Á. Gallego & D. Ott (eds.). 2015. *50 Years Later: Reflections on Chomsky's* Aspects. Cambridge, MA: MITWPL.
Available for purchase at http://mitwpl.mit.edu

PREFACE

NOAM CHOMSKY
Massachusetts Institute of Technology

Aspects of the Theory of Syntax was the first attempt to present a general framework for the work in generative grammar that had been developing since the late 1940s, with applications in a number of languages (in rough chronological order, Hebrew, English, German, Russian, Hidatsa, Turkish, Mohawk, Japanese, and some others, at varying levels of depth). It also discusses a variety of other problems of language, most of them still alive, even if in different forms. It was also one of several publications that sought to place these efforts within a larger context of earlier thinking about language and the mind that had been mostly forgotten or seriously misrepresented.[1]

This work was based on a number of assumptions that seem elementary but that departed from views that were prevalent at the time in the related disciplines. The first of these assumptions is that each language is a biological property of the individual language user, mostly parts of the brain. The biological framework was greatly enriched by Eric Lenneberg's seminal work (1967). A few years later the approach began to be called "the biolinguistic program," adopting a term suggested by Massimo Piattelli-Palmerini, who organized a major conference on these topics in 1974 bringing together biologists, linguists, philosophers and psychologists, held at MIT in conjunction with the Royaumont Center in Paris.

The second basic assumption had to do with the nature of language. Each language makes available an unbounded array of hierarchically structured expressions that have determinate interpretations at the interfaces with other internal systems: systems of thought and organization of action (the conceptual-intentional interface CI), and the sensorymotor system SM for externalization (production and perception); usually sound, though as is now known, other modalities are possible. We may call this core feature of language its Basic Principle.

The Basic Principle comprehends the entire computational aspect of language, syntax in the broad sense, including the narrow syntax that provides the expressions mapped to the interfaces and the mappings themselves, and of course the lexical atoms of computation and their various configurations.

[1] See among others Chomsky (1964, 1966, 1968), Bracken (1984).

At the time, each individual language viewed from this perspective was called "a grammar," in one of the uses of this systematically ambiguous expression. Adopting a later terminological suggestion, the system is an I-language, where "I" signifies individual, internal, and intensional (in that we are interested in the actual generative procedure, not some set of entities that it determines: a set of utterances, structures, etc.). The theory of an I-language is a (generative) grammar.

Languages can vary within the limits set by the genetic factors that specify the language faculty, called "universal grammar" (UG) in contemporary terminology. The general theory of language seeks to give an explicit account of UG, thus identifying the true nature of the Basic Principle.

Several goals for linguistic theory are discussed in *Aspects*. For particular grammars, the basic goal is essentially truth: to provide a correct theory of the I-language under consideration. To the extent that it does, the grammar meets the condition of descriptive adequacy. This is, needless to say, an extraordinarily difficult task, and even approximating it for a single language would be quite an achievement. We can regard a linguistic theory as descriptively adequate to the extent that it provides descriptively adequate grammars for every human I-language, real or attainable, and no others, thus determining exactly how the Basic Principle can be instantiated in human language. That is an even more far-reaching goal. But there is a deeper problem to be confronted as well: what is called *explanatory adequacy* in *Aspects*. A linguistic theory satisfies this condition to the extent that it shows how a descriptively adequate grammar for each I-language is selected on the basis of the data available (*primary linguistic data*), and thus shows how in principle languages can be learned, one crucial part of the biolinguistic program. From this point of view, the task of selection of a grammar is what came to be called "the logical problem of language acquisition."

As in every domain of science, we seek the simplest theory, the one with the most far-reaching capacity for explaining phenomena, rather than just describing them.[2] But within the biolinguistic program, there is an additional reason to seek the simplest theory of UG. It must have evolved, and the more complex it is, the greater the obstacles to some eventual theory of the origin of language – and the word *eventual* should be stressed, though there are a few useful hints.

As soon as the first careful efforts were undertaken to construct explicit grammars for particular languages, it was discovered that very little was understood, contrary to widespread belief. Major problems appeared everywhere, many of them never noticed in the thousands of years of intensive and productive research into language. It was also quickly discovered that from the earliest stages of acquisition, language resembles other biological systems in that what is attained is vastly underdetermined by the evidence available. This is understood to be the merest truism in other domains of biology, and is therefore not dignified by a name. In the study of language, even what seem to be truisms are highly contested, so there is a name: Poverty of Stimulus (POS). Many illustrations are given in *Aspects* and earlier publications. Most have been ignored by those who contest the significance of POS for language. A few have been addressed, mostly rather marginal questions, but with very little success, to my knowledge.[3]

In accord with normal scientific practice, the study of an I-language should seek to sort out the many variables that enter into what is directly observed in order to focus on what is relevant

[2] Surprisingly, this scientific truism is sometimes denied (Margolioash and Nusbaum, 2009), but on the basis of serious misunderstanding.

[3] One significant case has received considerable attention, structure-dependence of rules. Berwick et al. (2011) reviews every proposal that was clear enough to examine. All fail, irremediably, and in fact are asking the wrong question. More recent proposals, to my knowledge, suffer the same fate. There is a simple and well-confirmed answer, with interesting consequences, but it is never considered in this general literature. See Chomsky (2013) for a non-technical review.

for the purpose at hand. In the terminology of *Aspects*, research seeks to investigate an "ideal speaker-hearer," putting to the side non-linguistic factors that enter actual performance (memory constraints, lapses of attention, etc.). A related notion is that inquiry should distinguish competence (roughly, what the person knows) from performance (what the person does in particular circumstances).

All of this should be – in fact is – second nature to every researcher. Thus in field work with an informant, the first step is transcription of utterances, which already involves high-level idealizations, disregarding non-linguistic factors and adopting non-trivial assumptions about the linguistic factors, thus abstracting to a postulated ideal speaker-hearer and distinguishing competence from performance. Any further inquiry simply carries the process further. The same holds for investigation of any biological system, in fact anything in the empirical world. Every experiment, for example, is based on essentially these distinctions, seeking to extricate the matter of interest from what is background noise for the purpose of the inquiry.

An elementary illustration in the domain of syntax is the widely discussed property of embedding – often confused with recursion. English permits such sentences as (1) but not (2):

(1) a. If it's hot then Bill will be happy
    b. Either it's hot or it's winter
    c. Days that Bill likes are rare

(2) a. If it's hot or Bill will be happy
    b. Either it's hot then it's winter
    c. Days that Bill likes is rare

Furthermore, these can be embedded in one another, as in (3), where (1b) is embedded in (1a):

(3) If either it's hot or it's winter then Bill will be happy

Such embedding yields nested dependencies (if-then, either-or, plural-plural, etc.), and it is easy to show that unbounded nested dependencies cannot be accommodated by any of the theories of language proposed in the 1950s, the main reason why these structures received attention at the time.[4]

In 1963, George Miller and I observed that without external memory, people can understand repeated embedding up to about depth six. That is as expected, in the light of the general limits of short-term memory across species.[5] With external memory and time, of course, the limits can be exceeded, in principle indefinitely, much as in the case of arithmetical operations.

The rules of the I-language in the case of (1)-(3) are transparent. Plainly, it would be absurd to add to them a limit that happens to be the limit of short-memory, with some extra statement about how it can go on without limit when memory and other language-independent constraints are lifted.

Restating the observations in the terminology of *Aspects*, unbounded self-embedding (and nesting of dependencies) is part of the speaker's competence, but performance is constrained by language-independent memory factors.

[4] Another reason was that they fall within varieties of phrase structure grammar (PSG), though as the relevant publications of the time emphasized, all varieties of PSG fail on the crucial scientific grounds of explanatory adequacy, an observation summarized and extended in *Aspects*.

[5] Miller and Chomsky (1963), *Aspects* (p 195). Miller (1956).

All of this should be obvious. Nevertheless, these notions were highly contested at the time, and since.

The most far-reaching objection is to deny the Basic Principle, by now a very fashionable position, hence perhaps worth some attention. Perhaps the first explicit challenge was in a highly influential paper of Rumelhart and McClelland (1986), which introduced connectionist/neural net models of language. They deny that productivity is "of the essence of human computation" on the grounds that center-embedded sentences are hard to process – as indeed they are, just as large numbers are hard to multiply in one's head. The argument is by now common. For example, Margoliash and Nusbaum (2009) write that "In reality, humans rarely achieve, and only awkwardly, even a recursive level of three when using center embedded sentences in natural speech." They therefore propose that "embracing an infinite recursion from linguistics unsupported by corresponding data of actual human performance arises from an unfettered first-principle perspective" of the kind they condemn as ignoring the modern biological sciences. To take another case, Stephen Levinson (2013) reaches the same conclusion on the basis of the observation that center-embedding of even degree two is "vanishingly rare" in actual corpora, and degree three "almost never" appears.

The observations are correct, and explicitly refute the conclusions drawn from them. It is entirely true that for center-embedding beyond degree two, data are "vanishingly rare," in fact virtually non-existent. Nevertheless, as noted earlier, it has been known for half a century that people can process up to the limits of short-term memory (6 or 7) with no external aids. It follows at once that without exposure to relevant data, children know the obvious rules and in particular that they have no bounds. It follows further that that the models proposed by those who reject the Basic Principle are radically incorrect for natural language. The conclusion is indeed based on "an unfettered first principle perspective," namely the principles of rationality taken for granted in the modern biological sciences as elsewhere.

The analogy to arithmetic, just mentioned, brings out the essential point in an elementary and uncontroversial way. The arithmetical competence that appears to be a common human possession provides triples (x, y, z) where z is the product of x and y, but without external aids (memory, time, attention, life span,…) multiplication of large numbers is "vanishingly rare." The same is true of everyone's laptop. It's possible to provide it with a calculator, but without access to external memory, it will have bounded performance. Lacking any such access, the system has to be redesigned for larger calculations. With such access, it needs no redesign. That is the familiar distinction between strictly finite automata and Turing architecture.

All of this is elementary. It suffices to show that the favorite example of those who reject the Basic Principle and the competence-performance distinction refutes their assumptions, and all that goes with them.

At least some of the objections to the general framework outlined in *Aspects* make sense on the assumption that there is nothing special about human language: it is just "more of the same." In some versions, "the same" is human cognitive processes, in which case the properties of language reduce to the analogue of UG for these processes generally. In other versions, "the same" holds more broadly for non-human organisms as well. Under these assumptions, POS of course remains, as it must, but it is not specific to language.

Prima facie, this stand seems to be rather hard to sustain, if only because human language is so radically different from anything else observed in the natural world. Many more specific reasons were brought out in Lenneberg's work on biology of language, in particular, the (double)

dissociations between language capacity and other cognitive faculties. By now there is rich data on these dissociations.[6] Such facts seem hard to reconcile with the "nothing special" assumptions.

Another reason is that a new-born infant at once selects language-relevant data from the "blooming buzzing confusion" with which it is confronted. Certain brain areas selectively react to human speech rather than other noises, becoming more specialized during the first few months of life.[7] After that, language development typically follows a regular course, on to full mature competence. In contrast, an ape with about the same auditory apparatus (or visual apparatus in the case of sign), does not take the first step, let alone those that follow. Again, it seems fairly evident that "something special" is involved.

The crucial questions however have to do with explanation of the properties of language that have been discovered. This is, of course, not the place to review that the status of these considerations. I will only express my personal view that the "more of the same" conception, despite its popularity, has very little to recommend it and is largely a recipe for continued failure to learn about the nature, use, and acquisition of language.

With some qualifications, the conclusion about these matters in *Aspects* seems to me to remain valid: "there is surely no reason today for taking seriously a position that attributes a complex human achievement entirely to months (or at most years) of experience rather than to millions of years of evolution or to principles of neural organization that may be even more deeply grounded in physical law." The same holds if we replace "entirely" by "mostly." The major qualifications have to do with the length of the relevant evolutionary process. There is by now considerable evidence that the relevant time frame may be very brief and recent. And there is increasing evidence that general principles of computational complexity, perhaps grounded in physical law, are of fundamental significance in acquisition of I-language.

As I mentioned, *Aspects* was a first attempt to develop a framework for theories of language that accommodated the Basic Principle within an emerging biolinguistic framework. While the general approach seems to me essentially correct, there have, of course, been very substantial developments since.

One crucial move was to revise the *Aspects* approach to selection of grammar given data. The format-and-evaluation approach of *Aspects*, which derived from work of the late 1940s, suffered from a serious defect, as recognized there: unfeasibility. In principle, it provided an answer, but only with astronomical amounts of computation. A solution to the problem was offered by the Principles and Parameters framework that crystalized in the early 1980s, opening what proved to be a highly successful research program. It revitalized the study of language acquisition, which has made impressive progress since. It opened the way to a vast explosion in range and depth of descriptive work on languages of great typological variety. There has also been very illuminating work on the structure and variety of parameters, with consequences for comparative and historical linguistics. It also made possible a deeper inquiry that seeks to reduce the architecture of I-language and the principles specific to UG to very simple forms while maintaining (and in some interesting cases, enhancing) explanatory adequacy, thus approaching what should be true of UG for the reasons indicated above.

This is not the place to try to review these developments, or the range of approaches that have been pursued. My own sense, for what it is worth, is that the basic approach outlined in *Aspects* remains generally viable and can serve as a framework for carrying forward today's very lively and exciting research endeavors in the study of language and related domains.

---

[6] See Susan Curtiss (2012).

[7] For some recent evidence, see S. Shultz et al. (2014).

References

Berwick, Robert, Paul Pietroski, Beracah Yankama, and Noam Chomsky. 2011. Poverty of the Stimulus Revisited. *Cognitive Science* 35: 1–36.

Bracken, Harry. 1984. *Mind and Language*. Dordrecht: Foris.

Chomsky, Noam. 1964. *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1966. *Cartesian Linguistics*. New York: Harper & Row.

Chomsky, Noam. 1968. *Language and Mind*. New York: Harcourt, Brace and World.

Chomsky, Noam. 2013. What Kind of Creatures Are We? *Journal of Philosophy* 110: 645–662.

Curtiss, Susan. 2012. Revisiting Modularity: Using Language as a Window to the Mind. In R. Berwick and M. Piattelli-Palmarini (eds.), *Rich Languages from Poor Inputs*, 68-90. Oxford: Oxford University Press.

Lenneberg, Eric. 1967. *Biological Foundations of Language*. New York: Wiley.

Levinson, Stephen C. 2013. Recursion in Pragmatics. *Language* 89: 149–62.

Margoliash, Daniel and Howard C. Nusbaum. 2009. Language: the Perspective from Organismal Biology. *Trends in Cognitive Science* 13.12: 505–510.

Miller, George A. 1956. The Magical Number Seven, Plus or Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63: 81–97.

Miller, George A. and Noam Chomsky. 1963. In R. Duncan Luce, R. W. Bush, and E. Galanter (eds.), *Handbook of Mathematical Psychology* (vol. II), 269–322. New York: Wiley.

Rumelhart, David E. and James L. McClelland. 1986. PDP Models and General Issues in Cognitive Science. In D.E. Rumelhart, J.L. McClelland and the PDP Research Group, *Parallel Distributed Processing: Foundations* (vol. I), 110-146. Cambridge, MA: MIT Press.

Schultz, Sarah, Athena Vouloumanos, Randi H. Bennett and Kevin Pelphrey. 2014. Neural specialization for speech in the first months of life. *Developmental Science* 17.5: 766–774.