A letter from Zellig Harris to André Lentin

Explanatory notes

Bruce Nevin
Edgartown, Massachusetts

Toward the end of his life, Zellig Harris (1909–1992) wrote a brief account of the origins and development of his work to establish the foundations of a science of language on mathematical principles. A French translation was published in 1990, and in the same publication appeared a parallel appraisal from a mathematical point of view by André Lentin (1913–2015). Early the following year, having reread Lentin's essay several times, Harris wrote him the letter which is presented here. More than a gesture of thanks and appreciation, this letter further illuminates the comprehensive overview afforded by the two essays, both of which appeared in English in 2002.

1. *Introduction*

The purpose of this paper is to provide some background and context for the letter presented here, particularly as to terminology and technical concepts of mathematics which may not be familiar to all readers. Mathematicians and other specialists will of course be familiar with these matters in depth and detail that is inappropriate for this audience and purpose. I take my lead from the editor and contributors of *The Stanford Encyclopedia of Philosophy*, who have most capably balanced coverage vs. accessibility for many years.

Among those whom I invited to contribute to *The Legacy of Zellig Harris* (Nevin 2002; Nevin & Johnson 2002) was the French mathematician André Lentin (1913–2015). In April of 2000 he replied with regret that due to his advanced age he lacked the "energy" to write something new for this, but proposed that I should make an English translation of an earlier essay of his, "Quelques réflexions sur les références mathématiques dans l'œuvre de Zellig Harris". This essay (Lentin 1990) had been published a decade earlier in the journal *Langages*, in an issue titled "Les grammaires de Harris et leurs questions".

Lentin stated his reasons for recommending that I translate and publish his essay. First, in his view it couldn't be thought of as a repetition because "it did not get very wide distribution". Secondly (and he said this was his main reason),

https://doi.org/10.1075/hl.00077.nev | Published online: 2 April 2021
Historiographia Linguistica 47:2/3 (2020), pp. 303–321. issn 0302-5160 | e-issn 1569-9781

"Harris wrote me an especially friendly and warm letter to tell me that he was in perfect accord with what I had said, and that he appreciated the way in which I had expressed it." He proposed to send me the letter, an offer I of course was glad to accept. He provided a reference to it for me to place in the English translation of his essay (Lentin 2002) as its first numbered footnote, but out of evident modesty he asked that I not publish the letter itself until after his death.

Also published in that issue of *Langages* was Harris's own retrospective essay summarizing the origins and development of his work, translated to French by the editor of that issue, Anne Daladier, under the title "La genèse de l'analyse des transformations et de la métalangue".[1] His original English text[2] stands as the introduction to the first volume of *Legacy* and Lentin's essay (in translation) nicely complements it as the introduction to the second volume. It would be best to read these essays in company with Harris's letter, published here. Links to download them from the Internet are provided in the reference section.

2. *Foundations of mathematics and of empirical science*

2.1 Background

In his letter, Harris calls Lentin's essay "a masterly essay in metamathematics". In broadest terms, metamathematics is the application of mathematical methods to the study of mathematics itself, treating its language, axioms, and rules as formal mathematical objects in their own right. A major concern of metamathematics in the 20th century has been inquiry into how the foundations of mathematics could be consistently stated within mathematics itself. The demise of this grand quest, and the wonderful efflorescence of new mathematics that resulted from efforts to save it, may be said to have begun when Bertrand Russell (1872–1970) noted that the set of all sets cannot contain itself without paradox. To resolve such paradoxes and other problems, while advancing his thesis of logicism (the thesis that mathematics is in some significant sense reducible to logic), in *Principles of Mathematics* (Russell 1903) he introduced the theory of types, which restricts operations to terms of a specified type; in the instant example, the set of sets cannot be in the argument of a function which defines or generates the set of sets because it is of a higher logical type.

David Hilbert (1862–1943) undertook to find a complete and consistent set of axioms for all of mathematics, but in 1931 Gödel's second incompleteness theorem established that the consistency of such a foundation cannot be proven.

---

1. Original title: "The background of transformational and metalanguage analysis" (Harris 2002). These essays and more may be downloaded from the zelligharris.org website.

2. Thanks to the exemplary generosity of Dr. Daladier.

Although in consequence it is generally accepted that Hilbert's aim is unattainable, nevertheless the work continued in very productive "relativized" Hilbert programs, i.e. each relative to a given system and method of proof, and in these his program can be said in large measure to have succeeded.

This was all in very lively development as Harris was beginning the 'homologous' project (Lentin's term) of establishing the foundations of a science of language. Harris agreed that Lentin's pithy aphorism captures the essence of his long quest: "comment un peu de mathématique peut-il se transmuer en linguistique" (how a little mathematics might become linguistics).

2.2 Constructivism, intuitionism, and *tertium non datur*

Aristotle postulated that any proposition is either true or not true, there is no third (*tertium*) alternative—*tertium non datur* in Latin translation, also known as the principle of excluded middle. Accordingly, in classical mathematics, a demonstration that denying the existence of a given object results in contradiction is sufficient proof that the object must therefore exist. Hilbert said "Taking the principle of excluded middle from the mathematician would be the same, say, as proscribing the telescope to the astronomer or to the boxer the use of his fists."[3]

The pugilistic image is apt. He was defending the classical point of view from proponents of constructivism, who require an explicit demonstration of how to construct the object before they accept that the object exists. The classical formalism of Hilbert and others took for granted the Realist claim that mathematics discovers fundamental principles which exist in an objective reality. For L[uitzen] E[gbertus] J[an] Brouwer (1881–1966) and his followers, "mathematics is a free creation of the human mind, and an object exists if and only if it can be (mentally) constructed." On this intuitionist principle, a proof that it is impossible for a certain mathematical object not to exist is insufficient ground for asserting that it exists; one must also be able actually to construct it (Bridges & Palmgren 2018; Iemhoff 2019). Harris's comment in his letter on "proving X from non-Y" is in this spirit, at least for "any man-made or finite situation" with an empirical basis, as distinct from the problems of pure mathematics.

2.3 Finitism and recursive mathematics

An important crux of this disagreement concerns infinity and properties of infinite objects. A formalist conceives of an infinite mathematical object as an ideal object, distinguished from finite mathematical objects, which are concrete

3. In one of his lectures on the foundations of mathematics, as reprinted in Heijenoort (2002: 476).

objects. Finitism in a constructivist sense rejects infinite objects unless they can be constructed by finite means—for example, by the constructive recursive mathematics of Nikolai Aleksandrovich Shanin (1919–2011) and Andrey Markov Jr. (1903–1979).[4] We will touch on finitism again below when we consider the relevance of formalisms to grammar.

Errett [Albert] Bishop (1928–1983) brought a decisive end to the supposed indispensibility of *tertium non datur*. He was already highly regarded for his work in classical mathematics when, in the mid 1960s, he began reconstructing a large portion of its most important achievements. His book *Foundations of constructive analysis* (Bishop 1967) presents a "constructive development of deep analysis … without a commitment to Brouwer's non-classical principles or to the machinery of recursive function theory" (Bridges & Palmgren 2018).

Bishop's book had not been published when Per Martin-Löf used constructive recursive mathematics to create the formal logical system called intuitionistic type theory (Martin-Löf 1968; Dybjer & Palmgren 2020), which he subsequently developed as a philosophical foundation for Bishop's constructive formulation (Bishop 1967). As in Russell's type theory, every term has a type, and every operation is restricted to terms of a specified type. Martin-Löf (1975: 76) includes this informal statement:

> We shall think of *mathematical objects* or *constructions*. Every mathematical object is of a certain kind or *type* […and] is always given together with its type. … A type is defined by describing what we have to do in order to construct an object of that type. … Put differently, a type is well-defined if we understand … what it means to be an object of that type. Thus, for instance **N**→**N** is a type, not because we know particular number theoretic functions like the primitive recursive ones, but because we think we understand the notion of number theoretic function *in general.*

2.4 On the homology with empirical science

Lentin says Harris's work on the foundations of a science of language is 'homologous' to work on foundations of mathematics, meaning that there are parallels or systematic correspondences across their essential differences. Whereas the existence of mathematical objects can only be demonstrated by a logical proof (with contention, as we have seen, as to what constitutes sufficient proof), the existence of the objects of an empirical science is demonstrated by experiential evidence. Harris demurs from *tertium non datur* "in any man-made or finite sit-

4. In a parenthetical remark, Lentin (2002: 5) seems to tease holders of both views equally. Any pertinence to the Realist claims of Generative grammarians is unstated.

uation—other descriptions are always possible there. Indeed a major mistake in scientific articles is setting up an 'alternative' and proving X from non-Y." It need hardly be said that writings in linguistics commonly use this argument form. Hermann [Klaus] [Hugo] Weyl (1885–1955) likened such proofs to "a piece of paper which announces the presence of a treasure without disclosing its location" (Weyl 1921: 97).[5] However, both the existence and the locations of the data of language are empirically determined. Beginning with the phonemic contrasts, Harris "characterizes natural language as a system of sets of arbitrary objects, the sets being closed with respect to particular operations, with certain mappings of these sets into themselves or onto related sets" (Harris 1968: 1). It is their arbitrariness which "makes these elements available for treatment as mathematical objects" (Harris 1968: 9).

3. *Grammar and its formalizations: Realism vs description*

3.1 Properties of language, analytical methods, and formal descriptions

What is the relationship between the grammar of a language and a formalization of that grammar? This distinction is often obscured, and is even effaced when a formalization is regarded as the Reality underlying imperfect manifestations of language in speech and writing. Harris understood that grammar is immanent in the very data of language, its functional structure, and that a formalization of a grammar is a symbolic description or representation of it, more or less comprehensive, and more useful for some purposes than for others. In computational linguistics, for example, mapping task logic to program code and thence to binary machine code in digital computers places demands on what constitutes an adequate formalization, demands which cannot legitimately be presumed to be requisite for other purposes, much less intrinsic to language.

In the most general terms, to formalize a grammar is to state or restate its rules (or implied rules) in symbolic form. The objects and relations disclosed by Harris's methodology may also be presented as processes, or as rules operating upon symbolic (algebraic) representations of structures. Early formalizations in this general sense include the grammar of form-class expansions (Harris 1946), string-adjoining and tree-adjoining grammars (the history is epitomized in Joshi

5. Commonly misattributed to Weyl (1946). There, he says (p. 6) that in Russell's system "mathematics is no longer founded on logic, but on a sort of logician's paradise, a universe endowed with an 'ultimate furniture' of rather complex structure and governed by quite a number of sweeping axioms of closure. The motives are clear, but belief in this transcendental world taxes the strength of our faith hardly less than the doctrines of the early Fathers of the Church or of the scholastic philosophers of the Middle Ages."

2002), and homomorphisms among sets of sentences represented by sentence-forms and construction-forms (Harris 1957, 1965). The intimate connection of these to the data of language is evident: each of these was first presented as a method of analysis and a symbolic (algebraic) specification of the structures and relations which the method discloses in the data.

In Harris's naturalist science, the arbiter of truth is what is given in the data of language, and "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" (Lentin 2002: 6). Harris showed that each analytical method discloses a characteristic property of language. Language does have an endocentric constituent property, an exocentric adjunction property, a transformational property, and so on. These properties are discoveries in a science of language, no less objective and perdurable than the properties of elements in the periodic table.

Harris's openness to all sources of insight extended farther. Although the science of language that he developed differs from the many philosophical systems about language, primarily because its methodology arises from the insight that "the metalanguage of grammar is part of its subject matter, language itself" (Gross 2002: 67), he nevertheless did not inveigh against proposals that presuppose an *a priori* metalanguage for a science of language, such as 'Universal Grammar'. As a close colleague of Harris said, "Let many theories succeed! We will have gotten that much more insight" (Hiż 1983: 287).

3.2 From abstract algebraic system to applicative calculus

More strictly, a formalization organizes symbolic representations as a formal system in which a logical calculus of rules may be used to infer theorems from axioms. As we have seen, Hilbert proposed such a formalization of mathematics. In seminars in the late 1960s, Harris expressed hope that mathematicians might develop the 'abstract system' outlined at the end of Harris (1968) in this direction.[6] However, "—a banal observation—the facts that this science deals with are in space-time, partaking of the 'real world'" (Lentin 2000: 8). To normalize this mathematical object as an algebraic system requires partial functions (semi-functions) in order to remain faithful to the facts of language, a motivation "which had no counterpart that was more or less satisfactory much less elegant, in pure

6. In addition to André Lentin, colleagues in mathematics to whom he gave a copy of the book included Max [August] Zorn (1906–1993), best known for Zorn's lemma (in set theory), and M[arcel]-P[aul] 'Marco' Schützenberger (1920–1996), noted in linguistics for the Chomsky-Schützenberger hierarchy of context-free languages, who supervised Lentin's 1969 dissertation and was for many years his friend and working colleague.

mathematics. What to do so that a semi-function found its arguments in a natural way, mathematically speaking?"[7] (Lentin 2002: 7).

A way forward, emergent in the 1970s, was "typed operators … permitting the definition of the structure of a sentence as a clearly defined partially ordered structure". In operator grammar (Harris 1982), the type of each word in the primary (unreduced) vocabulary is determined by the type of the word or words which must be in its argument, that is, upon which it depends as a condition for word entry in the construction of a sentence. Harris refers to this as a "dependence on dependence" which imposes a partial order on the vocabulary.[8]

3.3 Typed lambda calculus

Of this partial order, Lentin says "Perhaps one could arrive at an object cryptically linked with a certain typed lambda calculus." His implicit reference is to two important papers by Anne Daladier (1990a; b) which appeared in the same volume of *Langages* with Harris (1990/2002) and Lentin (1990). In the introductory paragraphs of Daladier (2016) she recounts how she undertook a lambda calculus formalization of operator grammar (Harris 1982), presents additional examples demonstrating the difficulties she encountered, and further advances her argument.

Lambda calculus is a rewrite system invented by Alonzo Church (1903–1995) as part of his research into the foundations of mathematics. In simplest terms, a lambda function over an expression 'binds' a variable that may occur in the expression, and specifies the expression to write in its place.[9] Lambda functions may be typed, and functions in a second-order lambda calculus may bind and rewrite not only variables but also types. Montague grammar and categorial grammar are applications of lambda calculus in linguistics.

7. A *function* is a binary relation between two sets, its domain and codomain (formerly termed its range), which associates each element of the first set to exactly one element of the second set. Other terms often used synonymously are map, mapping, transformation, correspondence, and operator. A *semi-function* has a domain whose elements are functions, sets, or the like. In the definition of a *partial function*, a subset of the stated domain is intended.

8. Beginning with zero-order words or primitive arguments *N*, mostly concrete nouns, which have no dependency. In dependency formalisms the sense of 'dependency' is reversed, e.g. subject and object nouns are dependent on a verb.

9. For example, the λ in the function $\lambda x[x^2-2\cdot x+5]$ binds the variable x and in $(\lambda x[x^2-2\cdot x+5])2$ this function is applied to the argument 2. The effect is to substitute the argument for the variable x, yielding $2^2-2\cdot 2+5$, thence (by arithmetic) $4-4+5=5$ (see e.g. Alama & Korbmacher 2019). The use of the Greek letter λ to identify lambda functions is a completely arbitrary convention originating from a typesetter's error (Barendregt 1997: 182).

3.4 Inconsistency as a challenge to finitism

Daladier found that her axiomatization of operator grammar (Harris 1982, 1991) is inconsistent—that is, it leads to contradiction within that axiomatization.[10] She found that she had to use a second-order typed lambda calculus in order to account for, e.g., sentences which in her view could not be decomposed down to primitive arguments *N* (Daladier 1990b: 92) (one of several cases she advanced).[11] From this, she concluded that Harris's operator grammar must be extended with operators and reductions that are not finite.

Here, we return to finitism. "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" (Lentin 2000: 2). Among the mathematically relevant properties of language, Harris (1991: 150) affirmed, for a number of empirically observed reasons, that the grammar of a language must be finitary, having "a finite stock of elements and also of rules (constraints) of combination, with at least some of the rules being recursively applicable in making a sentence."

In his letter to Lentin, Harris was manifestly unperturbed by Daladier's result. After (1968) Harris did not himself try further to specify an abstract algebraic system for language. Beginning in the 1970s, his presentations are entirely in prose. Daladier did not undertake the task of formalizing operator grammar until the mid 1980s (Daladier 2016: 15), so this abandonment is not simply due to delegating it. Stephen B. Johnson[12] (Johnson 2002: 143) suggests additional reasons for "the complete avoidance of formalism" in his descriptions of operator grammar:

> His restraint may be partly due to a philosophy that seems to permeate his writings, that no formal system can provide the best characterization of language, and that each system is a tool that is useful for some purposes, and less useful for others. Another important methodological consideration for Harris was the search for a "least grammar"; any notation or formalism requires additional definition and explanation. (Harris 1991: 39)

10. Even if this were a proof that all axiomatized formalizations of grammar must be inconsistent, in a parallel to Gödel's proof in mathematics, it would not have a parallel significance. The principles informing the methodology of Harris's science of language are mathematical, but its foundation is empirical.

11. This empirical claim is of course subject to peer review by linguists well versed in Harris's methodology and fluent in the languages from which the examples are drawn. The difficulty of demonstrating a negative is well known. It is also subject to corpus tests (see below).

12. Because of a production error, the publisher substituted the name of another author, Stephen M. Johnson as the co-editor of Nevin & Johnson (2002).

Lentin emphasizes Harris's responsibility to the empirical data of language. After proposing that "perhaps one could arrive at an object cryptically linked with a certain typed lambda calculus" as Daladier had proposed, Lentin continues: "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." Language is what it is. This is not the first time that Harris's summary of results, or a formalism based upon it, has failed to completely 'capture' what is in language. To update Sapir's famous quip, it may even be true that "all formalisms leak," though it has not been proven.

4. *Formalization and finding further regularities in the data*

4.1 Testing an axiomatization of grammar

The adequacy of any axiomatization of grammar must be tested against large amounts of language data. This exposes essential properties of language that escape notice in the anecdotal examples and counterexamples which are so characteristic of the literature of linguistics. For example, in corpus studies of word and word-dependency frequencies it has been demonstrated repeatedly with collections even of enormous size that the ratio of *hapax legomena* to repeated words remains at about 50%, regardless of how large the corpus is made. This dooms any dictionary-based formalism. Any such corpus study requires that the formalism be implemented as a parser, raising issues of computability. Wells (1999) showed that typability and type checking in second-order lambda calculus (a.k.a. System F) are equivalent and undecidable, so any parser based on it will be computationally intractable unless it is statically typed. Moot & Retoré (2019) suggest that models of natural language semantics will almost certainly be NP-complete.[13] It should also be noted that the growing consensus in neuroscience rejects notions of symbolic representation and rule-governed information processing by the brain.

13. The computational decision problems which can be solved by a deterministic Turing machine in polynomial time are in the class P, for polynomial. (In a deterministic Turing machine, at most one action is permitted for each situation. Polynomial time means in a number of steps which is a polynomial function of the size of the problem input.) In the class NP (for Non-deterministic Polynomial) are those problems which can be solved by a non-deterministic Turing machine in polynomial time. Among these, the class of *NP-complete* problems has the peculiar property that if one of them is tractable all of them are, because any one of them can be reduced to any other. Despite being an intense focus of research, they remain computationally intractable.

4.2 Resolution of exception cases

Another reason for Harris's equanimity with respect to Daladier's result, surely, is his long experience finding unexpected resolutions of apparently intractable irregularities in language data, work which required not only persistence but great imagination and creativity. This is not an imposition of invented mechanisms, but rather a direct reflection (dare we say a representation?) of the innate capacity in language for analogic extension of established regularities, which is essential for historical change and indeed for the origin and development of language (Harris 1988: 86–113, 1991: 357–405). This admits marginal constructions—'infra-sentences', 'missing sources', and 'missing intermediates'—which are authorized by the grammar and required for derivational chains but which (for identifiable reasons!) are not supported in usage. There is even the (unexploited) possibility of factoring the selection sets of some words (Harris 1991: 281, 330, 375), which could yield homophonous words with more regular distribution, and subsets thus factored from the selection sets of different words might be recombined as suppletive allomorphs (perhaps of an existing word) if resulting regularities are advantageous. All of this kind of work obviously requires native fluency, which indeed Harris (p.c.) felt was prerequisite even for the excision operations of string analysis.

At a deeper level, Harris was moving from grammars of word-classes, which entail lexical insertion into abstract structures, to the grammar of words, in which the *N* and operator classes are emergent properties. The statistical nature of acceptabilities and selection, and the lexical specificity ('extended morphophonemics') of the domains and codomains (ranges) of the reductions, enable the self-organizing character of language that Harris describes in prose.

4.3 Cryptomorphisms subject to "pertinence and linguistic transparency"

What does it mean for an empirically grounded mathematical object to be "cryptically linked" to a formalism, as Lentin proposes? Any structure-preserving map from one mathematical structure to another one of the same type is called a morphism, and if their equivalence is not obvious, the relationship is called cryptomorphic. (The term, introduced by Birkhoff (1967), is little more than a pun on the naming of morphisms.) Lentin observes that "cryptomorphisms come under the theory of categories (algebraic functors)" and in his letter (3rd paragraph) Harris refers to category theory "as housing for cryptomorphism." Category theory is a metamathematical discipline formalizing mathematical structure and concepts. Lentin continues (2002: 6) "Some mathematicians draw the conclusion that while we are doing this it would be better immediately to undertake a formalization in the categorial framework", categorial grammar, which as noted is an application of lambda calculus to linguistics, and indeed a number of workers in Harrisian linguistics (e.g. Michael Gottfried, p.c.) have suggested the affinity of categorial grammar for representing Operator Grammar.

But an empirical science has a responsibility to the data of language that is not incumbent upon mathematics. For the linguist, a 'cryptic' commensurability of alternative systems entails judgment on empirical grounds as to which system "is 'less tortured' …, more natural and more apt to favor the progress of research" (Lentin 2002: 6). This may help to explain why Harris felt that the recurrent polemical contests over "the correct" formulation are a wasteful distraction from the work to be done.

Lentin (2002: 5–7) discusses epistemic problems with the cryptomorphic relation of alternative systems for characterizing a language, such as those generated by the several analytical tools that Harris developed and demonstrated. Harris deferred these epistemological issues to the theory of language and information that emerged later (1988, 1991), and handled the relationships among these alternative formal descriptions by his characteristic deference to their utility for one purpose or another.

5. *Beyond operator grammar*

"The way is opened to new developments in the future" (Lentin 2002: 8). Given the prior history of his formalizations of grammar, Harris would not be surprised to see Operator Grammar superseded by a more comprehensive formulation, and indeed work since his death has moved on from the typed operators and semi-functions of operator grammar.

A brief review will clarify the issues. Operator grammar has two aspects, word entry and reductions. As a simple example, in the sentence *I poured and drank water*, the operator *pour* enters on the pair (*I, water*), and *drink* enters on the pair (*I, water*), with both operators under the conjunction. As this is linearized to *I poured water and I drank water* the repeated same-positioned words are reduced to zero phonemic content.[14] As was evident even in the "report and paraphrase" analysis (Harris 1969), the word dependencies are easy to formalize, and all the notorious complexity of grammar is in the 'extended morphophonemics' of the reduction system. Inspired by the lexicon-grammar of Maurice Gross (e.g. Gross 1984), Johnson (2002) employed tables as an organizing framework in

14. The zero form of the second occurrence carries the metalinguistic information that it 'is the same' as the first (Harris 1991: 133). For simplicity, I have omitted Harris's reconstruction of this metalanguage assertion as e.g. *water (same as prior)*, for which the paired subscripts of logical symbolism would be merely an allograph.

which each lexical item has a row, and the columns specify reduction operations that apply (or not) to the given item. Gross found that "no two lexical items have identical lexical properties" (Gross 1979: 860), where for him the columns specify transformational rules and their conditions, and this is true also of Johnson's more fine-grained specification of reductions. But perhaps surprisingly the treatment as an array enables the easily formalized rules for word entry to be eliminated. There is no "insertion lexicale" in which a form-class label specified in the grammar is replaced by a lexical member of that form-class. Instead, the words themselves include among their attributes (indicated in the columns) the conditions for their entry and reduction. This easily accommodates the prevalence of incomplete sentences in spoken language (Miller & Weinert 1998) and in informal writing such as patient notes written by medical personnel (see e.g. sublanguage work of Sager and others). The compatibility with our subjective experience as language users is, I think, obvious.

What is a lexical item? On the one hand, as noted above, Harris proposed factoring vocabulary, differentiating superficially identical items to extend relations of homophony and suppletion that are recognized in traditional grammar:

> Characterizing words by their selection … might lead to … possibilities of decomposition (factoring) for particular sets of words … possible because in most cases the selection of a word includes one or more coherent ranges of selection (3.2). The effect of the coherent ranges is that there is a clustering of particular operators around clusterings of particular arguments, somewhat as in the sociometric clusterings of acquaintance and status (e.g. in charting who visits whom within a community). (Harris 1991: 330)

On the other hand, as Gross and his co-workers showed by the abundance of idioms and 'fixed phrases' in language, what superficially appears to be a phrase may function as a single lexical item. More especially, each discourse or text has a grammatical structure, disclosed by Harris's discourse analysis methodology, in which the lexical items (members of discourse equivalence-classes) are more often than not phrases rather than single words. When this is done for a number of texts in a restricted domain, a sublanguage grammar may be obtained, and the classifier vocabulary which is characteristic of the domain, especially in its common-knowledge discourses, and which is instrumental for formulating its definitions in an external metalanguage, may also be used as names of the form-classes of the sublanguage grammar. For example, the pharmacology sentence *digitalis increases the beating of the heart* is of the form DRUG AFFECTS SYMPTOM, as also is *simvastatin reduces inflammation*.

Practitioners of a technical domain such as medicine or a science sharply distinguish acceptable sentences from nonsense in the sublanguage for that domain. Their basis for doing so is articulated normatively in definitions, prescriptions for disciplined practice, and other domain-specific common knowledge, typically employing sentences from one or more epistemically prior domains, as e.g. physiology is a prior science to pharmacology. By like means, the metalanguage for a sublanguage is external to it, obviously while still within the language as a whole. In such a domain, labels for lexical classes and subclasses can constitute a formulaic representation of linguistic information (Harris et al. 1989).[15] The cross-linguistic universality of these results for a given domain gave Harris hope for a universal language of science (Harris & Mattick 1988; Bloomfield 1935, 1939) that would help ameliorate the human condition.

However, for less disciplined domains and for broad coverage, Harris's semantics is statistical in nature. This follows from his criterion for a mapping from one subset of sentences to another (transformation), namely, that relative acceptability or likelihood of sentence pairs in one homomorphous subset should be preserved for the corresponding sentence pairs in the other. Johnson's approach can specify likelihoods of word dependencies and the likelihood conditions for reductions. One way that Harris suggested to specify likelihood is to indicate for each dependency the domains in which a lexical dependency is fully acceptable. This can provide a vehicle for organizing relations among sublanguage domains, and the relation of sublanguages to their language as a whole.

This is one of many aspects of Harris's work that are supported by the impressive results of statistical learning theory (e.g. Gleitman 2002) and the achievements of 'neural net' computational treatments of language in which each word is represented as a vector of real values, and words that are close in meaning have similar vectors. A relatively recent mechanism of neural nets, tendentiously called 'attention' (Bahdanau et al. 2015), weights word-to-word associations within a sentence and across sentences. Resolution of anaphora, one of Daladier's concerns, is not programmed, it is emergent in the course of performing language tasks in specific contexts. Johnson's work in medical informatics since (Johnson 2002) has moved beyond the need for rules, treating directly of words in their alternative forms and conditions of occurrence. To reiterate what was said earlier, Harris understood that the grammar—the functional structure of the language—is immanent in the very data of language. Here, words themselves are the functions in the grammar. To identify and study mathematical objects in language from this point of view will require appropriate mathematical methods of a different kind than is familiar from the past.

---

15. For example, in an immunology sublanguage the information formula $A\ V_i\ T_b$ corresponds to English phrases like *antibody is found in blood serum* and French phrases like *des extraits de lymphocytes sont plus riches en anticorps* (Harris et al. 1989).

6. *Conclusion*

The procedural bent of constructivism in mathematics bears a 'cryptomorphic' homology to operationalism and method in empirical science, and in both there are disputes about Realism, but there are crucial disanalogies. First, as has already been noted, the data of empirical science are grounded in the observation of nature so their existence and locations are not in question. Secondly, whereas the construction of proofs is essential to logic and in mathematics, science can prove nothing conclusively; its reports of findings are always theory-dependent and provisional. The abstraction of logic and mathematics from experience, which Weyl (1946: 9–10) called the "original sin" of set theory,[16] led to the grand and productive failure to establish the foundations of mathematics in logic. In contrast, Harris demonstrated that it is possible to establish the foundations of a science of language on mathematical principles by rigorously following empirical methods that are dictated by the nature of the subject matter (Harris 1988: 1–4; 1991, ch. 2). Language is what it is, an observable natural phenomenon, and as this is an empirical science with a naturalist methodology, it neither results from nor depends for validation upon any formalization of grammars. Nevertheless, any effort at formalization may be useful, e.g. for computational treatments of language data, and all are to some degree revelatory, especially for relations of language to kindred systems of logic, mathematics, music, and other human achievements.

I adopt here Lentin's conclusion (Lentin 2002: 8–9):

> Harris … 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, […] *principles.*

This brief letter from Zellig Harris to his "dear friend" André Lentin provides a thread linking together their two essays, Harris's introducing the first volume of *The legacy of Zellig Harris* and Lentin's introducing the second. It opens to us a kind of binocular vision, one 'eye' seeing from within a science of language

16. We "see how far classical mathematics, nourished by a belief in the 'absolute' that transcends all human possibilities of realization, goes beyond such statements as can claim real meaning and truth founded on evidence. … [C]classical logic was abstracted from the mathematics of finite sets.… Forgetful of this limited origin, one afterwards mistook that logic for something above and prior to all mathematics, and finally applied it, without justification, to the mathematics of infinite sets. This is the Fall and original sin of set-theory, for which it is justly punished by the antinomies.…"

founded on "a little mathematics", the other comprehending this remarkable achievement more deeply and widely from within mathematics. We are fortunate that now all three documents can be read together. We may also note that at the time that Harris wrote this letter, 6 February 1991, there remained to him less than a year and a half of life. He surely knew that further work would be left to others, on the foundation that he had laid.

REFERENCES

Alama, Jesse & Johannes Korbmacher. 2019. "The Lambda Calculus". *The Stanford Encyclopedia of Philosophy* ed. by Edward N. Zalta. Stanford: Metaphysics Research Lab. URL =https://plato.stanford.edu/archives/spr2019/entries/lambda-calculus/

Bahdanau, Dzmitry, Kyunghyun Cho & Yoshua Bengio. 2015. "Neural Machine Translation by Jointly Learning to Align and Translate". *3rd International Conference on Learning Representations* (ICLR 2015). San Diego, CA [arXiv:1409.0473].

Barendregt, Henk. 1997. "The Impact of the Lambda Calculus in Logic and Computer Science". *The Bulletin of Symbolic Logic* 3:2.181–215. https://doi.org/10.2307/421013

Birkhoff, Garrett. 19673 [1940]1. *Lattice Theory*. American Mathematical Society Colloquium Publications, 25 (3rd ed.). Providence, R.I.: American Mathematical Society.

Bishop, Errett. 1967. *Foundations of Constructive Analysis*. New York: Academic Press.

Bloomfield, Leonard. 1935. "Linguistic Aspects of Science". *Philosophy of Science* 2.499–517. https://doi.org/10.1086/286392

Bloomfield, Leonard. 1939. *Linguistic aspects of science*. [Monograph.] *International Encyclopedia of Unified Science* 1.4. Chicago: University of Chicago Press.

Bridges, Douglas & Erik Palmgren. 2018. "Constructive Mathematics". *The Stanford Encyclopedia of Philosophy* ed. by Edward N. Zalta. Stanford: Metaphysics Research Lab. URL = https://plato.stanford.edu/archives/sum2018/entries/mathematics-constructive/

Daladier, Anne. 1990a. "Aspects constructifs des grammaires de Zellig Harris". *Langages* 99.57–84. https://doi.org/10.3406/lgge.1990.1592

Daladier, Anne. 1990b. "Une représentation applicative des énoncés et de leurs dérivations". *Langages* 99.92–127. https://doi.org/10.3406/lgge.1990.1594

Daladier, Anne. 2016. "L'inconsistance de la théorie des opérateurs de Harris : de la fin des grammaires finitistes à une typologie inductive". *Perspectives harrissiennes* ed. by Claire Martinot, Christiane Marque-Pucheu & Sonia Gerolimich, 13–35. Paris: Cellule de recherche en linguistique, (CRL).

Dybjer, Peter & Erik Palmgren. 2020. "Intuitionistic Type Theory. *The Stanford Encyclopedia of Philosophy* ed. by Edward N. Zalta. Stanford: Metaphysics Research Lab. URL = https://plato.stanford.edu/archives/sum2020/entries/type-theory-intuitionistic/

Gleitman, Lila. 2002. "Verbs of a feather flock together II: The child's discovery of words and their meanings". *The Legacy of Zellig Harris: Language and information into the 21st Century, Vol. 1: Philosophy of science, syntax, and semantics* ed. by Bruce Nevin, 209–229. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.228.17gle

Gross, Maurice. 1979. "On the Failure of Generative Grammar". *Language* 55:4.859–885. https://doi.org/10.2307/412748

Gross, Maurice. 1984. "Lexicon-grammar and the Syntactic Analysis of French". *Proceedings of the 10th International Conference on Computational Linguistics (COLING '84)*, 275–282. Stanford (Cal.): Association for Computational Linguistics. https://doi.org/10.3115/980431.980549

Gross, Maurice. 2002. "Consequences of the Metalanguage Being Included in the Language". *The Legacy of Zellig Harris: Language and information into the 21st Century, Vol. 1: Philosophy of science, syntax, and semantics* ed. by Bruce Nevin, 57–67. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.228.07gro

Harris, Zellig S. 1946. "From Morpheme to Utterance". *Language* 22:3.161–183. https://doi.org/10.2307/410205

Harris, Zellig S. 1957. "Co-occurrence and Transformation in Linguistic Structure". *Language* 33:3(1).283–340. https://doi.org/10.2307/411155

Harris, Zellig S. 1965. "Transformational Theory". *Language* 41:3.363–401. https://doi.org/10.2307/411782

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 Papers, 79.) Philadelphia: University of Pennsylvania. [Repr. in *Papers in Structural and Transformational Linguistics* ed. by in Henry Hiż, 612–692. Dordrecht/ Holland: D. Reidel, 1970; repr. and revised in *Transformationelle Analyse: Die Transformationstheorie von Zellig Harris und ihre Entwicklung / Transformational Analysis: The transformational theory of Zellig Harris and its development* ed. by Senta Plötz (=Linguistische Forschungen, 8.), 58–240. Frankfurt/Main: Athenäum-Verlag, 1972; repr. in a shortened version in *Papers on Syntax* ed. by Henry Hiż (=Synthese Language Library, 14.), 293–351. Dordrecht/Holland: D. Reidel, 1981.]

Harris, Zellig S. 1982. *A Grammar of English on Mathematical Principles*. New York: Wiley/Interscience.

Harris, Zellig S. 1988. *Language and Information*. (=Bampton Lectures in America, 28.) New York: Columbia University Press. [Revised version of the Bampton Lectures given at Columbia University, New York City, in Oct. 1986, URL= http://zelligharris.org/langinf.html.]

Harris, Zellig S. 2002 [1990]. "The background of transformational and metalanguage analysis". *The Legacy of Zellig Harris: Language and information into the 21st Century, Vol. 1: Philosophy of science, syntax, and semantics* ed. by Bruce Nevin, 1–15. Amsterdam: John Benjamins. (Translated by Anne Daladier as "La genèse de l'analyse des transformations et de la métalangue", *Langages* 99.9–19, 1990.) Available at https://www.persee.fr/doc/lgge_0458-726x_1990_num_25_99_1589

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

Harris, Zellig S. & Paul Mattick, Jr. 1988. "Science Sublanguages and the Prospects for a Global Language of Science". *Annals of the American Academy of Political and Social Science* 495.73–83 [Vol. Telescience: Scientific Communication in the Information Age.] https://doi.org/10.1177/0002716288495001007

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.

Heijenoort, Jean van. 2002. *From Frege to Gödel: A source book in mathematical logic, 1879–1931*. Cambridge: Harvard University Press.

Hiż, Henry M. 1983. Review: *Logico-Linguistic Papers* by Richard M. Martin, *Language and Society* 12:2.286–287.

Iemhoff, Rosalie. 2019. "Intuitionism in the Philosophy of Mathematics". *The Stanford Encyclopedia of Philosophy* ed. by Edward N. Zalta. Stanford: Metaphysics Research Lab. URL = https://plato.stanford.edu/archives/win2019/entries/intuitionism/

Johnson, Stephen B. 2002. "The Computability of Operator Grammar". *The Legacy of Zellig Harris: Language and information into the 21st Century, Vol. 2: Mathematics and computability of language* ed. by Bruce Nevin & Stephen B. Johnson, 143–160. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.229.11joh

Joshi, Aravind K. 2002. "Hierarchical Structure and Sentence Description". *The Legacy of Zellig Harris: Language and information into the 21st Century, Vol. 2: Mathematics and computability of language* ed. by Bruce Nevin & Stephen B. Johnson, 121–141. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.229.10jos

Lentin, André. 1990. "Quelques réflexions sur les références mathématiques dans l'œuvre de Zellig Harris". *Langages* 99.85–91. Available at https://www.persee.fr/doc/lgge_0458-726x_1990_num_25_99_1593. https://doi.org/10.3406/lgge.1990.1593

Lentin, André. 2002. "Reflections on References to Mathematics in the Work of Zellig Harris". *The Legacy of Zellig Harris: Language and information into the 21st Century, Vol. 2: Mathematics and computability of language* ed. by Bruce Nevin & Stephen B. Johnson, 1–9. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.229.03len

Mancosu, Paolo. 1998. *From Brouwer to Hilbert: The Debate on the foundations of mathematics in the 1920s*. Oxford: Clarendon Press.

Martin-Löf, Per. 1968. *Notes on Constructive Analysis*. Stockholm: Almquist & Wiksell.

Martin-Löf, Per. 1975. "An Intuitionistic Theory of Types: predicative part". *Logic Colloquium 1973* ed. by Harvey E. Rose and John C. Shepherdson, 73–118. Amsterdam: North-Holland. https://doi.org/10.1016/S0049-237X(08)71945-1

Miller, Jim & Regina Weinert. 1998. *Spontaneous Spoken Language: Syntax and discourse*. Oxford: Oxford University Press.

Moot, Richard & Christian Retoré. 2019. "Natural Language Semantics and Computability". *Journal of Logic, Language and Information* 28:2.287–307. https://doi.org/10.1007/s10849-019-09290-7

Nevin, Bruce E., ed. 2002. *The Legacy of Zellig Harris: Language and information into the 21st Century, Vol. 1: Philosophy of science, syntax, and semantics*. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.228

Nevin, Bruce E. & Stephen B. Johnson, eds. 2002. *The Legacy of Zellig Harris: Language and information into the 21st Century, Vol. 2: Mathematics and computability of language*. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.229

Russell, Bertrand. 1903. *Principles of Mathematics*. Cambridge: Cambridge University Press.

Sager, Naomi. 1981. *Natural Language Information Processing: A computer grammar of English and its applications*. Reading (Mass.): Addison-Wesley.

Wells, Joe Brian. 1999. "Typability and Type Checking in System F are Equivalent and Undecidable". *Annals of Pure and Applied Logic* 98:1–3.111–156. https://doi.org/10.1016/S0168-0072(98)00047-5
Weyl, Hermann. 1998 [1921]. "Über die neue Grundlagenkrise der Mathematik". *Mathematische Zeitschrift* 10.39–79. (Translated as "On the New Foundational Crisis in Mathematics", in *From Brouwer to Hilbert: The Debate on the foundations of mathematics in the 1920s* by Paolo Mancosu, 86–122. Oxford: Clarendon Press, 1998.)
Weyl, Hermann. 1946. "Mathematics and Logic". *American Mathematical Monthly* 53:1.2–13.

A LETTER FROM ZELLIG HARRIS TO ANDRÉ LENTI

[Transcript of the letter]
February 6, 1991
Dear Friend —

I have been in California this winter, and have only recently been able to read your article in Langages, which I have now reread more than once. It is a masterly essay in metamathematics, selecting and presenting the issues in a way that opens a path through them. "Comment un peu de mathématique peut-il se transmuer en linguistique" is precisely what I have been seeking, said better than I could have said. In general, every point you make here is just what I was after, except that the formulation is deeper, and the large picture you build sees further than I did.

About constructivism: I realize that finitary and constructive are inadequate for a total theory of mathematics if such can exist (and I appreciate what you say about a mathematician's task in lieu of such a theory), but in linguistics there is a special reason for a finitary metatheory and a constructive actuality, namely the finiteness of the human body and lifetime (and of the species to date)—unless one thinks that what drives the development [and structure] of language and its envelope structure is some relation or reality more general than man. I thought of my own attempts as being constructivist more than specifically intuitionist, because the reality and testability of the ultimate elements did not seem to me to be an issue in language (even if the ultimate elements are phonemic distinctions). But I do think that tertium non datur is untenable in any man-made or finite situation—other descriptions are always possible there. Indeed a major mistake in scientific articles is setting up an "alternative" and proving X from non-Y.

Partly because I did not study much mathematics in recent years (decades), but primarily because of course I am entirely no mathematician, I was working with too simple an understanding both of the theory of types and of category theory (as housing for cryptomorphism). The remarks you make enable me therefore to understand some aspects of my own work or results.

It was a rare experience for me how essentially my attempts could be understood; indeed my interests and intentions were always in some kind of applied mathematics (not in the usual sense) and not in linguistics for its own sake. But it was a greater experience to see what more one could make of it, as you did. It was worth doing the work, just to see such an analysis of it.

I was also not unaware that I was seeing here a piece of literary art constructed out of the scientific content itself. Science can support its own art.

As ever,
Zellig Harris

Résumé

Vers la fin de sa vie, Zellig Harris (1909–1992) a écrit un bref compte rendu des origines et du développement de son travail pour établir les fondements d'une science du langage sur des principes mathématiques. Une traduction française en a été publiée en 1990, tandis que dans la même publication paraissait un bilan parallèle, fait d'un point de vue mathématique, sous la plume d'André Lentin (1913–2015). Au début de l'année suivante, après avoir relu plusieurs fois l'essai de Lentin, Harris lui écrivit la lettre qui est présentée ici. Plus qu'un geste de remerciement et d'appréciation, cette lettre jette un éclairage supplémentaire sur le panorama complet offert par les deux essais, tous deux parus en anglais en 2002.

Zusammenfassung

Gegen Ende seines Lebens schrieb Zellig Harris (1909–1992) einen kurzen Bericht über die Ursprünge und die Entwicklung seiner Arbeit, um die Grundlagen einer Sprachwissenschaft nach mathematischen Prinzipien zu schaffen. Eine französische Übersetzung wurde 1990 veröffentlicht, und in derselben Veröffentlichung erschien eine parallele Bewertung aus mathematischer Sicht von André Lentin (1913–2015). Am Anfang des darauffolgenden Jahres schrieb Harris ihm den Brief, der hier vorgestellt wird, nachdem er Lentins Aufsatz mehrmals gelesen hatte. Dieser Brief ist mehr als eine Geste des Dankes und der Wertschätzung. Er beleuchtet den umfassenden Überblick, den die beiden Aufsätze bieten, die beide 2002 in englischer Sprache erschienen sind.

Address for correspondence

Bruce Nevin
Achumawi Database Project
The Endangered Language Fund
101 Chase Road
Edgartown Massachusetts 02539-7401
United States
bruce.nevin@gmail.com

Publication history

Date received: 30 April 2020
Date accepted: 14 February 2021
Published online: 2 April 2021