Reflections on references to mathematics in the work of Zellig Harris*

André Lentin
Université Paris V

Language bears, on the face of it, the promise of mathematical treatment.
— Z. Harris

The idea that language, by its very nature, requires the aid of mathematics for its study appeared as a leading idea very early in the work of Zellig Harris. As far back as 1946, in "From Morpheme to Utterance", by formalizing 'expansions' he introduced a hierarchical system of equations between linguistic categories, and he sketched the formulation of a grammar in terms of partially ordered homomorphisms. This was not the habitual way of seeing things at that time.

This leading idea only became stronger in subsequent writings. Recall that the adjective *mathematical* appears in the very titles of three major works: *Mathematical Structures of Language* (1968), *A Grammar of English on Mathematical Principles* (1982a), and *A Theory of Language and Information: A mathematical approach* (1991). The aphorism cited as an epigraph above is the beginning of a recent essay found in *Mélanges Schützenberger* (Lothaire 1990), entitled precisely "On the Mathematics of Language".

We see here then nearly fifty years during which, to realize the program that he established very early, Zellig Harris searched and found in mathematics some of his supports. This merits closer attention, and it is doubtless advisable to consider it without shutting it into the reductive box of "possible applications of mathematics to linguistics." Is not the question rather "how could a little mathematics transmute itself into linguistics?"[1]

---

*This is a revision of an essay that appeared in Daladier (1990: 85–91). Translated by Bruce Nevin with assistance of the author.

1. Harris subsequently affirmed the felicity of this expression of his aim (letter to the author, 6 Feb. 1991).

* * *

Mathematics — taking the term in a broad sense, that is, including logic — the mathematics with which Harris nourished his thinking was, quite simply, that of his times: no need to quest for a great ancestor.

From the beginning of the 20th century, logicians made an effort to give a solid foundation to the edifice of mathematics, which had been weakened by the discovery of many paradoxes. Now — chance or necessity? — this problem of *the foundations of mathematics* was more topical than ever just at the time when Harris took charge of the 'homologous' enterprise of establishing linguistics on a clear basis. Thus, the idea could be to attempt to transfer from one field to the other, certainly not methods put into practice, but their spirit, and, chief among these, the spirit of *finitism* and *constructivism.*

For the study of formal mathematics, Hilbert recommended the use of a *finitary* arithmetic as a metamathematical instrument. In this sort of arithmetic, one considers only a definite number of objects and functions that may be thought of and manipulated in an immediate and concrete manner. Because the severe restrictions imposed by finitism (this is not the place to go into technical detail) made an instrument that was too weak to reach the intended ends, subsequently the expanded point of view of *constructivism* was adopted (in which finitism appears as a special case).

In order to progress only on sure ground, constructive mathematics considered only entities constructed by means of explicitly stated rules and such that the existence of the entities in question could be held as intuitively assured. Intuitively — there you have it! The *intuitionism* of L.E.J. Brouwer (developed equally by A. Heyting) occupied among the constructivist schools a place that was important yet a bit apart from the others because it refused to define constructivity *a priori,* holding to a certain reference to intuition. Brouwer's (Kantian?) reference to intuition seems so obscure, only a philosopher would be able to examine whether Harris's thought owes something to this aspect of Brouwerian doctrine. On the other hand, it very well seems that Harris made his own the golden rule of intuitionism: its rejection of the unthinking use of *tertium non datur* (the principle of excluded middle) and its limitation to finite systems.

There is another direction of research that deserves special mention, that which was opened by the theory of types. Russell had given a first version of this as far back as 1903 in *Principles of Mathematics.* A second version, much restructured, appeared in 1910 in *Principia Mathematica,* written in collaboration with A.N. Whitehead. This theory was perceived as difficult to read, for good reason, with many commentators denouncing its faults and obscurities. However this may be for their implementation there, the ideas on which the theory was founded are solid and of real worth. They aim to prevent the paradoxes that appeared whenever one allowed oneself to apply "anything to anything else". They set up rules. For example, it is thus that a function necessarily requires an argument of a certain type. In other words, the arguments for which a function takes values — its domain of signification — lead to a characterization by a type. A type is a logical category. And since the values of such a function can serve as arguments for another, one knows that there must exist a hierarchy of types, and that this is not 'vertical' but 'branching' (whence an order). A theory of types, to be effective, must therefore provide an effective procedure for the calculus of types (and it is in regard to the details of this calculus of types that Russell's construction did not seem always satisfactory[2]).

The most difficult aspect of developing a fruitful interconnection of the two 'problems of foundations' seemed, *a priori*, to be in the notion of metalanguage. In fact Harris knew how to make good use of the very conflicts involved. Independently of Gödel, and somewhat before him, logicians of the Polish school (Łucasiewicz, Tarski) had elaborated an infinite series stacking up language (of mathematics) and metalanguages, where every step contains the syntax for the step immediately below. How could one avoid being enmeshed in a system of this sort? On what ground could one stand?
Inversely (if one could say that), Gödel, taking ordinary arithmetic as a language, succeeded in formulating in the language *certain procedures of the metalanguage*. This is how he was able to demonstrate the famous theorem which in the 1940s scintillated in its strange novelty.

One can imagine Harris meditating on, among other things, this sort of 'polarity', and drawing from it for his program this leading idea: to use in a positive way, and even in technical details, the fact that in linguistics one cannot conceive of a metalanguage situated outside of languages.

2. Russell's theory of types occasioned the glosses of innumerable commentators, among them Gödel himself, who reproached him for the lack of a rigorous statement of the syntax of the formalism. One of the more recent studies is found in Rouilhan (1996). Philippe de Rouilhan has had the merit, the courage, and the patience to invest his competencies in a difficult enterprise: he has entirely rewritten the theory of types in a rigorous language, reforming the letter while scrupulously respecting the spirit of the text. The Master emerges from the test cleansed of many accusations, such as that of having horribly confused type and order.

But there can be no doubt that Harris reflected most on algebra.

Born in 1909, Harris was the exact contemporary of many great algebraists, his peers, such as G. Birkhoff and S. MacLane in the United States or A.I. Mal'tsev in the Soviet Union: it was in this sphere of influence that he found 'his' algebra.

Algebra, as we know, can be defined as the science of *symbol manipulation.* In *classical* algebra, the symbols represent numbers, real or complex, whereas in the algebra not long ago called *modern*, they represent diverse axiomatically defined non-numeric entities.

Moderne Algebra, yes, that was the title (in German) of the two-volume work appearing in 1932 where B.L. Van der Waerden vigorously synthesized the achievements of the past and opened the way to new developments. Again, we must recall that this *Great Initiatory Book* dealt essentially with 'noble' structures: infinite commutative rings and their ideals, fields, etc. In the 1950s more 'common' structures appeared: lattices, for example. Bourbaki at the time refused to accept them, but many textbooks, of which the most famous remains *A Survey of Modern Algebra*, by the same Birkhoff and MacLane (1941), have brought them into higher education, granting them in the same stroke full citizenship. And subsequently the movement grew. In this way, during the 1950s, the notion of *algebraic system*[3] was born, which would become progressively richer in the 1960s. With this, Harris was able to express certain of his views of language in a framework familiar to mathematicians as well as to linguists.

What is an algebraic system? Perhaps it can be agreed to give the most elaborated definition (due to Birkhoff), although Harris did not *explicitly* make use of this definition.[4]

We give ourselves a family of disjoint sets, the *phyla* of the system, and a set of operations, nullary, unary, binary, tertiary (possibly partial). The nullary operations distinguish certain remarkable elements. The others each send their Cartesian product of phyla into such a phylum. Finally, we give ourselves a

3. It was probably Mal'tsev who coined the term *algebraic system.* It is attested as long ago as 1953 in his article "Ob odnom Klasse algebraiceskih sistem". To designate such an object, many authors wrote 'algebra' (homogeneous or heterogeneous) — implying 'in the sense of Universal Algebra'. For the concept, see, for example, G. Birkhoff, "The Role of Algebra in Computing", *Computers in Algebra and Number Theory*, SIAM-AMS Proceedings, Vol. IV, Providence, Rhode Island, 1971.
4. But see e.g. Harris (1991: 305 n. 16).

certain number of binary, tertiary, etc. relations. An algebraic system is called *homogeneous* if it contains just one phylum, otherwise heterogeneous. Finally, *relational system* is the preferred term for an algebraic system that involves only relations.

To firm up these ideas, we consider from this point of view an object familiar to everyone, constituting an oriented graph G. Classically, we would consider G to be a homogeneous relational system with binary relation(s) whose points are contained by the unique phylum S. But we could also define it as a heterogeneous system including a phylum V of vertices, a phylum E of edges, and two unary operations, $\alpha : E \rightarrow V$, $\omega : E \rightarrow V$, which provide each edge e with an origin $\alpha(e)$ and an extremity $\omega(e)$. This is not all, again we could consider G to be (among others!) a heterogeneous system with three phyla, understanding E and V as above and $T=\{-1,0,1\}$ with the function $\phi : E \times V \rightarrow T$ such that $\phi(e,s)=-1$ if S is the origin of e, 1 if it is the extremity, otherwise 0. This definition is particularly well adapted to algebraic topology. Rudimentary though it is, this example will suggest the idea that there are in general many very different ways to 'see' the same object as an algebraic system (the *same* object at least if one thinks that a mathematical object exists in its own right, independently of the procedures employed to grasp it).

We can see also how the notion of type could be introduced to clarify rules for calculation in a heterogeneous algebraic system: certain primitive phyla receive simple types. In fact, the types exist 'virtually' and they may be amenable to being made explicit, or not.

As we have seen, the conceptual framework of algebraic systems has nothing about it of a procrustean bed: when it is a matter of a given language, the linguist remains the absolute master of the choice of phyla, operations, and relations. But this extreme liberty is available to the linguist only at the cost of some extremely difficult epistemological problems. Suppose in effect that a linguist, studying a given language L, proposes first a system S_1, and then a system S_2. Are the systems in question different only by their way of defining one and the same 'mathematical reality' proposed as a model in one case or the other? Or on the contrary do they translate a change of model corresponding to a profound change of the linguistic 'content'?

In this matter, the mathematician can shed no more light on the debate than the following observation. The notion of isomorphism (thus also of automorphism) — meaning a biunique correspondence compatible with the operations and relations — this precious and fecund notion can no longer be applied when S_1 and S_2 have different structures, even if one has reasons to think that these systems proceed from some 'same mathematical object'.

Let us say, with Birkhoff, that when S_1 and S_2 proceed effectively from a 'same object', as in our example above, we pass from one to the other by a *cryptomorphism*. Here's the right expression: *e bene trovato!* Suffice for explanation that cryptomorphisms come under the theory of categories (algebraic functors). Some mathematicians draw the conclusion that while we are doing this it would be better immediately to undertake a formalization in the categorial framework.

So far as this concerns the linguist — but do we have the right to speak in his name? — without denying the epistemological problem, he perhaps says that a passage from S_1 to S_2 by 'simple' cryptomorphism could reflect, if necessary, a method that is rich and interesting from a linguistic point of view. For, even if S_2 is 'cryptically' equivalent to S_1, it can give to L a representation 'less tortured' than that which S_1 proposes for it, more natural and more apt to favor the progress of research. In brief, the specific task of the linguist now includes the determination of 'good' cryptomorphisms.

* * *

Harris wanted to give, even here, the leading thread appropriate to guide those who one day with a sense of responsibility will undertake to write the history of his work. Thinking of these future historians, one would suggest that mathematical considerations such as these which have just been discussed could have some usefulness. In effect, the evolution of transformational methods, when we attempt to follow it in its detail, turns out to be so complex that one must not neglect any instrument of analysis.

While awaiting the hoped-for grand essay, it seems possible to present a few introductory remarks.

The evolution of Harris's work answers to the classical schema of the spiral: it returns periodically, not of course to the point of departure, but each time to a corresponding point aligned above it.

Moreover, when Harris presented diverse theories in succession, he did not think that the latest necessarily outdated and excluded the first. In his evaluation, rather, all these theories were complementary in that they offered various points of view.

Finally, it is advisable to note that for Harris the study of numerous and diverse basic properties of sentences, together with the possibility of using selected ones of them as a central method for analyzing languages, enabled him to consider a theory of language without making appeal to a grammar of logical forms (see on this subject the end of Section 1 of "Transformational Theory" (1965)).

Is it improper to say that this 'articulation' is situated in linguistic theory at a level that is 'homologous' with that which the articulation of cryptomorphisms occupies in algebraic theory?

Besides, what does the notion of transformation cover? Nothing has evolved more in the course of time than the allocation of the stock of 'linguistic operations' between *transformations* on the one hand and *operations* on the other, the latter tending more and more to be understood as the action of some OPERATOR.

In this evolution, we see two great periods separated by an 'intermission'. The first period, which favored a set of transformations, culminated in the system proposed in *Mathematical Structures of Language* (1968). The 'intermission' is *Report and Paraphrase* (1969), the least 'mathematical' (in appearance at least) of Harris's works. The second period opens with *Notes du Cours de Syntaxe* (1976c); it strengthens the pairs (operator, operation) and ends with the specialization of transformations as reductions, proposed in *A Grammar of English on Mathematical Principles* (1982), with further developments in *Language and Information* (1988a) and *A Theory of Language and Information* (1991), not to mention *The Form of Information in Science* (1989).

But why this evolution? What prevented Harris from being satisfied with *Mathematical Structures* (1968)? Why did he pose problems again? Doubtless there were many reasons for this — such as these, among others, that in his fashion the mathematician sees.

The *abstract system* — these are the author's terms — which *Mathematical Structures* (1968) proposed was not defined canonically as an algebraic system, but only as an ordered 6-tuple comprising a set of base N and five kinds of symbols of functions. The normalization of this object into a heterogeneous algebraic system seemed not to present difficulties *in principle* but required great care in execution. The main task evidently would be the correct determination of adequate phyla. But even amended in this way, the system would still suffer from a defect due to the PARTIAL character of the functions (in fact: semi-functions). The linguistic reasons that explain this partial character had no counterpart that was more or less satisfactory, much less elegant, in pure mathematics.

What to do so that a semi-function found its arguments in a natural way, mathematically speaking?

It is in this connection that it would turn out to be useful to have read — as Harris had read — the theory of types of B. Russell. From *Notes du Cours de*

Syntaxe (1976c) there begin in effect to appear typed operators. The mathematicization moves toward an applicative calculus controlled by types permitting the definition of the structure of a sentence as a clearly defined partially ordered structure (see for example the *Grammar of English on Mathematical Principles* (1982)).

The way is opened to new developments in the future.

Extending what has just been said, from here it is possible to enrich and to specify the combinatorics of types and to make the applicative calculus benefit from the rich experience recently acquired in the domain of lambda-calculus. It is equally conceivable to reconsider the heterogeneous system considered above while bringing to bear the system of types. Perhaps one could arrive at an object cryptically linked with a certain typed lambda calculus. In this case, and for every other formalism that may be considered, it is evidently appropriate to judge as a last resort from the point of view of pertinence and linguistic transparency.

To conclude, it is perhaps appropriate to prevent a possible misunderstanding. Historical considerations have suggested a correspondence between two 'foundation problems', but it would be false to pursue the parallel too far. In the eyes of the mathematicians of today, the hope of *founding* mathematics seems to be a chimaera: mathematics is not to be founded. Does this mean that the immense efforts put out for the sake of resolving a problem now held to be impossible were all for nothing? Certainly not! Thanks to these efforts, we understand better what is and what can be the activity of mathematicians, our freedom and our responsibility, and the obligation to question ourselves periodically.

What are the integers, really? What is the continuum today? The idea prevails henceforth that the mathematician must work on the basis of *temporary agreement*, confronting the *horizon of the moment.*

In linguistics, the situation is necessarily different because — a banal observation — the facts that this science deals with are in space-time, partaking of the 'real world'. A banal observation from which the different schools draw different conclusions, even those that agree in placing a certain confidence in the utility of mathematics.

Harris, for his part, rejects any method that would fit the facts of linguistics into prefabricated formalisms, to which it is then necessary only to make a few adjustments. He believes on the contrary in the validity of mathematical structures progressively extricated from observables: such is his manner of 'founding'.

In brief, Harris does not require of mathematics anything off the shelf, but rather, as in the excellent title recalled above, *principles.*

References

Birkhoff, G. 1971. 'The role of algebra in computing'. *Computers in Algebra and Number Theory*, SIAM-AMS Proceedings, Vol. IV, Providence, Rhode Island.

Birkhoff, Garrett & Saunders MacLane. 1941. *A survey of modern algebra.* New York: Macmillan.

Harris, Zellig S. 1946a. "From morpheme to utterance". *Language* 22: 3.161–183.

Harris, Zellig S. 1965. "Transformational Theory". *Language* 41: 3.363–401.

Harris, Zellig S. 1968. *Mathematical Structures of Language.* (=*Interscience Tracts in Pure and Applied Mathematics*, 21.) New York: Interscience Publishers, John Wiley & Sons.

Harris, Zellig S. 1969. "The two systems of grammar: Report and paraphrase". Transformations and Discourse Analysis Project (TDAP) 79. Philadelphia: The University of Pennsylvania.

Harris, Zellig S. 1976c. *Notes du cours de syntaxe.* Translated and presented by Maurice Gross. Paris: Éditions du Seuil.

Harris, Zellig S. 1982a. *A Grammar of English on Mathematical Principles.* New York: John Wiley & Sons.

Harris, Zellig S. 1988a. *Language and Information.* (=Bampton Lectures in America, 28.) New York: Columbia Univ. Press.

Harris, Zellig S., Michael Gottfried, Thomas Ryckman, Paul Mattick, Jr., Anne Daladier, Tzvee N. Harris, & Suzanna Harris. 1989. *The Form of Information in Science: Analysis of an immunology sublanguage.* Preface by Hilary Putnam. (=Boston Studies in the Philosophy of Science, 104.) Dordrecht/Holland & Boston: Kluwer Academic Publishers.

Harris, Zellig S. 1990. "On the mathematics of language". In M. Lothaire (ed), *Mots (Mélanges offerts à M.-P. Schützenberger)*. Paris: Éditions Hermes.

Harris, Zellig S. 1991. *A Theory of Language and Information: A mathematical approach.* Oxford & New York: Clarendon Press.

Mal'tsev, A.I. 1953. "Ob odnom Klasse algebraiceskih sistem" ["On one class of algebraic systems"]. *Uspehi Matematiceskih Nauk* VIII.1(53):165–171.

Rouilhan, Philippe de. 1996. *Russel et le cercle des paradoxes.* Paris: Presses Universitaires de France.

Waerden, B. L. van der. 1932. *Moderne Algebra.* Berlin: Springer.

Whitehead, Alfred North, and Bertrand Russell. 1910–1913. *Principia Mathematica*, Vol. 1, 1910, Vol. 2, 1912, Vol. 3, 1913. Cambridge: University Press.