Theoretical linguistics today

Jan-Wouter Zwart

Keynote address T.W.I.S.T. Conference
Leiden, April 22, 2023

draft

In the grand scheme of things, theoretical linguistics is a relatively young scientific discipline. If its beginning is marked by Saussure's *Cours de linguistique générale* (1916) — as seems reasonable — then the field is just over a hundred years old. Starting out as 'general linguistics' (Algemene Taalwetenschap), many subdisciplines have now emerged, such as computational linguistics, psycholinguistics, neurolinguistics, applied linguistics, and communication studies, which to many seem more relevant, and consequently attract more students and young researchers, and stand a much better chance of gaining research funding and administrative support. Especially in the guise of Chomskyan generative grammar, the general sense appears to be that theoretical linguistics has run its course, and that we should be focusing on these other, more exciting fields of linguistics research.

As a case in point, we may consider the current excitement over the fruits of articifial intelligent research, such as ChatGPT. In a recent opinion piece in *The New York Times*, Noam Chomsky, together with fellow theoretical linguist Ian Roberts and philosopher Jeffrey Watumull, argued that ChatGPT is little more than a parlor trick, and that the research that went into its very successful development does not bring us any closer to understanding human nature ("The false promise of ChatGPT", March 8, 2023):

> The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and extrapolating the most likely conversational response or most probable answer to a scientific question. On the contrary, the human mind is a surprisingly efficient and even elegant system that operates with small amounts of information; it seeks not to infer brute correlations among data points but to create explanations.

To each his own, Chomsky seems to say, but if we are interested in the human mind, and if that is what theoretical linguistics is all about, then we should be doing something else.

Some of the readers' reactions, however, were dismissive of Chomsky's critique, in an *ad hominem*-style rebuke, such as:

> Chomsky and his followers are deeply worried about programs like chatGPT (…) because they represent something that their life's work has long predicted was impossible: learning a language fluently based purely on examples. Chomskyan linguistics has long held as its core belief (despite vanishingly little evidence) that only an entity with "universal grammar" (an innate, hard-coded knowledge of the core rules of all human language) can learn human language. That is, that human language is not a function of general human intelligence (like every other aspect of modern human achievement), but a specific function of "universal grammar". This was powerful hypothesis in the 50's when it emerged, but it has since morphed into an article of faith for Chomsky's linguistic followers with far too

> many counter-examples to remain a viable theory in today's day and age. ChatGPT and other simple neural networks, while extremely different from humans in many other dimensions, are doing exactly what this theory says isn't possible: become a perfectly native producer of English simply by observing an enormous amount of English data. Don't let the "moralizing" confuse you—this is the core of the hand wringing in this article and it's why two linguists and a philosopher are writing it.

There's a lot to unpack here, but the key point seems to be that Chomsky is acting out of sour grapes—ChatGPT proves that he was wrong all along, and he can't take it.

But it's important to point out that this is not a fight between disciplines (theoretical linguistics vs. Artificial Intelligence), but a debate on how to model human cognition, and this debate is also relevant within the AI-community itself. For example, Douglas Hofstadter has argued repeatedly that current practice in AI-research is taking us away from a better understanding of the

> mental mechanism that lies at the very center of human thought but at the furthest fringes of most attempts to realize artificial cognition. It is only thanks to this mental mechanism that human thoughts, despite their slowness and vagueness, are generally reliable, relevant, and insight-giving, whereas computer "thoughts" (if the word applies at all) are extremely fragile, brittle, and limited, despite their enormous rapidity and precision. (Hofstadter 2013:25)

Hofstadter argues for an approach to artificial intelligence that involves "basic scientific research about the nature of human thinking and being conscious" (Hofstadter 1995:1), and Chomsky simply argues for a similar approach to the study of language. The question that occupies us today is whether that approach to the study of language has run its course.

We should make clear what we understand by 'linguistic theory'. Science seeks "to formulate an increasingly comprehensive world view that is explanatory and predictive" (Hempel 1983:91). A theory (from the Greek *theōréō* 'view') is the formulation of such a world view, and it involves "a representation or description of a permanent structure which is responsible for the phenomena explained by the theory" (Harré 1970:2). The aim is for this description to be made in terms of universally valid propositions, or laws. However, it is important to realize that the 'permanent structure' we are after may be itself not available for inspection, in which case the object of inquiry needs to be modeled. In that case, the laws of science describe the *model* of the permanent structure which is responsible for the phenomena explained by the theory (Harré 1970:3).

That the object of inquiry cannot be directly observed, but needs to be modeled instead, is very common. As Lockhart (2012:295) remarks, "it would be hard to find any scientist (…) who is not in some way engaged in the process of modeling". We have a model of the universe, a model of the atom, and everything in between. The relation between the model and reality is an interesting question, to which we return, but modeling is an uncommonly succesful strategy in scientific theory, and has been so ever since Galileo's insistence on idealization in physics.

In linguistics, modeling began in earnest with Saussure. I say this because Saussure is responsible for a revolution in our conception of the object of inquiry in linguistics, which culminated in the Chomskyan notion of *I-language* or the Faculty of Language (grossly misrepresented in the *New York Times* comment section quoted above).

Within language (*langage*) Saussure famously distinguished between the language

system (*langue*) and individual linguistic utterances (*parole*)(see Saunders 2004:4-5 for terms, definitions and translations). The language system is a system of oppositions, and the elements featuring in that system obtain their value not from any inherent, physical properties that they may have, but solely from being different from other elements in the system. The point is this: the system of oppositions exists only in the mind of the speaker/listener, not in the sound waves that make up the physical reality of speech (Saussure 1916:27-28).

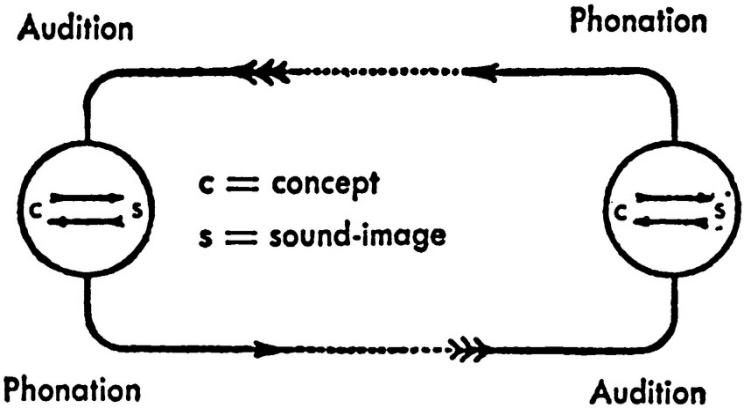


Figure 1. Speech model according to Saussure (1916:28)

In Figure 1, the linguistically relevant information is in the circles on either side of the diagram (corresponding to *signs* in the speaker and listener's language system). A sign (*signe*) is a unique and indivisible sound-meaning combination, that is construed (formed) by the speaker/listener out of the indistinct masses of sound and meaning out there. There is nothing in the properties of sound and meaning that motivates this particular formation. As Saussure (1916:157) points out explicitly, this is the deeper meaning of his famous concept of the arbitrariness of the sign (*l'arbitrarité du signe*):

> This perspective helps us better understand what was said earlier about the arbitrariness of the sign. It is not just that the two areas connected by the linguistic element are mixed-up and amorphous, but also the choice to appeal to this slice of sound for that concept is completely arbitrary. If not, the notion of *value* would lose some of its character, as it would contain an element that is imposed from the outside. But the way it is values remain completely relative, and hence the sound-meaning combination is radically arbitrary.

As we know, this was the origin of a new, structuralist way of researching language (as well as a range of other lines of inquiry), in which phonemes, morphemes, and higher order elements are analyzed not for their inherent, physical properties, but for their position in a system. Importantly for our discussion, that system is not in the data themselves, but in the mind of the speaker.

Saussure stopped short of explaining the system of language (*langue*) from any innate properties of the human mind. To him, *langue* was a social construct, although he admitted that there had to be a innate basis for establishing such a construct in the mind (*faculté de constituer une langue*, p. 26). But the important thing is that *langue* is presented as a newly defined object of inquiry in linguistics, which can be studied independently of *parole* (p. 31), even if it can itself not be inspected directly. Linguistic theory, therefore, is about the construction of a model of *langue*, which resides in the

human mind.

To Saussure, linguistic theory was part of a larger scientific enterprise, which he termed semiotics (*sémiologie*, p. 33), the study of signs in society. This was taken to be part of social psychology. To move from Saussure to Chomsky, all it takes is a different conception of psychology, from social to individual, and a further explicitation of the underlying biological basis. What remains is that the object of study in linguistics is the faculty of language, as a function of human cognition.

The Saussurian model of the faculty of language, a system of signs expressing oppositions, works well for (parts of) lexical semantics, phonology and morphology, as long as the observations to be explained revolve around semantic fields, phonological systems, and morphological paradigms. It works less well for those aspects of language that require productivity (see Harris 1987:150 for discussion). Foremost among these is of course syntax. What are the observations to be explained here?

Here, Chomsky (1964:7) offers the following compelling suggestion:

> The central fact to which any significant linguistic theory must address itself is this: a mature speaker can produce a new sentence of his language on the appropriate occasion, and other speakers can understand it immediately, though it is equally new to them. Most of our linguistic experience, both as speakers and hearers, is with new sentences; once we have mastered a language, the class of sentences with which we can operate fluently and without difficulty or hesitation is so vast that for all practical purposes (and, obviously, for all theoretical purposes), we may regard it as infinite.

This *creative* aspect of the faculty of language, I believe, is uncontested. Apparently, our faculty of language is or contains a system that can generate sentences ad infinitum. We therefore *have* to face the question: how does it work? In other words, what kind of model must we devise for the faculty of language, such that it can explain this 'infinite use of finite means'?

The idea that this should be a *generative* model was, at the time, pretty obvious (and it still is). A generative grammar describes the structure of a sentence as a set of instructions for building that sentence. The concept was familiar from Rudolf Carnap, Emil Post, and others, and played a central role in the burgeoning computer science of those days.

The *form* of the rules was also familiar from the Immediate Constituent analysis of Bloomfieldian American structuralism (going back to Wundt). These were the familiar phrase structure rules (rewrite rules) that outline the composition of a phrase. They could easily be made recursive, so as to capture the desired infinite use of finite means. What was revolutionary was that Chomsky proposed a *transformational* generative grammar, with construction-specific transformation rules supplementing the phrase structure rules, and expressing all kinds of purely linguistic information and generalizations that would otherwise be lost or have to be expressed in extremely uneconomical and nonexplanatory ways.

It has been argued (by Pieter Seuren, passim) that Chomsky stole the concept of transformation from his supervisor, Zellig Harris. This is false on two counts. First, the two-step generative process of generation and transformation did not originate with Harris; it can be found in Carnap already, and was a familiar concept in mathematical logic long before Harris adopted it (e.g. Carnap 1937:27f; see Tomalin 2006 for discussion). Second, Harris used transformations to express discourse connections among complete sentences; for Chomsky, transformations were an integral component

of the derivation of every kernel sentence (a basic sentence with inflectional morphology in place), and for the generation of the surface form of every derived sentence. Rather than expressing cohesion among completed sentences, they imbued the derivation with the necessary linguistic sophistication, as evident in his earliest analyses of inflectional morphology, passive, ellipsis, coordination, etc. (Chomsky 1957).

Transformations, in Chomsky's conception of generative grammar, were needed for two reasons. First, they expressed all kinds of generalizations of sentences, and hence served the purpose of explaining the observation that speakers appear to 'know' the regularities involved. Second, they served to simplify the model of grammar, which would end up being much more convoluted if rewrite rules were the only types of rules available. For instance, in spite of the obvious similarities between active and passive sentences, one would have to have two different sets of rewrite rules for deriving them, missing generalizations. So Chomsky was playing it the right way: modeling the faculty of language in such a way that crucial observations, reflecting the competence of native speakers, might be explained.

The connection between Chomsky and Saussure, both defining the object of linguistic inquiry as a system in the human mind, would have been obvious if structuralism had not been given a distinctly anti-mentalistic twist in its American instantiation, shaped by the formidable Leonard Bloomfield. Trained, like Saussure, in European comparative linguistics, Bloomfield was also inspired by the radically positivist philosophy of science emanating from the Vienna Circle (of which Carnap was a representative). In his scientific credo (Bloomfield 1926), Bloomfield defined language (and hence the object of linguistics) as "the totality of utterances that can be made in a speech-community" (p. 155). Although Bloomfield was aware that "the physiological and acoustic description of acts of speech belongs to other sciences than ours" (p. 154), he also explicitly wanted to stay away from "psychological dispute" (p. 153), and in true behaviorist tradition treated the human mind as a black box. Unlike Chomsky, Bloomfield felt no affinity with the writings of Humboldt (which he calls difficult and obscure, p. 153 note 4), and felt no need to account for the open-ended, creative aspect of language, stating that "the forms of a language are finite in nature" (p. 156). Bloomfield also deliberately distanced himself from the rationalist tradition of the Port-Royal grammarians and the medieval Grammatica Speculativa, ultimately going back to Aristotle, where he stated that "other notions, such as subject [and] predicate, (…) will apply only to some languages, and may have to be defined differently for different ones" (p. 162).

The term 'universal grammar' (scoffed at by the *New York Times* commenter) was coined by Chomsky in reaction to this anti-mentalistic, anti-rationalistic stance of the linguistics of his time. But it means nothing more or less than the language system in the human mind, already identified by Saussure as the only reasonable object of research in linguistics. This was much more of an uphill battle than getting the technical device of a transformational generative grammar accepted, as the Bloomfieldians had donned the mantle of responsible scientific austerity—even if, from a Saussurian perspective, they were simply wrong in limiting themselves to the observable data, and in ignoring the inobservable cognitive system underlying the data.

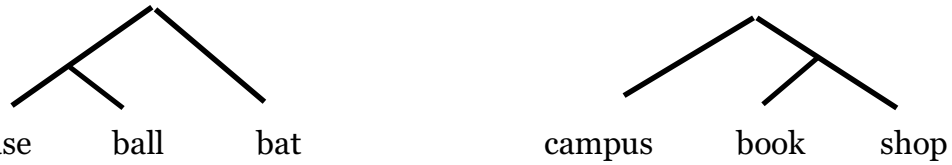
It should be obvious that universal grammar is not "an innate, hard-coded knowledge of the core rules of all human language"—how to model the faculty of language is an open question. There are no 'rules', except in the sense of 'steps in a procedure', as a generative procedure is still the most straightforward way of capturing the infinite potential of the system. But the model needs to develop on the basis of continued calibration with the observed data, i.e. what speakers can say and understand.

Ironically, transformations have not been able to survive this process of continually updating the model of grammar. Initially, transformations were just semi-formal notations of morphological and syntactic deviations from a presumed underlying basic sentence pattern. They were construction-specific, so that there was a transformation for, say, passive, and another transformation for wh-movement, etc. In and of themselves, these transformations were not that interesting, and certainly no candidate for hard-core rules of universal grammar. But having these transformations allowed us to detect common properties, and generalize over the processes they were intended to capture, so that a common core could be formulated that might be a closer approximation of what the model should look like.

We now know that speakers produce and interpret *constituents*, that constituents enter into asymmetric *dependency* relations, and that these dependency relations are limited to structurally defined *local* domains. All these observations call for explanation, and so what we should ask, in response to the current Chomsky-critique, is whether any progress has been made in the formulation of a model of grammar, such that these observations (of constituency, dependency, and locality) make sense.

Let us illustrate with a simple example, discussed in Wilhelm Wundt's *Die Sprache* (1900:670). Languages have compounds, often with two members, such as *baseball* and *bookshop*, but sometimes also with three members, such as *baseball bat* and *campus bookshop*. Wundt observed that it is always possible to subdivide three-member compounds in two pairs of constituents:

(1)



In (1), the pairs of constituents are (base, ball) and ((baseball), bat), and (book, shop) and (campus, (bookshop)). Apparently, phrases are invariably binary branching, as depicted in the graphic representations.

Note that 'branching' is a metaphor that we need in order to describe a feature of these structures. Like the coordinates on a map, this is a very useful device for the representation of reality, without necessarily subscribing to the view that these tree structures are 'real', any more than we would say that the coordinates are real. What *is* real is that speakers produce and interpret three-part compounds as involving pairs of constituents, and our model of grammar needs to account for this state of affairs.

As we know, the current conception of the model of grammar involves a generative procedure that derives these facts. The procedure is the process of Merge (first proposed in Chomsky 1995:226), which, in its simplest conception, combines maximally two elements in a group. The compounding facts suggest that the faculty of language is *limited* in the sense that it cannot create groups of more than two members. Of course *less* than two members would not work either, so that Merge, in its current conception, is the minimal and only structure generating procedure. This is both explanatory and attractive: apparently we may proceed on the assumption that the model of the faculty of language be maximally minimalist.

Another observation is that minimal structures are typically asymmetric, rather than symmetric. Even where it looks like we encounter symmetric pairs, such as with reduplication or coordination, on closer inspection it turns out that we are dealing with an asymmetric organization. Consider the examples in (2).

(2) money, schmoney
arma virum-que

Money, schmoney is a reduplication which shows asymmetries of a prosodic, morphological, and semantic nature. It means something like 'who cares about money', so the repetition signals pejorative predication (money = unimportant). In connection with this, the second member is marked by a higher pitch and by *sch*-prefixation.

In *arma virumque* 'arms and the man' we observe that the second member *virumque* is marked by a coordinating suffix *-que*. This is not a final coordinator, as the suffix is attached to the first word of the second conjunct, rather than to the right edge of the second phrase. This is evident from more complex examples such as (3), where *-que* is lodged behind the first word of the second conjunct *totius naturae capacia*:

(3) ingenia fecunda totius-que naturae capacia
'minds that are fertile and capable of grasping the entire universe'

This shows that the conjunction belongs to the second conjunct, as already observed by De Groot (1949:66), suggesting an asymmetry within even the most minimal coordinating constructions as well.

This suggests that Merge is somehow asymmetric. If so, we derive that every time the operation Merge takes place, we create (i) a constituent, and (ii) a dependency.

Dependencies may typically extend beyond a simple pair of sisters, such as when a subject binds an object or triggers agreement morphology on a verb, or when a fronted element determines the interpretation of a gap further down the road in the sentence. These dependencies require a concept like c-command, a relation that is typically defined in terms of the branching tree structure metaphor. However, as Epstein (1999) observed, the proper definition of c-command, and hence the explanation of the very existence of these dependencies (and not others) follows from the simplest conception of the structure building process Merge:

(4) If α merges with β, α c-commands β and everything contained in β

Note that if Merge were symmetric, (4) would express a reciprocal relation between α and β, which is not in line with observation. This confirms our earlier impression that every sister pair generated by Merge expresses a fundamental asymmetric dependency.

As Epstein (1999) shows, the description of typical dependencies that we observe in the interpretation of semi-referential elements, variables, and empty positions requires a definition as in (4), and there are no dependencies that require a different (say, inverse) definition. The simplest model of grammar succesfully accounts for the fact that these and only these dependencies can be observed.

This is not to say that the model is finished. In fact, the model as it stands performs rather poorly in relation to the third key property of human language, locality. But this is not to say that the model could not be improved to the effect that locality falls out naturally—just that currently we may be barking up the wrong tree.

Importantly, the physical properties of the acoustic signal as it travels from the speaker to the listener are vastly underdetermined with respect to these crucial properties of human language. We simply cannot identify constituents in the speech sounds, nor do the sound waves include any signals that mark the ubiquitous dependencies in the sentences making up human thinking and communication. We simply *have to* concentrate on human cognition, an unobservable structure that needs

to be modeled in order to be properly described, if we want to have any hopes of explaining these intricacies of human language. This is no less a powerful idea now, then when it was first conceived in the 1950s.

In fact, the conception of universal grammar that was proposed in the early days of generative grammar is a much less attractive and intuitively far less probable model of the faculty of language, than the Merge-based system being contemplated now. For one thing, the model of grammar initially had category-specific rewrite rules, defining different structures for noun phrases, verb phrases, etc. with little generality among them. This was only rectified in the 1970s with the development of X-bar theory, which prescribes a universal structure for all phrases, including sentences and phrases built on functional elements. But even this was not ultimately a plausible part of innate knowledge, and the X-bar theory had to make way for an even simpler concept, Merge, the application of which *derives* the properties of the phrase structures without detailing any of the specifics characterizing the X-bar theory or the category specific phrase structure rules. 'Merge' basically states as a property of human cognition that we are forced to break down any string of elements into hierarchically organized pairs of elements—hence the limits on the possible structures of compounds mentioned earlier.

Similarly, one feels intuitively that the construction-specific transformations of the early days of generative grammar are not good candidates for innate knowledge, and hence they should not be part of the faculty of language. Their elimination from the theory in the stage of generative grammar known as the Government and Binding Theory (Chomsky 1981), represents a significant improvement. But this came at the cost of introducing a rich arsenal of principles and requirements dealing with all aspects of grammar (argument structure, case, pronoun interpretation, etc., not to mention the 'cartography' of clause structure), that could at best be judged a step in the right direction, but still a far cry from a plausible model of an innate language system. The elimination of these principles and requirements in minimalism (Chomsky 1995) is again a move in the right direction, even if much work still needs to be done.

The point is that, if anything, theoretical linguistics is now in a much better position than half a century ago, and in spite of the inevitable difficulties associated with the study of human cognition and consciousness, the idea of a universal grammar (a model of the faculty of language) seems more within reach now than ever before in the history of science.

Meanwhile it should be clear that the research strategy to model the faculty of language does not imply the existence of a designated human language faculty as an article of faith. It's just that we have learned so much about language that this is a suitable, even ideal, point of entrance for studying human cognition. If the model we come up with turns out to have applications in other areas of human or animal cognition, so much the better. Such developments are to be expected, if theoretical linguistics (again, a very young field) is to grow more over time. If after centuries of isolation physics, chemistry and biology have recently started to show signs of convergence, there is no reason why the study of the language system should remain isolated for eternity. But in the absence of a clear time path, there is no reason not to keep focusing on the more tractable topic of modeling the faculty of language by itself.

In evaluating the merits of theoretical linguistics, it will be important to keep in mind that even a highly successful scientific model represents an idealization, with no claim of reality for any of the elements featuring in it. A system of coordinates on a map is very useful for describing the layout of the real world, but is in itself just a figment of the imagination. As Leng (2010) has shown, even the austere entities of mathematics, such as numbers, sets, and relations cannot really be said to *exist*, even if they are essential

to our understanding and analysis of the real world. And working with these fictional entities does not in any way invalidate the laws and generalizations they feature in. If we state that Calabria is located in the 'boot of Italy', that is a true statement in spite of its crucial use of a metaphor. In the same way, Leng argues, are numbers and sets metaphoric concepts, and we should always regard the tree structures and features and relations and operations that we use in our syntactic analysis as essentially metaphoric.

This means that we cannot use the criterium of 'psychologic reality' to evaluate our model of the faculty of language. Just like in physics, the only evaluation metric is the success of the model in explaining observations and making predictions. And this does not make it a fancy-free exercise: data are crucial, but since our object of inquiry is the human language system, and not the sound waves going from speaker to hearer, we have to tap into native speaker intuitions to find the data we need. This should put a damper on the resurging enthusiasm in our field for corpus studies, let alone corpus research conducted through artificial intelligence. A corpus, at best, is a reflection of the native speaker intuitions we are after, but just like the sound waves in the Saussurian model of communication, it rarely contains even the most basic information that is relevant to the study of the human language faculty, such as constituency and dependency.

Likewise, the circumstance that the model of the faculty of language developed in theoretical linguistics is not necessarily 'real' cautions against any feelings of optimism that the faculty of language may be approached equally well from brain studies, such as neurolinguistics. Not to detract from the phenomenal progress in these fields in recent years, it is at present still hard to conceive how events at the level of neural networks translate to speakers' robust intuitions of constituency, dependency and locality, But this is a topic best treated separately.

The commenter in the *New York Times* stoops to a common trope where he describes universal grammar as "an article of faith" among Chomsky's "linguistic followers" sharing his "core belief". It should be clear to anyone familiar with the basics the science of philosophy that a scientific hypothesis or even a scientific paradigm is not a "belief". The identification of 'something in the mind' of speakers as the object of inquiry in linguistics is almost trivial, if we follow Saussure's reasoning. Having suffered a strict religious upbringing, I know very well what it is like to find oneself in a community of believers, and from my experience in linguistics I can thankfully say that scientific communities are the direct opposite. Sticking to a paradigm, in the sense of Kuhn (1962), is not religion but good science. Nevertheless, the fact that the paradigm of generative grammar is to a large extent shaped by a single man, Chomsky, should give one pause, if only to reflect on the relative poverty in our field of independent thinking at the highest level.

A good illustration of this is the current conception of locality in generative grammar, as illustrated in pairs like (5).

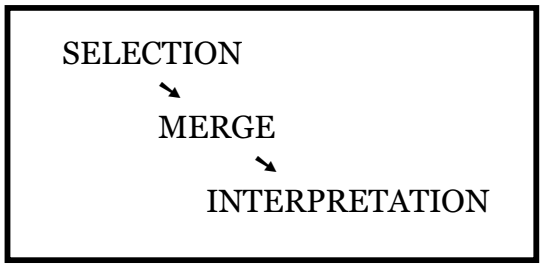
(5) a. Who do you think John was looking for [e]
b. *Who do you wonder why John was looking for [e] [e]

We can interpret *who* in (5a) in the position of *[e]*, but not in (5b)(regardless which *[e]*). This indicates that the dependency between *who* and *[e]* is subject to a condition of locality. The question is whether the model of grammar can explain, first, the existence of such locality constraints, and, secondly, their nature.

The model of grammar contemplated in minimalism is extremely simple. It contains as its core the operation Merge, combining two elements in a group, which can be applied iteratively or recursively, depending on implementation. Merge needs elements

to work with, so the model should also contain such a set of elements, often called Numeration. (I prefer the term 'Selection'.) Finally the model needs to connect with other components in (our model of) the mind/brain, having to do with externalization (preparing the structure for pronunciation and interpretation). Call this 'Interpretation'. This gives us the model in (6).

(6)



To account for locality effects, Chomsky (2001) proposed that Merge proceeds in *phases*, that is, after a limited number of operations Merge, the object under construction is already shipped off to Interpretation before the object is completed. Elements in the structure can only escape this fate by moving to the edge of the structure, which is somehow not transfered to Interpretation yet.

The facts in (5) now follow if the embedded clause is a phase, and *who* in (5a) is moved to the edge of the phase before the embedded clause is sent to Interpretation. After that it moves on to the edge of the entire clause. In (5b), this won't work because the edge of the embedded phase is already occupied by another element, *why*. If this sounds familiar, it should, because this is essentially the same analysis as that in Chomsky (1973).

The problem with this analysis is that the existence and properties of phases are not in any way derived from the operation Merge. These are externally imposed features that might have been different. Phases might not exist at all, or they might be characterized differently, or they might not have edges, or different edges, etc. Phases work, but in a clumsy, non-principled way.

Furthermore, there are many locality effects different from the type in (5) that do not fall out from this conception of phases, and have by and large resisted successful explanation in terms of phase theory. These include the phenomena captured by the Coordinate Structure Constraint (Ross 1967) and the Condition on Extraction Domains (Huang 1982).

(7) Coordinate Structure Constraint
a. * Who did John see [Mary and [e]]
b. * Who did John say [he met [e] and Mary went home]

(8) Condition on Extraction Domains 1, subjects
* Who did [that John met [e]] irritate Mary

(9) Condition on Extraction Domains 2, adjuncts
* Who did Mary go home [after John met [e]]

Chomsky defines phases as either a clause (CP) or a verb phrase (vP), these being the only 'propositional domains'. (A vP is propositional on the assumption that all the

arguments of the verb are merged within that domain, itself a questionable if traditional assumption.) If so, there is no phase associated with these particular opaque domains, or if there is, it is unclear why no escape via the edge of the phase is possible. Thus, coordinate structures are opaque regardless of their category. Subjects and adjuncts are opaque even if they are CPs with their edges unfilled. Even in its already ad hoc conception, phase theory is not equipped to deal with these locality effects, which are clearly demonstrable in all native speaker intuitions (these are not subtle facts).

The solution must be found in the nature of the Selection. It is commonly tacitly assumed that the Selection is homogeneous, in that it contains just a single type of elements, words. But this assumption cannot survive if we want to maintain a maximally simple definition of the operation Merge. Here's why.

Consider a simple sentence like (10).

(10) John and Mary left

If the Selection contains only words, it would have to look like (11).

(11) {John, and, Mary, left}

If we then merge the elements from the Selection one by one, as is the simplest process, the structure in (12) results.

(12) [John [and [Mary [left]]]]

In (12), the constituents are *left*, *Mary left*, *and Mary left*, and *John and Mary left*. But native speaker intuitions do not countenance these constituents, but rather *left*, and *John and Mary* and *John and Mary left*. That means that *John and Mary* must be in the Selection as a single item.

(13) {[John and Mary], left}

Starting from (13), we can simply merge *John and Mary* and *left*, yielding (14), with the right constituents.

(14) [[John and Mary] left]

But *John and Mary* is clearly structured, so (on the simplest conception of the faculty of language) it must also have been derived via Merge. That is, we also need a subderivation with the Selection in (15), or something similar, that yields (16).

(15) {John, and, Mary}

(16) [John [and Mary]]

This requires that even a simple sentence like (10) is the outcome of a *netwerk of derivations*, each structured as in (6), such that the output of a (sub-)derivation may be included in the Selection for another (sub-)derivation.

It doesn't take long to realize that derivations are virtually always such networks of derivations, and that coordinate structures and complex left branch elements, such as subjects and adjuncts, *must* be the output of a subderivation, and that certain phrases

that *need* not be the output of a subderivation, such as embedded interrogatives as in (5b), *may very well be* analyzed as the output of a subderivation.

All the locality effects now follow if we assume that whatever is merged in a subderivation, cannot be re-merged in another derivation. This is a familiar state of affairs from the opacity of complex words, like compounds. It is impossible to extract part of a compound, per the Lexical Integrity Principle (Lapointe 1980).

(17) Lexical Integrity Principle
It is impossible to extract part of a word

This now follows if compounds, as seems reasonable, must have been put together in a subderivation before being included in a Selection for a further derivation. And so with all complex words.

We can therefore generalize (17) to:

(18) Generalized Integrity Principle
Given two derivations D1, D2 involving two Selections S1, S2, and outputs E1,E2, such that E1 is included in S2, no member of S1 may be merged in D2

Now locality is a function of the structure building process, assuming (as seems inevitable) that derivations are invariably networks of derivations.

This goes to illustrate that theoretical linguistics, far from having run its course, can still make progress, by subjecting even the most venerated of concepts, such as locality, to a critical evaluation. It is precisely by focusing on the objective of coming up with an attractive and maximally simple model of the faculty of language, that we can expect to explain crucial observations in a "surprisingly efficient and even elegant" way.

References.

Bloomfield, Leonard. 1926. A set of postulates for the science of language. *Language* 2, 153-164.

Carnap, Rudolf. 1937 [1959]. *The logical syntax of language*. Paterson: Littlefield, Adams & Company.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1964. Goals of linguistic theory. In *Current issues in linguistic theory*, 7-27. The Hague: Mouton.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson and Paul Kiparsky, eds., *A Festschrift for Morris Halle*, 232-286. New York: Holt, Rinehart and Winston.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

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

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz, ed., *Ken Hale: a life in language*, 1-52. Cambridge: MIT Press.

Epstein, Samuel D. 1999. Un-principled syntax: the derivation of syntactic relations. In Samuel D. Epstein and Norbert Hornstein, eds., *Working minimalism*, 317-345. Cambridge: MIT Press.

de Groot, Albert Willem. 1949. *Structurele syntaxis*. The Hague: Servire.

Harré, Rom. 1970. *The principles of scientific thinking*. London: MacMillan.

Harris, Roy. 1987. *Reading Saussure*. La Salle: Open Court.

Hempel, Carl Gustav. 1983. Valuation and objectivity in science. In Robert S. Cohen and Larry Laudan, eds., *Physics, philosophy and psychoanalysis: essays in honor of Adolf Grünbaum*, 73-100. Dordrecht: Reidel.

Hofstadter, Douglas. 1995. *Fluid concepts and creative analogies: computer models of the fundamental mechanisms of thought*. New York: Basic Books.

Hofstadter, Douglas. 2013. *Surfaces and essences: analogy as the fuel and fire of thinking*. New York: Basic Books.

Huang, C.T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Dissertation, MIT.

Kuhn, Thomas. 1962. *The structure of scientific revolutions*. Chicago: The University of Chicago Press.

Lapointe, Steven G. 1980. *The theory of grammatical agreement*. Dissertation, University of Massachusetts, Amherst.

Leng, Mary, 2010. *Mathematics and reality*. Oxford: Oxford University Press.

Lockhart, Paul. 2013. *Measurement*. Cambridge: Harvard University Press.

Ross, John R. 1967. *Constraints on variables in syntax*. Dissertation, MIT.

Saunders, Carol, ed. 2004. *The Cambridge companion to Saussure*. Cambridge: Cambridge University Press.

Saussure, Ferdinand de. 1916 [1964[3]]. *Cours de linguistique générale*. Paris: Payot.

Tomalin, Marcus. 2006. *Linguistics and the formal sciences: the origins of generative grammar*. Cambridge: Cambridge University Press.

Wundt, Wilhelm. 1900 [1911[3]]. *Die Sprache, vol. 1*. Leipzig: Wilhelm Engelmann.