

A NOTE ON WEAK vs. STRONG GENERATION IN HUMAN LANGUAGE*

NAOKI FUKUI
Sophia University

The last section of Chapter 1 of *Aspects* states that “[p]resumably, discussion of weak generative capacity marks only a very early and primitive stage of the study of generative grammar. Questions of real linguistic interest arise only when strong generative capacity (descriptive adequacy) and, more important, explanatory adequacy become the focus of discussion” (p. 61). This is a clear and explicit statement of Chomsky’s perception of the linguistic relevance of formal and mathematical investigations of grammars and languages, which he had initiated and explored in the 1950s and early 1960s. Clearly the concept of weak generation (the size of the set of output strings) plays only a marginal role, if any, in the formal investigation of human language. Rather, the nature of human language lies in its strong generative capacity, i.e., what kind of structural descriptions it generates, and more deeply, what kind of generative procedures it involves in generating those structural descriptions.¹

However, this understanding has been largely ignored and forgotten in the literature of formal investigations of grammars and languages, particularly in fields such as computational linguistics, biology, or more recently, neurosciences.² Thus, many recent studies in the brain science of language (or in the study of animal communications as compared to human language) deal with the alleged formal properties of languages, e.g., local dependencies, nested dependencies, cross-serial dependencies, etc. However, as indicated in many places in the literature (see note 2), these studies suffer from two major drawbacks.

One, the object “languages” are all finite, showing no discrete infinity. Since finite languages are outside of (or “below”) the popular hierarchy (the so-called Chomsky hierarchy) assumed in

*I thank Noam Chomsky for his comments on an earlier version of this paper. The research reported in this paper was supported in part by a Core Research for Evolutionary Science and Technology (CREST) grant from the Japan Science and Technology Agency (JST).

¹ See Chomsky (1986), Chapters 1 and 2 for much relevant discussion.

² References are too numerous to mention. See Everaert and Huybregts (2013) for an illuminating review. See also Ohta et al. (2013a, b) and the references cited therein for a detailed discussion on the importance of hierarchical structures and the Merge operation in the brain science of human language.

the relevant literature, the finitary status of the object languages is crucial, rendering the comparison between, say, finite-state properties and context-free properties simply irrelevant. This drawback can be partially overcome by some clever experimental tricks, such as letting the subjects “suppose” that the object language exhibits discrete infinity – even though it is, in fact, necessarily finite in actual experiments. The possibility of such a treatment indicates something significant about human cognition. That is, when objects proceed from one, two, and three, humans are most likely led to suppose that the sequence of that object will go on indefinitely, rather than stopping at some arbitrary number. This is certainly true for natural numbers, and should also hold when the object is a linguistic element. While this is an important point, I will not delve into this issue any further here.

Second, the focus of discussion so far in the literature has been on formal dependencies defined on the strings, and little attention has been paid to the abstract (hierarchical) structures assigned to the strings. This is a serious problem – because, as indicated ever since the earliest phase of contemporary generative grammar (see above), the nature of human language lies in the way it assigns abstract structures to strings (of words, phonemes, etc.) and not the set of such strings *per se*. That is, one should look for the nature of human language in its “strong generation” (the set of structures) – or ultimately, its *procedures* strongly generating the structures – rather than its “weak generation,” which is related to the set of strings it generates.

From this viewpoint, it is not at all surprising that human language does not quite fit nicely into the hierarchy in terms of weak generation (the Chomsky hierarchy). Human language is clearly beyond the bounds of finite-state grammar, and is “mildly” beyond the scope of context-free phrase structure grammar, but perhaps stays within the weak generative capacity of a certain subclass of context-sensitive phrase structure grammar (Joshi 1985). However, these results may not imply any substantive point, because even if a given grammar is adequate in its weak generation, the potential inadequacy in its strong generation will be sufficient for its exclusion as an appropriate model of human language. This is indeed true in the case of phrase structure grammars, regardless of whether it is context-free, where it seems inadequate even in its weak generation, or context-sensitive, where the way it strongly generates the structures for human language clearly exhibits its inadequacy. In short, the hierarchy in terms of weak generation, despite its naturalness and usefulness for other purposes, simply cannot provide a relevant scale along which human language is properly placed.

Let us consider some of the well-known types of dependencies in this light. For concreteness, let us take up the artificial languages discussed in Chomsky (1956, 1957): $L_1 = \{a^n b^n\}$ ($n \geq 0$) (counter language), $L_2 = xx^R$ ($x \in \{a, b\}$, x^R stands for the reversal of x) (mirror-image language), and $L_3 = xx$ ($x \in \{a, b\}$) (copying language). All these languages are shown to be beyond the bounds of finite-state grammar, and L_1 and L_2 are shown to be within the bounds of context-free phrase structure grammar. L_3 , on the other hand, is proven to be beyond the scope of context-free phrase structure grammar, since it shows cross-serial dependencies; however, its sentences can be generated by context-sensitive phrase structure grammar.

How does human language fit in this picture? It is well known that human language exhibits nested dependencies (cf. L_2) all over the place. It is also observed that human language sometimes shows cross-serial dependencies (cf. L_3). If we look at the actual cases, however, the distribution of these dependencies (among terminal elements) remains rather mysterious. For example, consider the following schematic Japanese example.

(1) NP₁-ga NP₂-ga NP₃-ga ... NP_n-ga ... V_n-to ... V₃-to V₂-to V₁ (-ga = Nom, to = that)

If we have n number of NP-*gas* and n number of Vs in this configuration, we can only have NP-V matching, as indicated above, i.e., the nested dependency. Thus, “John-*ga* Bill-*ga* Mary-*ga waratta* (laughed) *to omotta* (thought) *to itta* (said)” can only mean “John said that Bill thought that Mary laughed.” It can never mean, for example, that Mary said that Bill thought that John laughed (a case of cross-serial dependency). In a configuration such as (1), a typical sentence-embedding configuration, the linking pattern forming a nested dependency is the only possible one.

On the other hand, if we have the following coordinate configuration:

- (2) NP₁-*to* NP₂-*to* NP₃-*to* ... NP_{*n*}-(*to*-)*ga*, (*sorezore* (respectively)) *warai* (laughed) (V₁), *omoi* (thought) (V₂), *hanasi* (spoke) (V₃), ..., *itta* (said) (V_{*n*})
 [*to* = and; it can be replaced by other elements with the same meaning, e.g., *ya* (and)]

it can mean that NP₁ laughed (V₁), NP₂ thought (V₂), NP₃ spoke (V₃), ..., and NP_{*n*} said (V_{*n*}). That is, a cross-serial dependency is certainly allowed here. In addition, other dependencies are also possible. Specifically, the so-called “group reading” is possible, where the interpretation is such that a group of people comprising NP₁, NP₂, ..., NP_{*n*} collectively laughed, thought, spoke, ..., and said. The nested dependency and other (mixed order) dependencies are probably impossible to obtain. The question is why this should be the case. Why is it that in a configuration like (1), only nested dependencies are allowed, whereas in (2), group readings and cross-serial dependencies (but perhaps no other dependencies) are allowed?

The answer to this question is difficult to obtain if we only look at terminal strings. However, if we look at the structures of these configurations, the answer seems obvious. In (1), neither of the sequences NP₁, ..., NP_{*n*} (subjects of different clauses) and V₁, ..., V_{*n*} (predicates belonging to different clauses) forms a *constituent*, whereas in (2), each of the sequences NP₁, ..., NP_{*n*} (NP₁ and NP₂ and ... and NP_{*n*}, conjoined NPs) and V₁, ..., V_{*n*} (V₁-V₂- ... -V_{*n*}, conjoined predicates) forms a constituent. Thus, the generalization here can be stated as follows: cross-serial dependencies in human language are possible only when the relevant terminal elements form a constituent. This constituency requirement strongly suggests that a transformation – a structure-dependent operation – plays an important role here. The actual way of deriving the cross-serial dependencies may be due to Copying Transformation, as suggested in Chomsky (1957) (p. 47, note 11, Chapter 5) – although in the cases considered here, “Copying” cannot be literal copying and should be characterized in a more abstract way, abstracting away from terminal elements and their immediate categorial status and focusing only on structural isomorphisms.³ Note that such a copying operation is more readily formulable in the current Merge-based system than in the classical theory of grammatical transformations. While grammatical transformations in the earlier theory operate on terminal strings with designated structures (structure indices), Merge directly operates on syntactic objects (structures), rather than strings, and with no direct reference to terminal strings which are not even “strings” at the point where Merge applies, since terminal elements are yet to be linearized. In this sense, Merge is even more structure-oriented than classical transformations, and leaves room for an operation that cannot be performed by classical transformations.⁴

³ See Stabler (2004) for some related discussion on copying operations.

⁴ In various other respects, Merge is much more restricted and thus much less powerful than classical transformations (and phrase structure rules). Although there has been much interesting work in the literature (see, among others, Stabler (2010), Collins and Stabler (2011), and references cited therein), the issue of the overall

The observed linking patterns can be accounted for along the following lines. In (1), as we mentioned above, the nested dependency is naturally obtained by applying Merge in a phase-by-phase fashion, a conventional way of embedding sentences. Since neither the sequence of NPs nor that of predicates forms a constituent, the group reading is impossible. Cross-serial dependency is also impossible, because there is no structural basis for such a dependency. By contrast, the NPs and the Vs in (2) each form a constituent, and since there is a “copying” relation between the two constituents (the NP_1, \dots, NP_n sequence and the V_1, \dots, V_n sequence), group reading, which only requires the matching constituents (with no requirement on the order of terminal elements), is readily possible. Cross-serial dependency is also possible under the assumption that the structures generated by Merge ought to be maximally preserved through the linearization process, i.e., the structural properties should be directly mapped onto sequences, yielding a kind of maximal “copying” relation between the two constituents created by copying Merge. Nested dependency would force a departure from maximal copying, and hence is virtually impossible. Other mixed-order dependencies are even more difficult to obtain, yielding unintelligible interpretations.

If this line of reasoning is on the right track, the abovementioned generalization can be stated as follows:

- (3) Cross-serial dependencies are possible only to the extent that they are “transformationally” derivable.

That is, item-by-item matching on terminal strings – which is necessary when the relevant item sequences do not form a constituent, as in case (1) – is not possible in human language. Consequently, context-sensitive phrase structure grammar is also disqualified in this regard as an adequate model for human language (cf. Chomsky (1963) for relevant discussion).

Given the basic properties of the current Merge-based system briefly discussed above, the same insight can be stated even in a more general (and stronger) form, which constitutes the main hypothesis of this paper.

- (4) Hypothesis: Dependencies are possible in human language only when they are Merge-generable.

This is a generalization that cannot be made when the rule system of human language is divided into phrase structure rules and grammatical transformations, and when only a weak generation of human language is intensively investigated.

generative power of the Merge-based system, particularly with respect to its strong generative capacity, seems to remain largely open. In the current Merge-based bare phrase structure theory, where no specification of linear order and labeling is made by Merge itself, the status of such classical concepts (and all the “theorems” and generalizations based on the concepts) as right- vs. left-branching structures, nesting vs. self-embedding structures, cross-serial dependencies, etc. does not seem obvious. Thus, Merge alone cannot of course distinguish right- vs. left-branching structures. The distinction should be made in the process of linearization (phonology). Merge, applying phase-by-phase, does generate nesting, as discussed in the text, but it underspecifies it, specifying no linear order. The distinction between nesting and self-embedding cannot be made by Merge, but it ought to be handled in the course of labeling. And so on. Also, even in narrow syntax, most of the dependencies are defined not (solely) by Merge, but via Agree, “predication,” and other miscellaneous relations/operations whose status is not crystal clear at this point. All of these problems should be cleared to seriously address the issue of the generative power of the Merge-based syntax.

Merge is a crucial operation (perhaps the only core operation) of the human language faculty, a biological endowment. Thus, when humans deal with a given dependency, the language faculty comes into play if the dependency is Merge-generable (directly or indirectly), but if the dependency is not Merge-generable, humans have to utilize other cognitive resources to handle it. More specifically, Merge does not specify linear order, does not count, and applies to structured objects (constituents) under certain structural conditions (characterizable by No-tampering, etc.). It follows, then, that L_1 (counter language) above cannot be dealt with by humans in terms of “finite-state grammar with counters,” since Merge does not provide counters. In this case, however, there is an alternative approach to characterize this language by Merge alone, without recourse to counters. Thus, this may well be the “right” account of L_1 if we are concerned with what is actually happening in the human brain. Note that in terms of weak generation, there is no right-or-wrong issue concerning the choice between the two ways to characterize L_1 . L_2 (mirror-image language) is also Merge-generable. Thus, dependencies observed in (1) – nested dependencies – are easily generated by applying Merge cyclically (phase-by-phase). This is why nested dependencies abound in human language. However, cross-serial dependencies are Merge-generable only if the relevant items form constituents (note that Merge does not provide counters, nor does it give us linear order). Therefore, the distribution of cross-serial dependencies is rather limited in human language. Such a dependency is possible in (2), where the constituency requirement is fulfilled, but is disallowed in (1), where it is not.

Dependencies observed/defined on terminal strings are, in fact, epiphenomena, obtained as a consequence of Merge. Merge-generable dependencies are handled by the faculty of language, while non-Merge-generable dependencies are processed, perhaps as a kind of a puzzle or an intellectual exercise, by other cognitive capacities. It is thus predicted that the brain regions responsible for syntax, such as the left inferior frontal gyrus (L. IFG) and the left supramarginal gyrus (L. SMG) (Ohta et al. 2013a, b), will be significantly activated when Merge-generable dependencies are being processed, whereas brain activation patterns will be quite different when non-Merge-generable dependencies such as non-constituent cases of cross-serial dependencies are being processed. Reasonable and well-designed brain science experiments are likely to demonstrate these points. In fact, an experimental research is being conducted in an attempt to shed light on how Merge-generable and non-Merge-generable dependencies are differentiated in the brain (cf. Ohta et al. 2014).

In closing the discussion on generative capacity of grammars in *Aspects*, Chomsky argues for the utmost importance of explanatory adequacy/feasibility requirement as follows.

It is important to keep the requirements of explanatory adequacy and feasibility in mind when weak and strong generative capacities of theories are studied as mathematical questions. Thus one can construct hierarchies of grammatical theories in terms of weak and strong generative capacity, but it is important to bear in mind that these hierarchies do *not* necessarily correspond to what is probably the empirically most significant dimension of increasing power of linguistic theory. This dimension is presumably to be defined in terms of the scattering in value of grammars compatible with fixed data. Along this empirically significant dimension, we should like to accept the least “powerful” theory that is empirically adequate. It might conceivably turn out that this theory is extremely powerful (perhaps even universal, that is, equivalent in generative capacity to the theory of Turing machines) along the dimension of weak generative capacity, and even along the dimension of strong

generative capacity. It will not necessarily follow that it is very powerful (and hence to be discounted) in the dimension which is ultimately of real empirical significance. (p. 62)

These remarks, which appear to have been mostly forgotten or otherwise disregarded in the past fifty years or so, seem to hold true almost verbatim even now, or since the current theory is supposedly trying to go “beyond explanatory adequacy,” perhaps the empirically most significant dimension is to be defined by factors that lie beyond the feasibility requirement. The status of evaluation procedure that is behind the notion of “the scattering in value of grammars compatible with fixed data” in *Aspects* has been blurred particularly after the principles-and-parameters approach emerged around 1980.⁵ Thus, it is presently not even clear how to address what seems to be the most important theoretical question for the mathematical study of grammars. Perhaps it is premature to tackle such a question until we come up with a reasonable mathematical theory of the Merge-based generative system, which in turn should be based on a fuller understanding of the properties of human language. What has been suggested in this paper is a much more modest proposal, i.e., that we should shift our focus from weak generative capacity to, at least, strong generative capacity, i.e., the matter concerning descriptive adequacy, hoping for future development of the formal study of generative procedures themselves (Grothendieck’s theory of schemes comes to mind as a promising framework).

References

- Chomsky, Noam. 1956. Three models for the description of language. *I.R.E. Transactions on Information Theory*, vol. IT-2, Proceedings of the symposium on information theory, Sept., 113-124. [Reprinted with corrections in 1965. In *Readings in mathematical psychology II*, ed. R. Duncan Luce, Robert R. Bush and Eugene Galanter, 105-124. New York: John Wiley & Sons.]
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Chomsky, Noam. 1963. Formal properties of grammars. In *Handbook of mathematical psychology II*, ed. R. Duncan Luce, Robert R. Bush and Eugene Galanter, 323-418. New York: John Wiley & Sons.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1986. *Knowledge of language: its nature, origins, and use*. New York: Praeger.
- Collins, Chris and Edward Stabler. 2011. A formalization of minimalist syntax. Ms., New York University and University of California, Los Angeles.
- Everaert, Martin and Riny Huybregts. 2013. The design principles of natural language. In *Birdsong, speech, and language: exploring the evolution of mind and brain*, ed. Johan J. Bolhuis and Martin Everaert, 3-26. Cambridge, MA: MIT Press.
- Joshi, Aravind. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In *Natural language processing: theoretical, computational and psychological perspectives*, ed. David Dowty, Lauri Karttunen and Arnold Zwicky, 206-250. New York: Cambridge University Press.

⁵ It is important to note in this connection that the logical possibility of eliminating the evaluation procedure is already hinted at (and disregarded) in *Aspects* (pp. 36-37).

- Ohta, Shinri, Naoki Fukui and Kuniyoshi L. Sakai. 2013a. Syntactic computation in the human brain: the degree of merger as a key factor. *PLoS ONE* 8 (2), e56230, 1-16.
- Ohta, Shinri, Naoki Fukui and Kuniyoshi L. Sakai. 2013b. Computational principles of syntax in the regions specialized for language: integrating theoretical linguistics and functional neuroimaging. *Frontiers in Behavioral Neuroscience* 7, 204, 1-13.
- Ohta, Shinri, Masatomi Iizawa, Kazuki Iijima, Tomoya Nakai, Naoki Fukui, Mihoko Zushi, Hiroki Narita and Kuniyoshi L. Sakai. 2014. An on-going research: the experimental design. Paper presented at the CREST Workshop with Noam Chomsky, The University of Tokyo.
- Stabler, Edward. 2004. Varieties of crossing dependencies: structure dependence and mild context sensitivity. *Cognitive Science* 28, 699-720.
- Stabler, Edward. 2010. Computational perspectives on minimalism. In *The Oxford Handbook of Minimalism*, ed. Cedric Boeckx, 617-641. Oxford: Oxford University Press.

