

# Simplicity and the form of grammars

*Noam Chomsky*

Massachusetts Institute of Technology

University of Arizona

## ABSTRACT

The goal of theory construction is explanation: for language, theory for particular languages (grammar) and for the faculty of language FoL (the innate endowment for language acquisition). A primitive notion of simplicity of grammars is number of symbols, but this is too crude. An improved measure distinguishes grammars that capture genuine properties of language from those that do not. The theory of FoL must meet the empirical conditions of learnability (under extreme poverty of stimulus), and evolvability (given the limited but not insignificant evidence available). Recent work provides promising insights into how these twin conditions may be satisfied.

*Keywords:*  
*simplicity,*  
*explanation,*  
*evaluation,*  
*grammar, faculty*  
*of language,*  
*learnability,*  
*evolvability,*  
*externalization*

There is a close relation between the two concepts in the title – which also happens to be the title of the first talk I gave as a graduate student and the topic of my first paper *Morphophonemics of Modern Hebrew* (MMH; Chomsky 1949/1951)<sup>1</sup> – concerns that have remained salient for me to the present. The relation becomes clear when we consider the goals of the theory of language. Pursuing the relation more

---

<sup>1</sup> An improved 1951 version was published in 1979. I bring up this text, a student paper not intended for publication, because it is the first extensive study of these topics, and the last at any such level of detail. It soon became obvious that the effort was far too ambitious though the general concerns persisted in new forms, even some of the measures of simplicity outlined, as discussed below.

closely gives a good deal of insight into the nature and development of linguistic theory, and also provides a more principled basis for elements of common practice.

As in other domains, the primary goal of theories of language is to explain in the best way the data that constitute the subject matter of the theory, along with determining just what is the relevant subject matter.<sup>2</sup> The concept “best way” is traditionally (and plausibly) understood in terms of simplicity/economy. And when spelled out, these notions are necessarily relative to the formal nature of the system under consideration.

In his investigations of these topics, Nelson Goodman – with whom I was studying at the time – observed that “The motives for seeking economy in the basis of a system are much the same as the motives for constructing the system itself”; “To seek truth is to seek a true system, and to seek system at all is to seek simplicity” (Goodman 1943, 1955).

From a somewhat different perspective, Herman Weyl drew essentially the same conclusions: “The assertion that nature is governed by strict laws is devoid of all content if we do not add the statement that it is governed by mathematically simple laws... That the notion of law becomes empty when an arbitrary complication is permitted was already pointed out by Leibniz in his *Metaphysical Treatise*... The astonishing thing is not that there exist natural laws, but that the further the analysis proceeds..., the finer the elements to which the phenomena are reduced, the simpler – and not the more complicated, as one would originally expect – the fundamental relations become and the more exactly do they describe the actual occurrences” (Weyl 1932, cited by Roberts and Watumull 2015).

In a similar vein, Galileo held that nature is simple and it is the task of the scientist to demonstrate that in particular cases – a quasi-empirical claim, but so powerfully verified over the centuries that it is fair to adopt the precept.

There are many similar observations by distinguished figures, for good reasons. If we are serious about linguistic theory we can hardly

---

<sup>2</sup>Not given *a priori*. Just what constitutes a language *L* is in part a matter of decision. What data belong to *L* is theory-driven, a familiar matter.

ignore the question of finding a way to measure its simplicity, which will, transparently, depend on the form it assumes.

For language, there is an additional reason to suppose that the basic system is quite simple. There is mounting evidence that the core elements of the faculty of language (FoL) emerged pretty much along with modern humans and haven't changed since, hence emerged rather suddenly in evolutionary time (Berwick and Chomsky 2016; Huybregts 2017). If so, one would expect that they would have assumed a simple form.

The task of finding the simplest theory for language is posed at two levels: for the theory of each language (its grammar), and for the theory of FoL (UG, in contemporary terminology). FoL provides the framework within which each language develops much as the general faculty of human vision does for each individual visual system, allowing considerable variation as classic experimental work has shown. FoL must satisfy at least what has been called the Basic Property of language: it must provide mechanisms for a language to generate an unbounded array of hierarchically structured expressions in a form that can be interpreted at two interfaces with external systems, at the conceptual-intentional level CI for expression of thought and at the sensorymotor level SM for externalization in some medium, typically sound. There are important asymmetries to which we return.

More generally, UG must satisfy the condition of “explanatory adequacy,” answering the question how a particular language can in principle be acquired from the data available (Chomsky 1965). To do so, UG must specify the “search space,” the class of possible languages PL, along with a selection procedure SP that selects the correct grammar (or set of grammars) for each language given relevant data.

These conditions become far more restrictive if we take a language to be a property of the organism in accord with the “Biolinguistic Program” BL, Massimo Piattelli-Palmarini's term for the evolving discipline.<sup>3</sup> This was a departure from standard views,<sup>4</sup> and partially remains so. While sometimes regarded as contentious, it seems to me that the legitimacy of the BL approach is obvious to the point of

---

<sup>3</sup> Ibid. Lenneberg (1967), the classic exposition.

<sup>4</sup> For a sample, see Chomsky (2013).

truism.<sup>5</sup> If so, adoption of it raises no issue of substance but only one of decision as to which concept of language we choose to consider, so we can put it aside.

Adopting BL, explanatory adequacy requires the further condition that PL-SP be *feasible*. They must provide a realistic abstract account of language acquisition on the basis of the Primary Linguistic Data. In particular, they must account for the huge gap between the data available and what the child knows. It was recognized from the early days of work on generative grammar that this problem of Poverty of Stimulus is enormous, and later investigations of what is known by a very young child along with statistical study of the sparsity of data available have revealed that the problem is far more severe even than what had been assumed.<sup>6</sup> Accordingly, PL-SP must be sharply constrained.

Whether our concern is feasibility and BL or the weaker notion of just explanation, the next problem is to spell out what we mean by “simplicity.” For Goodman (1943), as the quote above indicates, the answer reduced (mainly) to minimal number of primitives as the basis for the constructional system under consideration. MMH was an attempt to explore these ideas over a broader range. Language provides interesting cases, and the subject matter for MMH was a natural choice: the data are readily available and sufficiently intricate to require richer notions of simplicity. Much richer, it became clear as study of the dual problems of theory construction for language proceeded.

The form of grammars in MMH is a system of rewriting rules with the conventional interpretation: the rule  $X \rightarrow Y$  maps  $AXB$  to  $AYB$ .<sup>7</sup> Exploring ways to measure simplicity, we can begin with the most obvious idea: take SP to rank grammars by the number of symbols they contain. While that seems a natural measure, it quickly becomes clear that it is seriously inadequate. One reason is that the measure does

---

<sup>5</sup> Further, I think it can be argued that other concepts tacitly presuppose it.

<sup>6</sup> See Yang *et al.* 2017.

<sup>7</sup> MMH included a rudimentary syntax with optional unordered rules (basically what became phrase structure grammar) and a complex morphophonemics – part of externalization in current terms – with obligatory ordered rules. The reasons for such distinctions only became clear much later; see below.

not distinguish between rule systems that express legitimate linguistic generalizations from others that do not do so.<sup>8</sup>

Suppose for example that we have the rule sets (1), (1'):

- (1)     a.  $X \rightarrow YWB$   
          b.  $X \rightarrow YW$   
(1')    a.  $X \rightarrow BWY$   
          b.  $X \rightarrow YW$

(1) expresses an expected configuration: *B* is optional in the context  $YW\_.$ <sup>9</sup> (1') in contrast expresses no legitimate generalization. But the number of symbols in each is 7. Counting symbols is clearly too crude a measure. In MMH the distinction is captured with a notational transformation taken to be part of the simplicity measure of UG, mapping the rules of (1) to (2):

- (2)      $X \rightarrow YW(B)$

The notation is interpreted as: *B* is optional in the context  $YW\_.$  No similar notation is provided by UG for (1'), not considered a legitimate generalization – an empirical assumption about language, as noted, but well confirmed. Under this notational transformation, the simplicity measure of (1) is 4, capturing the intended distinction.

Consider a more complex configuration, very commonly found. Suppose that  $X \rightarrow Y$  before *A* and elsewhere  $X \rightarrow Z$ .<sup>10</sup> The set of rewriting rules is (3):

- (3)     a.  $XA \rightarrow YA$   
          b.  $XW \rightarrow ZW$  (for  $W \neq A$ )

---

<sup>8</sup>What is taken to be a “legitimate linguistic generalization” is an empirical hypothesis, subject to testing by examination of languages and by direct experiment.

<sup>9</sup>E.g., the parenthesized optional element in such phrase structure grammar configurations as  $VP \rightarrow V_{tr} NP (PP)$  “read the book (in the library).” In contrast, we do not expect to find the rule set  $VP \rightarrow PP NP V_{tr}$ ,  $VP \rightarrow V_{tr} NP$  (irrelevantly, such outcomes might result from a series of rules). Or more marginally, “all (of) the men,” “cyclic(al) rules,” [siŋ(g)r] (with *g* missing in some dialects, yielding a *singer/finger* contrast).

<sup>10</sup>Voicing assimilation of final consonants as in *wife-wives*, sets of irregular verbs, and innumerable other cases.

The parenthesized phrase must be spelled out, listing all cases of  $W \neq A$ . The list is infinite, but even if we impose some sharp restriction on  $W$ , the list is very long, and the number of symbols in the expanded version of (3) gives a completely wrong simplicity measure for a configuration that should be highly valued.

What is clearly the right answer requires several steps that are of more general significance.<sup>11</sup>

First, we have to distinguish obligatory from optional rules, and ordered from unordered rules. For the syntax – mapping to CI – the normal case is unordered and optional, if such rules exist at all; they may not (see below). For externalization to SM, the normal case of rule systems (which are quite complex) is ordered and obligatory. The configuration (3) falls within externalization.<sup>12</sup>

With these conventions in place, we can introduce the notational transformation of (4), interpreted as (3):

- (4)     a.  $X \rightarrow Y / \_A$   
          b.  $X \rightarrow Z$

The simplicity measure is small, as it should be for this legitimate generalization. The rules (3)–(4) state that  $X$  becomes  $Y$  before  $A$ , and becomes  $Z$  elsewhere. This device is the familiar “elsewhere condition”: first list the exceptions, then the general rule for everything else.<sup>13</sup>

From the early inquiries into generative grammar it was found that rule ordering was still more intricate: with cyclic application of rules and implicational relations, grammars are greatly simplified and (accordingly) yield deeper explanations, while also providing the ba-

---

<sup>11</sup> These steps were all taken in MMH and commonly adopted in later work in generative grammar.

<sup>12</sup> NB: “normal.” There are some exceptions on the periphery, like free variation. There are interesting questions about the tacit choices here but they do not bear on the main points about simplicity and general architecture of grammar, so I will put them aside.

<sup>13</sup> The elsewhere condition, which may trace back to classical India, has been widely used in practice. It is also a core element of Charles Yang’s tolerance principle, which has been highly successful in explaining when rules are productive and establishing a firm core–periphery distinction. See Yang *et al.* 2017, and for more extensive analysis Yang 2016.

sis for compositionality of semantic interpretation, matters I will put aside here (Chomsky *et al.* 1956; Chomsky and Halle 1968).

The notations and conventions in MMH, now common, provide a reasonable step towards a feasible evaluation procedure: the simplicity measure of a rule system is the number of symbols under the conventions and notational transformations that capture legitimate linguistic generalizations – all expressing empirical hypotheses about language.

In MMH, the main problem was to find the simplest ordering of rules, which was quite deep. In those hand-computation days, the task was impossible, so the analysis was restricted to finding a relative minimum: a particular ordering with a lower measure (higher valued) than any re-ordering of adjacent rules. The exercise illustrates some of the problems of constructing UG, tasks challenging enough that they have rarely been undertaken on any large scale.<sup>14</sup> Note that the complexities arise primarily (perhaps completely) in externalization, a matter to which we return.

All of this is only the beginning, however. Another aspect of the quest for feasibility is restricting the search space PL and constraining the selection procedure SP. These topics have been the main concern of the study of *narrow syntax*,<sup>15</sup> generation of structures at the CI interface. The topic is too rich to review here. I will briefly mention only a few stages, keeping to one course of development, which I think is on the right track.

Early generative grammar assumed that two systems of rules interact: Phrase Structure Grammar PSG and Transformational Grammar TG. Both were relatively unconstrained, yielding a huge search space, remote from any hope for feasibility. Serious efforts to restrict the search space began in the early 1960s. It was quickly recognized that PSG permitted far too many options; there was, for example, noth-

---

<sup>14</sup> One of the last cases I know of is *Sound Pattern of English* (Chomsky and Halle 1968). Later study of externalization, the primary locus of these issues, took a different course that ignores the questions, and as far as I can see, cannot accommodate them. See Chomsky 1995, p. 380.

<sup>15</sup> Broadly construed, syntax incorporates all internal symbolic computation, including externalization to phonetic form and logical syntax, often called formal semantics. For reasons discussed elsewhere, human language may not have semantics in the technical sense based on reference/denotation.

ing to bar the vast array of “crazy rules” such as  $NP \rightarrow V PP$ , and the symbols used were themselves illegitimate, tacitly incorporating structural relations that must be spelled out (why  $NP$ ?). PSG was therefore abandoned in favor of X-bar theory, sharply restricting the options for grammars.<sup>16</sup>

Though it wasn't recognized at the time, X-bar theory had rich consequences, some not explored seriously until recent years. Unlike PSG, X-bar theory (also TG as it developed) yields pure structures, without linear order or other organization. Hence resort to X-bar theory introduces a sharp distinction between (i) narrow syntax, consisting of X-bar theory and TG and yielding CI representations, and (ii) externalization of syntactic structure to the sensorimotor system SM (typically phonetic form PF). As noted, externalization appears to be the locus of the apparent complexity, variety, and mutability of language – not surprisingly. Externalization relates two systems that are entirely independent, both in character and evolutionary history: language proper and SM. Establishing that relation is a complex cognitive process that can be carried out in many ways. In particular, it must deal with the mismatch between narrow syntax, a system of pure structure, and SM, which imposes a requirement of linear order for reasons that have nothing to do with language.<sup>17</sup> There must, it seemed, be a “head parameter” that each  $L$  has to set one way or another (V-Object for English, Object-V for Japanese, etc.). Along with other work of the 1970s, including radical simplification of TG, that led to a new conception of the form of language, the Principles and Parameters (P&P) framework, with fixed principles of UG that determine PL and parameters that have to be set in acquisition of language, the latter restricted largely to externalization (perhaps completely, we might someday learn).

The problems of simplicity of grammars and of UG are accordingly reshaped. A crucial problem is to find a feasible search process through the set of parameters, and to determine their status: how did they evolve? How are they captured in UG and stored in the brain?

---

<sup>16</sup>Too far, it was later realized. See Chomsky (2013), opening directions I will put aside though they bear directly on explanatory adequacy and simplicity.

<sup>17</sup>Sign language, using the options available in visual space, permits somewhat different devices.



The “head parameter” suggests possible answers. It is, strictly speaking, not a parameter. It is not part of UG, did not evolve, and is not internally stored. Rather, it expresses a mismatch between two independent systems: language proper and SM. The mismatch must be resolved in acquisition, but is not part of grammar.

Recent work by Ian Roberts (Roberts 2019), supported by rich empirical evidence from a wide range of typologically different languages, suggests a radical solution to these problems. It provides a feasible search procedure and concludes that parameters altogether are not part of UG (hence did not evolve and are not stored) but rather emerge in the course of acquisition in predictable ways.

Meanwhile work in the “Minimalist Program” has subjected the principles of the P&P systems to much closer analysis, reducing generation in narrow syntax to the simplest combinatorial operation (binary set-formation, called “merge”). That step turns out to incorporate and unify earlier proposals and to yield solutions to long-standing puzzles and new ones discovered along the way, along with suggestions as to how language evolved.<sup>18</sup> One conclusion reverses the general view (mine included) concerning compositionality and displacement: that compositionality (provided by PSG and its descendants) is unproblematic and displacement (handled by TG) is a curious “imperfection” of language that has to be explained away somehow. It turns out that the opposite is true. Displacement is the simplest and unproblematic case, and composition beyond displacement requires an explanation in terms of special properties of language. All of these developments bear directly on our topic here.

Without further elaboration, even a brief review of the course of research in generative grammar since its modern origins reveals that the concepts of simplicity and form of grammar have been closely related throughout, that measuring simplicity is an essential task and is no simple matter, and that inquiry into this relation has led to substantial insight into the general nature of language, with a promise of more to come.

---

<sup>18</sup> See note 3. For more general discussion, see Chomsky (2015).

## REFERENCES

- Robert C. BERWICK and Noam CHOMSKY (2016), *Why Only Us: Language and Evolution*, MIT Press, Cambridge, MA.
- Noam CHOMSKY (1949/1951), *Morphophonemics of Modern Hebrew*, Master's thesis, University of Pennsylvania.
- Noam CHOMSKY (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge, MA.
- Noam CHOMSKY (1979), *Morphophonemics of Modern Hebrew*, Garland Publishing, Inc., New York and London.
- Noam CHOMSKY (1995), *The Minimalist Program*, MIT Press, Cambridge, MA.
- Noam CHOMSKY (2013), Problems of projection, *Lingua*, 130:33–49.
- Noam CHOMSKY (2015), *What Kind of Creatures are We?*, Columbia University Press.
- Noam CHOMSKY and Morris HALLE (1968), *The Sound Pattern of English*, Harper and Row Publishers, New York.
- Noam CHOMSKY, Morris HALLE, and Fred LUKOFF (1956), On accent and juncture in English, in M. HALLE, H. LUNT, and H. MACLEAN, editors, *For Roman Jakobson*, pp. 65–80, Mouton.
- Nelson GOODMAN (1943), On the simplicity of ideas, *The Journal of Symbolic Logic*, 8(4):107–121.
- Nelson GOODMAN (1955), Axiomatic measurement of simplicity, *The Journal of Philosophy*, 52(24):709–722.
- MAC Riny HUYBREGTS (2017), Phonemic clicks and the mapping asymmetry: how language emerged and speech developed, *Neuroscience & Biobehavioral Reviews*, 81:279–294.
- Eric LENNEBERG (1967), *Biological Foundations of Language*, volume 68, Wiley New York.
- Ian ROBERTS (2019), *Parameter Hierarchies and Universal Grammar*, Oxford University Press, USA.
- Ian ROBERTS and Jeffrey WATUMULL (2015), Leibnizian Linguistics, in Ángel J. GALLEGO and Dennis OTT, editors, *50 Years Later: Reflections on Chomsky's Aspects*, pp. 211–222, MIT Working Papers in Linguistics.
- Hermann WEYL (1932), *The Open World: Three Lectures on the Metaphysical Implications of Science*, Yale University Press, New Haven.
- Charles YANG (2016), *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language*, MIT Press, Cambridge, MA.


*Simplicity and the form of grammars*

Charles YANG, Stephen CRAIN, Robert C BERWICK, Noam CHOMSKY, and Johan J BOLHUIS (2017), The growth of language: Universal Grammar, experience, and principles of computation, *Neuroscience & Biobehavioral Reviews*, 81:103–119.


*Noam Chomsky*

Massachusetts Institute of Technology, Cambridge, MA, USA  
University of Arizona, Tucson, AZ, USA

Noam Chomsky (2021), *Simplicity and the form of grammars*, *Journal of Language Modelling*, 9(1):5–15

 <https://dx.doi.org/10.15398/jlm.v9i1.257>

This work is licensed under the *Creative Commons Attribution 4.0 Public License*.

 <http://creativecommons.org/licenses/by/4.0/>