Progress in syntax as of quite recently
Chomsky 2013 and "problems of projection"…

This is in some weird mixture of my words and Chomsky's, not really differentiated—some of this is quotes transcribed straight out of the paper, some of this is commentary or paraphrase by me. Read the paper to find out which is which if you are not sure.

1 History and science

In the introduction, Chomsky basically reviews preceding notions of language, taken to be idiosyncratic and infinitely variable. Hidden within here is some defense/critique that Chomsky goes on at more length about elsewhere, though the feel is essentially that those who are still operating under such a view have not really thought it through.

Parallel to biology: consensus understanding has moved from a view in which organisms may vary in practically unbounded number of ways, to one in which there is a small set of "building blocks" upon which biological organisms are based.

> From this perspective there is only a single multicellular animal from an appropriately abstract point of view.

This is something Chomsky says fairly often about language too, although not here. I had some trouble chasing down something to cite, oddly. But, the story is essentially that a Martian studying human language wouldn't see the differences, that it would all look like one language.

The "non-existence" thesis (referred to occasionally in what follows) is the idea that language isn't anything more sensible to study than today's weather is—it arises from other, studiable things that have independent functions that happen to come together in language. Induction, analogy. These are generally the approaches that "oppose" the idea of UG, but without credible proposals to handle the results that have been achieved under the UG assumptions, there's not much to evaluate. And, besides, "no one can rationally question that there is a genetic component to the language faculty"—which is what UG is—what are we even talking about?

> Each language incorporates a mechanism that determines an infinite array of hierarchically structured expressions that are transferred for interpretation to two interfaces: the sensorymotor system SM for externalization, and the conceptual-intentional system CI for thought (broadly understood).

It follows that each language incorporates a generative procedure GP of some sort that characterizes the internal expressions and provides the appropriate "instructions" for the interfaces, by means of its transfer mechanisms. UG determines what qualifies as a GP for some human language.

UG in the technical sense of the generative enterprise is not to be confused with descriptive generalizations about language such as Joseph Greenberg's universals...

2 Binding theory as an example of the scientific approach

Taking *each other* to have *they* as an antecedent: a general effect of trying to resolve ambiguities rapidly? That is, essentially, take the most recent antecedent you heard.

(1) do they expect to talk to each other?

But *John* can't be an antecedent for *each other*, so isn't *they* below still the most recent option? Yet it isn't good enough, the sentence (2) is bad. And (3) has (1) within it, but the interpretation has to be different: *each other* can't refer to *they* anymore but somehow has to refer to *who*.

(2) * do they expect John to talk to each other?

(3) who [do they expect to talk to each other]?

If we spell out the logical meaning of (3) as in (5) (below a bit, in order to preserve the numbering in the original paper), we see that there's a "second *x*" in the same place in (3) that *John* was in (2).

The difference between (4a) and (4b) is mysterious unless we start getting into the details of where PRO is, what controls it, how the verb type matters, etc.

(4) a. * They gave instructions to John to talk to each other

b. They received instructions from John to talk to each other

(5) For which persons x, they expect persons x to talk to each other?

At first, linear order was taken to be important for establishing these kinds of binding relations, but further study indicated that it's actually not linear distance but structural distance that matters. Which we can take even further as a hypothesis:

[Thesis] (T) Order and other arrangements are a peripheral part of language, related solely to externalization at the SM interface, where of course they are necessary.

This winds up meaning that language is meaning with sound, rather than sound with meaning, if we're designing bumper stickers. And the "non-existence" thesis isn't faring well. Though there is one methodological thing right about it, which is that we want to reduce our assumptions about what constitutes UG as much as possible. We just know that we can't get it down to nothing. But, the simpler the assumptions the deeper the explanatory force.

Three factors (which again recur in the discussion) in studying problems of growth and development:

1. Genetic endowment
2. External data
3. Organism-independent factors (natural law, physics, etc.)

> Inquiries into UG have to deal with several fundamental features of language. There are first of all, both continguous and non-contiguous relations, the latter including the ubiquitous phenomenon of displacement—the fact that a phrase is interpreted both where it appears in surface forms and in some other position—and morphological discontinuity (including agreement). Another prooerty is linear order. Still another is identification of the category of a phrase (projection, more recently called "labeling").

Projection/labeling is not directly evident, so care is needed.

> In recent years, work on these topics has often been called "the minimalist program (MP)." The term has been misunderstood. The program is simply a continuation of the efforts from the origins of the generative enterprise to reduce the postulated richness of UG, to discover its actual nature... The literature contains many criticisms of the MP, including alleged refutations, charges that it is not truly minimalist, and so on. None of this makes any sense. Research programs are useful or not, but they are not true or false. The program might be premature, it might be badly executed, but it is hard to see how it could be fundamentally misguided, since it hardly goes beyond holding that the study of language should keep to standard norms of science.

3 Minimal distance is always minimal *structural* distance

In our discussion of eagles, we narrow our focus to just those that fly. In order to do so, we use a relative clause, so we have a noun phrase with the structure [eagles [that fly]].

The question in (6a) is asking about swimming capacity, not flying capacity, though *fly* is closer in time to the interrogative signal (starting with *can*). The adverb in (6b) is associated with the more temporally distant *swim* than with the closer *fly*. If you put the question in the progressive form, *swim* gets the morphology (7) and not *fly* (8). So, temporal/linear distance can't be what's at issue here, at least if things are governed by principles of "minimal effort" or "minimal distance." But minimal **structural** distance, that gets these cases. *Fly* is buried inside *eagles that fly* and is structurally further from C and from that high-attached adverb.

(6) a. can eagles that fly swim?
b. instictively, eagles that fly swim

(7) are eagles that fly swimming?

(8) * are eagles that swimming fly?

Now, an interesting and relevant point: whenever one finds a kind of "minimal distance" effect in language, it is always the minimal structural distance that is relevant. Of course, it always could be. But what's odd is that it's never minimal linear distance even when that would otherwise be logically possible and as easy if not easier to make the generalization on. People's brains light up in a language-like way when learning artificial languages that conform to this, and in a puzzle-solving-like way when learning artificial languages that don't. So, thesis (T). SM considerations (like linear order) never play a role in the computation, only in the externalization to the SM interface itself.

4 Implementation: Merge

We need hierarchy so we need at least Merge to take two things we already have, X and Y, and make a new object Z. In the interest of minimal computation, we will suppose that neither X or Y is modified ("No Tampering Condition") in this process, and X and Y are within Z unordered. So:

Merge(X,Y) = {X, Y}

For example, *read books* as *read*, *books*. We can differentiate two sub-cases, depending on whether Y is a part of X (WLOG) or not. Where neither X nor Y is a part of the other, we have External Merge (EM). If Y is a part of X, then {X, Y} contains two copies of Y (one Merged with X, the other within X), and this is the case we'll call Internal Merge (IM).

Both operations (EM, IM) come for free: it would require stipulation to bar either of them.

To head off an immediate objection one might have to this, if "John left" is really something like "John John left" (because we moved the subject up), then what about "Who likes who?" (why isn't it "Who likes?") or "I told him that Pat likes him" (why isn't it "I told him that Pat likes"?)? Answer: Don't be ridiculous. Of course we need to distinguish copies taken out of the lexicon independently from copies that arise in the course of the derivation from a single lexical access.

> IM yields displacement (movement)—in fact, in a form appropriate for the CI system. Consider, for example, the sentence "which books did John read." Here the phrase *which books* has two semantic roles: it receives its role as object of *read*, just as in the case of *read books*. And it is an interrogative operator, binding the variable in the object position, so that the interpretation is something like "for which books x, John read books x." That is read off directly from the generated structure "which books did John read which books," the product of IM.

Well, yes, sure, ok. In that one case. We're going to lose the "in a form appropriate for the CI system" motivation, at least in this form, shortly though. The case of *wh*-movement (or covert quantifier movement) could almost be said to be required in order to get the logical form correct. We're about to run into cases where, although movement is driven by interpretation in a sense, it's driven not by the need to get something to fulfill all of its logical roles, but rather by the need to even be visible/usable by the interface at all. But continuing a bit anyway on this road:

(9) a. [which of his pictures] did they persuade the museum that [[every painter] likes best]?

b. [which of his pictures] did they persuade the museum that [[every painter] likes [which of his pictures] best]?

The structure in (9b) is externalized as (9a), where the lower copy is deleted, but *which of his pictures* is still understood to be the object of *likes*, just like *one of his pictures* is in (10). And, the reason this example is constructed this way is that there is a *his* here that is bound by *every*. So, on the surface, *his* is not in the scope of *every*, but the interpretation with *every* binding *his* (so, with a different picture for each painter) is still available. Something we only expect under c-command. So in terms of the logical interpretation, the *his* is still in the c-command domain of *every* (and there it is, in the second copy that we didn't pronounce), even though in the externalized sentence it is not there but only higher. So, this is evidence of the structure being basically in both places, higher (because we pronounce it there, and because it contributes to the quantification as an operator) and lower (because it contains something that can only be bound in that lower position). This is actually kind of a common trick when constructing examples.

(10) they persuaded the museum that [[every painter] likes [one of his pictures] best]

To finish off the agument that the c-command of the lower copy is relevant, if we construct an example where even the *his* in the lower copy (if there is one) would not be in the scope of *every*, you can no longer understand *his* as varying with the painter.

(11) [which of his pictures] persuaded the museum that [[every painter] likes flowers]?

Minimal Computation: pronounce as little as possible? So, for SM, this turns into: pronounce only one of the copies, the one most structurally prominent (highest).

Let's talk about fixing cars. Here's a sentence in (12). It has an embedded interrogative, the object of *ask* is a question (it relates an asker to an asked thing, that is, an individual to a question). This is an island, an interrogative island or "*wh*-island."

(12) they asked if the mechanics fixed the cars

If we want to ask about how many cars or how many mechanics are involved, it turns out that we can ask one (13) but not the other (14). This is a difference between being a subject and being an object, and the principle from the olden days that described/predicted this is the ECP. But so if you want to ask (14), you can find some other way to do it, but you can't do it that way. The computational system has impeded what would otherwise be simple communication.

(13) how many cars did they ask if the mechanics fixed?

(14) * how many mechanics did they ask if fixed the cars?

When computational economy and comprehensibility get into a fight, computational economy wins. Language is meaning with sound, not sound with meaning.

> In earlier years it was assumed generally (by me in particular) that displacement is problematic, an "imperfection" of language, to be explained by external factors, perhaps parsing considerations. A residue of that error is the belief that EM is somehow simpler and preferable to IM. There is no basis for that belief. If anything, IM is simpler, since it requires vastly less search than EM (which must access the workspace of already generated objects and the lexicon). But the question doesn't arise: both are freely available, given the simplest combinatorial operation, Merge.

This part has a fair amount hiding behind it, and I don't yet see quite how it's going to all fit in. The discussion from prior work basically had this character: Merge is cheaper than Move because Move is conceptually Merge *plus* something else (like Copy). So, at

any point if you have the option in a derivation to Merge *or* Move, Merge will always win. The evidence and argument for this centered largely around "expletive constructions"—in English, sentences like "there is a fly in my soup" (as compared to "a fly is in my soup"). They differ in that in the first case, *there* has been used as a subject and so *a fly* does not move to subject position (at least in any overt way), whereas in the second one *a fly* moves. This is seen as being predicted because if there is a *there* around available for Merging, you Merge rather than Move. Though it's more relevant for multiply embedded structures: "there seems to be a fly in my soup" vs "*there seems a fly to be in my soup." The second one would have been derived by Moving *a fly* first (despite the fact that there is a *there* around) and then only inserting *there* later, but that's not an option if you have to Merge if you can, rather than Move. What Chomsky's saying now, it seems, is that there is no preference between Merge and Move, you can do either one because they're equally complex, they're the same operation. Yet he did make passing mention of the possibility that Move (IM) might even be preferred over Merge (EM), but seems to dismiss this. The question now is, if we go down that path, how do we rule out "*there seems a fly to be in my soup"? Perhaps we'll find out. Likely we will. Chomsky will have known that once he said/wrote this part, everyone would have traveled down this same path I just traveled down, and would be raising their hand and waving it wildly barely able to restrain themselves from blurting out "What about there seems a fly to be in my soup???"

5 Phases and cycles

Some discussion of cyclic operations historically, but the basic idea wraps up to: once you form an object in the computation, you don't change stuff inside the object later. Computational economy again, you can kind of "set it and forget it." This also makes a lot more sense if you are building structure from the bottom up rather than from the top down.

There is a kind of a leap here where we arrive at the idea that there are certain positions in the building of structure that count as "strict cycles" or "phases"—the leap is that it isn't just everywhere, but rather at certain specific points. Early papers had identified these places as being CP and probably *v*P and probably DP. So, the idea is that once you reach the edge of one of these phases, you can't change anything inside.

Chomsky proposes that syntactic operations are all limited to these phase levels, apart from EM which builds the phases up to those levels. In particular, IM (movement), agreement, and "transfer" (Spellout—fixing the pronunciation and semantic interpretation of the chunk of structure that's been built) can only happen at the edge of phases. This in fact is perhaps part of the motivation for *not* saying that everything is strictly cyclic,

but only the phases are—if this is the place where transfer happens (and in particular the determination of the pronunciation), it could be considered to be an illegal internal change to mark a lower copy as unpronounced after it's already been transferred to the SM system. Seems to me there are other ways to think about it that wouldn't necessarily have this consequence, but we can follow this path anyway. This means that anything that's going to wind up being a trace (anything that is going to have to move on) needs to be out of range of the transfer operation. So, it will need to be "moved" (via IM) to the "edge" of the phase, outside the part that is being transferred. That way, what gets transferred to SM only has something in it that is a second copy of something that is more structurally prominent and so would be designated as not being pronounced.

So, take H to be the head defining the phase (the "phase head"), e.g., C. Its complement Z is what gets transferred and can undergo no further modification. H itself and stuff Merged with {H, Z} (specifiers of H) are not transferred at this point, and are available to be moved higher if the need arises.

While Z is immune from further changes, it does not disappear. If a SO (syntactic object) containing Z is raised itself, it'll appear in the new position it raised to. The example given is "The man who said that Z was elected"—that's easily mis-parsed, but the idea here is this: we have a passive verb *be elected* and *the man who said that Z* is the (deep) object that later raises to become the subject. In the construction of *the man who said that Z* we have at least one phase at the CP ("that Z"). So Z there would have been transferred. And probably also the whole thing would have been transferred at DP ("the man who said [that Z]"). But then when we get up to the main clause TP and we need a subject in the specifier, that deep object moves to become the subject. So we wind up ultimately with something along the lines of

[the man [who said [that Z]]] was elected [the man [who said [that Z]]]

And the point here is that even though those lower phases were transferred in the lower position, they're still available in the higher copy, we don't wind up (necessarily) saying "the man was elected who said that Z" or "the man who said was elected that Z". It's a narrow little point, presumably responding to a question/critique brought up in the past.

Then there is some discussion of the relation of X-bar theory to the Merge-based phrase structure system. We can say X-bar like things in Merge terminology: Merge (the first time) is always to a head, which projects (meaning: determines the properties of the combined unit), and one or more "specifiers" result from second and further Merge operations in which the same features (originally from the head) determine the properties of the combined unit. Chomsky is now confronting a quesiton which was kind of at the periphery of earlier discussions: When you do something like Merge a DP with TP (to

fill the subject position, say), why is the result a T-type thing (where T projects) rather than a D-type thing (where D projects)? Previously at least in the simplified version dealt with in Adger's textbook (and my LX422/522/722), we can say that it was the one that needed the other one that has the honor of projecting. Still a stipulation, though, but it seemed like a reasonable generalization. T needs the subject, so T is the one that projects.

(15) $[_C$ C $[_T$ DP $[_T$ VP]]]

Here's the hope Chomsky now expresses: it would be simpler not to have to stipulate things like Specifier and "endocentricity" (having an internal head). Again: destroy what people's analyses rely on so they're going to need to rethink everything. There's going to be something that takes its place, but it'll require some kind of reorienting that might yield new insights. That's my prediction.

But, the idea is that as far as Merging is concerned, we wind up with something like:

(16) $[_C$ C $[_\alpha$ DP TP]]

where $_\alpha$ is not necessarily TP (meaning that DP is not necessarily the specifier of TP). Supposing now that we have something essentially like {DP, TP}, where there is no particular property of the unit apart from the fact that it contains a TP and a DP.

Observationally, we know that C interacts with T. There is subject-auxiliary inversion, for one thing (where T moves to C). And if we don't prioritize the DP from the TP (that is, if we don't say DP is within TP insofar as DP is in the specifier of TP), then the DP and TP are equally far away from C. Why then do we not have the option of moving the head of the DP instead of the head of TP? (And here, my translation of NP into DP is running into a kind of an obstacle, his examples assume the N is the head of the DP, when he asks why one couldn't form a yes-no question like: "eagles young are flying?").

The conclusion Chomsky reaches here is that C and T interact before DP enters the picture. So, there is a point when we have [C {T, XP}] and at that point T is closer to C than X would be, I guess.

Although it kind of seemed as if Chomsky had abandoned labeling by this point, we now return to it, with the idea being that something can only be interpreted at the interfaces if it is known what kind of thing it is. So everything has to be some kind of thing. The simplest approach would be to say that it's the same kind of thing as one of the things it contains. So if you have {H, XP} where H is a head and XP is not a head, then the closest available "kind of thing" is going to be H's "kind of thing." So, H projects, H is the thing that labels the unit. That gets us to what used to be the "bar-level" of a projection. Once you've done one Merge though, the next Merge is going to be the Merge of two XPs, neither of them heads. This is the situation when the subject raises. {XP, YP}. Either X or Y are available as labels, neither is closer.

Chomsky suggests that if you have an {XP, YP} situation where the head is ambiguous, this is going to be a problem for interpretation. We need to be able to label this unit. So, one way to do this is to "remove" one of them so that the other one gets to be the label. Continuing further out along this limb, we could suppose that if we do an IM a bit higher up with XP, then XP is now "discontinuous" in a way that means that XP is no longer fully contained within the original {XP, YP}. The example is "XP copula $\{_\beta$ XP, YP}"—now XP isn't there and YP can provide the label for β.

(17) T $[_\beta$ (EA) [v [V IA]]]

This does seem to be kind of out on a limb, as Chomsky says next: it may seem to be a stipulation to take the whole chain to be the relevant SO. But, no, it's natural, what are you talking about? Not only is it natural, it does what we need it to do. So, whether it seems to be a stipulation or not, it isn't, stop bothering me.

A more familiar case is raising the subject to the specifier of TP. The "*v*P" under T has the form {XP, YP} as well, and to get the result and provide a label for (what will then be) *v*P, we need to move the subject. So the EPP there turns out not to be something that satisfies some need of T, but rather some need of the interpretive system to have a label for the thing that got Merged with T. Neat. A brief discussion of the possibility of raising IA (the deep object) instead is mentioned, and the claim is that (since IA then becomes invisible) will still yield the right result. I get the feeling here that you need to stand across the room and squint for this to work without making some very specific assumptions about what is "visible to labeling" but the general character of the analysis Chomsky has in mind makes sense. [Ah—and, actually, this is acknowledged somewhat in footnote 34.]

(18) XP copula $\{_\beta$ XP, YP }

Successive cyclic movement has things like *wh*-phrases stopping in intermediate positions, like "Specifiers of CP"—but this puts them in a position where, if they don't move on, the thing they're part of cannot be labeled, so it makes the next step obligatory. Though it's not clear to me yet exactly why this problem doesn't arise at the top as well, where the *wh*-phrase ultimately ends up. Slightly further down, after a little bit of discussion of quiz questions, the proposed answer arrives: there's a special case, which is that if in {XP, YP}, both XP and YP share the same "most prominent feature" then the label can be in a sense coming from both of them. In this case the interrogativity that is part of the *wh*-phrase and the interrogativity is that is part of the interrogative C counts as the same ("Q") and so we can just label the whole thing "Q". This seems a little bit fragile, for one thing it means that we need to suppose that the feature that makes C interrogative is the same as the feature that makes a *wh*-phrase interrogative, and we need to assume

that Q alone is enough of a label to satisfy the interpreting interfaces, but as usual, the endpoint of this problem that arose was anticipated.

Having anticipated and dealt with *wh*-movement, Chomsky returns to moving the subject to "specifier of TP." This has the same problem, it's an object of the form {XP, YP}, and—unless it can be supposed that the interfaces don't need to interpret this (and so there's no need for a label)—something needs to be said about how it is labeled. Vague suggestions are made that it does need to be labeled. And then a hopeful proposal that we can use the same kind of solution here, in the form of "subject-verb agreement." By supposing that the subject and the YP that it's Merged with share some features (specifically ϕ-features), we can suppose that the unit is labeled with those ϕ-features and that this is sufficient for interpretation. A note here attributed to Marcel den Dikken indicates that it's not going to be sufficient just to assume that we wind up with matching features and that provides a label, but rather we need to have "actual agreement"—something that Chomsky relegates to some other discussion some other time, having to do with clarifying what's different between matching and agreement and how Agree is technically implemented. I'm not entirely sure what the issues are here, I have some guesses.

And again, Chomsky returns to answer a question that arose earlier—labeling is something that cares about features not full lexical items. It is ok to label a syntactic object with something that is not a complete lexical item but only a feature of it. Agree also cares about features. There's a mention here of a "probe-goal relation" which is considered to be part of Agree, Agree from the perspective of a "probe" looks down into the structure for a "goal" which either matches or can be made to match by filling in feature values. In some technical ways.

So, we decided that Q and ϕ-features were sufficient as labels, but not every feature is good enough to be a label. And we start talking about coordination.

Briefly the idea of "pair-Merge" is introduced, which is structurally the asymmetric <X, Y> rather than the unordered {X, Y}, and Chomsky suggests this is useful for unstructured coordination like "John is unhappy, hungry, and bored with TV" and also for adjunction generally. This is mostly set aside at this point.

For "structured coordination" like "John and Mary" the proposal is that it starts as [and [John Mary]] and we do the now familiar trick of moving *John* out over *and*. So we wind up with

[$_\gamma$ John [and [John Mary]]]

and we need a label for γ. We know it is going to be that of "John," observationally. Conclusion: whatever *and* is (Conj, say), is not a valid label. So, if Conj is not available as a label, we have essentially {XP, useless-thing}, and the label will be that of the XP.

There is some speculative commentary about the situation in which IA raises (yielding something along the lines of "the book will John put on the table" though I don't

think this is a proposal designed for English), and how this won't really work unless we assume that "put on the table" winds up being a kind of complex verb, implemented as a verb with "on the table" adjoined to it, implemented as Pair-Merge. Then even more brief and speculative comments about how this might apply to double object constructions, but the basic subtext here I think is that Chomsky has thought it partway through but it would be a good project for somebody to take it and really work out the details.

And then a few remaining cases are rounded up. If we have {H, XP}, this works only if H is not complex, but what does that mean to be complex? A lexical item H has several features within it, and there's some reason to think that these features have a computational life of their own, independent from other features in the same lexical item. So, even H is complex in some sense that needs to be worked out properly. On the other hand, pronouns and single word nouns like *he* and *John* need to be complex, so there must be structure we can't see there. Touching on the possibility that in "I like what you wrote" we have *what* projecting its nouniness to make it an appropriate object, rather than winding up being a CP that doesn't seem like a good complement for *like*.

Next, a brief comment on one thing Chomsky had set aside a long time ago in the paper but said he'd come back to, the case where we have something like {X, Y} where both X and Y are heads. There's a kind of opaque reference to a Marantz/Borer view on this, which maybe I'll talk about in class, but it's basically one kind of interpretation of what *v* is actually doing (and serves as motivation for an *n*, and other similar syntactic heads). The idea though is that V itself is like Conj insofar as it can't provide a label, only *v* can provide the label.

ECM constructions: very cursory mention, but the idea is that in "I consider John to be intelligent" *John* moves up out of the embedded clause in order to allow the embedded clause to be labeled, which itself winds up causing another labeling problem, potentially solved by moving *consider* up even further to get it back to the left of *John* again. Loose ends remain.

ECP cases (discussed earlier but skipped here in my notes): you can't move a subject *wh*-phrase out, but can move an object *wh*-phrase. Perhaps this is due to the Q feature in C being inherited by T, thereby putting the subject *wh*-phase and the Q in T in their "criterial" position already (used in labeling the combination) and so preventing the *wh*-phrase from being moved away.

[I should go back and put the ECP examples in here, but if I don't, they're in the paper around (13) and (14).]

Same kind of explanation could help with the *that*-trace effect, if we make a string of assumptions that includes the idea that if *that* is removed, C is "weakened" to the extent that it no longer has a force feature and doesn't cause "criterial freezing" due to labeling.

End of paper.