ZELLIG HARRIS' METHODOLOGY OF LANGUAGE AND INFORMATION
Boston Colloquium for the Philosophy of Science
October 8, 1991[*]

T. A. Ryckman
Dept. of Philosophy
Northwestern University

For much of the last quarter century, most beginning American students in linguistics have been initiated into the recent history of their discipline by a dramatic, even compelling, narrative relating how structuralism, as a viable research program, was rapidly eclipsed in about 1965 by the 'revolution' -- in the classically Kuhnian sense -- of generative grammar (and its many offshoots). Never mind that American structuralism was not a unified program of methods, goals or approaches, as any reading of the works of its two founders, Leonard Bloomfield and Edward Sapir, will attest.[1] But the message of this disciplinary history is that, through trenchant criticism of allegedly outmoded metascientific presuppositions and attachments of the older generation, a new 'paradigm' of linguistic research was established by a small but brilliant group of younger linguists which promised to bring together the militantly anti-behaviorist outlook of an incipient 'cognitive' psychology with an impressive array of results and techniques from the recent metamathematical study of formal languages. The confluence of these two currents, it was affirmed, would for the first time make possible a truly scientific study of the nature of language and mind, a promise which quite naturally attracted the attention of the philosophers and indeed, the wider intellectual community as a whole. And over the years, while much of the outside interest in the new linguistics has waned, dismayed by an ever-increasing

---

[1] On the different traditions stemming from Bloomfield and Sapir, see Dell Hymes and John Fought, American Structuralism, The Hague, Mouton, 1975.

proliferation of alternative perspectives, arcane notations and an apparent lack of progress in narrowing the gap between striking claim and substantive result, the 'mentalist' conception of linguistics has largely reminded intact, if only for want of a better alternative. In this respect, the legitimating function of disciplinary history is demonstrated: there has been no (or very little) looking back.

In speaking tonight about the linguistic theory of Zellig Harris I will be canvassing for a candidate alternative. Harris' work is an outgrowth of structural linguistics, or rather the combinatorial (distributional) method he pioneered. His views have culminated in the last ten or so years in a comprehensive account of the nature of language that is both remarkable in its detail and beautiful in the simplicity and power of its conceptual structure.[2] In its barest essence, Harris' work has made possible a detailed mathematical consideration of language as a self-organizing system that expresses and carries information.[3]

---

[2] See the papers 12-16 (on Operator Grammar) in Papers on Syntax, Dordrecht, D. Reidel, 1981; A Grammar of English on Mathematical Principles, New York, Wiley-Interscience, 1982 (hereinafter, GEM); Language and Information (Bampton Lectures in America at Columbia University, 1986), New York, Columbia University Press, 1988 (hereinafter, LI); The Form of Information in Science: Analysis of an Immunology Sublanguage, co-authored with M. Gottfried, T. Ryckman, P. Mattick, Jr., A. Daladier, T.N. and S. Harris and with a Preface by H. Putnam, Boston and Dordrecht, Kluwer Academic Press, 1989 (Boston Studies in the Philosophy of Science, v. 104) (hereinafter, FIS); and the synthetic presentation, A Theory of Language and Information: A Mathematical Approach, Oxford, Clarendon Press, 1991 (hereinafter, TLI).

[3] The mathematical treatment of language arises from a fundamental relation of 'dependence on dependence' which partitions the set of words (rather word occurrences in utterances) into dependence (argument requirement) classes of operators and their arguments. Operator words have a direct analogy to sentence-forming functors in the categorial grammars of Lesniewski and Ajdukiewicz in that the satisfaction of any operator word's argument requirement yields a sentence (thus some words are only arguments, i.e., have null argument requirement). Two other primary constraints are defined in terms of the dependence on dependence requirement: 1) gross inequalities of likelihood obtaining between an operator word and the words of its argument class, and 2) reduction in the phonemic shape (perhaps to zero phonemic shape) of words which have a very high likelihood of occurring in a particular operator-argument environment (are 'expectable' in that environment).

1. I want to begin by examining what I understand to be the guiding motivation for Harris' approach in linguistic theory. I do this in part because this motication has been frequently misunderstood, hence overlooked, in part because of the intrinsic interest for philosophy of science of the intimate interconnection between his theory and methodology. The entry point for this discussion is his assessment that there is no standpoint outside the data of language from which to advance theoretical inquiry. This is not to say that the study of language cannot or should not be conducted across many and varied dimensions: social, psychological, historical, anthropological, acoustic, physiological, formal, semantic, political among unnamed others. Rather, the fundamental issue concerns the problem of investigating the structure of natural language given that there is no external metalanguage in which to define the elements of language and to characterize their possible combinations. Of course, we know pre-theoretically that such structures exist since not all combinations of linguistic elements are possible sentences or utterances in a given language; moreover, people can use and understand 'new' sentences and discourses of which they have had no previous experience. And in speaking of an 'external' metalanguage, the intent is to indicate any set of sentences, expressions, or quasi-linguistic apparatus which is taken as primitive, hence not in turn requiring explanation in the background vernacular of natural language. One does not have to be too well-versed in recent linguistics or philosophy of language to find examples which violate this prohibition: levels of "logical form", Montague's higher-order intensional logic and possible-world semantics, "language of thought" hypotheses, and so on. Obviously, were such a system adopted, the linguist's job would be easier only if one saw no obligation to carry out linguistic analysis at the meta-level. On the other hand, if one accepts this challenge, one is off and running into a regress of metalanguages. For Harris the threat of regress levies an obvious injunction against pseudo-explanation; moreover,

it leads to a demand for an appropriate methodology which nowhere -- even implicitly -- appeals to an external standpoint. Although dim appreciation of this peculiar methodological situation facing the linguist was something of an article of faith in previous structuralism, I believe that Harris alone has placed the proscription of an external standpoint at the center of the devlopment of a proper linguistic methodology.[4]

The methodological situation facing the grammarian resembles, therefore, what H.M. Sheffer once referred to, in a different, but related context, as a "logocentric predicament". Under full realization of the dimensions and ramifications of this "predicament", Harris pioneered the development of distributional, or, since this term led to misunderstandings and further misperceptions, 'combinatorial' (combinational) linguistics: the task of formulating the constraints upon those combinations of linguistic elements that can occur. What Harris saw before anyone else is that such a characterization of language structure in terms of redundancies of combinations of elements, shows -- given the conceptual connection of redundancy and information[5] -- that the intrinsic effect of these constraints is to create information. My present point, however, is merely to call attention to Harris' assessment of the peculiar situation facing theoretical linguistics in view of the perceived nature of the object it studies, and second, to underscore his conclusion that it is necessary to devise appropriate methods that nowhere depend upon an external standpoint.

For the moment the concern is 'logocentricity' and the predicament thereupon

---

[4] Anecdotally: Once in conversation, when I referred to him as a 'linguist', Harris demurred, disclaiming any title as linguist, and said he preferred to think of himself as a "methodologist".

[5] For example, a random set (where nothing is predictable), or a perfect crystal (where everything is predictable), intrinsically carry no information. On the conceptual connection between redundancy and structure, see H.A. Simon, The Sciences of the Artificial, 2nd edition, Cambridge, MA, MIT Press, 1984.

facing the linguist. What is the force of this perception? By way of illustration, we can return to the context in which Sheffer's remark was made, which concerned rather the plight of the logician: In a review of volume I of the second edition of **Principia Mathematica** in 1926, Sheffer wrote:

> ...the attempt to formulate the foundations of logic
>
> is rendered arduous by a ..."logocentric" predicament.
>
> **In order to give an account of logic, we must presuppose**
>
> **and employ logic.**[6]

As Thomas Ricketts and Warren Goldfarb have recently reminded us,[7] Sheffer is here expressing a view of logic as constitutive of **any** rational thought whatsoever; thus there is no vantage point outside of logic from which to isolate and elucidate, let alone to critique, those basic notions that enter into the formulation of logic. This is, of course, the conception of logic of the logicism of Frege, and in somewhat different fashion of Russell and the Wittgenstein of the **Tractatus**. This view of logic is of course no longer the modern one and just a few years after Wittgenstein had cryptically remarked that "illogical thought" was a kind of **contradictio in adjecto** and that logical form could only be 'shown', and that what can be 'shown' cannot be 'said',[8] such a conception was completely undermined by the well-known limitative results of Gödel and Tarski, the consequent distinctions between truth and proof, of object language and metalanguage, and the controversy with intuitionism over the interpretation and meaning of even the logical constants. But at the time, the force

---

[6] "Review of Principia Mathematica, Volume I, second edition", Isis 8, (1926), 226-31, 228.

[7] Thomas G. Ricketts, "Frege, the Tractatus, and the Logocentric Predicament," Noûs 19 (1985), 3-15; Warren Goldfarb, "Logic in the Twenties: the Nature of the Quantifier," Journal of Symbolic Logic, 44 (1979), 351-68.

[8] Tractatus Logico-Philosophicus, 5.4731 and 4.1212; cf. 6.123: "Clearly the laws of logic cannot in their turn be subject to laws of logic."

of the 'absolutist' conception of logic was directed against a myriad number of attempts to show that the laws and concepts of logic itself were rooted in psychological operations of the human mind, that the laws of thought were indeed the laws that empirical psychological investigation -- in those days, that meant by introspection -- might eventually discover about the operations of the mind. The logicist perception of a "logocentric predicament" has therefore an antipsychological or antimentalistic pedigree. Logic is not reducible to psychology. Thanks to Frege's conception of a completely formal ('gapless') proof procedure for mathematics which transformed logic (at least after Russell and Whitehead's work), psychologicism in logic never recovered and logic became fully symbolic and 'mathematical'.

How does this episode bear on linguistics? It will be pointed out that there is at least an initial disanalogy which points to a crucial methodological difference between logic and grammar: the former but not the latter can presuppose the resources of language (as Frege, Russell, and Wittgenstein each observed). So there remains a fundamental distinction between the 'art of reasoning' and the 'art of speaking'.[9] Still, one might well conclude that perceptions of 'logocentricity' are simply the concomitant of a militant anti-psychologism. In the case of structural linguistics, one can certainly read (parts of) Leonard Bloomfield in this way (whose anti-mentalism is coupled with, at times, an extreme behaviorism). Yet quite apart from the philosophical battle the logicists waged against psychologism, for American structural linguistics there was another, and quite different, avenue which led to a similar

---

[9] In comments during the ensuing discussion of this paper, Professor Burton Dreben pointed out another disanalogy: for the logicists the very concept of an external standpoint was impossible -- they lacked even the notion that logic **had** a grammar (here the contrast between the early and the later Wittgenstein is dramatic). Dreben suggested another analogy to Harris' prohibition of an external metalanguage in the central tendency of modern geometry, stemming from Gauss' investigations, of investigating intrinsic properties of figures and surfaces without reference to an embedding space.

assessment of being placed in a "logocentric predicament". This arose within the

Boas/Sapir tradition of anthropological linguistics and, in particular, from the study

of Amerindian languages. Here, the concern was, as Boas warned repeatedly, not to

prejudge or anticipate the description of these widely differing languages by forcing

their description into a Procrustean Bed of grammatical categories or paradigms

inherited ~~from~~ for the most part from the ancient Greek and Latin grammarians. Historians

of ideas have noted a legacy of Wilhelm von Humboldt's conception of the 'genius' or

character of languages in the 'particularistic' approach to the study of language of

the Boas school, and it is of course this conception which issues, in the hands of

B.L. Whorf, in the strong thesis of linguistic relativity. My concern here is not with

the claim of linguistic relativity (which certainly overreaches any evidence in its

favor)[10] but rather to call attention to the central figure of Sapir both in this

tradition and as influence on Harris, and to the force of his view that

> Language is primarily a cultural or social product and
> must be understood as such.[11]

In 1951, in a 45 page review of the publication of a selection of Sapir's papers,[12]

Harris pointed out that, for Sapir, the major fact about language was the existence of

patterns -- most importantly, sound pattern -- and that these could be understood as

conventionally fixed aspects of cultural behavior in which the individual participated

but which were, in an important respect, autonomous of the individual. (Sapir writings

on language "drift' are a well-known example.) To be sure, Sapir saw patterning in all

---

[10] Cf. TLI, p.387.

[11] "The Status of Linguistics as a Science" (1929), reprinted in D. Mandelbaum (ed), Selected Writings of Edward Sapir in Language, Culture and Personality, op. cit. (note 12), 166.

[12] "Sapir's Selected Writings", Review of Selected Writings of Edward Sapir in Language, Culture and Personality, (Edited by David G. Mandelbaum), University of California Press, Berkeley and Los Angeles, 1949, in Language 27, No. 3 (1951), 288-333.

aspects of culture and of individual participation in culture. But as Harris noted, the fundamental fact here was the discovery of the existence of such patterning; it was left to a later generation to establish more precisely the kinds of structure present and their relations to one another. The starting point, established by Sapir (and less clearly by Saussure) was the principle of the phoneme. The fundamental significance of the phoneme principle lies in a distinction between **imitation** and **repetition**. As sound spectography already showed in the early post-war years, it is unlikely that any two utterances of the same sound are ever physically identical, yet each language has a relatively small number of functionally distinct sounds (perhaps two dozen or a few more) from which all utterances of the language are formed. A phoneme, of course, is an equivalence class of physical sounds, whose members are perhaps widely differing in physical contours, yet any two members of which are accounted 'the same' sound among speakers of a language; i.e., any occurrence of one member is functionally a repetition of any of the others. The methodological import of Sapir's social and cultural conception of language can be located precisely here, in his emphasis on the expressly **social** character of the determination of what counts as a repetition among sounds. This point is of surpassing importance for Harris' own innovative development of methods for investigating word combinations. For these can be seen as successively formulating broader and broader equivalence classes, elements of which, by virtue of their common environments of occurrence, are regarded as 'saying the same'. In effect, the method of phonemic analysis has been generalized and extended up through studies of languages in restricted semantic domains (e.g., science sublanguages) where word classes and subclasses, sentence types and even sentence sequence types are the equivalence classes of interest.[13] Harris has here been unique in combining the formal approach of Bloomfield with Sapir's intuitive and perceptive

---

[13] See FIS, passim.

understanding of the ~~inexorable~~ ineluctable linkage of language and culture, and in actually

seeing no conflict between the explicit item-and-arrangement method of Bloomfield and

Sapir's seemingly teleological 'process' formulations. The larger point, however, is

the emergence of a method for investigation of language structure which sees this

structure as a social and cultural product both conditioning and conditioned by the

aggregate of language users at a given time. As I will urge in a moment, this is a

conception at considerable variance from the traditional view held in philosophy of

language (from Descartes to Frege and beyond) which regards language primarily as a

means for the expression of thought. And it indicates another route from the

"logocentric predicament" which stands independent of any particular anti-mentalist

platform, such as Bloomfield's or Quine's, whatever the merits or demerits (and there

are both) of such a critique. It is in this context then that we should view Harris'

proscription of an external metalanguage.

2. Thus far the motivation. We have now to take account of Harris' insight that

pattern itself is accountable as a hierarchical structure of constraints upon

combinations of elements, whose combined effect is accretive and that language

structure, as a result, carries or expresses information. In this, Harris links up

with another current in structuralism in viewing language as having a code-like

character. We shall first need to briefly consider the widespread metaphoric

employment in theoretical linguistics of terms from communication engineering, in

particular, of 'code' and 'message' (and in particular that a 'code' carries the

information or meaning of a 'message'); subsequently to see how Harris' conception is

at variance with much of this usage.[14]

---

[14] E.g., Roman Jakobson, "Results of a Joint Conference of Anthropologists and
Linguists (1952)," reprinted in his Selected Writings (The Hague, Mouton), 2, 1971,
554-67, 559; "Linguistics and Communication Theory," Proceedings of Symposia in

As is generally known,[15] Saussure had already introduced the term 'code' into structural linguistics by initially designating <u>langue</u> as the code utilized in the combinations of sound uttered in order to express the personal thought of the speaker. (CLG,31). In so doing, it is plausibly maintained that he was reviving, in rather explicit form, a traditional (and simplistic) psychological 'theory' of the relation between thought and language.[16] However, there seem to be two conceptually separate rationales underlying his choice of this term. On the one hand, as is evident from the notorious 'circuit diagram' of speaker and auditor (CLG, 27-8), Saussure does seem to have naively assumed, at least for purposes of illustration, a traditional perspective on linguistic meaning and communication, the so-called "translation theory of understanding".[17] On the other, the structuralist concept of the phoneme as a discrete, combinatorially treatable, unit of sound (in Saussure's famous phrase, an entity wholly "contrastive, relative and negative" (CLG,164)), naturally suggested an analogy to telegraphic codes. It is worth pursuing each of these rationales in a little more detail, for, as we hope to indicate, they are associated with different and conflicting views of the nature of language and the character of language structure.

<u>Applied Mathematics</u>, vol. XII, Providence, R.I.,American Mathematical Society, 1961, 245-252.

[15] For example, Georges Mounin, "La Notion de Code en Linguistique," in <u>Linguistique contemporaires. Hommage à Eric Buyssens</u>. Bruxelles, Editions de L'Institut de Sociologie, Universite Libre de Bruxelles, 1970, 141-9, 141.

[16] One need only recall Bloomfield's ascerbic comment in an otherwise positive review (1923) of the second edition of the <u>CLG</u> (in C. Hockett (ed.) <u>A Leonard Bloomfield Anthology</u>, Bloomington, Indiana University Press, 1970, 106-8, p. 107): "Now de Saussure seems to have had no psychology at all beyond the crudest popular notions...".

[17] G.H.R. Parkinson, "The Translation Theory of Understanding," in G. Vesey (ed.), <u>Communication and Understanding</u>, Royal Institute of Philosophy Lectures, vol. 10, 1975-6, London, 1-19; and Roy Harris, <u>Reading Saussure</u>, London, Duckworth, 1987, 204-18.

The terminology of 'coding' and 'decoding' is but a relatively recent manifestation of an antiquated but powerful metaphor that still tends to dominate much thinking about linguistic communication and a fortiori about the character and nature of language structure. This is the image of a process of translation between 'ideas' or, in more au courant versions, some form of 'mental representation' and the external physical medium of language (either speech or writing). On this 'model', communication occurs when a 'message' or 'belief' existing in some (propositional?) form of representation in the mind of one person ('the speaker') is 'coded' into a physical (acoustic, orthographic) form which is the medium for transmission to another person ('the hearer') who then performs (usually unconsciously, of course) a 'decoding' of the message into a constituent set of ideas or mentally represented meanings. Accordingly, understanding between speaker and hearer occurs just in case the idea or mental representation elicited in the mind of the hearer is the same as, or sufficiently similar to, that originally in the mind of the speaker.[18]

The roots of this "conduit metaphor" of communication[19] can be traced back to antiquity, to the Stoics and even earlier, but it was John Locke (building upon a doctrine of ideas common also to, e.g., the Port-Royal Logic) who gave it nearly canonical formulation in Bk. iii of his Essay concerning Human Understanding (1690). Ecumentical in its appeal, it was adopted or tacitly assumed in both empiricist and rationalist traditions from epistemology and philosophy of language through philosophy of psychology. More recently, this doctrine, in its various and insignificantly different forms, the family of views which for obvious reasons can be termed

---

[18] E.g. P. Denes and E. Pinson, The Speech Chain: The Physics and Biology of Spoken Language, Bell Telephone Laboratories, 1963, 6; J.A. Fodor, T. Bever, and M. Garrett, The Psychology of Language, NY, McGraw-Hill, 1974, 13-14.

[19] For the term, see the paper of Michael Reddy in A. Orotony (ed.), Metaphor and Thought, New York, Cambridge Unviersity Press, 1979.

'mentalism', has enjoyed a resurgence in linguistics and philosophy in part due to the stimulus of communication engineering and computer analogies, in part due to the widely visible posture of an overtly mentalist linguistics.

On the other hand, and quite independently of the mentalist coding/decoding or translational 'model' of linguistic understanding, the supposition that the elements of langue were solely discrete, contrastive, and only relationally individuated entities, led Saussure to repeatedly make the comparison of the system of language to telegraphic codes, such as Morse (CLG,36), to the game of chess (43,125,149), and to an algebra (168). Later structuralism, particularly in America, placed considerable emphasis on the "telegraph-code structure of language" without any accompanying commitment to underlying 'ideas' or mental processes of translation, indeed, while being antipathetic to talk of such processes. To some theoretically-minded structural linguists, the code analogy seemed especially suited to the characterization of language structure as comprised, of various levels of hierarchical constructions ("molecules") from, ultimately, a few dozen phonemes ("atoms").[20] But beyond this, Harris pointed out that in the discernible code-like properties of language structure lay the grounds of an explanatory account of its character. Observing that the elements of language are discrete, arbitrary, and preset by convention within a linguistic community,[21] and, as well, noting that such structural features are required for transmissibility without error compounding, Harris drew the conclusion

---

[20] E.g., Martin Joos, "Description of Language Design," Journal of the Acoustical Society of America, 22 (1950), 701-8. p.705:"(Linguists) say, in effect, that the design of any language is essentially telegraphic -- that the language has the structure of a telegraphic code, using molecular signals made up of invariant atoms, and differing, e.g., from the Morse code principally in two ways: the codes called 'languages' have numerous layers of complexity instead of only two, and in each layer there are severe limitations upon the combinations permitted."

[21] This, Harris stresses, is what makes the hearer's rendition of an utterance a repetition (not an imitation) of the speaker.

that language structure is thereby also explicable as an instrument _for_ transmission (not only communication) of information.[22]

From this second perspective, then, one can at most say that language is analogous to a code, not that language _is_ a code. For a code is a 1-1 mapping between already well-formed expressions (the 'message') and the elements of the chosen cipher. Thus the grammatical structure of the message is presupposed.[23] But this assumption, equivalent to that of a metalanguage _external_ to the language under analysis, is precisely one that should not be made in linguistics.[24] For in attempting to specify how language 'carries' information, a linguist writing a grammar does not have the resources to reduce language to a prior 'message' inscribed in some 'internal' language or universal semantics or logical formalism. Such a reduction can only (if at all) be carried out by employing what is, from the point of view of grammar, an external metalanguage. But (as we urged above) this is an illicitly priviledged standpoint that masks the methodological importance of the social character of language: linguistic classification supervenes upon the shared behaviors of a speech community. Moreover, an external metalanguage is an idle posit that -- being explanatorily circular -- falls outside the domain of accountability of grammar.[25]

I have tried to establish that there are two distinct strands or affliations for

---

[22] _Mathematical Structures of Language_ (hereinafter, _MSL_), New York, Wiley-Interscience, 1968,6-8; more elaborately discussed in _LI_, 87-113, and in _TLI_, especially chapter 11.

[23] This is also clearly pointed to by F. François, "Le langage," _Encyclopédie de la Pleiade_, Paris, 1968, p. 11.

[24] _MSL_,11; _LI_,113;"On a Theory of Language," _Journal of Philosophy_, LXXIII (1976),253-76, 273; partially reprinted in _PS_, 377-91, at p. 389.

[25] This point is strikingly made in a different context in Michael Friedman's "Theoretical Explanation" in R. Healy (ed), _Reduction, Time and Reality: Studies in the Philosophy of the Natural Sciences_, New York and London, Cambridge University Press, 1-16, 14-16.

the prevalence of 'code' metaphors in recent linguistics. Certainly under the influence of the terminology of communication theory, some prominent structural linguists, particularly in Europe (but including Jakobson in the USA), collapsed this distinction, speaking indifferently of the coding of thought in language and of the code-structure of language.[26] In several attempts to extend concepts from communication theory to phonemic analysis, the fact that no applicable non-statistical concept of 'information' is developed in this theory has often been ignored or downplayed.[27] Harris has proceeded in quite another direction. In lieu of assuming one or another notion of information and attempting to map the structural or "logical" form of sentences into such a framework,[28] Harris has shown that the (hierarchy of) restrictions on combinations of linguistic elements have an accretive effect which is one of creating information; taken as an aggregate, they comprise what can be said to be the informational structure of the sentence, discourse, or sublanguage so characterized. In so doing, Harris has shown that he is not only a theorist of language but also one of information.

3. The informational character of language structure, coupled with the prohibition of an external metalanguage, leads at once to the methodological requirement of what

---

[26] For example, É. Benveniste (1963), "Coup d'oeil sur le developpement de la linguistique," in Problemes de linguistique générale, Paris, Gallimard, 1966, 30-1: "...on peut espérer des théories de l'information quelque clarté sur la manière dont la pensee est codée dans le langage." and 23: "la langue étant organisée systématiquement et fonctionnant selon les règles d'un code...".

[27] A. Martinet, Elements of General Linguistics, Chicago, University of Chicago Press, 1964, 172 ff., and B. Malmberg, Structural Linguistics and Human Communication, 2nd revised edition, Berlin/NY, Springer Verlag, 1967, chapter 3.

[28] TLI, p.348: "We cannot in general impose our own categories of information upon language...We cannot determine in an a priori way the 'logical form' of all sentences...We certainly cannot map them in any regular and non-subjective way into any informational framework independently and arbitrarily chosen by us."

Harris has termed a "least grammar". In this, we again see how theory and method are inseparably connected. Given the association between redundancy and structure, it is obviously essential that the grammatical statement of each restriction be maximally efficient in the following sense: it must not contribute to the redundancies of combination of elements it seeks to describe. For grammatical inefficiency gives descriptive standing to what is only an artefact of method. Of relevance here is that the notion of redundancy, in this sense, can provide a means for distinguishing information in language. Somewhat like the situation in the theory of computational complexity,[29] where the information of a string of digits in binary notation may be defined as a particular function of the length of the shortest program (also a string of digits) that computes it, the information of a language, set of discourses, or sublanguage, is expressed by the minimal grammar that completely characterizes it in terms of its recurrent elements and their modes of combination. And since there is no external metalanguage from which to derive these elements and their permitted combinations, the elements must be set up purely combinatorially.

The statement of distributional relations is a presentation of linguistic observations, data displayed in a certain organization, e.g., a tabular arrangement. Misunderstood and inappropriately criticized at the time and subsequently, as both "merely taxonomic" (as if an adequate classification of observations is unimportant) and as "hocus-pocus linguistics" (as if increasingly general, i.e., mathematical, formulation is not widely desirable in science), mere statements of distribution were never intended as "the goal" of linguistic theory; indeed, Harris' early and most detailed presentation of distributional methods concludes in a proposal that a grammar

---

[29] As developed by Martin-Löf, Kolmogorov, Chaitin, and others.

be presented as an axiomatic theory.[30] Precisely what an axiomatic formulation of a grammar had to account for ("explain") was the observed range of co-occurrences of each element, its 'distribution' in the language. The second requirement is incumbent upon the first. Since language structure (whether the recursively specified structure of the well-formed sentences of the language or the structures of discourse and sublanguage) is a structure of restrictions on word combinations, it is imperative that the grammatical characterization of this structure not contribute to the redundancies of combination, the bearers of information in the language, to be described.[31] This is not just the general methodological virtue of economy of means. A "least grammar" is not a nicety, it is a necessity. Every restriction on combinations registered by the grammar must correspond or correlate with a difference in information, a distinction recognized and intersubjectively attestable by speakers of the language. Through a process that Harris has termed "regularization",[32] the task of characterizing a language (or some restricted use of language in sublanguage or discourse) is that of replacing elements with many apparent restrictions on their combinability with less restricted elements that 'say the same'. Already in his first book of 1951 this methodology is manifest in the attempt, through successive chapters, to continually seek more and more general classifications of linguistic elements.[33] Subsequently, grammatical transformations are developed as a kind of "extended morphophonemics", more powerful regularizing methods that enable even the derivation

---

[30] (Methods in) Structural Linguistics (hereinafter,SL), Chicago, University of Chicago Press, 1951 (manuscript completed in 1947), 372-3.

[31] GEM,10-11: "the grammatical description [must be kept] as unredundant as possible so that the essential redundancy of language, as an information-bearing system (...) not be masked by further redundancy in the description itself"; see also MSL, p.12, fn 16.

[32] MSL, Chapter 6.

[33] SL, especially chapters 7 through 19.

of tense and affixes. Throughout Harris' work, elements fall into the same equivalence class only if the individual members share a formally characterizable common environment of occurrence. On the hypothesis that each such formally stated environment is also recognizably distinct by speakers of the language,[34] grammatical methodology acquires something of an 'operationist' cast. But this is because the formal characterization of information-bearing elements in language requires empirical justification through correlations with speech community- recognized repetition: which elements are 'the same' and which are 'different'. Non-repetition is then a difference that makes a difference: this is to encounter language primarily as "an item of culture". Naturally, it is also because of the fundamentally social character of repetition that there can be no 'private language'.

4. I doubt that I am alone in marveling that in his new book,[35] Harris addresses the hoary philosophical question of how language 'connects with the world', here in the context of a Gedankenexperiment concerning the development of syntax. Characteristically, Harris offers a sketch of a constructive and developmental (one might also say, historical-materialist) answer that nowhere appeals, implicitly or explicitly, to a **deus ex machina**. We have, first, a pre-syntactic use of 'words' with primary referential meaning. Certain words (sound combinations) may be thought to have been said consequent upon the saying of others: e.g., run of deer, sleeps of boy and deer, red and perhaps sleeps of flower, but not sleeps of run, or boy of sleeps. An understandable need for efficiency of communication, for minimizing ambiguity, conventionalizes and institutionalizes these differences in usage and eventually they

---

34 H. Hoenigswald, "Review of John Lyons, Structural Semantics", Journal of Linguistics, 1 (1965), 191-6, 192, gives an explicit statement of this hypothesis; see also Harris, "Distributional Structure," (1954) reprinted in PS, 3-22, 13.

35 TLI, chapters 11 and 12.

are 'frozen out' as categorial differences. This is not to say that the concept of

predication must have already been present. As Harris notes with reference to Piaget,

one does not need to understand in order to do.[36] But the predicational partial

ordering of words -- which creates syntax as a dependence on dependence relation among

words -- can plausibly be imagined to have emerged from these initial dependences of

one particular word on another; once frozen as a difference in usage, this relation

among words can be readily extrapolated to express more and more complex sentences:

e.g., continue said of eats, swims, but not of boy; believe said of man and boy

sleeps, but not of continue, and so on. But this is to say that the syntax-creating

partial ordering constraint emerges from the reasonable conjecture that certain words

are initially said 'about' certain saliencies -- i.e, objects and situations --in the

perceived world. If one can speak here, as does Harris, of the "co-occurrence" of

certain properties and objects in a commonly perceived world, e.g., red berry, red

flower, large fish, and so on, then one can understandably see in the constraints on

word co-occurrences a reflection and recording of this experience, and no doubt very

soon, even a substitute for it.[37] The partial-ordering constraint is sufficiently

general to be readily extendible for expression of many more complicated types of

predications beyond the simple attribution of salient properties to salient objects in

a commonly perceived world; so much so, these humble origins are easily obscured in

the complexity of further emerging grammatical relations (in particular, likelihood

inequalities of word co-occurrences, and in processes of ellipsis, that is, reduction

in phonemic shape). Further, even the concept of 'information' about the world may be

---

[36] TLI, 369.

[37] Cf. Sapir, "Language" (1933), as reprinted in his **Selected Writings** (note 12),
p.11: "It is important to realize that language may not only refer to experience or
even mold, interpret and discover experience, but that it is also a substitute for
it."

constructively accounted for as arising from the meaning of the sentence-forming predicational constraint that itself is a conventionalization and institutionalization of directly referential word use.[38] From this vantage point, the notorious difficulties in attempting to answer the question of how language 'hooks up to the world' stem from the non-developmental manner in which the question has been traditionally posed and treated. But from this perspective it is only obvious that form and content, syntax and semantics must surely have developed hand in hand: in Harris' new formulation, "content follows upon form" and "form follows upon content".[39]

So there is, after all, a kernel of truth in picture theories of meaning, but only that. This is not an endorsement of realism in philosophy of language; however, in locating the origin of grammatical relations in an initial purely referential and non-syntactical 'word' usage, an attenuated realism enters, as it were, through the back-door.[40] Language does not, and cannot mirror reality; at best, it reflects and records a salient order in the perceived world: this perhaps was its original mission and, not surprisingly, this remains its primary function in research reports of an experimental science, as can be demonstrated in the information structures of the language of these reports. Harris' hypothesis is that additional structures of language emerge as further conventionalizations of usage from this original, and primitive, referential function, producing constraints that only indirectly, or distantly, or not at all carry referential meaning. Within the additional latitudes of expression provided by new syntactical constructions are opened up new possibilities for purely symbolic or abstract vocabulary. Not incidently, these further structures

---

[38] cf. TLI, p.354.

[39] TLI, p.354.

[40] Harris used this metaphor in a conversation in 1986.

provide as well the wherewithal both for the development of abstract thinking, and for the formation of "nonsense" (as opposed to ungrammatical) sentences. This is a plausible accounting of language as a self-organizing system developing in tandem with the complexity of thought.

In conclusion, I wish to briefly consider the bearing of Harris' informational interpretation of language structure on the topic of explanation in linguistics. In recent years, there has been much discussion of the necessity for linguistics, as a science, to proffer explanatory theories as opposed to 'mere' descriptions of linguistic data or behavior. Influenced by prevalent realist currents in the philosophy of science, those engaged in a quest for an 'explanatory' linguistics have urged that the discernible regularities and patterns in linguistic data can only be accounted adequately explained, in the last analysis, by reference to underlying psychological and biological structures. According to this view, linguistics, with its reliance on socially and historically contingent linguistic data and with its 'abstract' characterizations of these underlying realities, is ultimately to be subsumed in some future science of the biology of cognition. For the time being, however, linguistics is to push ahead, seeking 'deeper' and more abstract theoretical characterizations, lying at further and further remove from the observable data of language. This is indeed an audacious program of research that appears to be fashioned, to a very considerable extent, on the model of the recent history of fundamental physics. Whatever the internal difficulties with such a program, it should be clear that the work surveyed here is of a completely different theoretical and explanatory orientation; as Harris has remarked, "generality is not the same thing as

abstraction."[41] My remarks have been concerned to show that in upholding the autonomy of linguistic theory (as manifest in the prohibition of an external metalanguage), Harris has not depended on outmoded strictures of positivist metascience, whether against unobservables or in favor of the instrumentalist character of scientific theories. Rather he has pursued an explanatory account of language structure consonant with the view of language as a system, evolving through a continual process of institutionalization of usage, for transmission of information.[42] As a product of selective processes of social institutionalization, the uniqueness of language's development is reduced, and language's supposed 'mystery' -- so long a source of lofty speculation -- is correspondingly diminished. On this view, language is not, in any interesting sense, located in the human genome but is a shared social practice par excellance, a point that seems especially appropriate if we consider the particular languages of the special sciences. Language is but the paramount means we have, not of communicating meaning -- for there are many non-linguistic or quasi-linguistic ways of doing that -- but of articulating, delimiting, and transmitting meaning, as predication-structured information, between one individual or group and a wider community. Moreover, the very means by which language manages to 'carry' information is not something external to the language itself but is only the structure of constraints, each one acting on another, governing its elements. Hence it follows that the pattern existing in those social practices we term a language, whether a natural language as a whole or particular uses of language in subject-matter specific domains, is one of information.

[* This paper is a slight revision of a lecture given on October 8, 1991. I would like to thank the auditors, and in particular Professors B. Dreben and M. Gottfried, for their comments and criticisms.]

---

[41] "Introduction," p.v to PS.

[42] LI, 107-113.