Chapter V

Information, Meaning and the Representation of Information as Language Structure

5.1 An Overview of a Naturalistic Approach to Language Structure and Information. Whatever the difficulties (and we hold them to be severe) that emerge for the views of so-called "informational realism" [1], or, in speaking of information as an "objective commodity, as something whose existence (as information) is (largely) independent of the interpretative activities of conscious agents" [2], matters are compounded, it seems, when we turn our attention to language and attempt to consider language informationally: the often-used (but irrevocably metaphoric) phrase ~~'~~ as a vehicle or means of 'carrying' or 'bearing' information. With the focus on language, the issue extends further than the admonition that "bare matter is inscrutable" [3]. We have also to say that bare information, i.e., without any reference to a system of representation, is similarly so, and the problem then ramifies into that posed by the question: In speaking of language 'carrying' information, are we speaking of two systems of representation and the relations between them, or only one? To raise the problem of language and information

[1] E.g., Sayre (1976:156).

[2] Dretske (1983:55).

[3] A point which, we can agree with Quine (1969:50), "does not need making". The same point, in a more specifically linguistic vein, is made by Harris (1951b:730): "Except for relatively simple parts of the physical world (like the small numbers), or very explicitly described parts of it (like the set-up of a physical experiment), we cannot get a description of the physical world except as variously perceived by the speakers of one language or another. It is therefore not in general possible to see how two language systems depart from their common physical world, but only how they depart from each other."

in this manner appears to implicate more than a distinction between the notions of 'language' and 'theory' -- theory presupposes language and not vice-versa; it is to attempt to situate the discussion of language and information in a resolutely naturalistic setting. And this is to require that grammatical analysis, no less than any other empirical inquiry, should hew to a (perhaps, the) fundamental tenet of naturalism which, in Quinian construal, is that it is illicit to invoke, or tacitly rely on, a "first philosophy". [1]

No more than ~~in~~ any other science can linguistics assume that a correspondence obtains between its objects and constructions and an antecedent (or 'external') reality (whether 'mental' or otherwise). However, this limitative epistemological maxim is doubly significant for linguistic theory since the other sciences, but not linguistics, can take for granted, i.e., as not meriting further attention or justification, the resources of ordinary language for defining its objects and for achieving the communality of understanding requisite for advancing inquiry. [2,3] For, as Quine, among others, has taken pains

[1] A classic modern statement of the case for a prima philosophia is the "Introduction" to Husserl's Cartesian Meditations.

[2] Linguistics cannot take for granted any relations held to obtain between utterances, e.g., 'S_1 says the same as S_2'. Each such relation must be accountable for within the grammatical theory and empirically warranted by speakers' determinable recognitions of sameness or difference. See below.

[3] A telling illustration, from mathematics: Speaking in 1908 of recent efforts to resolve what he terms "l'antinomie du transfini", Borel (1928:160) remarks: "Je ne m'égarerai par en discussions métaphysiques sur le sens du mot 'indéfiniment'; que l'emploi de cet mot soulève des difficultiés pour les philosophes, c'est un fait sans importance pour les mathématiciens: il leur suffit de savoir qu'ils s'entendent parfaitement entre eux, sans craindre aucune ambiguïté. Lorsqu'un de nous dit qu'il considére la suite naturelle des nombres

to demonstrate, though indeed the regress of background languages for the theories and languages of the special sciences ends with ordinary language, it is precisely here that a philosophy of language proves or not its naturalistic mettle.

Quine has directed our attention to the regress of background languages in order to stress the relativity of reference (and thus of ontology [1]). Relativity results from querying the reference of terms in a particular theory or vernacular, a procedure which invokes inevitable recourse to a background language, launching a regress which in practice is ended only by "acquiescing in our mother tongue and taking its words at face value". [2] Assuming [3] that Quine's doctrine of "ontological relativity" has, among its targets, critical bearing on Carnap's distinction between "internal" and "external" questions about linguistic frameworks, [4] Quine's point is to reject as non-naturalistic the notion of a linguistic framework whose structure,

(continued from previous page)
entires, chacun comprehend, et est assuré de comprendre la même chose que son voisin; c'est évidemment là le seul criterium possible de la validité d'un langage, celui auquel on est toujours forcé de revenir. Car les prétendus systèmes entièrement logiques reposent toujours sur le postulat de l'existence de la langue vulgaire; ce langage commun à des millions d'hommes, et avec lequel il s'entendent à peu pris entre eux, nous est donné comme un fait, qui impliquerait un grand nombre de cercles vicieux, s'il fallait le créer ex nihilo."

1969a

[1] (1969:50):"What makes sense is to say not what the objects of a theory are, absolutely speaking, but how one theory of objects is interpretable or reinterpretable in another."

[2] Ibid., 49.

[3] As Romanos (1983:49-62) seems to suggest.

[4] (1950?:??)

or conceptual content or ontology, can be completely and a priori specified by an explicit set of rules, thus rendering all significant questions as "internal" to the chosen framework and amenable to resolution in the sweet spirit of Leibnizian reasonableness -- "Calculabimus!". At this juncture, Carnapian toleration of differing frameworks may be seen to link up with rationalism in sharing the perennial dream of an unconditional or self-sufficient instrument of communication uniting peoples and cultures.[1] But naturalism needs to remain more critical and more hard-nosed.

Starting from naturalism's unrelenting hostility to the traditional portrayal of meanings as hypostatized entities, a view which he encapsules metaphorically as "the myth of the museum", Quine's concern is to show that a naturalistic philosophy of language, as rooted in a naturalized epistemology, is incompatible with the vulgar prejudice that regards

> a man's semantics as somehow determinate in his mind beyond what might be implicit in his dispositions to overt behavior (1969:27).

So it is that Quine's own naturalistic slant on language has been preoccupied with the issue of indeterminacy: the only admissible evidence for stating relations among utterances or parts of utterances is behaviorist in pedigree ("dispositions to overt behavior") and this does not suffice to guarantee the kind of in principle determinacy of meaning (as to synonymity and analyticity) supposed by the museum myth.

[1] We are speaking here only of a similarity of ésprit between the "linguistic absolutism" (Romanos' term (p. 34)) suggested by Carnap's well-known essay of 1950, and the Leibnizian quest for a Characteristica Universalis. In particular, we are not suggesting that Carnap proposed that any empirical issues were 'merely' matters of language.

Hence the standing of determinacy of meaning qua myth and the veil of illusion cloaking all mentalistic semantics.[1]

Quine's preoccupation with indeterminacy stems from a long-standing (but as we argued in Chapter 3 §2, unwarranted [2]) refusal to countenance

[1] Quine has, of course,argued at length (see, above all, his (1970a)) that indeterminacy is additional to the underdetermination of theory by observation. His case, invoking the familiar Gedankenexperiment involving "radical" translation from a hitherto unknown language -- presumably to render plausible limiting evidence of translational correspondence to determinable dispositions to overt behavior -- might be roughly summarized as follows: Even in the face of all possible observations, including those couched in "fairly common-sense talk of bodies", the pairing of observation ("occasion") sentences can only be evidentially based on what Quine terms "stimulus synonymy" (i.e., equivalence of "stimulus meaning") of some of these and perhaps inductive ('smoothing out and rounding off') simplifications. But "stimulus meaning", which is gauged by assent/dissent responses -- indices of dispositions to overt behavior -- to the linguist's queries, does not suffice to eliminate the possiblity of empirically equivalent (in terms of compatibility with all possible observation sentences) but logically incompatible (in terms of stimulus synonymous sentences having differing truth values) translations. This is primarily because stimulus meaning does not provide a determinative basis for translating a predicate of identity and other "apparatus of individuation", i.e., pronouns, plurals, relatives, etc. By appeal to a 'slippery slope' argument against absolutely distinguishing 'observation' sentences from 'theoretical' ones, Quine concludes that the foreigner's physical theory, as translated, is to be viewed as not simply underdetermined by observational evidence, but underdetermined by all possible observation, hence, indeterminate. "Ontological relativity" brings the indeterminacy home to the mother tongue inasmuch as stimulus meaning, even as aided by ostention, does not insure that the reference of terms here is requisitely determinate; the terms of the home language are likewise infected with (referential) "inscrutability".

[2] We took issue there with Quine's view of the phoneme, in particular, that phonemic contrast is a recognitive behavior of speakers interpretable as involving the notions of sameness or difference of meaning and thus implicated in the general indictment of synonymy and cognate notions of the theory of meaning.

a grammatical analysis which can and does exploit behavioral evidence bearing on the relation between utterances[1] that is not admissibly behaviorist or that such an analysis can employ an explicitly statable and empirically evaluable methodology. Corresponding to his ideologically-inspired strictures upon the scope and methods of grammatical analysis, there is an enormous effort directed at showing that what can be naturalistically said about language, as to its structure and meaning, can be said, if at all, from the supposedly firm epistemological ground secured by behaviorism. But the impracticality of ever writing an empirically adequate grammar of a language (in the sense of Chapter 4 §2 above) from this armchair perspective need hardly be recalled to the working grammarian, which must be accounted a shortcoming for any doctrine purportedly naturalistic. It is the task of this chapter, and the remaining one, to attempt, among other endeavors, to provide ~~an actual instance of a~~ demonstration that there is no need, in fact, to concur with Quine's construal of naturalism in philosophy of

[1] Both speaker and hearer must learn to recognize a common set of grammatical elements -- phonemic distinctions, vocabulary, etc. -- in respect to which they speak and perceive speech. As Harris (1968:7) points out, "It is this that makes the transmission of an utterance a repetition, whereas the attempt to redo or transmit something whose elements are continuous or not preset is an imitation." But there seems no possibility of distinguishing between repetition and imitation, nor of identifying an unredundant set of elements if, per behaviorism, the linguist is forbidden from stating that some utterances 'say the same' as others, typically as can be determined from the linguistic behavior of language users. Cf. Hiż (1985:6): "The evidence for linguistics is not only what people say but also how people relate various sayings."

language and in epistemology as inseparable from a commitment to behaviorism, a gloss which, we can agree with Shimony [1], is "excessively narrow". As well, our goal is to show that the naturalist need not despair of employing a controlled use of 'meaning' (e.g., as information) nor need he restrict it (as "stimulus meaning") to a behavioristically acceptable analogue that merely provides grist for Quine's mill of indeterminacy.

To make a beginning, the thesis of naturalism in philosophy of language might more generally, ~~but~~ albeit proscriptively, be put as holding that the analysis of a natural language cannot warrantedly proceed on the basis of objects and relations among objects defined, or definable a priori, in a metalanguage which is external to it or any other natural language. [2] The grammar of a language, that is, the statements characterizing the words and sentence-structures of the language, must be given in either the same language or in another natural language (as a grammar of English, e.g., in Turkish), making use of the same kinds of objects and sentence structures which are to be defined. But this is also to say that the statements of a grammar, though metalinguistic, are already sentences of the object language (or another natural language), not only because recognizable as such by speakers of that language, but as well because a certain restricted (proper) subset of the objects and relations -- word classes, sentence structures, and

[1] (1981:110).

[2] "External" signifies that the structure of the metalanguage is not contained within (is not a subset of) the structure of the object language; see the footnote on the next page. Whether mathematics, i.e., logic and set theory, need be considered as external to natural language is perhaps a matter of definition. Bloomfield (1939a:55-6) argues, not completely convincingly, that logic and mathematics "presuppose linguistics". The injunction here is not against the mathematization of grammar but against the supposed sufficiency of model-theoretic semantics for natural language, or, indeed, the assumption of an "Ur-language", or a "conceptual code", e.g., Fodor (1978) and (1980).

transformations -- characterizing the sentences of the object language suffice to describe the metalinguistic statements (including the statements of the grammar).[1] The assumption of a background language whose structure is not thus definable (i.e., an external metalanguage) is not innocent, for it begs precisely the fundamental questions concerning language structure that linguistic theory has as its task to answer. What is available to an informational analysis of language structure is bounded by the fact that there is no external metalanguage. This limitation suggests the elements of empirically adequate grammars can only be set up by an analysis, the basis of which lies in the

[1] Each sentence of a grammar (although not necessarily each sentence of a grammatical discussion as actually spoken or written down (Harris (1968:152)))is metalinguistic since it says something about sentences of the language, or their parts, or classes of these. Each can be characterized as containing (or derived from sentences containing) members of a class of words that name these objects (word, sentence, etc.) and as comprising one of a restricted set of sentence-types in which these words occur, e.g., (with variable 'X','Y', 'Z' etc., for mentioned object language material) X is a sentence, The word class X contains the following members: Y, Z, W, ..., The word class Y consists of words occurring in the environment W - Z, and so on. To be sure, the grammarian, in his discussions, uses sentences which are not overtly metalinguistic; some of these are, as in all discourse, forms of argument or meta-discourse, e.g., It follows from ...X... that ...Y...., or More will be said about ...Z...below. It follows that it is possible to construct a grammar of a grammar of a language, and a grammar of this grammar continuing in this way as far as one desires. More generally, "At any moment in the history of a language, it is possible to make as complete a grammar as we wish. No item of the language need be left out as undescribable: any item which is not a case of existing rules of the grammar can be fit in (as a special case under special conditions) to some existing rule in respect to which it can be described" (Harris (1968:18)).

comparison of utterances and in observations of relevant linguistic behavior. And this means the elements are determined by characterizing redundancy: which sound or word occurrences can be considered (and can be 'recognized' as such by speakers of the language) repetitions.[1] In consequence, the products of grammatical analysis cannot be a priori expected or required to be 'mirrors' or isomorphs of the purported structures of an antecedent or non-linguistic reality.[2] Rather they are structures only of an experienced world, though indeed not of the world of pre-predicative experience in which is sought the apodictic foundation for transcendental phenomenology, or of the 'moment' of aesthetic experience, or of emotional apprehension. The experienced world (or worlds) has many aspects, and gestures, intonations, the products of art, handicraft, and technology, even situations, can all

[1] As Goodman (1970:22) reminds, determining what is a repetition is always a matter of theory: "Repetitions of the same behavior, such as hitting a tennis ball against a barn door, may involve widely varying sequences of motions. And, if we experiment twice, do the differences between the two occasions make them different experiments or only different instances of the same experiment? The answer, as Sir James Thomson stresses (Boston Studies in the Philosophy of Science, II, 1965, 85), is always relative to a theory -- we cannot repeat an experiment and look for a covering theory; we must have at least a partial theory before we know whether we have a repetition of the experiment." But it is also a matter of metatheory: given the connection between redundancy and information (see below and in §§5.2 and 5.3), to characterize redundancy is also to require that redundancy be eliminated from description.

[2] Cf. Lewis (1929:360):"The most that can reasonably be believed is that experience when caught in the net of our categories, will always afford some clue to an actually existent further uniformity of some sort."

be said to have meaning. But what is special to grammatical utterances (i.e., linguistic events whose elements are discrete, preset and arbitrary (see §6.2 below)) is not that they have meaning but ~~their~~ that they are social transmissibility.[1] Grammatical analysis, from phonemes on, can yield at most structures characterizing a socially perceived and experienced world, structures in whose terms individual experience is articulated within a shared system of recognizable distinctions, rendering this experience socially transmissible.[2] And it is only through the recognitive behaviors of speakers of a language that grammatical analysis can approach the problem of

[1] Cf. Harris (1968:7):"What is special to a grammatical utterance (...) is not that it has meaning, expresses feelings, communicates, or calls for a relevant response -- all these are common to many human activities -- but that it is socially transmissible." Ziff (1979:310) cites this passage in support of the point that "coherence rather than existence is the critical factor in matters of reference". "Social transmissibility" refers to the fact that grammatical structure is preserved in communication, thus rendering all occurrences (by various speakers) of an utterance characterized by this structure repetitions. That one speaker can repeat (as opposed to imitate) another indicates a language community, which is a social entity.

[2] The evidence for language structure is based upon the comparison of utterances, with respect to whether two utterances are repetitions, i.e., 'say the same', whereas the statements and objects of a non-linguistic science require another manner of empirical justification which cannot be completely rendered by a comparison of utterances. The tendency to conflate the task of a grammatical analysis and that of a 'classification of nature' arguably rests on illicit views of the nature of language, e.g., as a Leibnizian Characteristic. Cf. Granger (1968:127-8) whose point is unfortunately obscured (see §5.2 below) by Saussurian form/content dichotomies: "Le but d'une sémantique est la structuration du système des signifiants, pris en tant que tels (ou mieux: des fonctions significatives), et non pas du système des signifiès, lequel, d'une part, constitue l'objet lui-même, thème d'une science du premier degré et non pas d'une science du langage, d'autre part, en tant que signification, renvoie à expérience totalisante dont l'interprétation est philosophique. La tendence naturelle à confondre une classification et une analyse des signifiants correspond du reste exactement à la réalisation supposée dan le langage du voeu leibnizien d'une Charactéristique. Si la nature même des objets et des expériences (la position leibnizienne rend quasi superflue la distinction des deux) est adéquatement figurée par les articulations de la langue, la science de la réalité se confondra avec une syntaxe et une sémantique,..."

meaning or information.[1]

It is not that the specifiable regularities of combination of elements have nothing to do with the traditional problem of semantics, as it is often conceived in terms of truth or correspondence, But whatever can be said by a grammatical analysis concerning an extra-linguistic reality can be said only via the intermediary of the determinable recognitions of sameness or difference shared by speakers of a language. It is in this sense that we can say with Dewey (following Peirce [2] and following Dewey, Quine [3]) that "meaning is primarily a property of behavior". [4]

[1] Cf. Cherry (1966:262): "recognition is the setting up of a relationship between two people, or one person and an object, and the particular relevant attributes, the information-bearing elements, depend upon the individual recognizing the sign. "Information," in this sense, is information to someone -- to the recognizer, with his own peculiar experience and habits."

[2] Cf. (1976:493-4):"I do not deny that a concept, or a general mental sign, may be a logical interpretant; only it cannot be the ultimate logical interpretant. It partakes somewhat of the nature of a verbal definition and is very inferior to the living definition that grows up in the habit. Consequently, the most perfect account that we can give of a concept will consist in a description of the habit that it will produce." Peirce was well aware of the threatened regress engendered by the doctrine that the meaning of an "intellectual concept" resided in a further "mental sign": "if this sign be of an intellectual kind -- as it would have to be -- it must itself have a logical interpretant; so that it cannot be the ultimate logical interpretant of the concept. It can be proved that the only mental effect that can be so produced and that is not a sign but is of a general application is a habit-change; meaning by a habit-change a modification of a person's tendencies toward action, resulting from previous experiences or from previous exertions of his will or acts, or from a complexus of both kinds of cause (1934:327)."

[3] (1969:27), (1970b:7) and (1981:46).

[4] (1925:179).

We must now distinguish information, as intended here, from the same term in information theory, where it has nothing to do with meaning.

It is often said [1] that what is central to a concept (or even the concept) of information is the notion of reduction of uncertainty. In the mathematical theory of communication, the so-called "Information Theory" stemming from Hartley [2] and given a widely-adopted formulation by Shannon, one is concerned with what is there termed "the fundamental problem of communication", viz., that the message received and reconsituted at one point occasions the same "reduction of uncertainty" (hence, has the same "amount of information") as a message selected at, coded for transmission and transmitted from, another point. [3] In this context, the reduction of uncertainty pertains therefore only to a measure in terms of which to gauge the capacity of a channel or physical system to preserve a quantity termed "amount of information" in the transmission of a message. Since this measure applied indifferently to 'messages' ~~comprised of~~ comprising arbitrarily selected symbols as well as to messages that have meaning, [4] it is often (and justifiably) said that the Shannon conception of information (more precisely,

[1] E.g., Attneave (1959:1) and Sayre (1976:22).

[2] (1928); see the discussion of this paper in §2 below.

[3] Shannon (1949:3):"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

[4] Ibid., "Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages."

"amount of information" [1]) has "nothing to do with meaning" (see §6.2 below).

Notice that the notion of reduction of uncertainty, consequent upon the selection of a message from among a set of alternatives, clearly supposes a set of possible messages; i.e., uncertainty is reduced with respect to what message is selected, not whether a given selection, or sequence of these, yields a message. And this is to assume a prior designation of elements and, if the selection is not arbitrary, of their permitted combinations, as is indicated by mention of the 'encoding' and 'decoding' of a message already linguistically represented (as the successive selection of symbols). Now it might be thought that the notion of reduction of uncertainty, so understood, and any concept of information based upon it, is an unsuitable one for a grammatical analysis compatible with naturalism where, precisely, the assumption of prior designation cannot be made.

To see what is involved here, consider another notion linked to a concept of information, namely, redundancy. Just as 'reduction of uncertainty' presupposes a prior language in which elements and their possible combinations are designated, so a natural language presupposes redundancy, i.e., that there is less than a complete utilization of all possible combinations of its elements. [2] Randomness, or lack of

[1] As Walter Pitts commented (with Shannon in agreement) in 1951, the expression "amount of information" cannot be parsed into "amount of" and "information" (von Foerster (1952:219)). This raises the question whether "amount of information" pertains to a concept of information at all; see §2 below.

[2] Harris (1968:11-12) reminds that complete utilization is possible for a simple but not for an error-correcting code. It is instructive

structure, may be expressed as an absence of redundancy (or periodicity)[1], whereas without determinable redundancy, all that can be said or exhibited of the structure of an object or physical system is merely the selfsame physical system itself: no simpler description is possible.[2] Without structure-creating redundancy, no information can be specified as 'borne' by language (except as specified in an external metalanguage) whereas with complete redundancy, information is minimal or null.[3] To exploit the Hindu image, redundancy is both the creator and the destroyer of information.[4]

The conceptual understanding of redundancy in communication theory is reasonably straight-forward; a balance is to be struck

(continued from the previous page)
to consider whether such a language, if it could exist, would be learnable. For even if it could be determined (via a prior language) that redundancy was totally absent, this would not necessarily deter what Lewis (1929:388-9) referred to as our ability to obtain "relative simplicity of recognition" by imposing redundancy; i.e., "By ignoring a sufficient portion of the characteristics of experience as it came to us, we should arrive at such simplicity that, in terms of it, even the most disadvantageous sequence of the primary constituents -- e.g., a "random" order -- must afford some repetition and uniformity. Knowledge might be made difficult, but could not be made impossible." It might then be that the response of children (who presumably lack another language) exposed to a completely unredundant 'language' would be to learn a different language, i.e., the unredundant 'language' perceived as having redundancy.

[1] Kolmogorov (1968:663) notes: "the absence of periodicity is, according to common sense, a symptom of randomness." Kolmogorov (and independently, Chaitin (1974),(1975)) has proposed a precise definition of randomness in terms of computational complexity, which is said to provide a new logical basis for information theory.

[2] Cf. Simon (1962:221):"If a complex structure is completely unredundant -- if no aspect of its structure can be inferred from any other -- than it is its own simplest description. We can exhibit it, but we cannot describe it by a simpler structure."

[3] Cf. S.S. Stevens (1950:689-90).

[4] 'Redundancy' here thus has two senses, as structure and as artifact of the characterization of structure. In what follows, the context should make clear which sense is intended.

between reliability (which favors redundancy) and variability (which does not): How much redundancy is required to achieve a minimax solution? Of course for communication theory, a practical objective -- to minimize the cost of transmission of messages -- is foremost. This leads to a search for maximally efficient codes (e.g., Huffman codes [1]) and theorems relating the amount of information in a message to inherent limitations of channel capacities. Since it is our contention that the corresponding situation in natural language cannot be expressed in terms of codes and channels, the conceptual link with redundancy can be no more than analogous. But the general point pertaining to redundancy remains: how much? and of what kinds?

The determination of redundancy is a general problem for the analysis and characterization of phenomena; it can be said that the analysis of nature has as its goal, in each case of its application, to describe a complex phenomenon through the specification of the recurrent elements of which it can be shown to be constituted (e.g., by being 'generated' from these). [2] In terms of the specification of recurrences, one constructs a set of elements and relations among these with the objective of obtaining a precisely statable hypothesis as to closure: the notion of a

[1] See e.g., Pierce (1980:92 ff).

[2] Weyl (1949:145) notes: "For this analysis it is decisive to isolate simple occurrences within the complexity of facts, and to dissect the course of the world into simple recurrent elements."

useful handle: extrinsic vs intrinsic redundancy
(We want to eliminate extrinsic redundancy (redundancy in the descriptive apparatus) so that only intrinsic redundancy remains.
Eg-

'possible object of the theory' is circumscribed as what can be so characterized via the specified elements and the permissible operations upon them.[1] In the analysis of natural language, since the elements and operations are required to correlate with determinable contrast or difference of meaning, one seeks to eliminate apparent redundancies[2] (restrictions on combinations of elements) by creating the broadest possible equivalence classes of elements. This methodology, termed 'regularization' (see § 3 below), is not pursued merely because of the substantial interest of simplicity and economy in scientific theories. Even more fundamental is the fact that since there is no external metalanguage for natural language, language structure can only be a structure of determinable contrasts or meaning differences based on comparison of utterances. This is a requirement which insists that grammatical description formalize only restrictions on combinations which correlate, in stateable ways, with differences between utterances 'recognized' by speakers of the language. Any excess capacity of the descriptive apparatus for natural language -- where "excess" means 'not so recognized by speakers' -- distorts or falsifies the structural characteri-

[1] Thus quantitatively, i.e., mathematically, describing the phenomena; cf. Weyl (1941:116): "In our analysis of nature we reduce the phenomena to simple elements each of which varies over a certain range of possibilities which we can survey a priori because we construct these possibilities a priori in a purely combinatorial fashion from some purely symbolic material. ...Because of this a priori construction, we speak of a quantitative analysis of nature; I believe the word quantitative, if one can give it a meaning at all, ought to be interpreted in this wide sense."

[2] E.g., the distinction between 'regular' and 'irregular' verb forms, or that of the various 'moods' of sentences; see § 3.

zation by contributing to the redundancy.[1] Since the equivalences between utterances are established not a priori but only with respect to what speakers of the language can consider as repetitions, the structure of these equivalence classes and their relations has meaning; it is a structure of grammatically characterizable meaning differences.

Crucially, the relation of recognizable repetition must be determined over a specifiable domain. This may be seen with regard to both the grammar of the language as a whole (e.g., English) and grammars of restricted subsets of sentences within the language as a whole.

For the grammar of the language as a whole, the restrictions on combinations are sought which suffice to specify 'all and only' (with the provisos of Chapter 4 §2) the sentences of the language. This entails that the grammar cannot impose restrictions that rule out as possible sentences of the language, word sequences which have low or very low acceptability. One approach, adopted here and elaborated in § 3 below, characterizes the redundancy of word combinations sufficing to specify the word sequences that can occur as sentences in terms of a primary structure of three

[1] This point is succinctly made by Harris (forthcoming):"the grammar must predict the existing combinations on the basis of the fewest constraints possible", and again in (1968:12 fn 16):"The fact that particular kinds and amounts of redundancy are essential parts of language structure makes it important that a description of a language should not add its own redundancy to the picture. A theory of language should not contain elements of wide combinability and then specify which combinations are language. It should contain elements of just such combinability as appears in the language itself." Given the informational interpretation of language structure proposed here, and the connection of redundancy and information, eliminating redundancy from description (and thus not attributing it to what is described) is a paramount consideration; see also the first pages of § 3 below.

relations or constraints on word combinations.[1] First, a word dependence requirement which is a partial ordering of the vocabulary of the language, a partitioning of its words according to what other (kinds of) words must be 'present' for a word of a given class to 'enter' a sentence, i.e., to be predicated of them.[2] The two other relations are defined in respect of this predicational (and information-creating) relation: the relation of (gross inequalities of) likelihood of occurrence that an operator word (a predicating word) bears to the words of its argument (its predicand) class(es),[3] and, a paraphrastic relation of reduction in phonemic shape. These latter are mappings

[1] Harris (1982).

[2] For example, in Max drinks wine, drinks is predicated of (operating upon) the argument pair Max,wine (i.e., 'drinks operating on (Max, wine) becomes Max drinks wine'); also for drinks operating on (Max, chaos), (chaos, Max) but not (importantly, swims). In English, an operator word (usually) occurs after its first argument word. Similarly, in I disapprove of Max's drinking wine, disapproves operates upon the pair (I, drinks) with 's...-ing a morphophonemic change indicating this further predication, as also in Max's drinking wine is a fact with fact operating upon drinks. It follows that some words, e.g., Max, wine, must have null entry requirement.

[3] While one can think of likelihood inequalities as differences in estimates of occurrence of an operator on particular argument words as might be determined in a vast sample of sentences, nothing really turns on this virtually impossible task. On the other hand, the relative inequalities of likelihood an operator word may bear to its argument is more readily assessed, especially if, for example in an argument pair, one word is held constant: drinks (Max, wine), drinks (Max, cement), drinks (Max, carpet). All that is required is that the transformational mappings (reductions) do not alter the assessments of these relative inequalities of likelihood. 'Likelihood' may be thought a potentially misleading designation for the relation of co-occurrence inequalities obtaining between an operator and the different members of its argument class, and one may choose, as Hiż does in recent writings, to speak instead of a set of assumed true sentences in characterizing normal selection. Again, the condition on transformation as preserving the inequalites or differences is required.

(transformations) in the set of sentences from less reduced sentences (ultimately from unreduced "base" sentences where the partially ordering of words is inspectably satisfied) to more reduced sentences, mappings which preserve the word dependence, partial ordering and likelihood relations among the words, and whose condition is high likelihood (expectability) and hence low information. [1]

Beyond the minimal constraints sufficing to predict the word combinations which can occur as sentences of the language as a whole, the grammatically possible sentences of e.g., English, the actual occurrence of language, in discourse, is distinguished by additional constraints on word combinations. A discourse is not simply a bag of disparate sentences; analysis reveals constraints on word combinations extending beyond sentence boundaries. [2] And, for the language of discourses that arise around a particular, relatively narrow subject matter, such as ~~the language of~~ research reports in a subfield of a science, a corresponding 'grammar' seeks to unredundantly describe the constraints characterizing not just the corpus, the sentences of a particular discourse or set of discourses, but

[1] E.g., We expect Max with expect apparently operating upon (We, Max) can be taken as reduced from We expect Max to come (or to arrive, to be here, etc.) where expect has a three-place argument (We, Max, come, etc.) since we otherwise have to account for this occurrence of expect. The justification for the reduction of (here) an "appropriate" verb (and its arguments) to zero phonemic shape is on grounds that the reduced words have high likelihood (are inferable or expectable, that is to say, redundant in this environment) and hence contribute only low or no information to their sentence.

[2] As pointed to in Chapter 4 §2, there are even constraints upon which sentences can be conjoined.

as well what can be considered a possible sentence for these discourses. Such a (sublanguage [1]) grammar of constraints presents -- in terms of its specific word classes, subclasses and (sequences of) sequences of these -- a structure of residual redundancies, hence an informational structure, which is considerably more articulated than the structures of a grammar whose domain is the language as a whole. Hence, one is able to exclude from the sublanguage sentences which are possible in the wider language as a whole, and even sentences whose words are contained within the limited vocabulary of the science. For instance, a sentence like The antibody was inflamed is indeed a recognizable sentence of English, although speakers might well differ as to what it means, or whether it has a clear meaning. But it is not a possible sentence for the sublanguage of cellular immunology. To exclude such a sentence from the science sublanguage is not to make the trivial case that it may be exceedingly rare or unlikely to occur in the science. Rather, ~~More strongly, it is to maintain~~ it makes the stronger claim that if such a sentence did occur we should have to say that the science had changed ('was not the same science') or that ~~it~~ The sentence was a misprint.[2] Certainly, it is a delicate matter to rule upon what a science can and cannot say, but the additional constraints of sublanguage and discourse make a further restriction upon what is a possible sentence. And, as will be sketched below in § 3 and more fully displayed in chapter 6,

[1] A sublanguage is defined (Harris (1968:152)) as a proper subset of the sentences of the language closed under some or all of the operations defined in the language. In this regard, a grammar of a language, in the sense of p. 295 fn 1 above, is a sublanguage of the language. See § 3 and Chapter 6 §2.

[2] For reasons discussed in Chapter 6 §1.

the characterization of these constraints presents a structure which 'says the same' as those parts of the text from which it is derived and thus may be said to informationally represent.

In summary, there are two central issues which must be kept at the forefront of a naturalistic approach to the question of language structure and information. First, there is no (external) language in which to describe natural language or the information it 'carries' so that information, as language structure, can only be characterized in terms of redundancies of combinations of elements. Second, the metaphor that language 'carries' information is misleading in that it either wrongly suggests the model of message and cypher or code (where a code symbol 'carries' the information of its translation) or perhaps some other, vaguer notion which again implies information is around, pre-linguistically and pre-representationally, to be carried. The structure of redundancies in language is a construction of the linguist. It is 'in' the message, text, or discourse in as much as the elements of this characterizing structure may all be determined as having made a meaning difference through a user's recognitions that some sounds or word sequences 'say the same' as others. No other sense is available for interpreting the remark that the structure is 'in' or 'carried by' the message which does not

ultimately invoke the notion of a prior language (or conceptual system, the point remains the same). This is not to say that the elements are not preset -- as socially shared patterns, or, in Peircean terms, habits -- in both speaker and hearer.

The portage metaphor also masks an important and necessary conceptual clarification regarding a concept of information (as distinct from ~~that~~ the concept of "amount of information"): (a) information is properly understandable only as a property of representations, in terms of a selection from among representable alternatives, and subsequently, (b) no clear sense attaches to saying that information is somehow 'out there' independently of the representations of perceivers and language users. Moreover, the elements of an information representation cannot be specified in terms of some purely physical scale but only ~~to~~ in terms of the varying discriminative (or differential response) capacities of organisms.[1]

[1] Haber (1983) and (1974); see the discussion in § 2.

5.2 Information and Meaning. An initial task in attempting to think about information ~~lies in~~ is to situating this concept with respect to another to which it is, ostensibly, related: that of meaning. This is all the more necessary since we are concerned here with those aspects of meaning which are of relevance to linguistic communication. But this task is a formidable one since it cannot be simply assumed that in using either term we are dealing with concepts for which a common antecedent understanding can be readily supposed. There is notorious imprecision in speaking of meaning and, it will be argued in the sequel, no less is true in speaking of information. We have no alternative therefore to plunging in medias res.

Fortunately, we can begin our discussion by reference to others who have similarly sought to elucidate a relation between the concepts of meaning and information. In particular, our point of departure lies in confronting the view, more or less inherited en bloc from so-called "information theory" [1] that information (or the concept of information that is supposedly developed in this theory [2]) "has nothing to do with meaning", a view which, if nothing

[1] Less misleadingly referred to, as has been the practice in Britain, as communication theory (or the statistical/mathematical theory of communication) to distinguish it from e.g., theories of statistical inference where different concepts of information are employed; see ~~the reference cited in fn 1 p. 327 below~~ Schutzenberger (1956).

[2] Some writers appear to maintain that this is a spurious distinction, holding, in effect, that the only concept of information available with which to address the question of the relation of information to

else, implies a certain determinateness regarding the signification and mutual relation of the terms 'meaning' and 'information'. [1] Our problem may then be reproportioned as, first, that of examining whether in fact a concept of information has been forwarded by this theory and, second, that of considering how the relation to meaning is conceived from the perspective of information theory. Only then can we suggest a somewhat different view of information adequate to a naturalistic approach to language structure which, via the notions of constraint, selection from among alternatives. and redundancy, does have, ~~I has,~~ to be sure, ~~does have~~ some connection with the concepts of communication theory. This provides us then with a basis from which to critique several attempts by prominent linguists to employ code-analogies and information-theoretic terminologies and concepts in linguistic theory.

(continued from the preceeding page)

meaning is that provided in this theory. Dretske (1981:46) e.g., criticizes the following line of reasoning (which we generally endorse):

> Information is a semantic idea. Semantics is the study of meaning. Communication theory does not provide a satisfactory account of meaning. Hence communication theory does not provide a satisfactory account of information.

~~as indicating~~ For Dretske, this indicates the "fallacy" of "assuming that meaning is the only semantically relevant concept, (and) that if a theory does not provide a satisfactory account of meaning, it does not provide a satisfactory account of any semantic concept." Such an objection assumes, however, that a concept of meaning can be reasonably marked off from other semantic concepts, including, presumably, information. It goes beyond the scope of this essay to argue that Dretske (see especially 222-231) does not succeed ~~in carrying~~ out the required demarcation.

book? monograph?

[1] Weaver (1949:116), see below; Dretske (1981:44) writes that "The information embodied in a signal (linguistic or otherwise) is only incidently related to the meaning (if any) of that signal...."

'Our concern in this section is to motivate the view, articulated in more detail in §5.3, that language structure is a structure of the 'objective' meaning or information of sentences as this can be determined by methods which relate sentences to each other, preserving speakers' recognizable differences of meaning. It is not an attempt to provide a satisfactory explication of a concept of information which may have wider application. But it may, nonetheless, serve as a propaedeutic which may help to clarify the issues involved in such an explication, especially in the wake of a recent and widely influential effort to revive a concept of information from communication theory and to develop on its basis a "genuine theory of information" adequate for the provision of a secure epistemological foundation for semantic and cognitive studies.[1]

From the origin of the employ of the term 'information' in communication theory, reservations were present as to the relationship between the concepts of this theory and the ordinary language term with its panoply of semantic connotations. Hartley's (1928) original discussion does not define the term[2] but speaks of information becoming "more precise" with the successive selection of symbols from a specified repertoire,[3] with the intended implication of the "elimination" of

[1] Dretske (1981:4). For influences, see also the reviewers' comments in Dretske (1983) and Cummins (1983: vi-vii).

[2] Contrary to what is alleged by Cherry (1966:43-4).

[3] Hartley (see below), observing that a message of n symbols chosen from an alphabet of s symbols, has s^n possibilities, defined the "quantity of information" of such a message as the logarithm $H_n = n \log s$.

"subjective factor" of meaning (see further below). And a distinction between meaning and information was again emphasized by Shannon in his seminal papers on "The Mathematical Theory of Communication". As noted above in §1, Shannon's schematic conception of the problem of communication as that of "reproducing" at one location a message selected at another abstracts from any concern over what might be considered to be the meaning of such a message; this latter is not the province of the engineer. In this work, Shannon defined a quantity, termed "amount of information", which is a logarithmic measure of the statistical unexpectedness (reciprocal of probability) of a message.[1] As the unexpectedness of a message need have no discernible connection with whatever may be taken as its semantic content or meaning,[2] Shannon cautioned time and again that the concept of meaning lay outside the scope of this theory.[3] Perhaps even more

[1] Reinterpreting Hartley's definition of "quantity of information" based on the successive selection of symbols or words from a given list, Shannon defined the average information of a sequence of n symbols as $H_n = - \Sigma_i p_i \log p_i$, where p_i is the probability (i.e., estimated relative frequency) of the occurrence of symbol i (or, in the continuous case under integration, state of the wave form i). The minus sign stems from the condition that $\Sigma p_i = 1$ and $\log \frac{1}{p} = - \log p$. The formal resemblance to Boltzmann's formula for the entropy of a perfect gas (which Shannon had pointed out), led Brillouin (1951) to designate this this quantity "negentropy", which, since it does not refer to the state of a physical system, is deemed inappropriate by Tillman and Russell (1961).

[2] Contrary to what another patriarch of the amalgam of disciplines and approaches which collectively acquired the name "cybernetics", Norbert Wiener, maintained; Wiener writes (1950:8): "The amount of meaning can be measured. It turns out that the less probable a message is, the more meaning it carries, which is entirely reasonable from the standpoint of common sense:", a remark cited by Bar-Hillel (1955:288).

[3] See e.g., the remarks of Shannon in van Foerster (1952:219).

influential (being more widely read due to its non-technical character), were Weaver's remarks on the relation of the new concept of information to meaning, made in an essay accompanying the republication of Shannon's papers in book form and excerpted in Scientific American:

> The concept of information developed in this theory at first seems disappointing and bizarre -- disappointing because it has nothing to do with meaning, and bizarre because it deals not with a single message but rather with the statistical character of a whole ensemble of messages, bizarre also because in these statistical terms the two words information and uncertainty find themselves partners. (1949:116)

And, on account of the wide gulf apparently separating the communication engineer's concept and the semantically laden term of common usage, communication theorists such as MacKay denied that communication theory put forward a concept of information at all [1] whereas Cherry, in his oft-cited account of the historical and conceptual development of this theory, expresses the same point by regretting that the term 'information' had ever been adopted in this context since it so little accords with the presystematic notion.[2]

[1] (1954:56-7):"Communication engineers have not developed a concept of information at all. They have developed a theory dealing explicitly with only one particular feature or aspect of messages "carrying" information -- their unexpectedness or surprise value....Their measure of unexpectedness, the average logarithm of the improbability of the message, $-\Sigma\ p_i \log p_i$, is not therefore information but simply a particular measure of what they termed amount-of-information: (i.e.) the minuteness of the selection which the message makes from the set or "ensemble" of all possible messages."

[2] (1966:51):"In a sense it is a pity that the mathematical concepts stemming from Hartley have been called "information" at all. The formula for H_n is really a measure of one facet only of the concept of information; it is the statistical rarity or "surprise value" of a source of message signs."

Be this as it may, an enthusiastic reception greeted the publications of Wiener (1948) and Shannon (1949). This "heady draught of general popularity" [1] was certainly due to many reasons, not all perhaps readily apparent, but, in all likelihood, one of them was the supposed promise of so-called "information theory" to provide rigorously "objective" (i.e., in terms of physical quantities, purely quantitatively considered) measures of "information", thus resolving in one fell swoop the often-lamented imprecision or latent subjectivism prevalent in, especially, the psychological and social sciences. There followed a "bandwagon" [2] of attempted applications of information-theoretic concepts and terminology ranging from biology and psychology to economics, linguistics, and the theory of organizations.

However, the suggestiveness and promise of the new theory seems to have prevented many proponents of these applications from sufficiently clearly appreciating the legitimate scope and limitations of its concepts. It is not therefore surprising that a fundamental difficulty was encountered in many of these attempts to apply these concepts beyond their locus of origin in communications engineering, namely, the problem of associating 'meaning' or 'content' or some index of qualitatively graded discriminable response with the "measures of information" that were defined. With few exceptions in biology[3] and

[1] Shannon (1956:3).

[2] ibid. Shannon notes that "information theory has, in the last few years become something of a scientific bandwagon" and that "as a consequence, it has perhaps been ballooned to an importance beyond its actual accomplishments."

[3] One is Gatlin (1972) who provides an interesting application of information-theoretic measures of the capacity of DNA to 'transmit' information, in an attempt to operationally define life in terms of information.

psychology, this problem has stymied the applications of information-theoretic measures. Information-theoretic descriptions of perception, memory and concept formation (more usually lumped together currently under the heading of 'cognition') in the words of one reviewer, "briefly dazzled psychologists in the 1950's and early 1960's, and then simply faded away". The source of the difficulty was soon apparent:

> while it was generally easy to calculate the amount of information in a stimulus or in a response, such calculations did not correlate with any interesting or relevant behavior of real perceivers, rememberers, or thinkers. [1]

The measure of information content defined in information theory (specifying only statistically average quantities from statistically stationary sources) did not prove particularly appropriate or revealing. As an average quantity, H_n (which may be written avg (log p_i)) could hardly, on reflection, be thought ~~to be of~~ any semantic or cognitive significance, for ~~this latter~~ seems obviously to require that the information content of individual messages or situations be specifiable. [2] Moreover, it is the usual postulate of this theory that the information-generating source be ergodic or statistically stationary, that is, that estimations of the relative frequencies of occurrence of a given symbol (or quantitized state of a wave form) do not depend upon the time at which the estimate is made. But this is surely not a legitimate

[1] Haber (1983:71). See also his (1974) for a review of the problem.

[2] Cf. Dretske (1983:56): "insofar as communication theory deals with quantities that are statistical averages (...), it is not dealing with information as it is normally understood. For information as it is ordinarily understood, and as it must figure in semantic and cognitive studies is something associated with, and only with, individual events (signals, structures, conditions)."

assumption -- even as an idealization -- for the study of any source which possesses a differential response or learning ability, i.e., whose relevant behavior alters with passage of time. [1] To have empirical significance, any purely quantitative measure must be relativized

> to what the recipient of the signal already knows about the signal and about the circumstances of its reception. [2]

On the contrary, the measures provided by so-called information theory were, in Haber's words, "entirely independent of the recipient." What has remained, apparently, of value from the "bandwagon" of applications of information theory is more "its qualitative concepts than its quantitative measures." [3]

[1] Cherry (1966: 178-9).

[2] Haber (1983:71).

[3] Cf. the 1969 postscript to the reprinted version of MacKay (1950): "It soon became clear that the biggest problem in applying Shannon's selective information measure to human information processing was to establish meaningful probabilities to be attached to the different possible signals or brain-states concerned. After a flourish of 'applications of information theory' in psychology and biology which underrated the difficulty of this requirement, it has now come to be recognized that information theory has more to offer the biologist in terms of its qualitative concepts than of its quantitative measures, although these can sometimes be useful in setting upper or lower limits to information-processing performance." Whatever may be determined to be the "qualitative concepts" of information theory is not further identified. It is doubtful, in any case, that these are what linguists like Martinet (1964:172) refer to in saying that "the features of information theory which are of use to the linguist are in the main those which result from common sense"; see the discussion of "distinctive information" below.

Restricting our discussion hereinafter to the "qualitative concepts" of communication theory, it may be instructive to return to a consideration of Hartley's early paper, since, as Shannon himself remarks, his work should be seen as continuing and extending a line of inquiry into the problems of communication engineering whose conceptual basis was laid some twenty years earlier by Nyquist (1924) and Hartley (1928). [1] Whereas Nyquist's paper treats a number of problems of telegraphy,[2] we follow Bar-Hillel (1955) in viewing Hartley's as being of significant interest for an examination of the (or a) concept of information. Writing from the point of view of telephone engineering, Hartley sought to develop a theory treating the capacity of physical systems for transmitting, from a sender to a receiver, 'messages' comprising successively selected symbols. To this end, he proposed to evaluate this capacity in terms of a "quantitative measure of information" (alternately, "amount of information" and "information content"). A notion of information is introduced only indirectly, however, as resulting from the sender's successive selection of symbols from an alphabet or repertoire of possible symbols. Recognizing that 'information' is "a very elastic term",[3] Hartley's

[1] Shannon notes (1949:3):"a basis for (a general theory of communication) is contained in the important papers of Nyquist and Hartley on this subject."

[2] Nyquist proposed a logarithmic measure of the "speed" of "transmission of intelligence", showing that this quantity depends upon both the speed of the signal and upon the number of different signal elements employed. He also introduced in this context the term "redundancy".

[3] (1928:536).

choice of a starting place for speaking about information is the general situation of communication, schematically reviewed:

> In the first place, there must be a group of physical symbols, such as words, dots and dashes or the like. which by general agreement convey certain meanings to the parties communicating. In any given communication the sender mentally selects a particular symbol and by some bodily motion, as of his vocal mechanism, causes the attention of the receiver to be directed to that particular symbol. By successive selections a sequence of symbols is brought to the listener's attention. At each selection there are eliminated all of the other symbols which might have been chosen. As the selections proceed, more and more possible symbol sequences are eliminated, and we say that the information becomes more precise. For example, in the sentence Apples are red , the first word eliminates other kinds of fruit and all other objects in general.... In as much as the precision of the information depends upon what other symbol sequences might have been chosen it would seem reasonable to hope to find in the number of these sequences the desired quantitative measure of information. [1]

Hartley goes on to observe that the number of different possible sequences of n symbols chosen from an alphabet of s symbols is s^n and he therefore defines the "quantity of information" of such a sequence as the logarithm $H = n \log s$. [2]

Now this is a very curious passage for a number of reasons. Notice, first of all, that Hartley brings together under the head of "physical symbol" words as well as dots and dashes. But a word is, surely, a sequence of physical symbols in the sense in which

[1] (1928:536).

[2] Thus Hartley's measure is appropriate only if the successive symbols are chosen independently (e.g., the selection is not a stochastic or recurrent dependence process) and the symbols are all equipossible at each selection.

dots and dashes are physical symbols. Thus words may be said to be "complex symbols".[1] But in addition physical symbols such as dots and dashes do not, even under the license of metaphor, "convey meaning", nor do the individual letters of the alphabet they symbolize. It may be replied that a dash in Morse code "means" the letter E, but this is only to say that a dash is a physical symbol of a symbol, i.e., the fifth letter of the English alphabet. Secondly, Hartley speaks of the communication situation in general as characterized by a sender's "mentally selecting a particular symbol", yet his measure of information is defined only for sequences of equipossible symbols unconditionally selected. This is taken up directly below. Third, Hartley does not actually define 'information' but introduces it in the context that successive selection of symbols enables one to say that the "information becomes more precise". Now this is peculiar in the light of subsequent developments, since what is termed the "increased precision of information", resulting from a successive selection of symbols that narrows the set of alternative messages that a completed symbol sequence may finally realize, is therefore inversely proportional to the number of messages which still may be alternatives after each selection; whereas the defined quantity, termed "amount of information", varies directly with s, the number of equipossible

[1] Cf. Chao (1968) Chapter 12.

~~9 symbols~~ symbols available at any given selection, [1] a seemingly paradoxical result. For an obvious requirement of coherence which any adequate concept of information should meet would appear to hold that whatever may be termed informational precision cannot be reciprocally related to what is termed amount of (or: content of) information. [2] Finally, Hartley is guilty of a rather glaring species of use/mention confusion. [3] The first word of Apples are red does not eliminate "other kinds of fruit" but the

[1] Some twenty years latter, Warren Weaver, commenting on Shannon's similarly defined concept "amount of information", remarks (1949:108-9):"Information is,..., a measure of one's freedom of choice in selecting a message. The greater this freedom of choice, and hence the greater the information, the greater is the uncertainty that the message actually selected is some particular one. Thus greater freedom of choice, greater uncertainty, greater information go hand in hand." On Shannon's concept, and whether it implicates a legitimate concept of information, as Weaver's statement implies, see further below.

[2] Others have remarked on the "paradox" that meaning is inversely related to "information content", a variant of the view that "information has nothing to do with meaning" (see below), e.g., Granger (1968:127 fn 16) who notes that

> Il est sans doute paradoxal de voir les sens opposé au contenu d'information.

However, the supposed "paradox" dissipates upon examination, i.e., once it is realized that concepts such as "amount of information" or "information concept" do not really embody a legitimate concept of information at all, a view urged by Shannon himself (see below).

[3] As Bar-Hillel (1955:285-6) pointed out.

word "apple" eliminates (among other alternatives) that words naming other kinds of fruit, e.g., "peaches", "lemons" are the subject of the predication "are red".

In a widely-cited review of "information theory" Bar-Hillel (1955) suggested that the persistent misunderstanding and confusion surrounding the term 'information' has been generated by Hartley's initial use of the term, with all of its ordinary language connotations, to refer in the context of the expresssion "amount of information" or "information content" to a measure of the rarity of occurrence of a certain symbol sequence selected from among other possible sequences of the same number of symbols. As an illustration of this confusion, Bar-Hillel cites the apparently inconsistent remarks of Weaver who, on the one hand, asserts that information "has nothing to do with meaning" and on the other proclaims that Shannon's "analysis has so penetratingly cleared the air that one is now, perhaps for the first time, ready for a real theory of meaning". [1] Indeed, Bar-Hillel sees in Hartley's glaring use/mention confusion an indication of a psychological inability to keep separate in the mind the senses of 'information' pertaining to the measure of rarity of occurrence of a certain symbol sequence and the ordinary sense of the term:

> it is psychologically almost impossible not to make the shift from the one sense of information,...,i.e., information = signal sequence, to the other sense, information = what is expressed by the signal sequence,... [2]

[1] Weaver (1949:116).

[2] Bar-Hillel (1955:284).

Now Bar-Hillel had quite definite ideas about the two senses of 'information', i.e., the term 'information' as used in communication theory and a notion of information pertaining to meaning or semantic content ("what is expressed by the signal sequence") which is more in accord with the ordinary language sense; these ideas were, of course, put forward in the well-known Carnap - Bar-Hillel "Theory of Semantic Information". This theory deals exclusively with the concept of the semantic information conveyed by a statement, its semantic content, and various measures for this concept. As such, it is

> pragmatics free, abstracts from the users of the language and deals only with the relationships between linguistic entities and what they stand for, or designate, or denote.[1]

More particularly, then, Bar-Hillel had quite definite opinions about the relation between "the (statistical) Theory of Signal Transmission and the Theory of Semantical Content"; namely,

> I would now say that both of these theories can be regarded as different interpretations of a common formal system, the Calculus of Information.[2]

But there is no general Calculus of Information. There is no intrinsic relation between (A) and the logarithmic function

$$- \sum_{i=1}^{n} p_i \log p_i$$

which is only a particular [3] measure of various properties of the

[1] Bar-Hillel (1952:299).

[2] Bar-Hillel (1955:291).

[3] Satisfying the conditions: (1) if all but one of the probabilities are zero, the function has the value zero; (2)the function has a maximum value when all the probabilities are equal; (3) it increases monotonically with n. See Shannon's discussion (1949:18-19) of this choice of conditions on H; the derivation of H is given in Appendix 2 (1949:82-3). See also the discussion in Tillman and Russell (1965:128).

distribution of a set of probabilities $p = p_1, p_2, \ldots, p_n$ where $\sum_{i=1}^{n} p_i = 1$ and the concepts (including the term 'information') employed in communication theory. In communication theory, the significance of this particular measure is that it leads to the celebrated Channel Capacity Theorem; otherwise, it is "just another measure". [1] It would therefore seem that Bar-Hillel's criticism of, e.g., "Weaver's uneasiness" regarding the term 'information' might with equal justification be applied to himself. For Bar-Hillel is smitten with the same disease as Weaver, i.e., failing to see that à propos communication theory, one cannot speak legitimately of information in a way that implies an underlying concept of information since this does not exist; all that is warranted by Shannon's theory is speaking about a particular probability measure. [2]

There are also reasons to be critical of Bar-Hillel's too-ready willingness to assume that the meaning or semantic content of a statement has to do with the designata or denotata of linguistic entities. Hartley himself does not make such an assumption. Instead in a section of his paper entitled "Elimination of Psychological Factors", he alludes to the "psychological considerations" that constrain actual situations of communication, and according to which not all symbol sequences which

[1] Licklider (1956:24):"Shannon's measure H would be 'just another measure' if it did not lead to the Channel Capacity Theorem. The fact that H leads to that remarkable insight gives H a definite status. In problems concerning coding of information for efficient transmission through restricted channels H is the natural measure."

[2] Shannon, cited in von Foerster (1952:219):"This kind of information is an ensemble concept. It is not a statement about a proposition, if you like or a fact, but a statement about a probability measure of a large ensemble of statements or propositions or facts, It is a measure of a kind of dispersion of that probability distribution.
I think perhaps the word "information" is causing more trouble in this connection than it is worth, except that it is difficult to find another word that is anywhere near right. It should be kept solidly in mind that it is only a measure of the difficulty in transmitting the sequences that are produced by some information source."

can possibily be selected will have assigned meanings. Hartley's illustrative example involves the transmission of Morse code but the point may be taken more generally:

> The operation of recognizing from the received record the sequence of symbols selected at the sending end may be carried out by those of us who are not familiar with the Morse code. We would do this equally well for a sequence representing a consciously chosen message and for one sent out by (an) automatic selecting device ....A trained operator, however, would say that the sequence sent out by the automatic device was not intelligible. The reason for this is that only a limited number of symbols available to the sending operator at certain of his selections is here limited by psychological rather than physical considerations. [1]

text missing?

Hartley proceeds to observe that the psychological constraint that not all sequences of symbols can occur meaningfully is irrelevant to his overall concern to measure a physical system's capacity to transmit sequences of symbols, a capacity which depends only on being able to distinguish at the receiving end of the system the result of selections made at the sending end.

> Hence in estimating the capacity of the physical system to transmit information we should ignore the question of interpretation, make each selection perfectly arbitrary, and base our result on the possibility of the receiver's distinguishing the result of selecting any one symbol from that of selecting any other. [2]

In making each choice or selection "perfectly arbitrary" (as indeed is evident from the function $H = n \log s$) Hartley is thereby able to define a "definite quantitative measure of information based on physical considerations alone".[3]

[1] (1928:537-8).

[2] ibid., 538. Bar-Hillel (1955:284) is therefore not quite correct in stating that "Hartley goes on to assume silently that all possible signal sequences are <u>equipossible</u>,...".

[3] ibid.

It is surely Pickwickian to ~~speak, via~~ use the locutions "amount of information" or "measure of information content", to speak of information at all in referring to a physical system's capacity to transmit arbitrarily selected symbols (or better, distinguishable signals, since these need not be 'symbolic', i.e., mean or stand for anything). And, since as Hartley's remarks indicate, sequences of arbitrarily selected symbols are not to be recognized as meaningful in virtue of the arbitrary character of their 'selection', his expressions "measure of information content" and "amount of information" do not bear upon or implicate a concept of information which has, on the received view, "nothing to do with meaning", in as much as they do not bear on any concept of information.[1] But, in pointing to "psychological factors" constraining combinations of symbols in meaningful communication, Hartley had an insight into the character and essential role of redundancy in communication. This linkage of redundancy to meaning, i.e., that constraints upon combinations of their elements are a property of 'messages' (sequences of symbols) having meaning, raises the possibility that meaningful sequences of elements can be characterized in terms of these constraints. How this may be done, Hartley does not, of course, suggest but he does speak of "the information" becoming "more precise" as successive selections of symbols

[1] So Hartley's schematic depiction of the general situation of communication -- to which these expressions have presumed applicability -- as involving a sender's "mentally selecting" a symbol (which is hardly an arbitrary event) is not really appropriate to his concern with the capacity of physical systems to transmit messages.

narrows the number of alternatives that the completed sequence may realize. This is the sense in which it is often said that information means or can be defined as "reduction of uncertainty", a sense congenial with that accorded the term in theories of statistical inference.[1]

We have seen that "selective information" as understood in the sense made prominent from communication theory pertains not to a concept of information but to a particular probability measure which determines a quantity termed "amount of information". The notion of selection, as employed there, suggests that an ensemble of possible elements -- symbols, messages, sentences -- has been delimited from which selection is to be made. And it must be recognized that the notion of selection is here used ambiguously, as referring to the product or resultant of an information-generating "source" and to the occurrence of a particular message which can be represented as a choice or selection from a "preconceived ensemble" of messages. 'Selection', as with other terms characterizing a communication process between senders and receivers, belongs to the descriptive metalanguage of an "external observer".[2] So construed,

[1] An explicit comparison of the different concepts (actually, measures) of information in statistics and in communication theory is given in Schützenberger (1956).

[2] Meyer-Eppler (1959:5):"Die in einer Kommunikationskette sich abspielenden Prozesse können nur von einem ausserhalb der Kette stehenden externen Beobachter hinreichend exakt beschrieben werden, einem Beobachter, dem sämtliche Glieder der Kette zugänglich sind. Zur Beschreibung des Beobachteten und zur Formulierung von Gesetzlichkeiten bedient er sich einer wissenschaftlichen Metasprache, die nicht mit der zwischen dem Expedienten und Perzipienten vereinbarten Objektsprache übereinstimmt."

as presupposing a set of possible messages, it tacitly invokes what the linguist aims to characterize, i.e., the restrictions on combinations of elements which suffice to specify the notion of 'possible sentence' of the language. Can the notion of selection play therefore a role in explicating a concept of information adequate for application to natural language?

A way forward, implicitly suggested in these remarks of Hartley, lies in thinking of information (or: socially transmissible meaning) in language as reconstructable from or generated by a (hierarchy) of constraints upon combinations of elements of the language. Consider (as ~~e.g.,~~ was once the fashion) the task of the grammarian as that of producing a 'selective device' which 'accepts' or 'recognizes' all and only the word sequences which can occur as sentences of the language. Thus the job of this device is to select from the set of all possible sequences of word combinations the proper subset of this immense set that are sentences of the language. And this is to ask: What constraints upon word combinations suffice to specify this set?

To this, it was suggested (at the end of § 1 above) that the three constraints of the theory of language structure of Harris (1982) create a 'space' (a closed set) in terms of which grammars of particular languages can be constructed: a partially-ordered word dependence relation which partitions the vocabulary of the language and determines a "base" set of elementary, transformationless sentences , together with mappings

which from this "base" set produce all the remaining sentences of the language preserving the partial ordering of words. Each sentence of the language can then be represented by the constraints governing its word combinations -- the partially ordered entry requirement of its words and the domains of the mappings (reductions) they enter into. As we attempt to show in §6.3, this theory and the grammar of English exemplifying it realize the programmatic goal proposed in Harris (1951a): the provision of an axiomatic characterization of a language which is a compact 1-1 representation of its sentences, i.e., a representation of its sentences as comprised of discrete combinatorial elements, purely positionally defined, each representing a recognizable difference in meaning.[1] In the next section it is shown in some detail how this program has employed a particular methodology termed "regularization" which consists in redefining the elements or generalizing the operations of the grammar, thus eliminating redundancy from the description of the language; in effect not giving descriptive standing to what can be shown to be a case of something else. "Regularization" necessitates a specification of the domain over which the regularizing operations are defined. This can be seen by contrasting the task of a grammar of the language as a whole (in which avoidance of multiple classification and maximal derivability are stressed) from that of a 'grammar' of restricted parts of the language, i.e., of a sublanguage or discourse, where, due to additional constraints upon word combinations, the goal of a least redundant description is served by maximizing similarities with other word sequence occurrences.[2]

[1] (1951a:366-7 and 372-3); see the discussion in §2.6 above and below in §3.

[2] 'Grammars' of sublanguages require the grammar of the language as a whole; see the discussion of this point in Chapter 6 §2.

We remarked in § 1 above that the further restrictions on word combinations in a sublanguage of a field of a science enables one to say that certain well-formed word sequences in the language as a whole are not possible sentences of the science, even if comprising only ~~comprised completely by the~~ composed only of vocabulary of the science. Such a narrowing in the set of alternatives, of what can be a possible sentence, corresponds to the equivalent conception in Hartley's paper, that of making "the information more precise". If elements represent (or correlate with) determinable differences or specificities of meaning, the more articulated the characterizing structure of a sentence, the more specific is the character of this correlation. In a sense this is to make operational ~~ize~~ a notion of information as determinable language structure:[1] what can be represented as information, i.e., as 'housed' within the combinatorially delimited set of alternatives, is only what can be established as recognizably distinctive or contrastive. How

[1] We intend by ~~speaking here of~~ using the term "operational" to indicate only that distinctness of signs be based upon some observable or inferable differences of use, or of such behavior as is occasioned by their use in particular inquiries. This would seem to be in accord with the sense of "operational analysis" in Nagel (1942:188-189):

> (E)ven when they occur as parts of statments which characterize a given subject-matter, signs often do not signify "objects" in the explicit subject matter, but modes of conducting inquiries into that subject matter.
>
> That this is so will surprise only those who take language out of the context of overt operations of reflective thought. It is not unreasonable to maintain that every language, however much one may try to purify it of such elements, will inevitably contain expressions whose adequate understanding requires a consideration of the activities of those who use that language as much as it involves a reference to the ostensible subject-matter of that language. ...Because of this characteristic of language, operational analysis of what language signifies seems to me the most fruitful way of performing critiques of abstractions. If the distinctive philosophic task is that of criticizing abstractions, a philosopher is worthily employed when he studies the operations or behavior, both overt and symbolic, which are involved in using language." Cf. the slightly different rendering of this passage in Nagel (1944:240-1).

informationally articulated a given sentence is is not, therefore an inherent semantic property of its words, but is wholly a function of the determinable similarities and differences it bears to other sentences over a specified domain. And there is always the possibility that further regularization (~~via~~ by elimination ~~of~~ ing dissimilarities which are only apparent) of the described sentences can be achieved, leading to a yet more informationally articulated structural characterization.

The operational character of this specification of language structure as information may be viewed as a needlessly restrictive limitation. For example, it may be urged that there are clearly meaningful (under some criterion) elements which are not adequately structured in the informational representation of language.[1] But conversely, it may be countered that this limitation is rather a strength, given the connection between redundancy and information, since it guards against the danger ~~that~~ of the description contributing to the redundancy which it purports to describe. Where it is felt that the methods of regularization do not suffice, they nonetheless afford a basis upon which a considered comparison is possible and the costs of extending the description outside the otherwise sufficient structures they provide can be assessed.

[1] Thus Harrisian grammars do not "overgenerate". Any unused or infrequently used descriptive apparatus degrades the correlation of structure and meaning. This point may be stressed since it may not be readily understandable why there is no employment in these grammars of logical or set-theoretic apparatus to describe, e.g., quantifiers, or quantificational terms, relations of reference, or tense, etc. See further on this point § 3 below.

The constraints on word combinations (over the domain of the language as a whole) specifying the notion of a grammatically possible sentence of the language, or the additional constraints which permit the identification of particular word classes and sentence-types in sublanguages, are only the most ~~prominent~~ obvious indications that meaning or information can be analytically reconstructed in terms of a hierarchy of constraints on combinations of elements, i.e., "a system of contributory redundancies". [1] Not all combinations of phonemes are morphemes (in English, /sb/,/gb/, /ls/ after pause, etc), not all combinations of morphemes are words ((e.g., we have persist, desist, consist, resist, but not unsist), not all word combinations are sentences, and not all sentences can combine in discourse and sublanguage. Each successive constraint on combinations is defined in terms of elements of the level immediately 'below' it; correspondingly, each 'higher' level enables a more efficient or less redundant characterization of the language described.

Clearly any view that meaning or information can be reconstructed in terms of restrictions on combinations of elements relies -- in order to avoid a regress -- on a system of elements that are not in themselves meaningful, hence not themselves comprised of restricted combinations of more 'ultimate' elements. There is accordingly a system of elements (phonemes) which are not defined in terms of restrictions on lower level elements but which are identified by

[1] Harris (1968:12).

perceptual contrast. Contrastive identification (as "perceptual Gestalts") at the same time determines a system of contrasting elements, a point expressed by Saussure's famous dictum that phonemes are, above all, opposing, relative and negative entities.[1] Phonemic contrast is therefore the foundation for the whole of grammatical analysis and for the characterization of language structure as information.[2] Beginning with phonemic contrast, the

[1] Saussure (1916:164):"Les phonèmes sont avant tous des entités oppositives, relatives, et négatives."

[2] It is true that phonemic identification is operationally specified in terms of speakers' distinguishing 'higher' level elements, morphemes or words (see §3.2 above). This should not be taken, however, as indicating that phonemes are meaningful elements, but only that phonemic contrast determines a semantic datum (Hiż (1979:344)). The fact that phonemes are only identifiable by contrast of higher level elements (which are defined as composed of phonemes) appears analytically circular, but the circularity is not vicious since in practice the linguist has to assume certain clear cases of morpheme and word distinctions. These initial assumptions are of course susceptible to revision as the analysis proceeds. In practice, distributional determination of elements proceeds from "some arbitrary point of departure" but, analytically considered, engages in the fiction that distributional determination of elements is carried out on all elements simultaneously. Cf. Harris (1951a:7) who, considering "the problem of setting up relevant elements...on a distributional basis" notes:

> x and y are included in the same element A if the distribution of x relative to the other elements B, C, etc., is in some sense the same as the distribution of y. Since this assumes that the other elements B, C etc. are recognized at the time when the distribution of A is being determined, this operation can be carried out without some arbitrary point of departure only if it is carried out for all the elements simultaneously.

whole hierarchical construction of constraints (restrictions on combinations of elements) required for socially transmissible meaning can be generated.[1] by distributional methods

[1] It is instructive to consider how the doctrine that information "has nothing to do with meaning" has encouraged some prominent structural linguists (e.g., Martinet (1964:172 ff) and Malmberg (1963:31)) to attempt to apply notions from communication theory in linguistics by speaking of phonemes as being "units of distinctive information". Since phonemes do not designate nor can be considered to have a semantic 'content', this is in agreement with the rule that information "has nothing to do with meaning". Martinet, for instance, after observing that "the features of information theory which are of use to the linguist are in the main those that result from common sense" gives the following illustration whose supposed point is that phonemes, which do not have meaning, by reducing uncertainty (eliminating certain outcomes or alternatives), can nonetheless be said to possess information, abiding thereby by the rule regarding the distinction between meaning and information.

> Everything is deemed to possess information which has the effect of reducing uncertainty and of eliminating certain possibilities. If I hear /hi:hæz p.../, /p/has no meaning by itself, but it possesses information in the sense that it excludes all kinds of possible utterances such as he has given, he has seen. If /r/ is added to the utterance (/hi:hæz pr.../), uncertainty is further reduced since it excludes he has paid, he has pushed etc. and this shows that /r/ also possesses information. Information is therefore not an attribute of meaning since non-significant units such as /p/ and /r/ participate in it(173).

Note first of all that Martinet's argument does not warrant the claim that "information is not an attribute of meaning". In particular, it does not provide the required demonstration that significant units (i.e., elements having meaning) may be seen to not possess information. And, as Martinet's argument patently shows, phonemic distinctivenss is established by reference to occurrences in significant (and well-formed) utterances. Thus the distinctiveness of phonemes (correspondingly, the determination that they "possess information") is not based on contrasting sounds considered in themselves, but only of sound contrast as indicated by speakers' recognitions that two utterances are not repetitions of the same word (assuming homonymities can be distributionally distinguished). Since words are significant or meaningful elements, the argument supporting the claim that "information is not an attribute of meaning" is, in addition to the defect noted above, viciously circular.

Such a system of 'ultimate' elements must then be comparatively unrestricted, i.e., comprised of relatively few elements of relatively wide combinability. This indeed is the case with the few dozen phonemes of each language. It has been suggestively remarked [1] that if 500 - 600 distinct sounds were contrastively (i.e., phonemically) identifiable by the speakers of a language, a system would exist which could be sufficiently rich to have an internal structure of restricted combinations (and thus allow for meaning). [2] It seems reasonable

[1] By Z. Harris in conversation.

[2] One might then say that a systems of phonemes of this kind could not be considered 'ultimate' but only as comprised of some lower level of relatively unrestricted elements which were contrastively distinguished. Jakobson, who has often written of the "quest for the ultimate constituents" of language (e.g., (1979:80-121) and (1965)), has been a leading proponent of an effort to reduce phonemic contrast which seeks to classify phonemes as a "linguistic code" of physiologically or acoustically defined "distinctive features". Each of these is represented as a binary opposition along some sound perception dimension, e.g., for vowels and consonants -- grave - acute, sharp - non-sharp, compact - diffuse; for consonants -- nasal - non-nasal, strident - mellow; for vowels -- voiced - voiceless, lax - tense (see 1979 passim). However, despite a great deal of enthusiasm (even as to the supposed perceptual necessity of binary oppositions (e.g., (1979:25)), there has been little success in providing details of the proposed reduction. See the papers of Delattre (1967) and (1968) and, for a survey, Lipski (1974). Evidence, by no means compelling, is sometimes cited that humans possess specific "feature detector mechanisms", evolutionary adaptations that respond to some particular attribute or component of a speech signal (e.g., Lieberman (1984:chapter 8). On the other hand, a psycholinguistic reviewer of this literature recently concluded that "there is no evidence at all for specialized detector mechanisms tuned to the acoustic correlates of abstract linguistic features" (Studdert-Kennedy (1979:68-69). On the notion of language as a 'code' see Mounin (1970).

therefore to suppose that the small number of phonemes in each language reflects a limited human capacity for making the necessarily rapid perceptual discriminations required by the flow of speech, which thus sets an upper limit on the number of "recognizably distinct" sounds in each language.[1] Beyond this, the molecular biologist and historian of biological thought, F. Jacob, has suggested that the constraint of temporal linearity is imposed on language by the physical structure of the vocal and auditory apparatus extant in the mammalian evolutionary line leading to man and spoken language. Temporal linearity, he further

[1] It has been found that phonemic perception by the listener blends discrete (categorial) perception of some phonemes (e.g., of voiced stops /b/,/d/, /g/) with continuous (non-categorial) perception (which permits a higher degree of perception of intra-phonemic difference) of vowels. The combination of categorial and non-categorial phonemic perception "provides an effective basis on which incoming speech sounds can be sorted rapidly and unequivocally into the appropriate 'phoneme bins'". In this way, a phonemic system meets

> the psychological necessity that most (phonemic) distinctions be made quickly: if the rate of flow of phonemic information is not above some rather high minimum, the organization of the phoneme units into morphemes, words, and sentences becomes psychologically impossible. These considerations mean that some reasonable fraction of the phonemes must be highly distinctive. (Lisker, Cooper and Liberman (1962:104-5))

speculates, is required in any combinatorially productive system of spoken communication.[1]

'As against the notion of 'selective information' arising from communication theory, where a prior set of symbols and their associated values (a priori estimates of frequencies of occurrence) are assumed, and where the notion "amount of information" really doesn't implicate a concept of information but designates a quantity that may be assigned to a message and which varies inversely with the probability of occurrence of that message, we have attempted to delineate a concept of information which admits of precise determination as a structure of constraints on combinations of linguistic elements and where reduction in the set of alternatives determined by these constraints leads to (as Hartley originally put it) informational precision and a specifiable structure for meaning.

The notion of 'selection' is, of course, metaphoric (as is clear from its first use in ~~contemporary~~ discussions of information by Hartley). The redundancies of combination are the linguist's a posteriori construction using methods which specify only elements

[1] Jacob (1977:202):"It is evident...that the nature of this equipment had to impose constraints on the very structure of language, and the principal one of these constraints is the temporal linearity required for the emission as well as for the receipt of messages. This is true since a temporal sequence represents the only way to combine infinitely the short sounds that are produced and received by the majority of mammals and to transform thereby the signals into language. It is, therefore, the physical structure of the vocal and auditory apparatus that has probably imposed linearity on language."

having determinable correlates in the recognitive behaviors of language users. Still, it can be maintained, with Weaver, [1] that the notion of selection as it bears upon the concept of information pertains "to the situation as a whole". This means: with respect to the domain of constraints specifying the notion of 'possible sentence'. The constraints are 'in' the sentence when this sentence can be represented in terms of its similarities or differences with other sentences over a specified domain. Information, as we develop the concept, as a structure of constraints on combinations of elements, is therefore contextual. This inherent contextuality of information stands in direct opposition to traditional views about meaning, which at base consider meaning in an essentialist manner, as context-independent and sempiternal.

[1] (1949:100)."This word information (sic) in communication theory relates not so much to what you do say, as to what you could say. That is, information is a measure of one's freedom of choice when one selects a message. ...The concept of information applies not to the individual messages (as the concept of meaning would), but rather to the situation as a whole...."

5.3 Regularization. The objective of attaining a least redundant system of classification of some domain of phenomena has been likened to the traditional quest for a natural classification, a classification which, in Aristotelian terms, "divides nature at the joints". [1] Of course, the modifier "least" requires an important qualification: it is not to be understood as implying any reference to some mythical absolute economy or simplicity of theory or description, but signals only the comparative difference between alternative classifications of the same phenomena. It may then be said that the reduction of redundancy of one classification with respect to another over a given domain can be regarded as ~~comprising~~ an acquisition of information about this domain. [2]

It was suggested (in § 1 above) that the program of eliminating redundancy from the grammatical description of a language presses the point as to the importance of minimally redundant description even further than perhaps is recognized in the familiar strictures on simplicity and economy of theories in science. This additional insistence is inspired by the

[1] Hawkins (1968:44). Another, perhaps less metaphysically-laden view of the same point is provided, classically, by Mill (1879: 549): "What are the fewest assumptions, which being granted, the order of nature as it exists would be the result? What are the fewest general propositions from which all the uniformities existing in nature could be deduced?"

[2] Hawkins (1968:47).

naturalistic tenet that the information 'carried' by language cannot be characterized except as distinguished by [means of] the hierarchy of constraints upon combinations of the linguistic elements themselves, where each identified redundancy requiredly correlates with some aspect of determinable recognitive behavior of users of the language.[1] In so many words, language structure, as a structure of recognizable differences, is a structure of information. Accordingly, it is necessary to keep

> the grammatical description 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. More generally one must recognize that every new term or category or subclass that is not derivable from the primitives of the system, including every limitation on the carrying-out of a rule, and every ad hoc explanation is a redundancy of description. [2]

The methodology of eliminating redundancy from linguistic description, or "regularization", was already the prominent theme in Harris' early major work (1951a) in structural (and pre-transformational) linguistics. In §6 of Chapter 2 above, it was shown how the impetus throughout this work is to repeatedly obtain new elements from elements already constructed, defining on their basis 'higher' level elements with fewer restrictions on combination: from phonemes to phonemic long components to morphemes, morpheme classes,

[1] The operationalist cast of this methodology was remarked upon in § 2 above. A strong case can be made that such operational strictures need not and perhaps should not be imposed on physical theory.

[2] Harris (1982:10-11).

morpheme class sequences and morphemic long components. The goal of this analysis is to establish a 1-1 correspondence between determinable contrast (or increments of) meaning and defined elements, thus eliminating by successive redefinition of elements, restrictions on combinability that are only ~~apparent as~~ artifacts of the description. The culmination of this procedure was envisaged as an axiomatic grammar, a characterization of the sentences of a language in terms of basic elements (primitives) and a derivational apparatus:

> The work of analysis leads right up to the statements which enable anyone to synthesize or predict utterances in the language. The elements form a deductive system with axiomatically defined initial elements and with theorems concerning the relations between them. The final theorems would indicate the structure of the utterances of the language in terms of the preceding parts of the system.[1]

While there is no overt mention of transformational relations between sentences in this early book (ms. completed in January, 1947), grammatical transformations may be regarded as another step in the regularizing procedure of constructing 'higher' and 'higher' level elements to minimize redundancy of description. It has often been remarked that the concept of grammatical transformation was developed by Harris in extending the methods of descriptive (structural) linguistics beyond the boundaries of a sentence, to discourse. Here[2] transformations are introduced as operations that facilitate the distributional division

[1] (1951a:372-373).

[2] Harris (1952a).

of the sentences of a text into intervals, each of which is a sequence of equivalences, i.e., a statement of which elements (or element sequences) in a given sentence of the discourse have identical or equivalent environments of occurrence to those of other elements (or sequences) in other sentences of that discourse.[1] For example, in a discourse in English one can obtain from a sentence having an $N_1 V N_2$ structure (Max read a book), another transformed sentence whose nouns are in reverse order $N_2 V^* N_1$ (A book was read by Max), where the * indicates a morphophonemic change[2] (in this early formulation noted as a "change of suffixes") around the verb. The regularization or "normalization" of texts in discourse analysis is thus aided by transformations (here, the passive) which make possible further applications of the methods of discourse analysis. Transformational analysis

> merely transforms certain sentences of the text into grammatically equivalent sentences (as $N_1 V N_2$ above was transformed into $N_2 V^* N_1$), in such a way that the application of the discourse-analysis method becomes more convenient, or that it becomes possible in certain sections of the text where it was not possible to apply it before.[3]

In view of some long-standing misconceptions about the origins of

[1] (1952a:347 fn 6):"(The) equivalence (of particular sentence orders to each other) can be rediscovered linguistically by finding that the distribution of each sequence is equivalent to that of the others."

[2] I.e., a 'conditioned' or automatic change in phonemic composition.

[3] (1952a:316).

transformations, [1] it may be appropriate to reiterate that transformational relations between sentences were established as supplementary aids to the purely distributional methods of discourse analysis; in enabling a statement of equivalence relations among sentences, transformations made possible a less redundant characterization of regularities of word co-occurrences.

The first relatively detailed presentation of an axiomatic grammar [2] of the sentences of a language incorporating this view of transformations was Harris (1957). Here transformations apply to a restricted set of basic, non-transformed sentences, thereupon deriving the remaining sentences of the language. These basic sentences are called kernel sentences, where 'kernel' has the standard algebraic meaning of a set of elements which, as the residue under a specified mapping, are carried into themselves. Thus, to say that the set of sentences is closed under transformation, i.e., that the set of transformations partitions the set of sentences (is a quotient set of the set of sentences), requires that the mapping carrying the set of sentences into the set of transformations carry these basic sentences into the identity transformation. The basic sentences are therefore the kernel of the set of sentences with respect to this mapping. [3]

[1] E.g. Lyons (1968:155) writes: "The term 'transformation' was also used by Harris in roughly the same sense as it was used by Chomsky."

[2] Cf. Harris (1968:153): "The axiomatic view of grammars is that a grammar constructed for a language (a set of sentences) consists of a set of word and morpheme classes (and subclasses), a set of well-formed sequences of these (elementary sentence structures), and a set of transformational rules which derive one sentence structure from another." See also the formulation in Harris (1966:608).

[3] Harris (1957:456 fn 61) and (1956:387-388).

In this paper the status of transformations as additional to and outside of the purely distributional methods of descriptive linguistics was pointed to; at the same time it was again noted that transformational regularization furthered the applicability of the constructional (distributional) methods in extending them to sentences which could not (without excessive cost) otherwise be so analyzed.[1]

Transformations are viewed as equivalence relations in virtue of their preservation of word co-occurrence relations; more precisely, a transformation is established by determining that co-occurrence ranges are identical (or: "about the same) for a given n-tuple of words satisfying two word class constructions. This distributional condition on transformation has a semantic {interpretation / value} in that "some major element of meaning seems to be held constant under transformation", an element which is also referred to as "information content".[2] The "special meaning status" of transformations

[1] (1957:447-8):"Transformations cannot be viewed as a continuation of (this) constructional process. They are based on a new relation, which satisfies the conditions for being an equivalence relation, and which does not occur in descriptive linguistics. All sentences which are described in constructional terms must have a specific constituent analysis, since the constructional analysis proceeds in terms of immediate constituents (component subconstructions). This is not necessary, however, for all sentences in transformational analysis. Some of the cruces in descriptive linguistics have been due to the search for a constituent analysis in sentence types where this does not exist because the sentences are transformationally derived from others. For this and other reasons a language cannot be fully described in purely constructional terms, without the transform relation."

[2] ibid.,396-7:"transformations seem to hold invariant what might be interpreted as information content."

was problematic, of course, pending some convincing evaluation of meaning. But however meaning is to be evaluated, given the presumed correlation between relations of co-occurrence and meaning, transformations, which preserve word co-occurrence relations, are likely to preserve meaning as well:

> Meaning is a matter of evaluation and cannot be fitted directly into the type of science that is developed in structural linguistics or in transformational theory. Still, for various purposes it may be possible to set up some practical evaluation of meaning; and with respect to most of these evaluations, transformations will have a special status. That many sentences which are tranforms of each other have more or less the same meaning, except for different external grammatical status (different grammatical relations to succeeding sentence elements), is an immediate impression. This is not surprising since meaning correlates closely with range of co-occurrence, and transformations maintain the same co-occurrence range.[1]

In a later formulation of the kernel and transformation model of an axiomatic grammar, the co-occurrence condition was refined to preservation of acceptability ordering.[2] As opposed to the co-occurrence criterion for transformation, which requires (by implication) stating that a certain n-tuple never satisfies (i.e., does not occur at all in) a given sentence form (word class construction, e.g., N V N), no implied or tacit appeal to a fixed scale of acceptabilities is involved in stating that transformations maintain relative acceptability orderings. While it makes little or no sense to consider the acceptability of a given sentence within one or another ordering of acceptabilities, one can say that the

[1] Ibid (1957:449).

[2] Harris (1965); see fn 16 at pp. 573-4.

relative acceptability of a sentence (vis-a-vis other sentences) is preserved under transformation, e.g., that John eats meat and Meat eats John differ in relative acceptability and that this difference is preserved under transformation: e.g., Meat was eaten by John and John was eaten by meat.

From the point of view of eliminating redundancy from linguistic description, the primary significance of the transformational invariance of co-occurrence or relative acceptability ordering (indicating that many transformations are paraphrastic [1]) is that transformations permit co-occurrence ranges to be stated just for kernel sentences rather than separately for different sentence-forms (word class constructions) satisfied by the same word n-tuples. However, the hypothesis that meaning, as correlating with word co-occurrence ranges or relative acceptability ordering, is preserved under transformation remains at most a suggestive

[1] An important exception to transformational paraphrase are the so-called "incremental" transformations, e.g., the negative and question forming transformations. But these were held to add only constant differences or increments of meaning to all the sentences to which they applied, hence they altered meaning only in an a priori specifiable way; Cf. Harris (1964:476):

> There are, it is true, transformations which bring in a large difference in meaning, For instance, the question and the negative are transformations, since they simply permute some words of the sentence, (and) add constants, in the same way for all the sentences of a given form; and this without changing the difference in acceptability. But the difference in meaning which is due to the transformation is the same for all sentences, and does not affect the relative acceptability of the sentences. (Differences of truth result from negation, differently for different sentences, while the question (transformation) eliminates any property of truth; but truth is not directly involved in defining transformations.)

I think it needs saying someplace how ZSH's K & T model defines "Kernel S" differently than in GG, and that the "obsolescence" of kernel S in GG is irrelevant to their status here.

but rather impressionistic generalization pending a determination of why this is so, as might be provided by an actual syntactic specification of meaning. Moreover, the kernel and transformation model is ~~~~ as an explanatory account ~~~~ not completely satisfactory because it assumes ~~since~~ a purely distributional (i.e., non-transformational) syntax ~~is assumed~~ for sentences of the kernel.[1] And, in actuality, the desired factorization of the set of sentences by the set of transformations, which requires that the kernel sentences be mapped onto an identity transformation, contributes to the redundancy of language description since, in an explicit axiomatic presentation, this transformation would have to be listed separately for each kernel sentence.

There are additional shortcomings of the kernel and transformation model of grammar which can only be briefly considered here. There is, most notably, the difficult problem of specifying the domain of the various transformations, which is compounded in this model in that transformations are sui generis operations on sentence forms, i.e., they are specified by determining which word n-tuples of a given sentence form satisfy another sentence form. For example, since both John reads and John is reading exist (and are paraphrastic in some environments of occurrence), a transformation between the 'present' tense and the 'progr.ssive' may be defined:

$$N\ V \rightarrow N\ \underline{be}\ V\underline{ing}$$

or

$$N_1\ V\ N_2 \rightarrow N_1\ \underline{be}\ V\underline{ing}\ N_2$$ [2]

[1] Cf. Harris (1972:248 fn 2).

[2] E.g., John reads history → John is reading history.

While a large number of verbs undergo this transformation (e.g., study, runs, eats , thinks etc.), many do not: e.g., John knows → John is knowing.[1] Another, well-known problem of specifying domain, is the passive. For many years the passive (N_1 V N_2 → N_2 was Ven by N_1) was considered a paradigm of transformational analysis, despite the fact, already recognized in the earliest work on transformations[2], that many exceptions to the passive exist: that certain verbs did not passive (e.g., # Twenty dollars was cost by the book) or did so only under special conditions, such as involved restrictions on the N co-occurrents of V (e.g., inhabits which does not passive with some N_1 e.g., # Manhattan is inhabited by John, does if N_1 is plural: Manhattan is inhabited by greedy art dealers).[3] Each exception to a posited transformation (and this is the important consideration here) must be listed; since transformations are sui generis operations on sentence forms, this adds to the redundancy of language description.

The passive points up another shortcoming of the kernel and transformation model: nothing is said concerning the relations between obviously related transformations, for instance, where similar morpho-

[1] See the derivation of the 'progressive'(is...ing) tense from the entry into a sentence of aspectual operators that carry an implication about time, e.g., in the process of, which otherwise occur and hence must be accounted for, in Harris (1982:265). Harris derives tense from the partially-ordered word dependence requirement and system of reductions, thus eliminating tense as a primitive in language description.

[2] E.g., Harris (1954:794 fn 22) poses the additional problem of distinguishing the by of the passive, a morphophonemic change accompanying the passive, from the prepositional by, as in The letter was finished by noon which does not have the 'active' form Noon finished the letter.

[3] For some discussion, see Gross (1979:863-4).

phonemic changes are involved.[1] This bears upon a more general difficulty with the kernel and transformation model considered as explanatory of language structure, namely, no account is provided[2] showing that the transformational decomposition of a sentence (in effect, undoing the transformations and resolving ~~it~~ the sentence into one or more elementary "base" sentences) could simply be the ordered inverse of the transformations involved in synthesizing the transformed sentence by successively applying transformations to the elementary sentences.[3]

The same theoretical and methodological objective of eliminating redundancy from linguistic description that motivated the development of transformational analysis, briefly surveyed here, also spurred a subsequent major reformulation of transformational analysis in which

[1] For example, the passive is obviously related to what is termed a by-nominalization, a transformation taking a sentence into a sentence fragment; e.g., Mitterand nationalized the banks The nationalization of the banks by Mitterand. See the discussion in Harris (1982:364).

[2] Although Harris (1965:557) suggests the possibility and then demonstrates how the proposed transformational analysis falls short of this goal.

[3] This is the problem of "derived constituent structure". Little effort, beyond the important work of Emonds (1976), has been devoted to addressing this problem in generative grammar; for discussion of this point, see Bowers (1981) Chapter 1. In the continual reformulation of the generative model, the details and difficulties of constructing explicit derivations from base forms have often been overlooked.

these shortcomings are addressed. In his (1969), which is a stage on the way to the operator grammar of his (1982), Harris presents an axiomatic model of a grammar in which there is a "base" of transformationless "source" sentences formed by a system of predicates and from which, by further operations (predications) on these elementary sentences and attendent morphophonemic changes (which are paraphrastic), all other sentences of the language may be derived. Each of these elementary sentences is constituted by a "predicate system (which) carries all the objective information in the sentence" and of which "the most natural interpretation of its structure is that of giving a report".[1] The set of these elementary predication-created sentences comprises a "sublanguage (which) carries all the objective information, or report, which is carried in the language"; as informationally sufficient, these sentences "can be used without the rest of the language".[2] Distinguishing "objective information" as predication-created report may be viewed as culminating the program of regularizing linguistic description by eliminating variant forms which 'say the same'. This reformulation of transformational analysis provided the explanatorily required (and up to this point lacking) syntactic specification of the meaning that is preserved under transformation. As fully presented in his (1982), the specification of the objective information of a sentence is given by a partial order of word dependences,

[1] (1969:613).

[2] (1969:614).

Explicitly Distinguish "base" sentences from "Kernel" sentences here, or your audience will assume they are the same.

which has the semantic interpretation of predication ('what can be said of what'). In this theory, transformations are reconstructed as reductions (in phonemic shape) consequent upon entry of certain further words into the sentence. As reductions can be said to be, in an extended sense (see below) of the term, paraphrastic, [1] their functional role in that they enable the derivation of the remaining sentences of the language from unreduced "base" sentences, where the information-creating predicational relations are explicitly displayed, is analogous to transformations.

According to this theory, the base of a language consists of sentences (or sentence-like formations; see the discussion of "daggered" sentences, below) to which no reduction has applied. Each element of a base sentence is analyzed as "operator" or "argument" according to the same schema, a partially ordered word dependence relation. [2] There are words without dependence requirements in their environments of occurrence (e.g., Max, France, car, book, etc.). The class of these words, termed "elementary arguments" is marked N. In addition, there are other words ("operators") which have an argument requirement expressed in terms of categories. [3] Every word of a base sentence is assigned

[1] Sentences related by reduction, as reduction preserves the partially ordered word dependence requirement, are therefore a specifiable proper subset of the set of all paraphrastic sentences.

[2] (1982:4):"The dependence (i.e., requiring) relation is a partial order: if X > Y,Z, it may be that Y > Z, Z > Y, or neither....In a base sentence in which X > Y and there is no W such that X > W > Y, we say that X in that sentence is the operator, or next later entry, on Y and that Y is an argument, or immediate prior entry, of X; also that Y was free for X."

Define > in metadiscourse on this quote

[3] ibid., 4:"(E)ach word in the base sentences is associated with particular (linearly) ordered word sets of which its immediately prior entries must be members; we call these sets the argument requirement for that word."

a unique argument requirement which it has in every base sentence in which it appears.[1] All of the operator words are sentence-forming, i.e., the resultant of their entry is a sentence. Thus O_N (which often is written O_n) is the designation of an operator word which together with an N (as its argument) forms a sentence. Correspondingly, O_{nn} designates an operator forming a sentence together with two N's; examples of O_n are run(s), swim(s), cough(s); of O_{nn} are read(s), love(s). Further (second-order[2]) operators may operate on the sentence which is the result of the application of an operator to its required argument. Thus O_o: suddenly, as in John coughs suddenly, continue, as in Max's swimming continues (where the 's and -ing are so-called "argument indicators", morphophonemic changes induced by the entry of the higher operator continue into the sentence); and O_{oo}: since, entail, cause, and, etc., as in John runs and Mary swims. Other operator types include O_{nnn} e.g., give, introduce as in Max introduces John to Mary; O_{no} : wish, believe, suppose, as in Max believed the ice would hold; O_{on} : mainly prepositions, such as in, on, to, around, etc., Max skated to the crack in the ice[3]; O_{nno} : e.g., tell, promise as in Max promised us the ice would hold; O_{noo} : e.g., attribute as in Max attributed the ice's cracking to Monday's thawing temperatures. The variety of operator types may vary from language to language but their number is fairly restricted in each. For grammatical analysis, we take

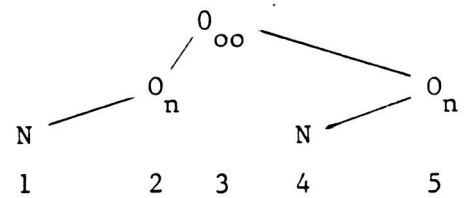
[1] (1982:34):"All base words of the language have only one argument requirement, which is satisfied in each base sentence in which they appear."

[2] An operator word is second-order if one or more of its arguments is itself an operator. There are no further higher types.

[3] to operates on skate (O), crack (N); in operates on the nominalized sentence crack which Max skated to (O), ice (N). Although the discussion here is of base sentences, the past tense used in the examples is derived by reduction from base sentences.

the relation between an operator and its arguments as fundamental, and not, e.g., the order of recitation or the order of appearance of the words in a sentence, Thus, 0_{no} designates the word believe, e.g., in Harry believes that Max is alive, where Harry is an N, and Max is alive is an 0.

By the order of entry of words into a sentence, we understand a sequence of words $\alpha_1,\ldots,\alpha_k$ in which, for each α_j $(1 \leq j \leq k)$, there is some preceding $\alpha_{j_1},\ldots,\alpha_{j_i}$ $(j_1,\ldots,j_i < j)$ which are words of the categories required by α_j and hence constitute the arguments of α_j, provided that no α_{j_g} $(1 \leq g \leq i)$ is argument to more than one operator[1] and provided that only one operator (the last entering the sentence) is not an argument to another operator.[2] For example, the 0_{oo} operator but is the last entering, or highest (least upper bound in the oriented semilattice given by the partial ordering) operator in



Mary is nice but Jean is nasty.[3]

[1] (1982:35): "When an 0_{yz} type operator,..., has entered a sentence on the basis of the presence of an operand YZ, no part of YZ is available for any further entering operator (...); hence no word in a sentence can have more than one operator directly on it in that sentence."

[2] ibid., 20: "For every operator in it the sentence may contain an operator with one 0 in its requirement, or for every operator pair (...), it may contain an operator with two 0's in its argument requirement. Thus each sentence has precisely one (latest entering, 'free') operator that has not become the argument ('bound') under some later entering operator."

[3] The example can also be given in a parenthesis notation:

N O_n O_{oo} N O_n

((Mary) nice) but ((Jean) nasty)

Both forms bring out the fact that is is here an "operator indicator", a morphophonemic change resulting from the entry of the 0_n operator upon its N argument, and not thus a "word" in this environment.

If, in addition, we require that the first argument of an operator enters before the second argument, and so on, then the order of entry is unique (i.e., strict); the words of a sentence are linearly ordered by their entry.

Of course, the order of entry induced by the partially ordered argument requirement and the linear ordering of a given operator's arguments from the order of appearance in a sentence may be quite different. Many languages distinguish between the entry order of arguments by inflectional markings or cases (e.g., Latin, Slavic languages). English, on the other hand, distinguishes entry order of arguments by e.g., prepositions, a very restricted case system (for pronouns)[1], and order of occurrence: an operator is normally "said" after its first argument. He loves her has nominative and accusative indicators of argument status, but John loves Mary, where there is no case morphology, must rely on order of argument occurrence to be distinguished from Mary loves John. The supposition is therefore that sentences in the base are not ambiguous and are completely determined by the order of entry of words. This requires that John loves Mary and Mary loves John be different sentences, the O_{nn} operator loves having a different first and second argument in each. If, in addition to the argument requirement, operators impose an ordering upon their arguments, the words of a base sentence comprise a full (linear) order, or chain,[2] not just a partial order.

[1] Cf. Sapir (1921), Chapter 7.

[2] For this terminology, see Grätzer (1978).

Therefore, the set A of words of a base sentence may be partially ordered by the entry requirement relation; call this partial ordering A_o. Then there is an automorphism A_1 from the partially ordered set A_o to a linear order for the sentence which results from imposing a linear order on the arguments of each operator. And there is another automorphism A_2 from this linear order that puts the words of the sentence into the linear order of their appearance in a recitation of that sentence. A_o shows the grammatical dependences. It also shows that, for instance, John loves Mary and Mary loves John have the same grammatical structure. A_1 exhibits the information of the sentence. A_2 is the chosen manner of speaking. The automorphism A_2 is based upon A_1. For instance, in English a fragment of the automorphism $A_1 \rightarrow A_2$ is the 'rule' that an operator normally occurs just after its first argument. Note that in A_o it does not make sense to say that an argument of an operator is first, second, etc. since this relation is defined only by A_1. A_2 allows, e.g., both Max reads a book and A book Max reads. [1]

There is, however, a useful way of looking at the same phenomena which makes the words in a base sentence an only partially ordered set. Loves, e.g., takes two arguments which are N's. It does not take operators as arguments. [2] Accordingly, loves takes any N as its first argument and any N as its second argument; more generally, an O_{nn} operator is not particular as to which N is first and which is second. Of course,

[1] The difference between A_o, A_1, and A_2 was pointed out by H. Hiż.

[2] So, e.g., I love singing ~~will not be considered~~ is not a sentence of the base, but rather the resultant of a derivation through several reductions from a more complex base sentence. See (1982: ___).

the result of an operator's entering on a different ordering of its arguments (which still satisfies its argument requirement) is semantically, and moreover, informationally, different, but for O_{nn} (as well as O_{n}, O_{nnn}, etc.) operators, any N can serve as one of the basic arguments.[1] Similarly, an O_{no} operator does not require this or that N to appear as its first argument, nor that a particular sentence be its second argument.[2] In the same way, O_{oo} can join any two sentences, and O_{o} can occur on any sentence, and so on.

Considering a given sentence of the base, the argument requirements of each word may be stated according to the following procedure: N's do not require another word, O_{nn} requires two N's, O_{oo} requires two O's, O_{no} requires any N as first argument and any O as second, and so forth. In an abstract way, each operator requires each of its arguments equally. Thus in John loves Mary, loves requires both John and Mary. Neither John nor Mary require anything. (A more algebraic way of stating this is to say that John, Mary require themselves only.) The partial order of word entry is thus formed by the set of words of base sentences[3] and the relation of requiring. The set of words of base sentences together with the requirement relation forms a categorial grammar in the sense of

[1] Cf. (1982:9):"The partial ordering of word entries that creates the sentence specifies its informational content at the same time." But, as the John loves Mary example shows, we must also note the linear ordering of the arguments of each operator where a permutation of this order still satisfies the operator's argument requirement.

[2] ibid., 34:"...what is required is just N and O and not particular words."

[3] As noted below, words containing affixes (e.g., -ment, -ly, -hood, etc.) are derived by reduction from free-standing words and so will not occur in base sentences; similarly for wh- relatives and tense (except "present").

Lambek [1] with only the several categories noted above: N, O_n, O_{nn}, O_{nnn}, O_{no}, O_{nno}, O_{on}, O_o, O_{oo}, and perhaps a few others. Since in the base sentences one need not distinguish between elements of the same category, if $O_{nn}\ N_1,N_2$ is (i.e., formulaically states the structure of) a base sentence, then $O_{nn}\ N_2,N_1$ is also a base sentence. And generally, if $A_{BBB}\ B_1B_2B_3$ is a base sentence, then $A_{BBB}\ B_2B_3B_1$, $A_{BBB}\ B_2B_1B_3$, $A_{BBB}\ B_3B_1B_2$, $A_{BBB}\ B_3B_2B_1$ and $A_{BBB}\ B_1B_3B_2$ are also sentences of the base where A and B are any categories. This closure axiom extends considerably the set of base sentences, leading from fully acceptable and attested sentences to sentences of low or questionable acceptability, for the base may be said to contain, e.g., not only John needs water but as well Water needs John.

Thus there are two relations between words of a base sentence. One, the requirement which gives the partial order, the other, the linear order of arguments of each operator. We may think of the structure of a base sentence as a mapping between the two sets of words, a mapping of the partial order onto the linear order. This mapping amounts to deciding the order of arguments of each operator.[2] Sentences related by reduction are considered informationally equivalent, information being 'created' in the base, unreduced, sentences upon the satisfaction of a word's entry order requirement and ordering of the arguments of each operator (by A_1 not A_o). This information is preserved through all reductions. The base can be considered as a distinguished subset of the sentences of the language in which is contained

[1] Lambek (1958) and (1961).

[2] Harris (1980:4); Cf., (1982:4 and 36).

all predication-created information that may be expressed in the language.[1] A similarly explicit specification of the remaining sentences of the language requires an enormously detailed and painstaking inquiry into the conditions under which particular operator-argument pairs fall in the domain of particular reductions. Such an undertaking can provide, at any particular point in time, at best only a proximate characterization of the set of all mappings from base sentences onto reduced sentences and vice versa. Nonetheless, it still may be said that associated with each sentence of English is one or more base sentences, each of which is a set of words under the A_o and A_1 orderings, and that this structure is preserved under all reductions and reconstructions.[2] The mathematical characterization of the base sentences means that the information of a sentence can be grammatically reconstructed in terms of a succession of predications (and initial words N with null entry requirement), each of which, at the point of its entry into the sentence, may be represented as a selection from among permitted alternatives of word combination. The result is what has been termed a "regularization beyond language",[3] i.e., the reduction of language to its information-creating structure.

[1] As participation in a relation (A_o) in respect to one another may be taken as the defining characteristic of the words of the base sentences, any other properties these objects may have (in particular, their 'meanings') can be considered as incidental or arbitrary. So, Harris concludes, the set of these objects may be said to form a mathematical object; see (1983:632 ff), (1978:13), and (1980).

[2] (1982:22): "The operator-argument relation is not only found in the base sentences after undoing all the reductions but is also preserved in the sentences under the various reductions that take place."

[3] (1968), Chapter 6.

Reformulated in terms of entry and reduction, transformations need no longer be seen as sui generis[1] relations between sentences, or sentence-forms, but rather as effects of a basic, information-creating operation -- word entry into a sentence -- together with a semantic property according to which an operator word, entering a sentence and having a relation of high likelihood (or expectability) of occurrence with certain[2] words of its argument class, may be reduced in phonemic shape (perhaps also changing the phonemic shape of its argument), possibly to zero, since it contributes little or no increment in meaning to the sentence.

This result enables the difficulties adumbrated briefly above, with previous models of transformational grammar, to be seen in a new light and accordingly addressed.[3] And it has this

[1] Cf. Harris (1982:21): "Transformations are not a set of word manipulations coming full blown and are not a grammatical process at all; but an effect."

[2] Certain operators, termed "broad selection operators" (1982:60 ff) have very high likelihood to many members of their argument class and are consequently almost always reduced.

[3] The problem of specifying the domain of transformations is eased by the explicitness with which the domains of reduction can be

consequence: in eliminating redundancy from language description, this reformulation of transformations serves the goal of establishing a 1-1 correspondence between the constructed elements of description and behaviorally determinable meaning. Redundancy is further eliminated in that the major categories of traditional grammar -- tense, aspect, mood, affixes --are all derived under the entry and reduction system.

(continued from the previous page)

> formulated in terms of the high likelihood of co-occurrence of an operator with various words of its argument class. The domain of reduction is therefore a subset of the word dependence class of that operator. Further, resemblances between transformations, e.g., that cited above between the passive and by nominalization, are shown to be due to the result of identical processes of entry and reduction. Hence the domain of the passive is the logical product of the domain of its component reductions (one of which produces the by nominalization; see Harris (1982:362 ff)). Accordingly, there is no passive transformation; the passival form of sentences is the resultant of the successive effects of entries and reductions. Finally, since reductions take place upon entry of an operator, and since each reduction leaves a "trace", which consists in an apparent deviation from the word dependence requirement of the operator or argument, the analysis and synthesis of sentences are explicitly inverse operations. Cf. Harris (1982:19):
>
> > In the present theory a sentence analysis is a derivation of the given sentence via reductions from sentences in the base set. It thus serves as the basis for a recognition method, applied to given English sentences. Because of the explicitness of the reductions and their domains, however, these derivations can also be used directly to produce the sentences of English from sentences in the base set, always keeping in mind the optionality of the great bulk of the reductions. Given this information, an effective procedure for analyzing sentences is possible in principle because the entries into a sentence are ordered and the reductions will take place in an entering word, in the argument of an entering word, or in a stated earlier entry when a later condition related to that entry is satisfied. Each reduction leaves a trace: the trace of a zeroing is a recognizable emptiness in matching the argument requirements of operators to their required arguments. No proposed reduction that left no trace was admitted. (Footnote suppressed.)

However, the more compact and efficient statement of regularities of combination of elements which is possible in operator grammar has not been achieved without a certain 'cost' which involves a rather different conception of the notion of paraphrase. Since the object of description, 'the' set of sentences of a language, is not given in advance (e.g., by acceptability considerations), problems are obviously presented in attempting any description, let alone a least redundant description of that object. On the other hand, this situation also affords certain opportunities which may be exploited. Thus the possibility exists of achieving greater regularization of description by 'extending'[1] (what is usually referred to as) the set of sentences of a language in a specifiable way, i.e., by allowing as grammatically possible (grammatically well-formed) sentences,[2] certain word sequences which cannot be ~~maintained~~ لستدى to occur normally (as too unwieldy or awkward), or which occur only as nonce forms, or which are no longer said though historically attested,[3] and which correspondingly

[1] Cf. Harris (1968:158): "Regularizing the grammar without changing the set of sentences which the grammar describes means replacing a grammatical or dictionary difference by a morphophonemic operation. ...Regularizing the grammar by extending the set of sentences to include nonextant (source and intermediate) sentences which are implicit in the transformational structure of the extant ones is different, but does not change appreciably the informational capacities of the language."

[2] Thus these sentences must be 'sayable' or be 'useable communications'; cf. Harris (1972:246). See also the qualification on 'sayability' of certain of these sentences in which metalinguistic material is reconstructed (1982:90).

[3] E.g., the derivation of childhood from a source like the state of of being a child; see (1982:175): "...-hood as in childhood, is from the Old English had as in child-had, 'the state or condition of being a child', papan had 'papal status', ...".

are of low acceptability. These 'occur' primarily in derivations, either as base sentences or intermediate sentences between base sentences and occurring (attested) sentences; the sentences of the language, including those of normal acceptability, being either base sentences or reductions from base sentences. They are, therefore, mainly reconstructed sentences which belong to ~~the~~ that sublanguage of the language as a whole (e.g., English) which is its (homophonic) metalanguage, i.e., its grammar. And, in providing the steps necessary ~~in~~ for the derivation of all attested sentences, where each successive step in a derivation exhibits a stateable change in phonemic shape from the sentence of the preceding stage but preserves the word dependence relation of all words, these sentences make possible a less redundant, which is characterization of the sentences of the language ~~through the~~ because it utilizes ~~ation of~~ a smaller number of primitives. In the trade-off between ease of derivability and fewer primitives, a characterization seeking to be least redundant will opt for fewer primitives at the cost of ~~maximizing derivability~~ more complex derivations.[1] OK?

[1] See the section entitled "Note on Method" (1982:26) which begins:

> The central problem...was to find what objects and relations could effectively characterize with the least redundancy those combinations of words that occur as English sentences against those that do not; to find the simplest system adequate for the task, with as little as possible unused capacity in the apparatus of description. It dictates a minimum of multiple classification of words (...) and maximum derivability: the theory has one primitive relation (argument requirement) and not many derivational steps (reductions), although the chains of derivation for a given sentence may be long.

These reconstructed sentences whose special status may be indicated by a dagger (†) (and are thus termed 'daggered' sentences), provide regularized source forms for many of the attested sentences of the language which may not otherwise be easily or legitimately fitted into an axiomatic grammar in an nonredundant manner. A daggered sentence may, e.g., be used to explain the difference[1] between Mozart and Verdi wrote operas (which is reduced from the non-daggered source ← Mozart wrote operas and Verdi wrote operas) and Gilbert and Sullivan wrote operettas (which does not have a source Gilbert wrote operettas and Sullivan wrote operettas but instead the daggered A team -- the team contains Gilbert and the team contains Sullivan -- wrote operettas → A team which contains Gilbert and Sullivan wrote operettas). Several additional points about daggered sentences may be noted. (1) The unwieldyness of these sentences stems in many cases from their containing free-standing words to indicate meanings which are normally only indicated by reduced forms of these words, e.g., words reduced to affixes, as in tense (e.g., John went ←†I say that John goes; said going is before said saying that John goes[2]) or nominalization, or to zero (phonemic form) as in the Gilbert and Sullivan example above. Some of the daggered sentences are suggested by actual historical sources, e.g., the case of childhood, noted above, or in the derivation of the have...en of the so-called "perfect" tense from the aspectual operators state, situation, condition and the like. Thus I have caught the fish

[1] (1982:16).

[2] (1982:103).

has a daggered source I have the fish in the situation of one's catching it (where caught is the reduced form carrying the characteristic stative aspect of the perfect) since I have the fish caught meaning, roughly, 'I have the fish in caught-state' is historically attested.[1] (2) Another source of low acceptabilty of some daggered sentences comes from their incorporation of explicit metalinguistic word sequences which are rarely if ever spoken outside of grammatical discussion, e.g., the derivation of questions (including intonation) from metalinguistic operators (Is he here? ←† I ask whether he is here). Incorporating metalinguistic material into sentences is also used to do the work of specifying grammatical reference, e.g., conjoining S_2 to S_1 by wh- requires that a word in S_2 be identified, in a metalinguistic sameness statement, as the same as a word in S_1, whereupon the word in S_2 may be reduced to zero or to a proform (He sold the land, which I would never do ← † He sold the land; sell the land (prior is same as penult) I would never do [2]). (3) Further reasons for the low acceptability of some daggered sentences may be seen in: (i) the fact that some reductions are obligatory (or the reduced sentences are much more comfortable); this is the case with the transposition of adjectival modifiers (introduced by a wh- appended sentence) to the left of their noun host (e.g., big red car ← red car which is big ← † car which is red; said red car is big, where said abbreviates the metalinguistic sameness statement); or (ii) that one wishes to

[1] See (1982:291-2) for details.

[2] (1982:87 ff) and especially pp. 90-91.

avoid including among the primitives of the grammar forms that are lacking in other languages (e.g., the definite article, derived in English from propositional that which is [1]); or (3) avoiding appeal to structures (such as logical or set-theoretic) which are outside those definable in terms of the relations which suffice to describe the rest of the language. As an example, one may cite quantifiers, e.g., many, few, some, etc., which modify plurals (themselves derived from finitely many conjoined sentences) and collective nouns, and which are obtained as noun second arguments of the zeroable operator mounts to (e.g., Many books fell ← † Books fell; which were many ← † Books fell; which mounted to many ← † Books fell; books falling (prior is same as penult) mounted to many [2]). However, since "the daggered sentences consist of English words that appear in positions held by these words in the attested English sentences" and since they differ from the attested sentences only by listable changes, "no clear line can be drawn between them and the normal sentences" [3] and they may be taken as sentences of the language.

[1] (1982:237 ff).

[2] (1982:262 ff).

[3] (1982:18).

Any criticisms which might be generally directed at the extension of the set of sentences to include 'daggered' sentences (as opposed to criticisms of particular derivations) should be balanced by an assessment of the measure of success which their employment brings in carrying out an axiomatic characterization of the sentences of the language; i.e., whether all the sentences of the language can be demonstrated either to be base sentences (which have the simple information-creating predicational structure) or to be derived from base sentences using the system of reductions.[1] And to be sure, to speak of informational equivalence as paraphrase is to adopt a specialized meaning for the term, which may conflict with its presystematic relations to traditional notions like synonymy or to more empirical notions like speaker acceptabilities. But the fact of the matter is that 'true paraphrase', whatever it may be, is on all accounts an exceedingly rare phenomenon[2], not to mention the fact that there have yet to be proposed any adequate operational tests of paraphrase in terms of acceptability or in any other terms. It appears, then, that

[1] (1982:15):"(The daggered sentences') importance lies in this: if we include them in the set of English sentences, whether actual or possible, the reductions...will suffice to derive all English sentences from a subset of sentences (including daggered ones) that have the simple "base" structure..."

[2] The linguistic use of paraphrase and consequence has been most extensively treated in the writings of Hiż who (n.d.:10) points out that while paraphrase may be considered to be an identity relation, nonetheless "in empirical sciences an identity of measurements must accept a degree of accuracy, or rather an acceptable degree of inaccurary". In an earlier paper (1964:97-8) Hiż notes that to say any two sentences are paraphrases does not implicate any"thing" such as meaning or content:"Whether there are meanings of sentences...is of no concern for grammatical considerations. What matters is only that the speakers recognize a sentence as saying the same as another sentence. This "saying the same" is just a relation between two sentences and does not presuppose something else, the "thing" said in each sentence." The information-creating word dependence relation now permits a refinement of this initial formulation.

this usage is not so much a 'deviant' extension of the meaning of the term as a delineation or restriction in meaning of a term formerly possessing only an imprecise meaning. [1]

It was remarked in §§ 5.1 and 5.2 above that additional word recurrences were characteristic of sentences connected in discourse or in a collection of discourses pertaining to a particular subject matter. In §5.2 we claimed that the extent of the articulatedness and detail of informational structure describing a sentence is wholly a function of the characterizable similarities and dissimilarites it bears to other sentences over some specifiable domain. For a grammar of a language as a whole, which (as Harris (1982)) is a grammar of its sentences considered independently of context of occurrence (but nonetheless as ~~comprised~~ including, perhaps, ~~of~~ other sentences), the sole basis of comparison with other sentences is ~~the commonality of being~~ that they are all compositionally characterized as well-formed by the objects and relations which minimally suffice to define the set of sentences of the language. The particular categorization given a word of the language in the predicational partial ordering is based on two fundamental considerations: (1) an assessment of the most widespread environments of occurrence of the word and (2) the system of reductions at hand which provides a criterion for choosing a given environment of occurrence as

[1] On acceptability tests for paraphrase, see e.g. Quirk and Svartnik (1966), and further Nolan (1970).

its 'regularized' environment in the base set of sentences, and from which its other occurrences may be derived. Accordingly, the categorial assignment given a word is neither inherent nor unrevisable if sufficient adjustments are made elsewhere in the system. In this respect categorial relations among words reflect not so much ontology as ease of derivability of the entire range of environments of occurrence of a word from a stipulated, 'regularized' occurrence.[1]

However, the additional constraints on word combinations of sublanguage and discourse make possible further regularization of description beyond the partitioning of the vocabulary of the language into operator - argument word classes. Discourse and sublanguage provide specifiable domains over which to define additional regularizing operations which eliminate variant forms that 'say the same'. In discourse there is a discernable patterning of word recurrences as, e.g., in the word repetition which raises the acceptability (that is, raising their position in a relative acceptability ordering) of certain conjoined sentences.[2] Already the word repetition constraint governing which sentences can be conjoined extends beyond the constraints on word combinations stateable for each conjoined sentence separately.[3] Beyond conjoined sentences, the word

[1] The wide range of occurrences of prepositions present a particularly appropriate illustration of this point; see Ryckman and Gottfried (1981) for details of the categorization proposed in Harris (1982).

[2] See the discussion in Harris (1968), pp. 132-135.

[3] Harris (1981).

recurrence that may be identified in a set of texts dealing with a particular subject matter, in the case briefly considered here and in more detail in Chapter 6, in a specific research area of of science, indicates still more possibilities for regularization of description. It can be shown that the constraints on word combinations in these texts make possible the delimitation of a particular sublanguage, i.e., a set of sentences closed under some of the operations defined for the language as a whole. Within the domain of this sublanguage, whose vocabulary and possibilities of word combination are much more restricted than in the language as a whole, regularizing operations (including transformations) play a somewhat different role than in the description of the entire language; ~~i.e.~~ that is, the regularization is carried out with the objective of structurally representing the sentences of the sublanguage so as to maximize their similarities. In establishing repeating sequences of word classes (each sequence forming a sentence type), transformations may be applied to a text sentence so that the words in that sentence may be shown to have the same grammatical relations to each other as they do in other text sentences. Here, as in the earliest work on discourse analysis, transformations facilitate the purely distributional identification of word classes by regularizing the environment of some word occurrences so as to place them in inspectable conformity

explain what this means for some of your audience

with that of others. As a simple illustration, the word sequences antibody production by lymphocytes in S_j, the production of antibody by lymphocytes in S_k, and lymphocytes' production of antibody in S_l can all be represented by the same left-to-right "normal linear form":

[of] antibody | ['] [the] production [by] | lymphocytes [←]

which receives the (word class) formulaic index $A\ V_p\ C_y$ mirroring the linear order (here, the subscripts to V and C designate word subclasses). In this representation, the vertical bars indicate the segmentation of the various word sequences into the established word classes and subclasses. The square brackets -- adopted for the purposes of this example only -- enclose the actual morphophonemic variants of the different word sequences, while the bracketed arrow is a scanning instruction to read the segments of the representation right-to-left so as to obtain the linear order of the words in the actual texts. With the use of this leftward pointing arrow and a "relinearization transformation", this representation permits the three readings corresponding to the occurrence in the text of S_j, S_k and S_l:

i) antibody production by lymphocytes corresponding to S_j

ii) of antibody the production by lymphocytes " " S_k

iii) lymphocytes' production of antibody " " S_l.

Further examples and details of the use of transformations and other regularizing operations will be given in the next chapter, together with an account of how they are employed in obtaining formulas of information (e.g., A V_p C_y) for the sublanguage. It may be remarked that as opposed to regularization over the domain of the language as a whole, which is conducted axiomatically in deriving all attested sentences from simple reductionless "base" sentences, regularization over the domain of a sublanguage proceeds not axiomatically but rather by transforming sentences (if necessary) to align them into a "normal form" which, with respect to the stated methods, is proposed as a compact and maximally efficient structure in which to 'house' the transformationally relatable sentences of the texts. Moreover, the formulaic representation of this "normal form" can be read as stating what is informationally equivalent or distinct (e.g., by change of subclass or "local modifier" designation; see Chapter 6 §3) in the sentences thus related.