = $\{X_j, \{X_j, Y\}\}$][5]. At this point, the SO can be labeled by the raising of the H {X} of the **Phrase** (P) [XP {X {X, Y}}]. Ordering as seen within sets {X, Y} has now emerged as an ordered pair <X, Y> (x comes first, then y).

[2] [*The intermediate steps are of the form* $\{_\alpha XP, YP\}$*, where XP can be for example a wh-phrase with YP a CP*].

In this phase of derivation, $\{_\alpha XP, YP\}$ again are sisters, and properly make-up the equivalent membership set of (XP, YP) with no order. Both are potential (independent) projections which must then get later defined based on the actual (spell-out) projections of their contained lexical features. So, to suppose that XP is a 'wh-phrase' and that YP is a CP (which houses the wh-phrase) is tantamount to saying the 'projection' is one thing, and 'that which projects' is another—though the two are intertwined as specified by what type of P can project what kind of X (where X = H(ead), and XP is P(hrase) headed by X. Phrases are projections of Heads as Heads are bundle of features). For example, Chomsky addresses the fact wh-expressions can also remain *in-situ* without raising (and therefore may in fact not constitute CP), as in the example: *They thought JFK was assassinated in which Texas city?* (so-called 'quiz-show' structures) (p 44).

But the problem here is about projection of both [wh-expression & CP] when they form part of a raised constituency, as is found within an intermediate step along successive cycling: viz., if both {XP, YP} are sisters of a set at some intermediate step in the derivation, then how does labeling generate necessary projection? How does the SO get labeled? In order for a P to project, it must be defined by a set of features specific to its H. So, a potential symmetry of {[[wh-Q]…],

[5] Chomsky tends to distinguish the site of the labeling of a syntactic object (SO) as either being placed at some point after movement (IM) (as with successive-cycling movement), or at the place of a merge-based/base-generated point of the derivation (EM)—with the former (IM) XP bearing the subscript {β}, and the latter (EM) {α}. E.g., where raising of XP is involved and where a copy of itself is found in the lower structure, the use of β is employed $\{_\beta XP, YP\}$ = [XP [copula [$\{_\beta XP, YP\}$]]], as opposed to $[_\alpha N\ TP]$ where NP is the product of simple merge. Take {α…β} to form a chain of [SO {α…β}] (or {γ.. {α.. {β}}} that spans three domains). For example, $\{_\alpha NP_i, \{T, \{_\beta NP_i, \{v, VP\}\}\}\}$, the highest copy/instance of the NP is in the domain of the entire SO that is labeled α, since every instance of the NP is within this domain. The lower NP instance, however, is not in the domain of β since not every instance of this NP is within this domain.

& [[CP]…]} = {XP, YP} cannot stand. One of the two Ps must raise in order to form DA (as discussed citing Moro, Ott). One question here is why shouldn't YP raise if both {XP, YP} are equal-distant sisters. The raising conditions behind DA should allow either of the two terms to raise. But it seems only XP (the wh—phrase) raises. (Chomsky expresses this concern in his footnote 36: 'One may ask why YP doesn't raise'). Perhaps the notion of 'computation atom' expressed within the lexical item (LI) has a determining factor behind why the XP (=wh-phrase) of the set {XP, YP} raises leaving {YP} (=CP) without recourse of IM. The CP may be a vacuous phrase (without a head) unlike the wh-expression which is headed—creating an inherent DA within the set even before potential IM takes place for labeling of SO.

Chomsky cites Moro here regarding copular/Small Clause (SC) constructions—e.g., [be [lightning, the cause of fire]] where one of the two terms of the SC must raise so that the SO can be labeled. Hence, DA comes out of a need to label the SO projection.

A very nice example of this is what we find in otherwise symmetric {X, Y} configurations. For example, consider the SO to be of the two lexical items, the membership set: {boat, house}. As it stands, the SO cannot be labeled. What is needed is **raising** thus providing a syntactic hierarchy via a DA-set. (Raising thus renders an unordered membership set into a recursive set). As Moro claims, one of the two terms must raise. If {boat} raises from {X, Y}, we get {boat, {boat, house}} = {$_{\beta}$X, {X, Y}} where the lower X copy is invisible to pronunciation, it being a discontinuous element. Hence, the SO gets labeled and projects as [boat [~~boat~~ house]] (and where hierarchical syntax is generated that allows us to interpret 'boat-house' as a kind of 'house' and not a kind of 'boat'). Such N-compound structures are likened to Adjectival Phrases where the H of the P labels the projection—e.g., [AdjP [Adj black] [N bird]] whereby 'black' is the Head of the phrase, so that the interpretation is that a 'black-bird' is a kind of 'bird' and not a kind of 'black'.

[3] *The syntactic object α cannot be labeled, but it must be interpreted, if only for theta-marking.*

[4] *If XP raises, then α will be labeled Y, as required*

One of the questions regarding 'performed operations on syntax' is essentially 'how does the algorithm operation work? (Chomsky refers to the operation as a labeling algorithm (LA). The problem of labeling is inherent in language, bottom- up vs. top-down solutions have been jousted from the very conception of the problem: viz., (i) in order to label a word, it must be embedded in a top-down structure (language is structure-dependent), (ii) in order to generate a structure, there must be a well-defined status of word. (The catch-22 is similar to what we find in our

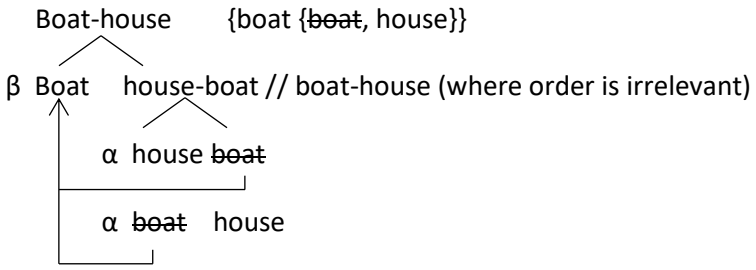
approximate understanding of the way genetics works: proteins create amino acids and amino acids create proteins, etc.).

Here is how the problem is stated. A labeling algorithm (LA) must scan a syntactic object (SO) and provide a label for one of the two (within a binary selection). So, suppose SO = {H, XP} (H = head, XP= non-head) is rendered via the **external merge** (EM) of two items. Then LA will select H as the label. In other words, the feature specificity for labeling is already encoded in the pair, and projection is straightforward since the H is visible to the LA—e.g., H = verb, and XP = noun complement, [VP [V, NP]]. But what if LA scans two sisters {X, Y}, or {XP, YP}, where both terms are non-heads? Here, the search is ambiguous. It is in this case, when the label {H} is not made available in the search, that the MOVE-property of raising (**internal merge** (IM)) must establish an hierarchical configuration, breaking symmetry. So, consider {XP, YP} as a mere membership set. In order to break with symmetric sister-relations, (since both reside within a flat structure), the raising of one of the two terms must proceed, yielding {XP, {XP, YP}} => [XP [~~XP~~, YP]]. A pair is created with derived syntax (order) <XP, YP>.

Antisymmetry in Compounds

Consider {X, Y} as two lexical items: e.g., the twin-noun formation of [N, N] leading to the compound {boat, house} [N boat-house]. In order for LA to label one of the two SOs as the (adjective-like) **modifier** (M) (in marking the compound 'boat-house' vs. 'house-boat') one of the two SOs must raise in order to break the symmetric reading. In other words, the labeling {N, N} would force one of the two Ns to raise thus allowing the M to project, and as a result, allow the H to be labeled accordingly: {N {~~N~~, N}}, where the moved item acts in an adjectival M capacity and where the residual unmoved item becomes H by default. As we see, what is different about lexical items/*compounds* as compared to English *Phrase* types is that Heads of an **English Phrase** are **Head initia**l e.g., [VP V [~~V~~, N]], whereas Heads of English lexical **compounds** are **Head-final** (as opposed to Spanish compounds where they are **Head-initial**—e.g., *comi-ratas* (eater-rat) (= rat-eater). In the examples below regarding 'boat-house' what needs to be selected is not H, but rather the modifier (M) of H. The item which stays in place, not a result of MOVE, will thus be labeled as Head. The item which raises will become the Modifier of H.

[5] boat-house (a kind of house)



Within the flat structure [5α], order is symmetric and irrelevant. MOVE (IM) (raising) is a unique displacement property which allows for recursive structures found in [5β] above, thus providing a mechanism for labeling the head of a phrase. A 'boat-house' is a kind of 'house' (and not a kind of 'boat'), where order becomes crucial to the compound's reading.

Where X modifies the head—This same formation is exactly what happens in Adjective Phrase (AdjP) structures: So, compound antisymmetry looks something like the following {X {~~X~~, Y}} where {X} gets labeled as M, and as a result, where in-situ Y takes on a head status. In the N-compound 'house-boat', the H of the compound is the N 'boat' (= a kind of boat).

[6] red car.

Where the two lexical items {X, Y} [X, Y] = [red, car]. In this formation, no word order is realized since both items are sisters and are thus ambiguous in their reading. It is not until some raising takes place via IM that order can be assumed: [red, car] > [red [~~red~~, car]] = [AdjP red [~~red~~, car]][6].

***Compounds: Root vs. Synthetic* (cf. §1 [15]).**

[7] a. coffee-maker = (maker of coffee) => [coffee-[maker of ~~coffee~~]]

cigarette smoker = (smoker of cigarettes)

=> [cigarette-[smoker of ~~cigarettes~~]]

b. chain-smoker (not *smoker of chains)=> *[chain-[smoker of ~~chains~~]]

[6] We may freely substitute {}-brace brackets with []-square brackets here as the latter is typically used in syntactic tree diagramming e.g., [VP [V, N]].

Where examples in [7a] (referred to as **Synthetic compounds**) require *double merge* via displacement: [coffee-[maker of ~~coffee~~]].

And where example [7b] (referred to as **Root compounds**) is the result of a *single merge*: [chain-smoker]

In [7b], the adjectival root is not derived via displacement (notwithstanding the fact that some first-order displacement had to have been initially carried out in order to derive the order {chain, smoker}, since the two items must derive an ordered pair: we don't equally get *{smoker, chain}. Hence, root compounds begin their journey with at least a first-instance merge of a recursive set, but they do not go beyond that in forming further displacement up the tree.[7]

In sum, it does seem that at least some amount if movement is involved with the formation of all types of compounds. For one, the placement of the Head of the compound follows similar parameterization constraints which are found in syntax (e.g., the Head Parameter [+/- Head initial]). As well, the fact as shown above that *synthetic compounds* are the result of MOVE (which contrasts with *root compounds* which are not derived via MOVE) suggests that displacement in language is of a morphosyntactic nature.

(See web-link no. 27. for movement in compounds).

(See https://www.academia.edu/34403441/Working_Papers_4).

[7] But this is based upon a theory-internal stipulation which states that 'everything must move at least once in order to become visible as part of labeling'. Accepting this, there still remains the basic assumption that root-compounds are indeed base-generated and are not derived via movement. In Galasso (2016, p. 79), drawing on the '*everything must move at least once*' stipulation, this first-instance of move—as shown in root compounds, as well as all phrases – is referred to as **first-merge (Move-1)**. A reference to **second-merge (Move-2)** then is made for subsequent second-level raisings which yields syntactic compounds.

Dynamic Antisymmetry

Following a version of Andrea Moro's dynamic antisymmetry (DA) (as presented in Chomsky's 2013 *Lingua* paper), let's consider that we have reached a point in the derivation where [XP, YP] have already been formulated, leavening us with only two Heads {if, how} to contend with, both having identical labeling. What DA stipulates is that in order for one of the terms [XP, YP] to be labeled and associated with the right head, one of the two [XP, YP] terms must raise.

So, if Syntactic Object (SO) = {XP, YP} where both Ps are identical in terms of labeling (both are equal phrases), in order for labeling>projection to be properly headed, one of the two terms (X, Y) would have to raise, thus breaking a symmetrical flat-sister relation.

Consider the [copular-small clause] structure below, where the small clause is of the form [XP, YP] (cited in Chomsky 2013 p. 43 (taken from Andrea Moro's work on *dynamic antisymmetry*)):

[copular (H) [small clause {XP, YP}]] , SO = [XP, YP]]

[8] [be [lightning, the cause of the fire]]

It seems the structure above would yield an ambiguous result if search were applied as is. In other words, in order to label SO (small clause), one of the two terms must raise, breaking its symmetric/sister relation. In this projection, the copular 'Be' triggers one of the two terms of the small clause to raise from out of the identical [XP, YP] labeling:

[8'] a. [XP lightning *is* [XP ~~lightning~~, YP the cause of the fire]]

a. [YP the cause of the fire *is* [XP lightning, ~~YP the cause of the fire~~]]

Another salient example to this structure would be the even more stark case of how one of the two terms in an otherwise identical [XP, YP] has to raise.

Consider the structure below:

[9] a. **If** *you keep eating ice-cream*, **how** *can you lose weight*?

b. How can you lose weight if you keep eating ice-cream?

The two heads (H) involved in the labeling of [XP, YP] are [H {how, if}]. Where XP, YP are given:

[XP you keep eating, YP can you lose weight]]

Noting here how both clauses {XP, YP} can be inverted as long as they remain headed by their proper **polarity-item** head (H)—in this case, the H 'if' triggers the [XP you keep eating ice-cream], while the H 'how' triggers the [YP can you lose weight]. Here's how antisymmetric raising might work in such examples (which show an inverted [YP, XP] for illustrative purposes only, since the order of XP, YP is irrelevant within its flat sister-relation [X, Y]:

[10] [[XP you keep eating] [YP can you lose weight]] (where both terms [XP, YP] are identical in label).

Notice their *symmetric* quality: (how can you lose weight if you keep eating? > if you keep eating, how can you lose weight?)

Now, what must happen to break this symmetry is that the head must select one of the two terms [XP, YP] to raise/MOVE, thus allowing for the proper labeling of the phrase—where the two Hs are {H {**if**}} and {H {**how**}}.

(a) [**if, how** [XP you keep eating, YP can you lose weight]]

(a') [**if** [XP you keep eating **how** ~~[XP you keep eating~~, YP can you lose weight]]]

(b) [**how, if** [XP you keep eating, YP can you lose weight]]

(b') [**how** [can you lose weight] **if** [XP you keep eating, ~~YP can you lose weight~~]]].

Of course, we quickly notice that the order of the Heads {how, if} too is irrelevant, with the same projections that follow:

(c) [**how, if**] [XP you keep eating, YP can you lose weight]]

(c') [**how** [YP can you lose weight] **if** [XP you keep eating, ~~YP can you lose weight~~]]

So, it appears based on our discussions of dynamic antisymmetry DA so far, that it is the head (H) which defines and projects all syntactic objects (SO). All relevant information about an SO, a lexical expression, will be provided by a single designate element within it—call it a 'computational atom', which defines the H. So, what is the triggering mechanism involved? It seems the H 'if' has encoded in its lexical head a collection of *computational atoms* (i.e., features) that requires 'if' as a polarity item to trigger raising of only the declarative [XP you keep eating]=> [if [XP you keep eating]], while preventing the projection of *[if [can you lose?]] (where only the H interrogative item 'how' can trigger an auxiliary inversion derived from the lower matrix phrase [<u>how</u> [YP <u>can</u> you…?]] ([_ _ [you can how?]].

Connected to this, consider how one might parse an already formulated expression such as the now infamous Pink Floyd saying (song 'Another brick in the wall' from the album 'The Wall'). Imagine there are two ways to say the saying, of course, depending on which H you select first:

> *<If you don't eat your meat, you can't have any pudding! How can you have any pudding if you don't eat your meat?>*

Again, notice how each polarity expression Head (if, how) selects its own complement phrase via antisymmetric raising. Consider how DA might go about determining the sequence of the Pink Floyd saying. Recall, at this point, the two Hs in question are symmetric (due to their sister relations) and either one or the other could be selected to start the complex sentence: {H, H}, or {if, how // how, if}.

So, the question we turn to now is to ask: What is the mechanism involved whereby the H selects its appropriate XP?

Head as Computational Atoms

What Chomsky suggests here is that each specific H carries a 'collection of atoms', minimal elements of the H which enter into a computation—'relevant information about SO, a lexical item, will be provided by a single designated element within it, call it a 'computation atom' which defines H. Hence, a lexicon is a class of atoms to be computed (p. 41). Let's consider that the two heads {H, H} [if, how] carry their own special properties of atoms which engage in a computation. Suppose the H {how} carries an anaphoric copy of itself which must be found somewhere within the search of [YP], as in [How…copy]. So, one such atomic feature of {how} can be expressed as follows:

[11] $\{how_j \ldots copy_j\}$ => [how……[if…..how]]

(along with other traditional features which go into the projection of the wh-element {how}, such as {+interrogative}, {+ aux inversion}, {+polarity express}, {+adverbial/manner}, {CP-projection}, etc.).

Consider below the search mechanism for {how$_j$….copy$_j$} => [how……[if…..how]]:

[12] **{how}**

[if, **how**$_j$] search => [XP [you ~~can~~ have any pudding ~~how~~ $_j$]]

How => [XP can [you __ have any pudding___]]

How => *[YP you don't eat you meat]

***how** you don't eat your meat **if** you can have any pudding.

The H {how} must select appropriate XP {how$_j$….copy$_j$}.

Consider now {if}

[13] **{if}**

[**if**, how] search=> [YP you don't eat your meat]

If => [YP you don't eat your meat]

*If => *[XP can you ~~can~~ have any pudding ~~how~~ $_j$]]

In this case, the H {if} can't select an XP which contains the anaphoric copy {how} found within the scope of search, therefore, the XP becomes what is referred to as *discontinuous* and is not visible to labeling, discontinuous due to it already being either selected by H 'how', or by the fact that the most prominent feature [copy] can't enter into an Agreement (AGR) checking relation of {how$_j$….copy$_j$}.

So, what we are supposing is that since the H {if} carries no such anaphoric copy, but rather only features associated with H of C (e.g, the polarity condition of (if> then), as well as CP-projection), then part of the LA search becomes modified in order to secure SO labeling. What we can say is that while part of the formal checking requirements of {how} must search for a copy

of itself, the computational atoms for H {if} require no such copy. It is in this sense that DA might be applied:

{if, how} [XP, YP] (where [XP, YP] are sister relations in the same sense that Moro talks about his small clause formations).

Notice again how the order doesn't apply to sequence of {if, how}. Since the XP, YP is already formulated (as part of knowing the Pink Floyd phrase), the only selection that needs to be made is which H you select in order to trigger the appropriate XP/YP raising:

[14] a. [*If, _ how* [*XP can you have any pudding [YP you don't eat your meat,*]]

b. [*If* [*YP you don't eat your meat,*] *how* [*XP can you have any pudding [YP* ~~*you don't eat your*~~ ~~*meat*~~]]

In sum: H {How} searches for a copy of itself, finds it in XP, then proceeds to raise XP. {If} doesn't raise XP because the copy of {how} would go unchecked. Hence {if} is rather forced to raise the other sister relation. It is in this case that {if} is less restricted than {how}, and where {how} must restrict itself to search out an anaphoric copy within the matrix clause. In other words, {how} requires less search ambiguity over {if}. Heads {if, then} (cited below) however would seem to be more on a par with each other (since neither contain a copy of itself in XP). What then drives a particular labeling perhaps has to do with logic.

[15] What of the Hs {if, then}?

The [if…then] checking sequence would work as follows:

e.g., **If** you eat your meat, **then** you can have pudding.

But notice how the two heads can't seem to be reordered:

*then you can have pudding, if you eat your meat.

This would mean that the feature attributed to the H {if} requires it to be an antecedent (in first position) to any anaphoric expression [if…then] => [if..[then]], where the two Heads are not symmetrical/identical in nature (they are not sisters). This is different from what we saw with the

Hs {if, how} where both were symmetric sister relations [if, how] (though where {how} contained an anaphoric feature {+copy}.

e.g., [if, [then]] [XP you eat your meat, YP you can have pudding]

It may be understood that the H {then} of the {if, {then}} sequence searches for the prominent element {can} within XP, where {can} contains a conditional element related to the lexical item {then}.

If X, then you can Y: {if X, then Y}. => logical structure

It was this same kind of underlying thematic/logical structure which motivates Move to take place on a Semantic/Thematic level. Recalling that Move/displacement up until now has had two motivations:

(i) syntactic (S), and (ii) phonological (PF).

Well, it may be the case that Move also caters to two sub-types of S structure: (i) syntactic (surface-level, PF) and (ii) thematic/semantic (underlying logical form, LF).

[16] ***Reasons for Move/displacement:***

(i) Phonological (PF)

(ii) Syntax (S)

a. Syntactic (CP)

b. Semantic/Thematic (vP)*

---

*Note. See the end section of this note 'Family of Merge' for a brief discussion of the **Duality of Semantics**. Following Miyagawa (2000), we consider the two Phases of **CP**, and the light verb **vP** to be the phases which serve this duality of semantics—viz., with the highest *functional* projection of CP dealing within the probe-goal relation of AGReement, and the lower *lexical* projection of vP dealing within the probe-goal relation Case. We'll come to consider Case as being somewhat lexical/semantic in nature, presumably loosely associated with theta marking, while we'll consider AGR

as being quintessential formal in nature. (Recall that in a phase-base theory (Chomsky 2008), TP is not considered a phase. But we may entertain the notion that a T-feature [T] may either serve as a featural {F} or Affixal {Af} adjunct which can adjoin to either TP or CP (along with affix lowering onto the V), and that as T(ense)/TP is not phasal in nature, the T-feature itself may be free to *percolate* up or down the syntactic tree). We'll propose, loosely following e.g., Miyagawa (2010) and Radford (2009, 2016) in a number of respects, that the duality of semantics, as pegged to Chomsky's notion of Phase-theory (where only CP & vP are considered phases), that the light verb **vP assigns Case** (Case being lexical semantic in nature) and **CP assigns AGR**.

Let's consider below this duality with respect to 'reasons for MOVE': semantic or syntactic?

Semantic Move.

Consider Chomsky's (2013) remarks below:

[17] a. Which books did John read => S/PF

b. 'For which books X, John reads books X' => semantic/thematic.

One of the reasons for Move/displacement here seems to be to get the semantic roles correctly assigned. For example, here the XP {which books} has two semantic roles: it receives its role as object of {read} (as in the expression 'read books'), and it also serves as a distinct interrogative operator, binding the variable in the object position, so that the interpretation is something like 'for which books X, John reads books X'. Showing base-generated structure, then movement, the structure looks like this:

b. ['which books did [John read ~~which books~~]]?

In addition to two types of Move: *Internal* & *External* Merge, Chomsky identifies two types of merge: *Copy* & *Repetition.* Consider the **repetition-structure** of the phrase 'What hit what'?

[18] What hit what?

It seems that the lexical item 'what' takes on two independent arguments {X, Y}, they are repetitions of the PF spell-out, but both are independent: consider the argument structure:

a. [HIT [x, y}]] = X hit Y…e.g., 'Boy hit ball'

Compare (a) above to the **copy**-**structure** below:

[19] What was hit (what)? (John hit what?)

In [19], 'what' is a copy derived by [__ was hit what] (=~~What~~ was hit what?) with Internal Merge (IM) raising of 'what' to the surface subject position. Consider the underlying structure:

a. The ball was hit
   i. __ hit the ball. (John hit the ball).

In terms of IM, the *raising of a copy* seems to be a **syntactic effect,** whereas *raising of repetition* seems to be a **semantic effect.** This same semantic/thematic vs. syntax split regarding merge can be seen in:

b. Which books did John read?

Here in (b), 'which' takes on two independent semantic roles (via repetition):

(i) Object of 'read'
(ii) Interrogative operator

For 'Which books X', 'John read books X'.

A further complication might actually be suggested that Copy (IM) serves a two-fold operation:

[20] (i) can serve **semantics** when dealing with **Case** (argument/thematic marking), as shown in (b) above, and/or when dealing with a light verb (vP) projection…, or

(i) It can also serve **syntax** when dealing with **Agreement,** when dealing with a CP projection.

Hence, what we get out of Internal Merge (IM) is a potential **double**-**merge** operation of (i) semantics, followed by (ii) syntax.

Copy as IM:

(i) Semantic (vP): Case

(ii) Syntax (CP): AGReement

In sum, what the above dual treatment of merge (IM) details is that while Repetition is triggered by a single-merge (exclusively for semantics), Copy may trigger double-merge (both for semantics (vP-Case), and for syntax (CP-AGR).

(See Miyagawa (2010) for such an account of a Case/AGR double-merge. In fact Miyagawa goes even further and suggests that the entirety of Move is based solely on AGR).

Syntactic Move.

Let's follow-up on Move as having a syntactic effect. The quintessential notion of syntactic move has to do with Agreement (AGR) (following Miyagawa here)—which is a movement-based operation triggered for no other apparent reason other than to check off a formal AGR feature. This is what we find in most instances where the higher non-argument position of CP is involved.

AGR-based Move.

T(ense) only carries the set of (person, number) AGR features in a clause (XP) where T is selected by C of a (C(omplementizer Phrase) (Radford 2009, p. 397).

What this means is that the Head C (as part of its computational atoms) carries and 'hands-over' AGR features from H of C to H of T.

[21] (a) I am hoping [CP [for] [TP him to win]] [-Tense/-Case]

(b) I am hoping [CP [C that] [TP he wins]] [+Tense/+Case]

Let's flesh such a treatment out in terms of DA and H computational atoms as discussed above.

Suppose the two Heads are {H, {for, that}}, and the {XP, YP} are [TP {him to win}] and [YP {he wins}].

What the search reveals in labeling the SO is that the H {for}, carries no Tense/AGR features (from C) and so only the default non-finite verbs along with the default accusative case get selected. In (b), it is the well-formed H of CP {that} which selects and labels for AGR (Tense/Nominative Case). What we have regarding the labeling of XP in terms of the H is precisely the same type of selection as discussed in our treatment of dynamic asymmetry:

[22] {for, that} [XP him to win, YP he wins]

Both {XP, YP} are identical. So in order to label the SO, one of the two terms must be modified by raising.

[for, that} [~~him to win~~, he wins]

(i) [for [him to win]] => search [<u>for him to win</u> [XP ~~him to win~~, (YP he wins)]]

(ii) [that [he wins]] => search [<u>that he wins</u> [XP (him to win),YP ~~he wins~~]]

H {for} carries computational/atomic features [-T, -Case] => [-AGR]

H {that} carries [+T, +Case] => [+ AGR].

Dynamic Antisymmetry & Recursiveness in Possessives {'s}, {of}

John's book vs. Book of John's

Consider the structure {John {Poss, book}}, [John [Poss, book]]. Here John [J] can raise either of the two items [X, Y] (X Poss, Y book]. If J selects (via search) {Poss}, leaving {book} in place, we get the underlying structure [John ['s book]] => [John's [_ book]] where clitic {'s} raises and attaches to the stem (a result of DA movement due to PF considerations, (as presented above in §1 [2], but see §4 [10] where PF clitics don't apply).

If, however, J selects to raise {book}, leaving {Poss} in place (so that it remains frozen in-situ and therefore can't become a clitic), we then get [book of [John ['s, ~~book~~]]. The double possessive elements found in expressions such as *I am a friend of John's* can be traced back to how the H John selects either *Poss* or *Book* to raise, following similar steps as laid out in our discussion of Dynamic Antisymmetry which shows DA merge as [H [X, Y]].

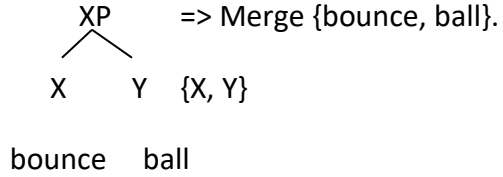
Dynamic Antisymmetry of Internal Merge: Closing Remarks

Various approaches have pinned DA to models which stipulate that 'every element must move at least once' in order to be visible to a labeling algorithm (LA). This reduces to meaning that every right-branching structure must end in a trace (See Kayne (1994). Hence, complements (object DPs) must vacant their base-generated positions. Added to this movement stipulation is that subjects (DP) too must move out of VP, coupled with other language-specific notions determining whether or not a verb must raise to H of T, (or if T affix lowering down to H of V is required, as in English). All in all, the notion of movement, as it entails all the aforementioned operations, is without doubt the crucial mechanism at play in determining fully-fledged human language capacity.

A Note: Family of Merge

As part of our closing remarks, let's recap how the step-sequences of Merge (both internal and external) come together to form a fully articulated syntactic tree, particularly focusing on **Case** and **Agreement**, (or semantics & syntactic, which make-up the so-called **duality of semantics**):

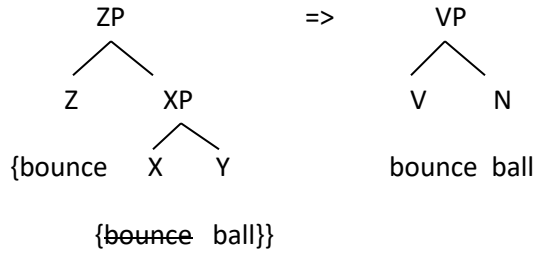
(1) Step-1: External Merge (EM)



At step-1, there is no hierarchical word order. XP could yield either 'ball bounce//bounce ball'. Step-1 is exclusively semantic, whereby two items have simply been pulled from out of the lexicon.

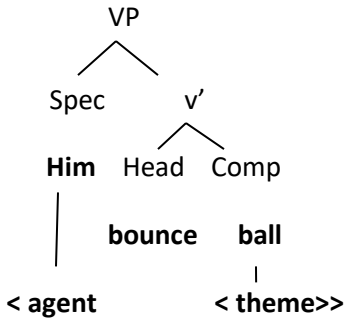
In order to create a hierarchical syntactic phrase, Move is required (our step-2):

(2) Step-2: Internal Merge (IM):



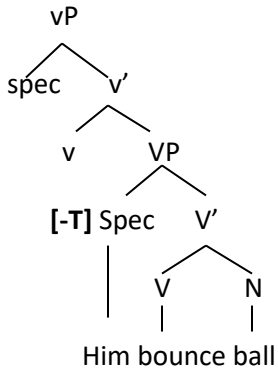
At step-2, we get Internal Merge (IM) (with a non-pronounced copied movement of 'bounce') which now allows order via dynamic asymmetry (DA). It is here that we arrive at a properly formed Spec-Head-Comp. This is a prosaic lexical/semantic projection still without Case, Tense or Agreement.

(2') Spec-Head-Comp configuration (Theta-marking):



Theta marking is a semantic assignment.

(3) Step-3: Light Verb (vP)



The [-T] tense feature may assign the sort of tense we find with [–Finite] imperatives, subjunctives.

Note no Case, AGR at this step.

At step-3 above, what this next merge projection allows is an added available Spec position for the subject of the lower VP to raise into. Subject/Spec of VP, once inserted into Spec vP, gets assigned Case [+Nominative]. Light verbs can assign Case, since vP is semantic in nature and Case is a residual effect of semantics. (vP is said to straddle both lexical and functional projections. Unlike VP which can only map a theta-grid, the light verb vP projection takes the derivation one step further in securing Case. While light verbs such as *make*, *do*, project alongside the main verb e.g., 'John *makes* roll the ball' (= John ø rolls the ball), such light verbs, when stranded within vP, can't project Tense). Light verb formations arguable give us the antiquated **subjunctive** mood (which particularly contrasts with the **indicative**) whereby *nominative Case is assigned without finite Tense*—e.g., the subjunctive sentence: '[I'd suggest that [**he study** for the exams]]') and not (*'He studies…').

(We'd suggest that subjunctives have a fully functioning vP with a somewhat defective TP, whereby case gets assigned by the above [featural T-feature] within T, but where the [affixal-T] feature of the Head is unspecified.)[8]

But before Case can be assigned, it must be handed-over by a [+T] feature (Case is marked via Tense—[+Finite] tense marks [+Nom] case, [-Fin] tense marks [-Nom] case (see below).

Case marking via Tense:

(i) Finite tense assigns [+Nominative] case:

e.g., I think [<u>he walks</u> often] : He => walks

(ii) Non-Finite tense assigns [–Nom]/Accusative case:

e.g., I saw [<u>him walk</u> often]: Him => walk

Since vP can't mark for [+Tense], in order for vP to assign case, a TP must project whose Head projects finite tense. Consider the next step-4 below:

[8] One distinction that can be made between featural vs. affixal is that featural may encode tense as incorporated in the verb stem itself, perhaps as in French, (or English auxiliary and irregular verbs), whereas the affixal-feature is a decomposition of stem+affix.

(4) Step-4: Tense Phrase

(i) TP — Featural-T 'hands-over' case marking feature onto spec vP.

Spec T' — [+T] is an adjunctive in nature and may attach to any phrase.

[+T] vP — Pronoun 'he' enters into a Case-marking configuration with T.

spec v' — 'him' receives [+Nom] case 'He' via T marking.

he v VP

Spec V'

V N

[-T]

~~Him~~ bounce ball

Tense treelet structure: (past tense).

When Tense projects [+Past] {ed}, along with Case marking, we now find affixal-T lowering, as [Stem + [affix]].

(ii) T' (with light verb vP not shown)

[+T/Aff] VP

[{ed}] V — Affixal Tense{ed} lowering.

[bounce]ed]

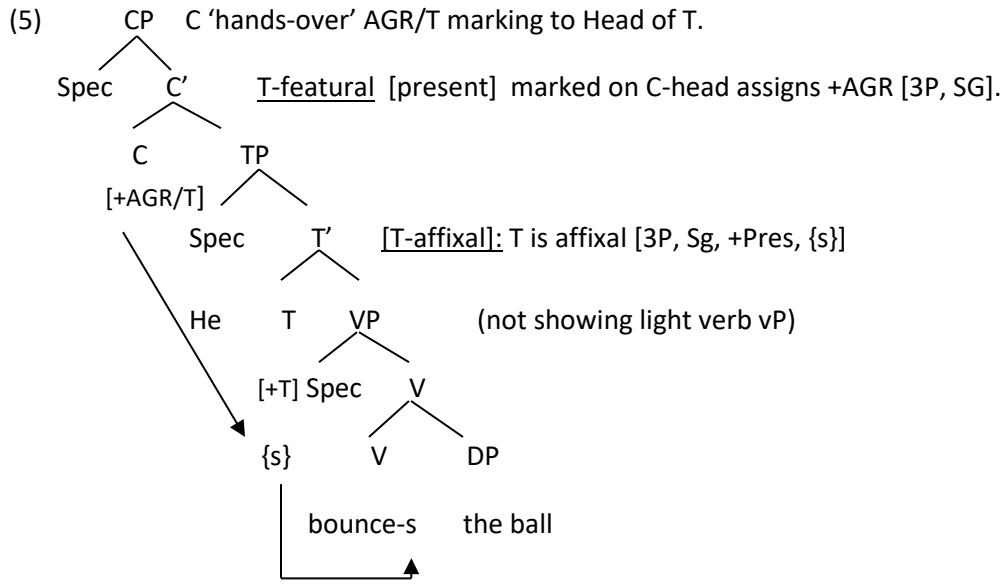
Once Case and Tense are projected via vP, TP respectively, the question now turns to the more formal projection of AGReement (since the two lexical/semantic assignments have both been satisfied). Following some of the remarks made in Radford (2009) about AGR being assigned via C-head of CP, as well as his more recent remarks that C-head may in fact house Tense (with the T-feature assigning AGR down to TP), we similarly suggest the extension, that, like what we find of **wh-subject constructs** (e.g., 'Who found it'?), all declarative TP projections are in fact fully

extended CP projections. By proposing that CP assigns AGR, coupled with vP which assigns Case, we arrive at the **duality of semantics** with the added feature that a **phase**-**base** theory (MP) also incorporates these two phases: CP & vP—with (i) CP being formal/syntactic in nature, and (ii) vP being theta/semantic in nature.

The duality this way is captured by the lexical vP vs. functional CP split, with Case being a residual lexical phenomena and AGR being quintessential formal. In fact, Miyagawa (2000) goes even further by speculating that it is AGR which is solely responsible for MOVE.

Let's consider the last step showing a full CP-articulated structure:

'He bounces the ball' (showing 3P, singular {s} (AGR)):



Subject Agreement (Radford 2016, p. 338):

When T agrees with its Subject/Spec:

(i) The person/number features of the subject are copied onto the Aux/affix T. (We are adding that the AGR projection is instigated by C).
(ii) The tense feature [T] on T is copied onto the subject.

With this treatment, CP 'Wh-subject' constructs of the 'Who bounces it?' type (where the wh-word 'who' must raise to spec of CP) mimic what we find with declarative TPs ('He bounces it'): both are full CP projections since, in this analysis, the AGR of the verb e.g., 'bounces' (of 'He bounces it') can only be assigned via 'hand-over' of C-head [+AGR/T] to T.

Featural vs. Affixal T.

The assumptions made above is that T(ense) comes in two forms: (i) featural and (ii) affixal. When featural, since T is adjunctive in nature, theoretical considerations could be made that T can be found adjoined to any appropriate phrase (not just TP). (Radford (2016, p. 338) goes further to suggest that even DP/Nominals (Nouns) can theoretically take-on a tense feature). Tense has been problematic in the MP literature for quite some time: for instance, it seems to be a [+Interpretable] feature, thus semantic in nature, while, at the same time it seems to be implicated both in Case and Agreement. And Chomsky considers Tense not to be a phase.

What this may call for is the notion that T, when functioning as [+featural], a featural-T can percolate up and down the tree without impunity (when inserted at the **edge** of any phrase) and enter into **probe**-**goal** relations of various kinds (as determined by the nature/feature specificity of the host head). For instance:

(i) when T-featural is adjoined to C, it seems to be implicated in securing AGReement of the subject-verb accord,

(ii) when T-feaural is adjoined to T, it delivers two functions:
   a. it can deliver Case (as found with light verb vP),
   b. it can deliver a non-affixal Tense (such as with bare verb stems, auxiliary do>did, have>had, is>was, as well as irregular verbs go>went, etc).

(iii) when T-affixal: {ed}, {s} [+/-past] affix lowers onto the verb stem.

Note 3

A Note on 'Proto-language': A merge-based theory of language acquisition—Case, Agreement and Word Order Revisited

·Language = Recursion, which is 'recently evolved and unique to our species'—Hauser et al. 2002, Chomsky 2010.

·If there is no recursion, there can be no language. What we are left in its stead is a (Merge-based) broad 'beads-on-a string' sound-to-meaning recurrent function, serial sequenced, combinatory non-conservative and devoid of the unique properties of recursion which make human speech special. It may be 'labeling' (see Epstein et al.)—the breaking of 'combinatory serial sequencing' found among sister-relations—that constitutes the true definition of language since in order to label a phrase one must employ a recursive structure— JG.

·If Continuity is allowed to run freely, in all aspects in respect to biology, and is therefore the null hypothesis, then what we may be talking about is a 'function' that matures over time, and not the 'inherent design' (UG) which underwrites the function, since, given strong continuity claims, the design has always been there from the very beginning. It may be that the (Move-based) function 'Recursion' may mature over time, in incremental intervals, leading to stages of child language acquisition, and in the manifesting of pidgin language. But when all is said and done, strong continuity claims don't necessary span across other species or even intermediate phases of our own species. In fact, strong evidence suggest the contrary—that the unique recursive property found specific to our own species, early Homo Sapiens (Cro-Magnon) has in fact no other antecedent that can be retraced past a date of approximately 60kya—JG.

[1] **Introduction**

Before advancing theories about the nature of protolanguage, it would seem that what we now know of the 'brain-to-language' corollary would help inform our understanding of critical issues on the topic. Along with an 'ontogeny-to-phylogeny' trajectory, perhaps indicative of how the stages of early child language seem to unfold and mirror what we know of language evolution in general, the best heuristic tool we have to solving the puzzle of language

emergence, growth, and mastery is two-fold in nature, which approximates answers to the questions: (1) what type of linguistic processing seems to be unique only to our species (species-specific)?, and (2) what levels or areas of neuro-cortical substrates seem to underwrite these unique processes? The former question is perhaps best articulated in the Hauser, Chomsky and Fitch (HCF) paper which first appeared in the journal *Science* in 2002. The latter has been addressed in multiple sources, the first of which drew my attention was the Fisher, Marcus review which first appeared in 2006 in the journal *Nature Reviews Genetics*, and others including most recently in Larson et al. (2010) *The Evolution of Human Language* (see both Lieberman and Stromswold chapters).

[2] What seems to be the locus to the question surrounding the nature of protolanguage hinges on our understanding of, first, how we should go about exactly defining 'Language'—not say language with a small 'l', (such as *French, English, Japanese*), but rather language with a capitol "L" (what we mean of *Language in principle* (question 1 above)). What has come out of the second half of the last century, in terms of our Chomskyan framework, is an attempt to demonstrate, *empirically*, the quite distinct notion of what had henceforth been assumed *a priori* by theory-internal devices—namely, that '*No syntactic principle or processing applies directly to words, or to superficial word ordering*' (Piattelli-Palmarini 2010, p. 151)[9]. Rather, what the Chomskyan view grants us is a language borne of a *categorial nature*, abstract and seemingly defiant of communicative functions and unrelated in critical ways towards any strict interphase with the environment.

[3] It seems theory had to move away from traditional 'word-based' constituencies (such as Noun Phrase, Verb Phrase (NP, VP)) and move towards more abstract constituencies dealing with inherent features of the H(ead) of a word, along with a H's relationship to other Heads. (For instance, the INFLectional Phrase (or IP) came out of this tension between head of words versus heads of features (see fn. 1)). A cursory look at the chronological record of the generative grammar enterprise takes us from early 1950s T-makers of transformational grammar (TG), to recursive phrase structure grammars (PSG), to X-bar theory (of the

[9] Hence, linguistic theory had to move away from a traditional 'word-based' Phrase-constituencies (VP, TP, CP) to more abstract 'feature-constituency' (Distributed Morphology of an INFLectional Phrase (IP) (Halle, Marantz). Most recently, 'phrase' has been replaced with 'phase' (an alignment which maps onto the so-called 'duality of semantics' (vP, CP)—a move in keeping with what had been articulated by prior notions of scope, c-command, Head-to-Head/Comp movement, dynamic anti-symmetry (Moro), as well as probe-goal relations: all of which have become central, abstract tenets of the theory.

principles & parameters framework (P&P) of the 1980-90s which delivered a 'Spec-Head-Comp' configuration (the holy grail of the 'Spec-Head' relation), only to be overturn most recently within the minimalist program (MP) by the prosaic 'Head to Head' relation whereby the simple 'Merge' of two Heads is now the driving force behind all syntactic operations. (See 'Note-2' for full discussion).

[4] Much of our discussion related to the 'four-sentences' section of this book is deliberate in showing just how rather byzantine constraints on abstract syntactic structure defy what would otherwise be intuitively expected of a simplistic means of functional communication: e.g., why should the clitic formation of [that's] in 'sentence no 4' (found in our 'four-sentences' analyses)—a clitic formation that is legitimately pervasive otherwise—not be allowed[10]?

Sentence-4: 'I wonder what *that's/that is up there'

Clearly, either form [*that's]* vs. [*that] [is]* should share in the 'equal status' of plainly being able to communicate the simple proposition; however, the clitic [that's] in this syntactic structure found in sentence no. 4 is ungrammatical. (What we could say is that while [that's] communicative value is plus [+Com], its syntactic value is minus [-Syn], demonstrating that there is disassociation between (formal) syntax and (functional) communication. (See Piattelli-Palmarini (ibid) for other such data and analyses).

[5] For example, it is the internal categorical structure of the 'H(ead)-features of the word' which is now seen as projecting the outer phrase constituency: e.g., for the lexical item V (verb), it may be the categorical features related to T (tense), a 'finiteness effect', which determines its syntactic valence of how it might select for a determiner (Subject, Object, Case). The dual probe-goal phases of CP & *vP*, as currently understood within the Minimalist Program (MP), may similarly assign H-features which map onto the so-called **duality of semantics**: where phase/CP is responsible for scope and discourse-related material as well as the functional projection of AGReement (and presumably Tense), and where the phase/vP maps onto argument structure (and presumably Case).

[10] See 'Four Sentences' [Sentence #4] for a full analysis.

[6] Recently, it has been proposed (HCF) that it may be the sole, unique properties of recursion which is behind the very underwriting of this 'categorical nature', and that more specifically, these features reside as 'edge-features' of a phase[11]. If we assume that Chomsky (in particular) and HCF (more generally) are right, within the linguistics context, namely, that 'language = recursion' (as defined in his terms by a **Language Faculty narrow** (LFn), then our question becomes: What types of neuro-substrates serve recursion, which would be separate from other cognitive/motor control function? A second follow-up question, relevant to the question at hand, would be what a proto-language might look like stripped of this narrow language faculty, where a language may only show evidence of linear sequence [A], [B], [C]…, a **recurrent** but not a **recursive** structure, as found in [A [B [C]]]…[12]

Broca's area/Wernicke's area revisited

[7] Thought it still makes for a nice pedagogical device, we now realize it is quite over-simplistic to talk about a compartmentalized 'seat of language' in this way that straddles the classic Broca-Wernicke divide. Although it continues to feel natural in wanting to map etiology of language-specific diseases to specific cortical regions of the brain—e.g., how Parkinson's disease (PD) presents differently from Alzheimer's disease (AD), or how with the Autism-spectrum Williams' syndrome (WS) suffers from unique processing deficits distinct from Asperger's syndrome (AS), etc.), and, furthermore, how these distinctions might in fact show up in specific areas of the brain (Broca (= PD, WS)) vs. Temporal-lobe (AD, AS) respectively)—what we have rather discovered is that the surface areas which we call Broca's area (BA) & Wernicke's area (WA) are merely terminus levels found on the cortex. (The notion of terminus nodes which surface on the outer-cortex was presented as early as 1885 when Lichtheim

[11] In a [Spec [Head-Comp]] configuration, the so-called 'edge' would be the Spec position, away from the core inner working of the phrase/phase. Spec is often defined as an 'elsewhere category' which allows for MOVE to take place (whether the Spec is serving as a host for the moved item, or is instigating the move in the first place in accordance to a Probe-Goal relation).

[12] Recall, a linear recurrent model would show a potential two-word utterance as [[drink] [water]] without the necessary syntactic/recursive properties which would allow for a full expression behind the notion of *someone drinking*. In other words, a flat combinatory sequence would only yield two items in isolation [x], [y]. What is lacking is the recursive syntax of: [drink [~~drink~~ water]], which shows MOVE allowing for a hierarchical expression. (See §[22] 'A Summary of Labeling and how "Merge vs. Move" affects Word Order'. Also see 'Note 2' for fuller discussion of 'dynamic antisymmetry').

analyzed interconnected neuro-pathways between BA & WA). If BA & WA are just termini which are *post hoc* defined merely by the gathering place where certain types of neuro-bundles that gather together fire together (under specific tasks, language tasks, etc.), then, the more critical question is: What actually underwrites such specific neuro-bundles?

[8] In other words, what we must reconsider is the possibility that perhaps it is not the cortex at all that is doing the underwriting of the neuro processing (not BA, WA), but rather the processing is being guided by more robust and underlying subcortical-clusters which precisely bundle and target specific areas of cortical mapping. In other words, if BA and WA do subserve specific types of language tasks (as classically assumed), they do so due to their mapping of subcortical-neural-circuit (SNC) triggering. The best case scenario here for such SNC processing is what we have learned over the past twenty years regarding the functions of the **basal ganglia** (a group of structures found deep within the cerebral hemispheres, which includes the relay-connectivity of the *Putamen* and *Thalamus*, both working in tandem which form a cortical 'feed-back' loop). Recent studies have now shown (see Cummings 1993 for review, as cited in Leiberman's, p. 167 found in Larson et al. (2010) that distinct regions of the frontal cortex indeed connect with their basal ganglia and thalamic counterparts, constituting largely segregated basal ganglia-thalamo-cortical (BTC) neuro-circuits.

[9] The main 'SNC-processing' which roughly maps onto BA is **movement** (MOVE), the unique ability (perhaps motor-control related) to displace an item in the surface-level (phonology) to some other place in the underlying (syntactic) structure, *inter alia*. The basal ganglia, with its 'looping effect', bringing subcortical neuro-circuitry to percolate up to the surface cortex, seems to be the best-case cerebral candidate to serve the unique phenomena of MOVE, where recursion is required to break with a flat sister-relation otherwise found of surface phenomena (See § 'Note 2' below regarding **recursion** & **dynamic antisymmetry** (DA)—processes which extend otherwise flat sister-relations to having hierarchical status). Recursion has the property which allows cortical mapping of two language-specific tasks (both seemingly BA-related): viz., that of *phonology*, and that of *inflectional morphology*. While phonology as recursive is still hotly debated[13], (and Chomsky doesn't appear to be

[13] Syllable structure of <onset, <nucleus, code>> might be recursive due to its inherent hierarchical structure. For review, see Schreuder, Gilbers, and Quene's paper 'Recursion in Phonology' *Lingua 119* (2009). It also bears keeping in mind that MOVE-related diseases such as PD do seem to impact both phonology and syntax, while other studies also suggest that MOVE correlates to mouth movement, planning and articulation of speech as well as syntax.

swayed by such arguments), to the contrary, inflectional morphology which is defined by movement (displacement) is clearly quintessential recursive in nature.

[10] In this note, I focus only on the displacement properties of recursive syntax found in inflectional (INFL) morphology (as present in morphological **Case** and **Agreement**, both which are INFL-related, and see what proto-language absent such INFL/Recursion might look like (comparing data results to that of pidgin language and/or even Chimp ASL e.g., Nim Chimski, (Terrace 1979)).

Proto-language and Derek Bickerton

[11] I know of no more passionate advocator for a protolanguage than Derek Bickerton. His and his colleagues' tireless work examining Hawaiian pidgin—as a heuristic model for what linguists should look for towards a proto-language grammar—has brought the once taboo topic to the fore of current linguistic theory and debate. Today, the theoretical notions leading to any understanding of a putative proto-language have suddenly found its underwriter by the larger, and perhaps even more ambitious, interdisciplinary field of *Biolinguistics*. This brief 'Note-1' is in response to some thoughts on what has been laid out in Derek Bickerton's 2014 paper 'Some 'Problems for Biolinguistics' (*Biolinguistics* 8).

Having set-up some discussion regarding the current state of the 'biolinguistics enterprise', and some non-trivial problems pertaining to its research framework, particular to the *Minimalist Program* (MP) (Chomsky 1995), Bickerton goes on to express his long-held views on the nature of a Protolanguage (§4.2)—namely, *pace* the given Chomskyan account, that there should be <u>NO inherent contradiction</u> between the coexistence of the two statements (below):

Broca's aphasia may impact both speech as well as syntax. What we could then say of PD is that it affects the basal ganglia along with its SNC-processing leading to the inability to exact MOVE-based recursion, as found both in phonology and syntax.

Statements:

[12] (i) 'Statement-1': That language is to be properly defined, very narrowly, within the terms of a **Language Faculty**-*narrow*, an **LF***n* which, by definition, *exclude*s most of what is typically accepted within the linguistics community (outside MP) as defining what normally constitutes a language (e.g., vocabulary, idiomatic & encyclopedic knowledge (= the lexicon), phonology (syllabic constructions), and some particular aspect of morphology (e.g., derivational processes, compounding, etc.). A layman's classical definition of what constitutes 'language' is intuitively very broad in nature. But Chomsky's definition of a language faculty (LF), to the dismay of many, is exceedingly Narrow (n): <u>That LFn is the sole property of recursion</u>: that language is exhaustively defined by the exclusive and very narrow property of recursion.

(ii) 'Statement-2': That a putative **protolanguage** theoretically exists and could **serve as an intermediate step between a partial language and a full-blown** LFn—viz. an intermediate language phase which would find itself tucked-in between what we know of pidgin languages (an L2 attempt to formulate a rough grammar for functional communicative purposes), and perhaps chimp sign-language and other animal cognitive-scope features (of the type taught to the chimp named Nim Chimpski (Terrace 1979)), along with other communication systems which are not on equal par with LFn, of what Chomsky refers to as **Language faculty-***broad* (**LF***b)—viz.,* 'broad' factors which *include* the aforementioned lexical-item development sensitive to frequency learning, formulaic expression, and other similar 'frequency-sensitive' morphological word-building processes such as compounding and derivational morphology.

[13] In other words, Bickerton's claim here is that we can accept both statements as true—they are not mutually exclusive:

(i) (LFn) Yes! 'Language-proper' is to be narrowly-defined as pertaining to the sole (and, as it turns out, quite a unique) property of 'recursion', and,

(ii) (LFb) Yes! There too could be a protolanguage (by definition, an LFb) without 'recursive operations'—a language just shy of maintaining the status of a ''full-blown language' along the language spectrum[14].

The two claims appear to reflect on larger dichotomy issues. Let's flesh this out below in the way of the dichotomy debates: 'form vs. function', 'continuity vs. discontinuity', 'nature vs. nurture…'

The dichotomy debates.

[14] Taking the former 'recursive property' (=syntax/LFn) as a critical aspect of a dichotomy-debate (say, of continuity), one would most certainly claim the emergence and development of recursion (MOVE) to be *discontinuous* in nature from all other non-human primate communicative systems, perhaps accepting Gould's version of recursion as 'exaption' at one end of the spectrum with Chomsky's single-mutation-event leading to a 'pop hypothesis' on the other[15]. In any case, both claims would be consistent with what Gould calls a 'punctuated equilibrium' hypothesis—i.e., that recursive language (LFn) emerged in one fell swoop, either exaption from prior material (his 'spandrels') or a completely novel structure[16]. The features of the latter (LFb), 'from prior material', most certainly would maintain at least some level of *continuity* assumptions, as widely expressed in the language-evolution literature (e.g., Pinker & Bloom (P&B), among others).

P&B may be correct in assuming that there is 'somewhat' continuity regarding the articulation mechanism of sound/phonology (i.e., the chimp's ability for syllabic pant-hoots, and other primate syllable-vocalization capacities—though it must be said that human speech is indeed quite unique and highly specialized due to the lowering of the larynx), as well continuity in what we would find regarding the idiomatic 'one-to-one' associative-learning mechanisms behind the mapping of 'sound/gesture to meaning' (manuofacial expression),

[14] Bickerton has long sought to advance an intermediate stage of language, 'a proto-language', as a grammar just shy of maintaining a fully-fledged recursive grammar. What so-called 'flat- recurrent' (non-recursive) grammars would not be able to do is creatively generate and parse constructions beyond a preconceived semantic/canonical specificity. See endnote of this section for discussion of recurrent versus recursive grammars.

[15] See Jean Aitchison (1998) for a review of the 'slow-haul' vs. 'pop hypothesis' in this context.

[16] See Crow (2002) for a sudden 'genetic mutation' hypothesis (which would be akin to Gould's 'punctuated equilibrium').

'cue-based' representation (in the 'here & now'), and other non-formal constructions leading to compounding and even limited syntax (lexical-root phrases such as [NP [N] + [N]] constructs which approximate possessive structures e.g., [NP *daddy book*] (= daddy's book) or prosaic [VP [V] [N]] constructs which approximate Tense/Agreement e.g., [VP *daddy drink water*] (=daddy drinks water), etc.

But I suppose, for Chomsky, the question is: Can we really get there from here? Can FLb turn into FLn?: Really, can broad-communicative features (attributed to non-humans) as laid out in (HCF) evolve into (human) FLn? (For Chomsky, the answer is an unequivocal NO! and hence another dichotomy debate). Chomsky's now famous analogy lends us to imagine a sudden mutation (or catastrophic event) devoid of any 'bottom-up' Darwinian selective pressure for FLn:

> *'We know very little about what happens when 10^{10} neurons are crammed into something the size of a basketball…'* (Chomsky 1975: 59).

These open lines of a much longer paragraph on the topic fully commit to a top-down 'form-precedes-function' analysis regarding FLn. Bickerton has carried on with the same theme arguing against any 'non-human to human-language continuity' when he claims that:

> '*[T]rue language, via the emergence of syntax, was a "catastrophic event", occurring within the first few generations of Homo sapiens sapiens*'. (Bickerton 1995: 69).

Child-to-adult-Continuity.

[15] Let's remind ourselves that Chomsky believes in 'child-to-adult' continuity (if not in 'function', in 'form') given that language, per se, has potential drop-off way-stations on its way to a full target-grammar projection. So, for early child language utterances, the nature of their errors (functions) is rather epiphenomenal since the underlying grammars (forms) which *underwrites* the syntactic templates must be (if we assume a UG) the same 'all the way up/down' between child and adult. Though, perhaps a better way to view Chomsky's remarks

is to suppose that we can still tease apart '*form* from *function'* (yet another dichotomy)[17]. For instance, assume that Chomsky agrees with the assertion that children first exclusively *function* with Merge (and not Move)—while still maintaining that the *form* of UG is the same as consistent with continuity. Well then, there could be space within such an argument for an emerging grammar. The hypothesis would be that young children (at the low Mean-Length of utterances (MLU) stages) would be *forming* the same UG as their adult counterparts while their *functions* would be immature, following a protracted maturational scheduling of function: UG…(stage-0)…..stage-1 (Merge)….stage-2 (Move)…on their way to a full Target language (stage-T). What could be claimed then is that it's the function 'MOVE' which matures and eventually comes on line.

True, the capacity for MOVE was always there (UG), it's just that the hidden processes which map 'form to function' followed a protracted schedule. This is not unlike what we would find for the maturational development, say, of **functional categories**—viz., while their form is intact, as part of UG (DP, TP, CP), their mappings of 'form to function' are delayed (see Galasso 2003). Again, this form-to-function disparity could be one way to reconcile Chomsky's strong stance calling for a non-developmental UG (since the empirical observation is valid: that language, at any given stage of development, never exhibit UG violations, nor do they ever exhibit 'wild grammars').

[16] Lastly, the above notion that MOVE is maturational-driven (within our Homo species) seems to nicely correlate with what Chomsky himself claims of language evolution (within our Homo species): that 'Every inquiry into the evolution of language must be an inquiry into the evolution of the computational brain machinery capable of carrying out edge-features operations' (Chomsky MIT lecture, July 2005, cited by Piattelli-Palmarini (2010, p. 151)). Recall that what we mean by 'edge-feature' operations are those syntactic operations which can only be handled by the unique recursive property of MOVE. Also recall that there is also

[17] 'Function-to-form': so, think baseball glove: catcher's-glove is padded due to repeated fast balls ('catch the ball softly!'), outfielder's glove is light due to its having to be held while running for the fly-ball ('catch the ball running!'), first-baseman's glove is extended due to a race between the ball and the bat's-man running to tag first base ('catch the ball quickly!'). This here is 'function defines form' (or function precedes form). But language seems to be the reverse (form precedes function), where the form of the (internal) mental template seems to shape the potential (external) function of language. For example, an Arabic speaker sticks out his phonological-perception glove to catch an (external) English '/p/-ball' (say, the word /P/olice) (using a phonetic-pitch metaphor to baseball), but (internally) catches it as a '/b/-ball', where /polis/ (police) gets caught as /bolis/. (Arabic has no /p/ phoneme in its phonological inventory, and /p/ vs. /b/ does not make up minimal pairs). See also discussions surrounding the dichotomy between 'functionalism vs. formalism'.

a high level of inherent abstract symbolism involved in any MOVE-related/edge-feature operation since such principles of MOVE (i.e., syntax) do not map onto words *per se*, in an iconic 1-to-1 manner (as might be intuitively imaged of language), nor is there any surface word-order mapping (which might be expected of surface phonology). Rather, MOVE inherently requires the mental manipulation of categories (=symbols)—these are categorical concepts such as Verb, Noun, or constituency structure which breaks with surface word order. (See our 'Four-sentences' analysis for full discussion of recursive constituency).

The implications here is that in order to question the nature of language evolution, and all of its complexity, the first order of business is to address the question of determining when the first evidence of MOVE appears in the early Homo species, and if, as a result of MOVE, other spin-off exaptions (or so-called 'hitch-hiking' free-rider adaptions) can be explained as being *bundled with recursion* (perhaps even neurologically bundled): I am thinking about theory of mind, shared attention, symbolism and displacement which contribute to so-called 'detached representations', altruism, dance, ceremonial practices & taboos, and other perhaps niche motor-control abilities such as tool-making capacities (which demonstrated a mental template for the design of the tool), throwing capacity , so-called 'remote threat', sheltering, cooking of food, etc.

[17] What all of the above features have in common—as a unifying thread which can lead to recursion/MOVE—is the ability to project an item, project oneself, away from an icon, an index, and to become symbolic and categorial, both in nature, in index, and in design. What we currently know—out of all archaic homo species (Homo-Habilis (*Australopithecus africanus*), (early Africa)-Ergaster, (late Asia)-Erectus, Heidelberg, Neanderthal)—is that only Cro-magnon[18] (our early homo sapiens-sapiens ancestors, say at around 40-60KYA) had emerged onto the scene (seemingly top-down) into being *categorial* in nature, gaining a rich symbolic system first drawn from an inner mental language (MOVE), with subsequent bootstrapping to be applied to other non-linguistic, cognitive, motor-control tasks. Once a full-blown symbolic inner-language system emerged (either via a catastrophic mutation or via exaption), what came with it was all the 'bells and whistles' of being a member of a unique symbolic club, what today we call the 'homo-sapiens-sapiens club'). It is now well recognized that by the time Cro-magnon comes on the scene, having evolved in whatever which way,

[18] The brain-size trajectory most certainly would be a major contributing factor with regards to any such evolutionary-based theory for either a 'gradual development' (bottom-up) or 'sudden emergence' (top-down) of FLn. To be considered is a respective brain-size spectrum that would begin at around 450cc with Australopithecus, Erectus at 1000cc, to roughly 1500cc with Neanderthal, followed by a very slight decline in Cro-magnum stabilizing at 1300cc).

they came on the scene drenched in symbolism (White 1989). If they evolved at all (bottom-up), what we can say is that they evolved from an earlier time/species of not yet having recursion, to a later time/species[19] when they have it, or if you prefer, using our current linguistic terminology, from an earlier time of having FLb, to having FLn.

[18] All of these FLb linguistic factors—which, as suggested above, may have hitch-hiked from categorial symbolism in its purest form (viz., *sensori-motor control, mapping of sound-to-meaning, lexical retrieval, word-building, compounding, and even derivational morphology*)—are all found in the very early stages of child language acquisition (Radford & Galasso 1998, Galasso 2003, 2016), and have antecedents which can be traced back to pidgin grammars (Bickerton), and, to a large degree, even further back to what we have gleaned from non-human primate communication systems (the use of ASL by Nim cited in Terrace). Of course, it goes without saying that notions of any putative '"somewhat"-continuity' between human and non-human as it regards cognitive scope, theory of mind, altruistic features, etc., must be taken with a grain of salt—viz., there really is not much continuity to speak of, in these realms, and the very fact that non-human primates lack what would in humans be such simple operations surely present us with the 'smoking gun' of discontinuity through and through. (The underlying question to ponder here is whether there might be a 'singular, unifying mode of processing which underwrites these realms—and, neurologically, might it be related to MOVE?).

In any case, for what it's worth, this dual acceptance allows for both *continuity (LFb)* and *discontinuity* (LFn) to flow from out of ontogeny and phylogeny trajectories—ontogeny in terms of 'critical period' cases in which a proto-language (Bickerton's 'bioprogram-hypothesis') may fare no better than the 'end result' of a trajectory of an individual's growth and plateau of syntax (leading to a pidgin language—in that a pidgin is in many ways discontinuous from a target L1, just as early child language shows discontinuous properties due to their lack of full recursion of syntax).

[19] In terms of phylogeny, we could assess claims which speak to how 'language-broad' evolution might be continuous in nature, with antecedents which harken back to animal cognitive capacities. In other words, Bickerton claims we can find an intermediate phase along these dichotomy-spectrums, in one sense leading to a human-language (immature) capacity which would solely incorporate LFb-features—including *inter alia* a limited lexicon with perhaps a

[19] One possibility implied here is that (early-FLb) Homo erectus evolved into (late-FLn) Cro-magnum.

maximum 'mean length of utterance' (MLU) count of below 3 (i.e., no more than three words per utterance), along with the complete absence of Inflectional morphology; what we would expect of pidgin-language capacity. Though it is the latter statement that Chomsky rejects, I, along with Bickerton, see no reason at all, at least conceptually, why there couldn't be a syntactically, albeit robust, LFb-phase of child language (ontogeny) on its way to a fully-fledged LFn, and if so, why this intermediate phase that the child passes through couldn't constitute what we would at least theoretically claim of a protolanguage (phylogeny). In one sense, this kind of argument mimics the old adage 'ontogeny recapitulates phylogeny' (first cited by Ernst Haeckel)[20].

[20] Chomsky's insistent belief is that there could be, conceptually, no intermediate step shy of a full language; if you have such a step, then it's merely a function of a communicative niche (as expressed above), and that such a deprived system (deprived of recursion) would by its fixed nature need to remain there, as a non-evolving, non-human communicative system (=LFb). This is tantamount to saying that **LFn cannot arise from LFb**—viz., that there can be no continuity between LFb and LFn (not in a phylogenetic way 'evolution', nor in an ontogenetic way 'child maturation' as cited above). I believe (and I assume Bickerton would agree with me here) that Chomsky's assertion is too strong. Chomsky has been quite consistent ever since our reading of the 'Fitch, Hauser, and Chomsky paper' (2005)—on the topic of the nature of LFn and of language evolution—that a definition of 'Language' (a language with a capitol "L") can only be purely based on one essential property, namely the property of **recursion**.

For Chomsky et al. (2005), [language = recursion]. This very narrow definition is perhaps the only way that Chomsky can maintain his long-held notion that language is biologically modular and human species-specific (modular in that it functions like any other organ, e.g., the liver, stomach, lungs) and species-specific in that its operation is uniquely situated in the human brain/mind (presumably Broca's area, a region which seemingly only serves recursive operations such as (inter alia) the planning of articulation leading to mouth movement, and the movement involved with syntax).

[20] For example, see Dan Slobin (2004).

The Minimalist Program (MP) Enterprise: **Resolving a dichotomy**

[21] But there does seem to be way to reconcile both statements within the MP enterprise. Within MP, there are two types of **movement** (mapping with what one finds regarding the 'duality of semantics'):

(i) Local-move (= Merge, the merging of two Heads) is based on the merging of two items (two Heads (H) within a Phrase (P))—e.g., such as what we would primitively finds in H-H compounding (e.g., Adj+N sequences such as [black]-[bird]=> [blackbird]), a simple base in-situ Verb Phrase such as [VP [V bounce] [N ball]], and non-formal sentence constructs such as 'Me go' or 'Him do it' which show a complete lack of inflectional morphology (a lack of Case and Agreement). All these constructs do show up in impoverished pidgin systems as well as in very early MLU stages of child language, and can be attributed to the kinds of features we find in non-human communicative systems of the sort famously demonstrated by the chimp named Nim (ibid).

But note here that in order to know where the H of the P is, one must involve a second (later) merge operation coming on the heels of the first. In order to reach the VP derivation of the unordered set {V, N}, locate the Head {V} and **label** the P accordingly, the speaker must utilize what is referred to as Internal Merge (IM) (an instance of distant-**Merge/MOVE**), so that the unordered set {bounce, ball}, becomes an ordered pair: syntactically deriving the mere twin lexical items [bounce, ball], to a fully-fledged VP [bounce [~~bounce~~, ball]]. So, we would speculate that at any impoverished mere 'local-move stage', we should find instances of mixed word order and lack of inflectional morphologies leading to the absence of Case and Agreement. This is indeed what we find not only of child language (See Radford & Galasso 1998), but also what we find regarding pidgin formations, and finally what the extremely curtailed limits of Nim's speech range.

(ii) Distant-merge (= MOVE) is based on the subsequent move (a second-order move) which, as a result, breaks the flat symmetry of an unordered set and allows the labeling of a Head of the phrase to be defined. In contrast to local merge, distant merge (= move) allows for a portmanteau of features and phenomena, among them the syntactic operation of movement which break base in-situ constructs and allows for the lexical item to percolate up the syntactic tree in order to check-off +Formal features (in current MP terms, as guided by the 'probe-goal' relation). Other consequences of MOVE would be the projection of

Case (+/-Nominative), AGReement (Person, Number) as well as Tense, all of which are found in higher phrasal projections above the base-generated VP. Again, any lack of these higher, formal projections would have syntactic consequences. (A brief summary follows. But see Note 2 for full discussion of Merge vs. Move and 'problems of projection').

A Summary of Labeling and how 'Merge vs. Move' affects Word Order.

[22] First-order /local merge—the simple assembles of two lexical items in creating an unordered set, say a Phrase (P) {a, b} out of the two items. Yet, there is no recursion; hence, there can be no labeling of what would constitute the Head (H) of the P. In order to derive H of P, a second-order /distant merge must break with the set in creating an ordered pair {α, {α, β}} = P (where α = H). It is via this second-order merge (which constitutes as a recursive property) that we can derive order within the P—an order which comes about as a result of the ability to label which of the two items is rendered as H.

[23] Consider, at least theoretically if not empirically[21], a young child's inability (at the early mean length of utterance stage (early-MLU) to derive second-order merge labeling, thus being incapable of understanding labeling of H, rendering such otherwise adult unambiguous structures ambiguous: e.g, [*house-boat*] is read and interpreted as a kind of boat (and not as a kind of house). But, if we first examine the base-structure of the two lexical items {house, boat}, there is no way we can glean from a flat, unordered structure what the Head word of the compound [N+N] would be. This problem is in fact what we find in very early instances of

[21] See the monograph *From Merge to Move* (Galasso, 2016).

child language[22]. Carol Chomsky[23] first found the lack of recursive operations regarding passive formations—that when young children were faced with (improbable) irreversible passives (e.g., *The ball was kicked by the boy/*The boy was kicked by the ball*) they scored quite well. But when children were presented with reversible passives—passive interpretations which must exclusively rely on 'syntax', as opposed to irreversible passives which were actually acquired quite early in development since 'semantics' can serve to help with the only probable interpretation—the children tested were unable to correctly demonstrate that type of movement necessary for a passive interpretation. In other words, children had a hard time with (e.g., *The man was killed by the lion/The lion was killed by the man*) where both reading are probable and reversible.

[24] It is interesting to note here that Grodzinsky (1986, 1990, 1995) similarly finds in Broca's asphasia subjects an inability to handle 'distance of movement' in embedded subject-relative clauses, where (i) **local movement** had an 'Above chance' level of acceptance/reading and where (ii) **distance movement** had only a 'Chance level'—e.g.,

a. 'The cat_t [that [$\_\__t$ chased the dog]] was very big.

(local move = Above chance)

c. cat_t [that [the dog chased $\_\__t$]] was very big.

(distant move = Chance).

[22] And as discussed herein, such an inability for labeling would force a flat reading of the two items [drink] + [water] as two separate interties without the luxury of syntax—viz., an individual with a recursive grammar can reconstruct the two items syntactically, within a VP, such that a proposition can be generated: that 'someone is drinking/wants to drink/should drink water', etc. The same two items, recursively, get structured as [VP drink [~~drink~~ water]] where the Verb 'drink' now dominates the Noun 'water' in a *mother-daughter* hierarchical relation [x [x,y]]. As long as the two items stand in a flat non-recursive, recurrent manner [x, y], all one could gleam from the utterances is that 'drink' and 'water' have been combined, 'stacked' in sequence, and where perhaps word order has no bearing on structure. The fact that a person (with full adult syntax) can reconstruct a meaning out of a simple two-word utterance such as 'drink water!' suggests that such a bootstrapping relies on a matured mental syntax in supporting the reading. (See Note 2 herein for Dynamic antisymmetry and Problems of projection).

[23] Chomsky, Carol. 1969. The acquisition of syntax in children from 5 to 10. Cambridge, MA: MIT Press.

[25] In other words, the greater difficulty in comprehending sentences in situations where syntactic form is not supported by semantic content suggests that the semantic component of grammar may play an important role in the young child's acquisition of syntactic comprehension—the latter 'sematic-content' interpretation being a product of local merge, viz., the yielding of the lexical items and how each item plays a thematic role in the sentence. In the case of 'distance-moved' (cf. Grodzinsky), it seems that adjacency of linear order (surface phonology) takes prominence over hidden structure at a distance. (Also see Galasso 2016, chapter 8 for treatment of Broca's aphasia data).

[26] Distant Merge (= Move) has something to say about how we glean a Head for a given phrase. In the case of 'house-boat' ('a kind of boat', not 'a kind of house'), in order to derive the head of the Compound (C) (heads are *right-branching* in English Compounds) we must employ second-order distant-merge. Following Moro's work on dynamic antisymmetry, accordingly, in order to label a H of a P (or C), we first must break with flat/sisterhood relations—of the kind typically associated with 'logical and' (e.g., I need to buy: 'a and b and c and d' whereby comma-insertion allows for displacement and rearrangement of 'a-d' in any order since sister-relations are symmetrical and hold no hierarchical order—and create an antisymmetric hierarchy, such that from out of a sister, first-order/local-merge set {α, β} <house//boat, boat//house> (showing symmetry), we derive {house {~~house~~, boat}}.

In this second-order 'Move-based' structure, notice how the Noun 'house' has risen up to a higher functional node within C. It is this movement that breaks flat sister relations and creates, as Moro puts it, dynamic antisymmetry in labeling H of C. For phrases, that is at work. Take for example the VP [VP [{V bounce}, {N ball}]]. In order to derive the H verb 'bounce' of the P 'bounce ball', (English H of P are left-branching, just the reverse we found with C), there needs to be second-order distant move, such that the H becomes labeled as distinct from its complement: first-order local merge: {bounce, ball} (showing no order) becomes second-order distant merge/move {bounce {~~bounce~~, ball}} => [VP {bounce {~~bounce~~, ball}]. It is clear that there are instances in the child language literature where young children cannot yet discern their proper word-order, e.g., a child may utter VP *bounce ball, or ball bounce* with identical intentionality (see Galasso, 2001, https://www.csun.edu/~galasso/worder.pdf).

Distant-Merge related structures missing in pidgin and which constitute the basis for Proto-language.

How 'Merge vs. Move' affects Case Assignment.

[27] Let's begin this section on Case with some basic assumptions, some of which are theory-internal:

(1) Case marking is a 'functional-category enterprise'—viz., a formal projection which requires movement of the case-marked item to raise out of the base-generated VP and insert into a higher functional phrase.

(2) That there are three distinct (and overt) Case markings in Standard English (SE): Nominative on subjects [+Nom] (e.g*., I, he/she, we*), Accusative on objects [-Nom] (e.g., *me, him/her, us*), and Possessive/Genitive when used as prenominal [Poss+N] [+Gen] (e.g., *my, his/her, our*), and when used as pronominal [N-{Poss}] (e.g., *mine, his, hers, ours*). Also [+Gen] is morphemic {'s}, {of} in examples [Tom's [house]], and 'The house [of [Tom] ~~'s~~] ~~house~~]]' (*The house of Tom*). The morpheme {to} also serves to case mark [-Nom/Obj] (e.g., *give it <u>to him</u> /*give it him/ give him it/*give him to it*).

(3) The default, base-generated order of Double-arguments is [Indirect Object + Direct Object] [IO, DO]. (For theoretical discussion, see Boeckx (2008).

(4) That Case can't be doubly marked from a single verb—in this sense, case is of a 'Probe-Goal' (PG) relation instigating an upward projection of a targeted item. Once a probe has located its goal and relevant features have been checked, the probe is no longer active in the syntactic derivation.

(5) There are at least two mechanisms for Case marking:

a. via **Structural/Configurational** with a lexical item (in local domains)—so-called **lexical Case**:

i. Verb-complement, PRN => Object [-Nom]

ii. Preposition-complement, PRN => Object [-Nom]

b. via **Morphemic assignment** (probe-goal) with clitics {-'s}, {-to}, and {-m}

c. Otherwise, via **default** (or **inherent case**).

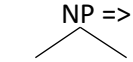
A Theory

[28] One very persistent characteristic of any putative protolanguage would be its lack of MOVE (movement operations which are motivated by (*inter alia)* **functional features** which make-up a **Probe-Goal** (PG) relation: e.g., Case, AGReement & Tense (and Word Order is most often a result of some movement operation whereby the surface-structure phonology is derived from an underlying hidden structure). Let's just briefly examine how MOVE might correlate to the functional feature of CASE (nominative, accusative, genitive) assignment.

Theory Internal consideration: <u>All functional features/projections must involve movement</u> from out of the base-generated VP/NP, (the VP/NP being a first result and product of simple merge).

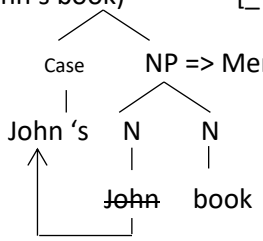
[29] Let's begin with two lexical items (they can be both Heads (H) for the time being). Consider the merging of [[John] [book]]. In this simple [N]+[N] merge operation (absent of any MOVE), the two Hs are considered flat-sequenced, base-generated and thus cannot generate any formal functional features such as Case.

(1) NP => merge



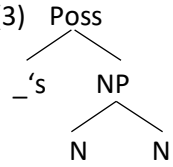
In this instance of merge, genitive/possessive case can't be assigned.

(2) Poss => Move (John's book) [_'s [John, book]]



It is rather 'Move', a recursive structure, which generates Genitive/possessives:

(3) Poss



[30] What we find here, theoretically, is that MOVE is responsible for triggering possessive (Genitive) case marking. One further speculation (see analysis below) is that such functional case marking: {-'s}, {-to}, and {-m} function as *clitics* (bound morphemes) which directly insert, perhaps directly pulled from the lexicon, (sometimes merely as a feature, but often as a lexical item itself, as in the case of 'to') which in turn motivate the raising of a lexical host (as specified by the Head information to search-out a PG relation up from a lower position within VP/NP.

[31] What we know of pidgin syntax, word order is often variable (e.g., Bickerton 1990, also see Galasso 2003, 2016, 2018 for accounts and analyses of early-child mixed word order). Theory internal considerations speculate that at the exclusive merge-level—what we would find of proto-language—no word order can be fixed since both Heads (H) {x, y} or H and Phrase (P) {x, yp} would serve as flat sister relations with no hierarchical dominance. In other words, in brief, fixed word order must be the result of MOVE {x {~~x~~, y}}, a recursive property, a property only seen in full-blown human language.[24]

[32] Given that pidgin, as well as very early child utterances, lack a fixed word order (at the early multi-word stage) (e.g, *me car, car me*, (= 'my car'), *mommy sock, sock mommy* (= 'mommy's sock'), etc.), this seems to suggest that pidgin, early child language would fall somewhere on the spectrum close to a protolanguage, <u>if what we mean by a proto-language system is that which is devoid of any formal movement operations</u>, and is a system which only employs merge.

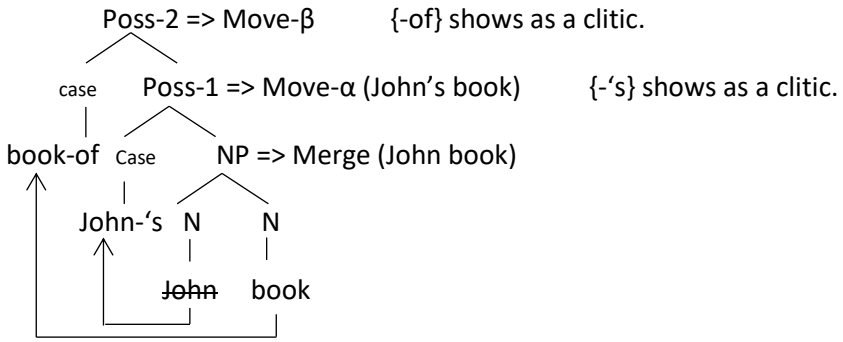
[33] What's also very interesting about the analysis above (and explicitly advanced in Bickerton's 1990 syntax) is that this would explain double-possessive markings found in examples such as [The book of John's], where possessive Case for 'book' seems to be marked twice. Let's see how this might work:

(a) [John 's [~~John~~ book]]

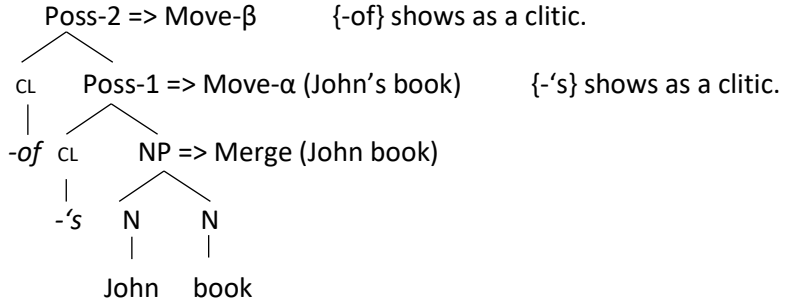
(b) [of [John 's [~~John~~ book]]]

(c) [book [of [John 's [~~John book~~]]]]

[24] See web-link no. 28. On Merge vs. Move in child language



[34] So, the **clitic-climbing** is expressed as shown in below, with lexical items raising up to attach to the CLitic (CL) as in a PG relation (in this sense, clitics and/or features of clitics force raising).

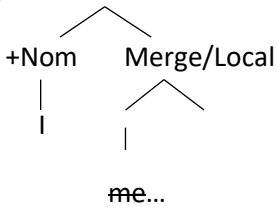


Let's consider further examples of how MOVE & CLitic PG-relation triggers and projects POSSessive case.

In sum, our analysis of how MOVER triggers Pronoun (PRN) Case assignment shows as follows:

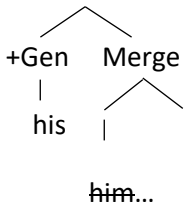
[35] **Case**

(a)



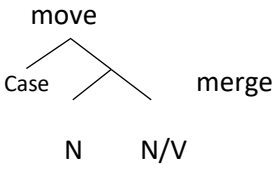
(= I like John) where PRN *I* is nominative case.

(b)



(=His book) where Genitive Prn *his* is POSS case.

So, Case marked syntactic tree looks like this:

(b) 

[36] Move-based Case: (Where accusative is default case):

a. From accusative to nominative via Move: [I [~~me~~ do it]], [He [~~him~~ do it]],

b. From accusative to genitive via Move: [my [~~me~~ dolly]], [his [~~him~~ car]].

Case in Double Object Constructs

In double-object constructions, when PRNs are employed—which must present overt case marking—e.g., I, me, my / *he, him, his / they, them, their (*Nominative-subject, Accusative-object, Genitive-possessive)—we see how MOVE can trigger Case assignment. Consider the distinction between the two sentences:

[37] (a) John gave…[Mary] [money] (b) ….*[money] [Mary]

NP NP

(Indirect Obj Direct Obj) (DO, IO)

[38] a. 'Give [him money/it]!', whereas b.*'Give [money/it him]!' is *unacceptable due to there being no 'PG-relation' to enable the Case-marking to be checked-off on the pronoun 'him' (noting that pronouns in English need the checking-off of the overt Case-marking feature, unlike Nouns which require no Case). It can be argued that only in (a) does the PRN *Him* remain in a PG configuration whereby it can receive and check-off the accusative [-Nom]feature structurally via the verb 'give' (cf. [27, (5a,i)]. But nothing hinges on that treatment: otherwise Case is acquired via default. Also note that there seems to be a preference for the structure in 'Give him money' over 'Give him it'*?, suggesting that the PRN 'it' is more sensitive to the right configurational position leading to Case-marking than is its Noun counterpart 'money' (again, since Nouns in SE don't require case-marking). (Below we note that in order to save the derivation found in b., 'him' must raise to be structurally adjacent to the verb, a position that would be forced by a PG-relation in any event).

'Him raising'

[39] In addition to a potential default setting (where configuration of word order is no matter), another work-around may be to assume that in order to save the ungrammatical derivation in *(b), movement of the Noun 'money' found in (a) optionally can be employed, allowing for the case-marking morpheme {to} to attach to an appropriate stem—now, the case-marking clitic {to}, attached to the N, serves as the probe of a PG-relation, attracting the goal PRN 'him' to raise in a local/adjacent domain in order to receive Case.

'Give [money-(to) [him ~~money~~]]!' Probe-Goal relation.

[40] So, restating what was said in (5), we have three ways in which the PRN 'Him' gets case marked:

(1) Structural via the verb {give}: a. John <u>gave him</u> money (to) ~~him~~. = 'Him' raising

(2) Morphemic via the clitic {to}: b. John gave <u>money-to him</u> ~~change~~. = 'money' raising

(3) Lexical via [Verb + Preposition {to}]:

c. John <u>gave (to) him</u> money. = (similar to (a))

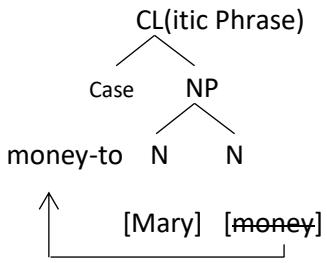
d. *John gave money him. = *(unacceptable)

Lexical Case must derive from a [Verb+Prep] configuration: *John [V gave [PP to him]] money*, as separate from [Noun + Clitic], as found in the structure: *John gave [N/CL money-to [him]].*

(4) whereby the ungrammatical sentence in d. **John gave money him* is due to there being no case marking PG-relation for the PRN 'him', it being stranded in a non-configurational manner without local domain to check-off its Case feature (i.e., neither morphemic, structural, nor lexical configuration is available, and a putative ACC default can't be employed within a structural configuration—viz., an accusative 'default status' can only emerge as a result of a non-configurational environment (i.e., 'no structure').

This is exactly what we would expect if Case were a functional projection triggered by MOVE (and an operation lacking at the early stages of child language, pidgin, and certainly not found in non-human primate communication). So, with raising (MOVE) we derive DO-IO and render the sentence:

[41] a. John gave money-to Mary.



With the underling structure showing PG relation between {to} & Mary, and raising of money to serve as a host to the clitic {to}:

[42] John gave money-to [Mary ~~money~~].

(Recall, as noted in our **lexical case-marking** treatment above, another work-around involving Case via prepositional {to} is to analyze the Verb Phrase (VP) [give+N] as the right kind of projection that can allow prepositions as its complement, thus allowing case marking to be applied lexically (**lexical case** given that Heads of PP can only assign Accusative [-Nom] Case in SE).

(Again, where {to} now serves as a case-marking clitic, this brings to the number of **morphemic case-marking** clitics to three: *to, of, 's*).

The PRN 'Her' *Mary* must be case marked by clitic {to}, just as in the other cases involving {of} and {'s}.

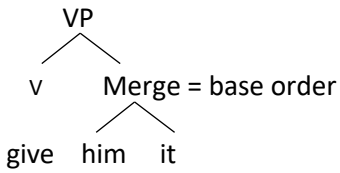
Note again how we cannot say *'*I want to give money Mary'.* This sentence is unacceptable since neither the proper name 'Mary' along with its object/pronoun counterpart 'her' wouldn't be case marked. But both *I want to give money to Mary/I want to give Mary money* are fine.

Note: The sentence *I want to give Mary money* brings our attention to the problem of how *Mary* gets case marked—since no necessary movement has been employed out of base-generated [IO, DO], and lexical {to} Case checking is absent. Well, Mary is first of all, not overtly marked, (only Pronouns in SE get overtly marked, perhaps explaining why we can't say *'I want to give <u>to Mary</u> money'), and secondly, we have a mechanism easy enough that can take care of Case, that of Structural case via the verb 'give'. [Give [Mary]]

is in the structural local domain where the verb ‘give’ is allowed to case-mark its complement/object ‘Mary/her’. Since lexical case-marking is present, the extra insertion of case-marking {to} would be redundant, and thus ungrammatical as seen in the contrasts between the two structures *‘I want to give <u>to Mary/her</u> money’ vs. ‘I want to give money to Mary/her’).

In summary of this section, a structure such as ‘give it to him’ derives the following steps:

[43]



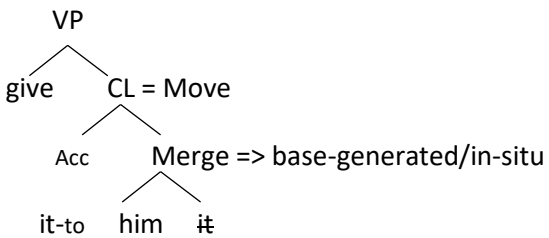
(1) Note how *‘give to him it’ is unacceptable, while ‘give it to him’ is acceptable.

(2) The ProNouns ‘him’ is case-marked via base-generated/in-situ verb ‘give’ (or otherwise by default). But the PRN ‘it’ must also be case-marked (overtly so). So the PRN ‘it’ must raise to clitic ‘to’ (morphemic case marking). Note the grammatical contrasts of

(a) **give it him.* (where PRN ‘him’ is left stranded without a case-assigning configuration)

(b) *give it to him.* (where PRN ‘him’ now receives a proper case-assigning configuration via CLitic {to}).

(c) *give him it => base-generated order [I0, D0].*



In the illicit example found in (a), case marking cannot be doubly assigned to PRN ‘it’ because the dative verb ‘give’ has (earlier) already assigned case to ‘him’ via base-generated/in-situ order (‘Give him it!’). (Ex. C).

[44] Given that the morpheme {to} now has a dual status (preposition and case marking), let’s consider the contrast regarding phonology/stress (schwa reduction) between the following clitic case-marker {to} versus the prepositional {to}—two very different items:

(a) John gave (*to*) him coins. (base order): [V+PP] => {to} is a Preposition (no schwa reduction)

(i) *John gave him them, (ii) John gave them to him

(In (a) above, it appears that when the Prep {to} is removed, and case of ‘him’ can be assigned via the verb, what happens is that the PRN ‘them’ is then left stranded without a case-assigning configuration. The structure in (ii) corrects this).

(b) John gave coins-*to* him __ : N+CL => {to} is CLitic case marker (schwa reduction is allowed).

In (a), ‘to’ is a preposition and no schwa reduction is observed; ‘to’, if pronounced, must either be pronounced with stress /tú/ or may be deleted (whereby ‘him’ receives structural case). (Recall, the Noun ‘coins’ doesn’t have to raise to be overtly case marked, and so it may remain in-situ in base-order with default case. Only Pronouns in SE get overtly case marked). In (b), the stress is weakened to a schwa.

(c) (i) John gave it-to him. {to}-Clitic

(ii) *John gave to him it. {to}-Prep

Recall, that for the example (c) above, ‘him’ has already been case marked structurally (in base-generated order) via the verb ‘give’, (cf. [40 (1)]) (*and so the prepositional structure in (c, ii) above is redundant for case marking). But the PRN ‘It’ still must be overly case marked, so raising to the clitic {to} (as in a PG relation) is forced. Note that {to} in this respect found in (c, i)

is a ‘case-marking’ clitic, and not a preposition, noting the possible phonological/stress ‘schwa reduction’: e.g., ‘John gave it /tə/ him’ as opposed to ‘John gives to/tú/ him so much money’.

Agreement

[45] Another formal feature we could consider is Agreement, which contains the dual features of Person, and Number. For instance, if the subject ‘He’ is [3rd person, singular], then the verb ‘speaks’ must match (so-called subject-verb AGReement): e.g., ‘He drives’. We could speculate that the default setting here is the non-affix Infinitive-verb stem of ‘speak’, so that a child at the very early multi-word stage might say ‘Daddy drive’ whereby no AGR features is present. Coupled with a default Accusative case (as suggested above), we might expect to find utterances of the example ‘Him drive car’, and we do. Let’s consider some token examples of such utterances dealing with the absence of Case and AGR below, as we draw our attention to what a speculative Proto-language would sound like.