CHAPTER 2 SCIENTIFIC SUBLANGUAGES

0. Introduction. The definition of cross-reference in the preceding chapter is tested in a description of referential relations in a research article ("Influenzal") drawn from a sublanguage of cellular immunology. In the first section of the present chapter some general considerations are introduced regarding the notions of 'discourse' and 'sublanguage'. A number of these are later developed in connection with a statement of specific results obtained in previous work on the immunology sublanguage (sections 3 and 4). While we lack fully-articulated theories of discourse and sublanguages, the study of sublanguages presents a vantage point in terms of which several topics hitherto regarded as peripheral, e.g., those in the hyphenated field of language-in-culture, can be integrated with grammatical investigations, and many, more global, questions, concerning grammar, e.g., the notion of ambiguity, can be formulated (or: reformulated). Some of these questions are addressed in section 4.

As with previous work on this sublanguage, the present investigation is concerned with regularizing texts in a way which permits a representation of their information. In this task, it is based upon a theory of language structure exemplified in the operator-grammar of English presented in GEMP (section 2.1). Discourse- and sublanguage analysis use transformations (section 2.2) to align sen-

tences in a text, in a manner which assists in establishing word-classes specific to the subscience (section 3.1). The methods of analysis are presented in section 2; results of these methods applied to a body of research articles (concerned with the site of antibody formation) are surveyed in section 3. Section 4 examines the larger issues involved in the grammatical specification of this sublanguage and comments upon those results of the prior investigation which bear upon concepts of synonymy and ambiguity. A number of hypotheses proposed in this examination of cross-reference are intended to test and extend these and other results.

1. Discourse and Scientific Sublanguages. In this section the notions of discourse and scientific sublanguage are introduced. Discourses serve both as data in respect to the establishment of linguistic elements, e.g., phonemes, morphemes, and as linguistic domains in their own right in which regularities of occurrence among certain of these elements can be examined (section 1.1). Scientific sublanguages can initially be characterized as a set of discourses in particular field of research (or theory) and are part of the 'sociolinguistic division of labor' obtaining in many speech communities. In section 1.2, scientific sublanguages are briefly characterized and contrasted with the neighboring notions of 'style' and 'dialect'. The question of grammatically delimiting a scientific sublan-

guage is addressed in section 4 after presenting a sketch of the grammar of the immunology sublanguage (section 3).

1.1 Discourse. Discourses are firstly data for grammatical investigations. Construction of the various units of linguistic analysis -- phonemes, morphs, morphemes, etc. -- proceeds from a comparison of discourses.[1] For instance, the basis upon which sounds are collected into phonemes is the relation of "being a repetition of": that [f] and [n] contrast in English follows from the fact that, e.g., His name is small is not a repetition of His fame is small.[2] Similarly, the determination of morpheme boundaries is made in respect to distinctions between discourses. A less familiar case is presented in transformational analysis. This analysis makes use of the result that within a particular sentence S_i, a sentence S_j is identifiable together with some material X. To obtain this result requires analysis and evidence: one must confirm that the S_j-portion in S_i is the same sentence as an independent S_j and not just the same words as S_j otherwise brought together. The existence of this other possibility is exemplified by ambiguous sentences such as They appointed a reactionary secretary of defense. Moreover, the evidence for the identity of the S_j-part of S_i with S_j requires an examination of discourses, either by a comparison of neighbors of S_i and S_j in sets of discourses or by relating inequalities of likelihood between S_i for each material X and those of S_j.[3]

Several important conclusions can be drawn from such cases. Phonemes, defined in terms of differences (distinctions) between word-repetitions, permit the representation of utterances in terms of discrete elements.[4] The distinctions among these elements and among others, viz., morphemes, are made in respect to their distribution, i.e., possible combinations with other such elements in discourses. It is in terms of such occurrence-restrictions that grammatical structure can be stated: one need not impute a substrate or locus, e.g., mind, to the distinctions or structures. Finally, these examples give some preliminary indication of the way in which distributionally-marked distinctions relate to information in language -- the relation of "being a repetition of" (or: not being a repetition of) introduced with phonemes is a "semantic" one identifying and distinguishing various utterances.

A grammar of a language is a statement of these elements and their regularities of occurrence. Regularization is a procedure by which more freely combining elements are defined or the operations of the grammar are generalized. Examples from structural linguistics include the redefinition of phonemes in terms of simultaneous components, and of morphemes as particular phoneme sequences in terms of various changes in the composition of phonemes.[5] Within the procedures provided by structural linguistics, a discourse -- that is, a connected piece of speech or writing -- is characterizable as a sequence of sentence structures.[6]

Discourse-analysis, as presented here (section 2.2), describes the additional structure presented by the recurrence of words in particular positions relative to other words which recur.[7] Regularization of a discourse consists in paraphrastically realigning (transforming) certain sentences in a text. By means of these transformations, dissimilarities in the environments of particular words or word sequences can be reduced; the realignment of sentences into a designated normal form assists in establishing equivalence-classes of word-occurrences which are positionally similar to one another. The structure of a particular discourse can thus be presented as a table in which each row consists of a particular sequence of these equivalence classes and each column consists of the successive members of each class.[8]

The principal interest which attaches to these methods of analysis is as a means of representing the information in a discourse. A theoretical explication of the informational import of these regularizing operations is given in section 2.21 and is demonstrated in results of a previous research project in the subfield of cellular immunology (section 3).[9]

Other studies can, of course, be developed on the basis of these procedures. As Harris notes, "Although the specific word-recurrences in the successive sentences of a discourse are unique to that discourse, various types of recurrence patterns seems to characterize various types

of discourse" (1982:233). These types of discourse, e.g., narratives, scientific observations, are akin to what is referred to as "text-types" and "genres". Some indications of these types appear in the results reported for the immunology sublanguage in The Form of Information in Science (hereafter referred to as FIS), where, for instance, sentences under the heading of "Discussion" differ in clearly stateable ways from those under "Results". Again, 'style', e.g., Ciceronian, telegraphic, is often used to describe features of particular discourses (as opposed to stray sentences). While certain elements of style are tied to the specific vocabulary used, other features of style may be related to the types and sequences of transformations applied to a discourse in order to obtain some designated "normal form".[10] Another interest, exemplified in the early work on discourse analysis, is in the critique of discourses whose patterns of recurrence may point to various ideology-linked distortions (Harris 1952a:342-43; 1952b).

1.2 Scientific Sublanguages.

1.21 Social Setting. Among the tasks assumed in sociolinguistics is a determination of speech-communities along various social dimensions (class, age, etc.) and an examination of variation within the speech of communities. Most all speech communities characteristically show a "sociolinguistic division of labor",[11] shared habits of

word usage connected to the special activities of its various sub-groups. Bloomfield, speaking of an ideal record of linguistic exchanges among members of a speech community, noted:

> We believe that the differences in density of communication within a speech-community are not only personal and individual, but that the community is divided into various systems of sub-groups such that the persons within a sub-group speak much more to each other than to persons outside their sub-group. (Bloomfield 1933:47)

Certain of these sub-groups have their own distinctive speech forms -- "Occupational groups, such as fishermen, dairy workers, bakers, brewers, and so on have...their own technical language" (Ibid., p. 50).

Scientific (and technical) sublanguages should be differentiated from the neighboring notions of style and dialect. While occasionally one finds references to 'scientific style' and 'scientific dialects', these terms need not coincide: scientific material can be presented in a decidedly non-scientific style (e.g., D'Arcy Thompson's On Growth and Form). The term 'dialect' is perhaps better reserved for the local and regional forms of speech to which it is typically applied.[12] A lecture on population ecology, for instance, can be delivered by a Chicagoan or a Brooklynite. All of these notions -- dialect, style, and science languages -- can presumably be characterized as sublanguages. In Mathematical Structures of Language (hereafter called MSL, p. 152), sublanguages are taken to be "certain proper subsets of a language [which] may be

closed under some or all of the operations defined in the language". One can suggest some rough criteria which serve to distinguish these various forms of sublanguage. In respect to regional dialects, one point of distinction is the phonetic differences relevant to its characterization. Moreover, whereas regional dialects share in the main the gross grammar, i.e., the major word-classes, formulated for the entire language, sublanguages of science can be said to have grammars distinct from those constructed for the language (see below). Styles, as suggested earlier, can be described (at least in part) in terms of particular sequences of transformations, and perhaps frequency of vocabulary under some sorting.

A question may be raised as to delimiting the boundaries of a particular scientific sublanguage. While the question is not answerable apart from some reasonably full descriptions of such sublanguages, it should be noted that the situation with science sublanguages in this respect is much the same as with languages themselves: there may be elements not fully integrated in the system, e.g., restricted to particular vocabulary items (as with /zh/ in measure), or others which are in the process of change and so describable in differing ways (cf. section 4).[13]

1.22 Overview. In this section, I present a broad linguistic characterization of scientific sublanguages, with examples drawn from some of the few which have been studied[14] (the immunology sublanguage receives fuller ampli-

fication below). These sublanguages can be initially, and provisionally, identified as a set of discourses within particular journals (cf. section 4). The restriction to such sublanguages as are written discourses may call for revisions at a later date.

Whereas a grammatical description of a large set of "randomly" selected sentences in a language approaches that constructed for the entire language,[15] a study of restrictions on word combination within discourses of a sharply delimited field yields various word-classes not constructible for the entire language. Thus, for the sublanguage of pharmacology investigated by Sager and her colleagues, a word-class G with members such as drug, glycosides, can be formed on the basis of their occurrence as the subject of such verbs as penetrate, diffuse into, is located in, and as the object of lose, resist, is treated with. A word-class M (for 'membrane') can be defined as the object of, e.g., penetrate (with members of G as subject), and of permeate. The verb-classes of the sublanguage are defined in respect to their possible subjects and complements. In the cellular immunology sublanguage (section 3), members of a word-class A (antibody, protein) occur as the subject of is present in, is contained in, is absorbed into (the occurrence of these words in subject position may have been effected by a transformation cf. section 2.2). These verbs in turn with subject A can have as complements other words, e.g., lymphocytes,

reticulum cells (members of a word-class C), lymphnodes, spleen (of a word-class T) and, in these combinations, form a word-class V. In the language as a whole, a sentence such as antibodies contain lymphocytes is counted as grammatical whereas the restrictions within the sublanguage exclude it (cf. section 4.4). Conversely, there are rules within the entire language -- e.g., for the formation of imperatives, which are not required in describing the discourses of this subfield. It follows that the grammar formed for a given sublanguage of science is not the same as, i.e., it intersects, the grammar formulated for the entire language.

The grammar of a particular sublanguage of science is constructed in part by operations which regularize the discourses in that sublanguage (section 2.2). Thus, in accord with the quote from Harris before, the set of discourses is closed with respect to these regularizing operations (other closure operations are discussed in section 4). Given a text-sentence such as lymphocytes contain antibody, a passive transformation may be applied, yielding antibody is contained in lymphocytes; the latter sentence, as the former, is within the sublanguage. Similarly, permutation of the PN phrase in In lymphnodes antibodies are produced yields the sublanguage sentence Antibodies are produced in lymphnodes. The role of these operations in establishing a sublanguage grammar is addressed more fully in section 2.2.

In terms of these restricted word-combinations in a science sublanguage particular sentence types can be formed. For instance, in the immunology sublanguage, the last two sentences above are of the sentence type AVT (which can be read as a sentence: antibodies (A) are produced in (V) tissue (T)). For some of the sublanguage word-classes, subclasses can be formed on the basis of further restrictions in environment. For instance, are produced in (with subject antibody) does not occur with a member of T, liver, as object. Given such restrictions a subclass of V, V_p, can be defined (also a subclass T_v for liver). The word-classes of a sublanguage of science are what Harris terms "locally closed" (FIS:112). That is, for a particular subfield at a particular time, the list of word classes and of their members is closed, though extendable with an expansion of the field (see sections 3.4 and 4 for important qualifications to this).

Natural languages such as English and scientific sublanguages also differ in respect to the status of their metalanguages. Importantly, natural languages lack a metalanguage external to them in terms of which their elements and operations can be identified. Thus, a grammar of a natural language is characterizable only in respect to restrictions of word combination within that language (or: some other natural language, as a grammar of English in Spanish, for which the same issue arises). Scientific sublanguages do not contain sentences of their metalanguage.

1.23 Related Topics. The features mentioned above are by no means exhaustive of scientific sublanguages (others are noted in section 3) nor are they, in the absence of other descriptions, clearly to be taken as somehow jointly defining them. Nonetheless, various connections can be traced between the study of scientific sublanguages and topics in adjoining disciplines.

For example, the word-classes and sentence types which are obtained through an examination of discourses within a scientific field can be said to "reflect" the relevant objects and relations within that field.[16] In this, the methods and results of studies of scientific sublanguages invite comparison with the efforts of various philosophers in the earlier decades of this century to establish what Carnap called "the logic of science".[17] Such a comparison would need to consider rather broad questions concerning the character of formalization and the difficulties supposed in describing natural language, e.g., ambiguity, vagueness.

In connection with formal languages it should be noted that the "fragments" of natural languages investigated in model-theoretic semantics, e.g., Montague grammar, could be considered sublanguages (proper subsets of the entire language), but in a rather Pickwickian sense. As there is, as yet, no grammar forwarded by these theories for an entire natural language, one cannot properly speak of a proper subset of the language closed under some or all of the

operations defined in the language. Also, the fragments studied are not restricted in respect to subject matter (thus they are not subject-matter sublanguages) and the word-classes are established not in respect to paraphrastic transformations but are stated in a formalized metalanguage (the metalanguage of which is generally some natural language).

The study of scientific and technical sublanguages is also pertinent to a more precise characterization of speech communities. That scientific and technical sublanguages may be said to 'arise' in connection with specialized activities is itself a fact regarding the use of language. Particular features of sublanguages or the various styles in which they are written (or: spoken) may relate to specific requirements of that activity. The sublanguage of aviation maintenance manuals, discussed by Lehrberger, shows, for instance, a high frequency of imperative forms, and omissions of the definite article, as in, e.g., Check indicator rod extension. Such comparisons -- of particular sublanguage features and the conditions and character of certain actions -- would be of relevance to "pragmatics". Results concerning the 'distribution' and 'range' of various subject-matter sublanguages are important to sociolinguistics; certain instances of borrowing might, for example, be analyzed as occurring between particular sublanguages.[18]

Other questions, more immediately relevant to the present work, have been raised by the previous investigation of the immunology sublanguage reported in FIS. Chapter 3 (section 2) discusses 'language-like' systems, i.e., tables and graphs, which appear in the "Influenzal" article. Certain of these can be transcribed into sentences which are analyzable in terms of the grammar of this sublanguage.[19] Other research articles contain pictures. While pictures evidently cannot be accommodated within a grammar of this sublanguage, it is of interest that the captions provided for them generally give the information which is used (in discussion, in argument) in the article. Questions concerning relations among scientific sublanguages, ambiguity, and sublanguage grammaticality are taken up in section 4.

2. Methods of Analysis. In this section and the succeeding one, the methods and results of an investigation of the cellular immunology sublanguage are presented. The corpus consists of sixteen research articles, the earliest dating from 1935, the last published in 1970, which are concerned with the by-now resolved question as to the cellular site of antibody formation. Analysis of the regularities of word combination in these articles yields a set of word-classes specific to the sublanguage. Aside from some sentences, chiefly occurring in sections marked "Methods" (or: "Materials and Procedures"), which contained only one of

the word classes (and thus did not exhibit sentential relations among the word-class members), each text-sentence has a structure representable as particular sequences of these word-classes together with conjunctions and meta-scientific material (see sections 3.4-5) which could not on the basis of this small sample be organized further. The formulas of the sublanguage are obtained by adding to these sequences, i.e., sentence types, subscript and superscript symbols for, respectively, distributionally distinguished subclasses and modifiers. We could test the adequacy of these methods as a means of representing information in the articles by noting whether there are appropriate changes in the formulas at points where in retrospect it is known that new results, methods, and understandings were introduced.

Details of this investigation are presented in The Form of Information in Science (FIS), especially chapters one through five. The present essay on cross-reference analyzes one of the articles in this corpus and draws upon the results of earlier research in framing some of its principal hypotheses (see section 3). Some of the methods and results of this research receive a somewhat more extensive discussion than in FIS (e.g., the discussion of word classification in sections 2.22, 3.3), while others, in particular the survey of transformations employed in regularizing the texts (chapter 5, FIS) are perforce slighted.

Above it was claimed that the methods discussed here serve as a means of representing information in the sublanguage. 'Information' as a central concept in this work awaits a thorough explication.[20] As used here, the term refers to a statement of the restrictions on word-combination in a given domain (a natural language, a sublanguage of science) in which (a) each element represented in the grammar is established in respect to formal differences of environment, and (b) those restrictions on elements which do not correlate with differences of meaning are removed.[21] Various points of connection between language structure and information are addressed in the brief presentation of operator grammar in section 2.1 and again in sections 2.22 and 3.3 with respect to the immunology sublanguage.

2.1 Operator Grammar. The structure of a given natural language can be presented as a succession of constraints on combinations of its elements. Operator grammar as presented here is a particular characterization of how words combine to form sentences[22] in which the syntactic elements and operations defined in the grammar have a definite informational value. The discussion is based upon presentations of operator grammar in various publications of Harris[23] and its exemplification in the detailed grammar of English (GEMP).[24]

The word sequences which form sentences in a particular language are characterized, i.e., analyzed or derived, in

an operator grammar by means of two mechanisms. One is the composition of words in a sentence in accord with the partial order of word dependencies within each sentence -- that is, a word of a given class A occurs in a sentence given the occurrence of ordered words of particular other classes B,...,D. The word A is said to be the operator on other words, its argument, in the classes B,...,D. In:

> Margie's selling the house entails Susan's weeping

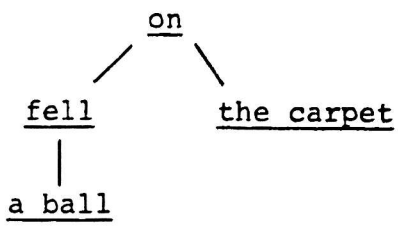
entails is an operator taking as its arguments the words sell and weep. Each of these words is an operator: sell has Margie and the house as its ordered arguments; the argument of weep is Susan. The argument requirement of a word consists of particular ordered word-sets, one member of which is requiredly a prior entry (in respect to that word) in the partial order within each sentence.[25]

Elementary arguments are those words which have a null argument-requirement -- for instance, dog, Richard, table, lamp. Those words whose argument requirement only consists of elementary arguments are elementary operators; these operators are distinguished by the number of arguments which they take. Walk and old require a single elementary argument (Philip walks, A man is old), buy requires two (Tony bought a share), and put requires three (Felix put a hat on the table). Finally, there are operators which have at least one operator in their argument demand. These non-elementary operators include such words

as likely, with a single operator as its argument (That a melt-down occurred is likely). Other non-elementary operators have as their argument an ordered pair consisting of an elementary argument and an operator -- e.g., suppose in Sally supposed that the trial would last a week, astonish in That Hildy filmed the race astonished Sylvan. Cause, as entail in the example above, has as its arguments two operators.

Distinguishing the entries in the partial order within each sentence requires careful consideration. In English sentences, there are words and affixes which are not entries in the partial order; rather they indicate that another word in the sentence is an operator or, as with the affix -s, "carry" the operator.[26] For example, that in That a melt-down occurred is likely indicates the argument status of a melt-down occurred under the operator likely (is is a carrier of that operator). In That fell, that is an elementary argument. Again, on in the sentence Sam relies on Juan is a required indicator of the argument status of Juan in respect to rely, whereas on is itself an operator in A ball fell on the carpet. Common argument indicators include (in English) that, to, -ing (GEMP, section 2.04). As entries are generally uni-morphemic (i.e, not composed of other morphemes), many other affixes, e.g., -al, -ment, mal-, are obtained by means of reductions.

The partial order among word-entries in each sentence can be presented as a semi-lattice; the sentence above can be indicated as:



As to information: an operator asserts something about, i.e., predicates something of, its arguments.

The second mechanism of sentence derivation is the system of reductions. Reductions are changes in the phonemic composition of words (rarely, they are changes in their relative position) which words have low information upon their entry into a sentence. These reductions are by and large optional; since they preserve, if only in a derived manner, the operator-argument relations in the sentences upon which they operate, the reductions are also paraphrastic. In the decomposition (analysis) of a reduced sentence, the inverse of a reduction is referred to here as a "reconstruction". Some of the illustrations of reconstruction provided are taken from (often excerpted) sentences of the immunology articles examined in FIS (citation numerals refer to Appendix I of that work).

In many reductions a word (or: word-sentence) is changed in phonemic shape to zero. Repetitionally-based zeroings involve reduction to zero of the second occurrence

of a word (-sequence) when it occupies particular positions in respect to the prior occurrence. In both lymphocytes produce antibodies and plasma cells produce antibodies, the latter occurrence of produce antibodies occupies a position parallel to the former[27] and is zeroable. The residue and plasma cells is then transposed to the first word not serving as a basis for the zeroing, yielding Both lymphocytes and plasma cells produce antibodies (from 3, 128.9.1). Another case of repetitional zeroing is the zeroing of the subject of a secondary sentence (along with is) under particular prepositional and conjunctional operators when that subject is the same as an argument in the primary sentence. Thus, the sentence Sera from non-immunized mice were negative when tested for antibodies (from 2, 297.3.6) is seen to be a reduced form of (or: is reconstructed to) Sera from non-immunized mice were negative when sera from non-immunized mice were tested....

A few words can be described as occupying unique positions in particular constructions and hence are zeroable. In the relative clause construction discussed in chapter one (section 8.1), the wh- pronoun (and the operator-indicator -s attached to the carrier be) are often zeroable: Pathogenic bacteria which were carried on the lymph stream are often arrested in the glands is reducible to Pathogenic bacteria carried on the lymph stream... (from 1, 783.1.7). If the residue of the relative clause upon zeroing which is,

etc. is an adjective or adjectival compound, the modifier can be fronted to a position before its host: Agglutinins were taken out of the blood by the inflamed nodes (from 1, 792.4.0) is a reduction of Agglutinins were taken out of the blood by the nodes which were inflamed.

In many reduced sentences, the reconstruction of a zeroed word (-sequence) can be distinguished as highly likely, i.e., as appropriate. For instance, under many prepositional operators which have an operator and an elementary argument as their respective first and second arguments, the first argument is zeroed. In Bill is on the line, the likely operator is located (or the near synonymous present).[28] Again, expect takes an elementary argument as its first argument and an operator as its second: given the reduced form I expect Edward, the likely operator is come (I expect Edward to come). A discourse context can effect the likelihood of a particular reconstruction: if the sentence above is preceded by Who posed this reformulation?, the assured reconstruction is I expect Edward posed this reformulation.[29] In the present work, it is proposed that most, if not all, of the reconstructions distinguished as appropriate to the sublanguage (cf. FIS chapter 5.4.4) can be reformulated as tacit referentials (cf. chapter 3).

Finally, as noted before, many affixes are products of reductions.[30] For instance, various suffixes in English can be described as the head of a compound: -er in baker

is a variant of one (from one who is in baking); -ly in delicately is a variant of in a way (from in a delicate way). Other affixes which yield for instance particular verbal or adjectival nominalizations, are formed from operators: -hood in childhood is a variant of state (or: condition) operating on a predicative occurrence of a child (the condition of being a child). Certain of these reductions are of especial interest in respect to the immunology and other medically related sublanguages. Thus, the prefix hyper- (as in hyperimmune) can be obtained from an adverbial modifier, e.g., more than normal; -itis in lymphangitis is a reduction of inflammation of (inflammation of lymph vessels.)

Nearly all of the reductions are optional. Some, however, are not -- for example, the required transposition of and plasma cells to a position after lymphocytes in the parallel-zeroing discussed above. That most reductions are optional (and paraphrastic) is important -- it implies that the unreduced source form is attested in the language. In respect to these required reductions, certain intermediate sentences are thus marked with a dagger. Other reconstructed sentences, particularly those established for tense and other complicated areas of English grammar (GEMP 1.5), are likewise daggered: these reconstructions are made to establish a base set of sentences (see below). Such daggered sentences, while not normally sayable, can be considered as grammatically possible. Their inclusion in the

grammar of English allows for a statement of reductions from a base set of sentences with the partial order noted above (importantly, these reconstructions satisfy the partial order constraint). Together with normal sentences, the daggered forms constitute "an extended set of 'grammatically possible sentences'" (GEMP:18).

Another constraint on word-occurrence is presented by various linearizations of the partial ordering of word-entries in a sentence. In English the operator is "conventionally" said after its first argument: bought operating on Ethel and nails (as its respective first and second arguments) is realized as Ethel bought nails. However, other linearizations are possible -- for instance, the second argument of bought above, nails, may occupy "front position": Nails Ethel bought. Similarly, the third argument of put -- the table (with I as first argument and a lamp as second) may be placed in front position along with the argument-indicator on: On the table I put a lamp. These alternative linearizations have the effect of giving focus (or: topic) status to the argument occupying front position. The possibilities of linearization in English are various and quite involved, meriting further study.[31] In the research articles, the text-sentences are, of course, already in a particular linearization; the various reorderings of words employed in regularizing the texts are re-linearizations (FIS, chapter 5.2).

The set of "grammatically possible sentences" mentioned is demarcated by the argument-requirement of words and the various linearizations of the partial order determined by them.[32] From this base subset of sentences reductions derive the remaining sentences of the language. As the reductions are paraphrastic, and for the most part optional, the base subset can be said to contain all the information contained in the whole set of sentences. A combination of two sentences of the base, e.g., by conjunction with and, yields again a base-sentence; hence the base subset of sentences is a sublanguage (cf. section 1.22) of an extended English, i.e., the 'normal' sentences of the language along with some daggered ones.

Each of the elements and operations defined in an operator grammar has a precise informational value. The relation of operator to argument is a predication. Reductions are paraphrastic and it may be conjectured that the varied likelihoods of particular reconstructions serve to differentiate the particular operators and arguments.[33] Linearizations of the partial order in a given sentence may introduce shifts in topic or focus but do not otherwise alter the "substantive" information in that sentence.

As Harris has stressed,[34] the grammatical description of a particular natural language must of necessity be couched either in terms of that language which is described or another natural language (e.g., a grammar of Danish in English). In the first case, the metalanguage, i.e.,

grammar, is a sublanguage of that language.[35] Not all combinations of elements occur in a language; the non-occurrence of particular combinations enables one to construct linguistic elements in respect to restrictions on their co-occurrence.[36] Finally, the information carried in the language is characterizable, not by recourse to some abstract structures for which there is no clear evidence, but in terms of those various restrictions on word combination stated in the grammar.

2.2 Methods of Discourse- and Sublanguage-Analysis. The (partial) grammar of a sublanguage presented in FIS is a description of restrictions on combinations of word-occurrences within a restricted subject-matter, i.e., a set of discourses concerned with the site of antibody formation. A characterization of these restrictions directly -- in terms of the actual text sentences -- would prove a complicated task due to the often dissimilar environments of particular word-occurrences. Many of these restrictions on combination can be eliminated by setting the sentences of a discourse into maximal similarity with one another, i.e., regularizing the text by transformations (reductions and their inverses). In the regularized texts, occurrence-restrictions are stateable in terms of word-classes of word- or word-sequence occurrences which are positionally similar to one another. The grammar is a statement of these word-classes and their combinations.

The regularizing operations do not correlate with differences in meaning -- that is, they are paraphrastic. The elements of the sublanguage grammar presented in section 3 -- the word-classes and sentence-types -- are seen in FIS to correlate with changes in the results, discussions, and methods introduced in the course of the articles, and so are claimed as an informational representation of the articles. In section 2.21, the course of analysis is described in some detail. The following subsection focuses on particular aspects of these methods, e.g., the construction of word-classes.

2.21 Course of the Analysis. The articles described in FIS were analyzed in order of their appearance in various journals. For each text-sentence within an individual article, its subject, main operator, complements, and modifiers are identified. In many, the effects of particular reductions are undone. Thus, to use an earlier example, Both lymphocytes and plasma cells produce antibodies is reconstructed to Both lymphocytes produce antibodies and plasma cells produce antibodies. Inasmuch as these reconstructions are used to set the sentences of a discourse into maximal similarity with one another, only a partial decomposition is needed. In the example it suffices to note that lymphocytes is the subject of produce and antibodies its complement (and likewise with the second sentence of the reconstruction). The sentence is not further decomposed in respect to, e.g., the plural or

the modifier plasma on cells. With this proviso, produce is taken as an operator -- in its first occurrence, with lymphocytes and antibodies as its first (subject) and second (object) arguments respectively; in its second, with plasma cells and antibodies as arguments.

The discourse property of word repetition noted before leads to the establishment of word-classes: within the operator-argument relations determined by the gross grammatical analysis, repeating combinations of word-occurrences are sought. The word-classes are constructed so that there is frequent recurrence of members of one class with those of another. For instance, as subjects of text-sentences with the main operator is injected into, occurrences of such terms as diphtheria toxin, an antigen, and influenzal virus are found. These occurrences are tentatively grouped into a word-class G. As complements of the operator, there are such terms as the rabbit, mice (forming a word-class B). Whether G (and B) are useful word-classes in the analysis is tested by noting whether other occurrences of these terms can be set into the same grammatical relations as those which form the tentative word-classes by recognized transformations. The text-sentence Both legs were injected with the same antigen (5,205.1.7) for instance, can be passivized, yielding the same antigen was injected into both legs. In the transformed sentence, the same antigen is subject of was injected into and both legs is the complement. If such

transformations are available which so align the word-classes, G and B are set up as word-classes, defined intensionally as subjects and complements of the operator injected. On the basis of its various subjects and complements, operators (e.g., injected) are assigned to sublanguage operator classes (J).

Several points in the analysis merit further notice. Firstly, the establishment of the word-classes yields at the same time a sentence structure, e.g., GJB, insamuch as these classes are defined on the basis of the occurrence of their members in particular sentence-forming operator-argument relations. Secondly, the classification is made on word-occurrences. Thus, strictly speaking, it is not diphtheria toxin for instance that is assigned to G, but rather a particular occurrence of it. With all of the word-classes, the argument word-classes in particular (e.g., G, B), we try to avoid instances of class-cleavage, e.g., the assignment of different occurrences of diphtheria toxin to differend word-classes (see section 3 on homonymities in the operator classes and subclasses). Given the definition of these word-classes -- for instance, of G as the subject of J (was injected into, etc.), occurrences of other words are assigned to these classes on the basis of their environment (perhaps altered as the result of a transformation). For example, B. enteritidis suspension is assigned to G in the text sentence In the right ears 0.02 cc of B. enteritidis was injected;

the right ears is assigned to B via a permutation of the initial PN phrase.[37] Finally, as will be discussed in greater detail below, alternative word-classes and sentence-structures as well as alternative definitions of much the same word-classes can be found. For instance, B could be defined as the subject of injected with on the basis of sentences such as Rabbits were intravenously injected with killed cholera spirilla (from 1.801.4.1); use of different transformations could lead to setting up a sentence type BJG.

Other word-classes can be formed in a similar manner. As subjects of is present in, is contained in, is synthesized by, occurrences of words such as agglutinin, antibody are noted and are assigned to the class A. The objects of these verbs include lymph, lymphnodes, spleen and serum which are grouped in a "tissue" word-class T. If these verbs are then classed as V, one obtains, given a sentence Antibody is present in efferent lymph, an instance of the sentence structure AVT (from 7,2.2.2). In the text-sentence which contains pathogenic bacteria are carried on the lymph stream (1,783.1.7), the subject pathogenic bacteria is of word-class G; the lymph stream is in T. The operator is carried on with subject G and complement T is assigned to a new word-class U (a superscript y is attached to indicate the preposition which introduces the complement. The sentence structure,

established in respect to these word-classes with members in particular operator argument-relations, is $GU^{Y}T$.

Given an indication of the recurring word-class combinations, each text-sentence is segmented in such a way that (1) each segment is grammatically a component sentence of the text-sentence and (2) the segment is composed of one of these word-class combinations. Transformational reconstructions are often employed in this task. For instance, the sentence pathogenic bacteria are carried on the lymph stream is a component of a larger sentence -- Pathogenic bacteria carried on the lymph stream are often arrested in the glands.... The component-sentence is indicated as a secondary-sentence on the primary (GUT) sentence pathogenic bacteria are often arrested in the glands; the host (symbol) of the secondary sentence receives a w-superscript to indicate the attached relative clause. The sequence is written $G^{W}UT$, $GU^{Y}T$.

Within each segment transformations are also used so that the various word-classes in different segments are aligned with one another. For instance, the segment (1,891.2.1) the rapid lymphatic distribution of antigen contains -- in order of occurrence -- words in the classes T, U, and G. Denominalization yields the lymphatic (system) rapidly distributes antigen. This sentence can then be passivized, resulting in the GUT sequence Antigen is distributed rapidly by (or: along) the lymphatic system.

The sentence Injection of horse serum was intravenous can also be denominalized into the GJB sequence horse serum was injected intravenously.

Some of the work of standard transformations, e.g., passive, denominalizations, can be achieved by re-linearizations, which may not involve the complex statements of domain presented by certain reductions. For example, rather than denominalize the segment the formation of antibody in lymphnodes (from 6,157.1.1) to obtain a sentence of the type AVT, the segment can be re-linearized (cf. section 2.1) to of antibody the formation in lymphnodes. In other sentence-segments, the effect of transformations, i.e., in aligning the sentences into a standard order of word-class members, can be obtained by use of an arrow which indicates the order in which the sentence is to be read. For example, the sentence Lymphocytes contain antibody contains in order of occurrence a member of the "cell" (C) word-class lymphocytes followed by a member of V (contain) and the A (antibody). This sentence can be aligned with others of the type AVC by passivizing it. Alternatively, it can be written: antibody/ contain/ lymphocytes ← where the word-class members appear in the prescribed order AVC. The arrow here indicates the order of recitation from right to left; the single slashes separate word-class members.

The result of these methods -- the regularized texts with a statement of word-classes and sentence-types -- is

not given by some discovery procedure. As word-classes are defined in respect to one another, the starting point of the analysis is perforce somewhat arbitrary; the alternative is to assume that all of the word-classes are established simultaneously (for a statement of the same point, see Harris 1951:7).[38] This allows for differing initial classifications which can be revised as other regularities in combinations of word-occurrences are established. Other classifications will entail the use of different transformations in regularizing the texts: the "cost" (that is, complications) of these alternative regularizations will vary.

In addition, the analysis of text-sentences is subject to several objectives which may in certain instances compete with one another, permitting alternative analyses of sentences with differing "costs". One obvious objective, noted before, is avoidance of class-cleavage, particularly in respect to the argument word-classes. Further, in the text-sentences we seek (1) repeating sequences of word-class combinations such that (2) each sequence covers as much of the text-sentence as possible. In FIS (chapter 1:21-22), Harris discusses one instance in which these objectives are at odds. The text sentence (1,796.4.3) contains a component which can be reconstructed to <u>The lymphnodes from the side injected with that antigen and the nodes from the other side were equally inflamed</u>. This

sentence is representable as a case of a TYT sentence type if the reciprocal (reflexive) status of equal is used to transform the sentence to The lymphnodes from the side injected with that antigen were inflamed equally with the nodes.... Establishing a TYT sentence type houses the entire segment within a single word-class combination and would be favored by the second objective cited.[39] However, the analysis would also entail a redefinition of the word-class Y which generally takes as its argument the pair of word-classes C, C. Moreover, there are other occurrences of the lymphnodes were inflamed, analyzed as a combination of the T word-class and the W word-class. The TYT analysis would thus be exceptional. Transforming the segment so that equally as appears as a conjunction between two TW- sentences is in line with the first objective and in FIS is the analysis chosen.

In the following text-sentence, slightly altered to focus on the point at issue, an analysis in line with the second objective is chosen:

> We therefore consider it improbable that the lymphocytes present in the fat of the renal sinus...would give rise to the antibody protein concentration of the extract of this tissue. (3,128.3.2)

The sequence We therefore consider it improbable that is assigned to a meta-scientific word-class (see section 3.4). In terms of the recognized word-classes, two analyses of the remainder of the sentence are available. In one,

would give rise to appears as a conjunction, not assigned to a word-class (cf. section 3.4), between a sentence of the type CWT -- (C) lymphocytes (W) are present in (T) the fat of the renal sinus -- and one of the type AVT (antibody protein concentration of the extract of this tissue). An alternative analysis would treat would give rise to as the main operator of an AVT sentence -- the lymphocytes would give rise to antibody protein, with lymphocytes are present in the fat of the renal sinus and antibody protein (appears in) concentration of the extract of this tissue as appended (reconstructed) secondary sentences of the respective types CWT and AVT. The latter analysis involves reconstructions not entailed by the former. However, it is in line with both objectives and so is preferred (cf. section 2.2.2 below on other objectives).

Formulas of the sublanguage are obtained by a further classification of words occurring in the segments presented by the sublanguage sentence-types, e.g., AVT, AVC, GJB, GUT. Subclasses of the word-classes are indicated as subscripts attached to the word-class symbols. They are identified, by and large, by noting possibilities of word-combination within the word-classes. For instance, lymphnodes occurs as the subject of the intransitive W-operator were enlarged; lymph, on the other hand, does not. Lymph (and: lymph stream) occurs as the complement of the verb are carried (on) unlike lymphnodes. Such restrictions on combination establish lymphnodes as T_n, lymph as T_l, were

enlarged as W_g, etc. Detailed subclassification yields subclasses whose members are synonymous in respect to the sublanguage, e.g., lymphnodes and lymph glands as T_n, synthesize and produce as V_p (synonymy in the sublanguage is discussed further in sections 3.3 and 4.4).

What in FIS are termed "local operator modifiers" (chapter 4.2) are distinguished as well by the environments in which they occur. These modifiers are indicated as superscripts adjoined (to the right) of the noun or operator word-class (symbols) upon which they operate. Modifiers on noun word-class members, e.g., large on cells, also occur as members of the W word-class when they are the main operator in a segment, e.g., the lymphocytes are large is indicated as C_yW_g. If another operator is the main operator, e.g., the large lymphocytes are found in the spleen, the modifier (large) is indicated as a superscript (C_y^g). Another superscript on these categories, w- for an attached secondary sentence (relative clause) was noted above.

Modifiers of operator word-class members include the following: (a) indicators of quantity, e.g., increase, decrease, (b) operators upon these indicators of quantity, e.g., comparative forms, (c) prepositions which mark argument requirements for operator subclasses, e.g., 'y' above for on in are carried on. Included here as well are various aspectual operators, which do not introduce a new subject into the sentence in which they occur, e.g.,

the lymphocytes began to proliferate in significant numbers (6,164.4.1). The important 'r' superscript (is responsible for, constitutes a factor in, etc.) is discussed in FIS, chapter 4.2, 4.7.

As a result of these procedures a set of word-classes, subclasses, and sentence types is obtained. The (transformed) sentences of the text can then be mapped onto the closed set of word-class symbol sequences (FIS, chapter 4.1). The formulas comprise the sentence types together with subscripts indicating subclasses and superscripts indicating modifiers, and can themselves be read as sentences: $AV_pC_y^q$ as "antibody is produced by large lymphocytes", T_nW_f as "the lymphnode is inflamed". The sequence of formulas along with meta-scientific segments marked 'M' (section 3.5) and conjunctions cover the succession of sentences in the texts. In conformity with the major hypothesis of this research, the different views and results within the articles are represented by appropriately differing formulas. Thus, the formulas can be said to represent the structure of information within this subscience (FIS, chapter 3.1).

2.22 Discussion of the Methods. Analyzing a sublanguage involves regularizing a text in order to establish word-classes of positionally similar terms, combinations of which recur frequently in the text. Relatedly, as much of the text-sentence as possible is brought within the confines of a single formula. Thus, modifiers on a particular word

in the text are, when possible, represented by adjuncts (superscripts) on the word-class symbol for that word, rather than reconstructed as a secondary sentence conjoined by wh-. Reconstruction of secondary sentences does, in other cases, assist in maximizing recurrence of the established sentence types (cf. the discussion of pathogenic bacteria carried on the lymph stream are often arrested in the glands). This objective also leads to the establishment of word-classes some of whose members are composed of two words, e.g., lymph stream, lymph glands. The inclusion of pluri-word segments within a single word-class is supported by the fact that these pluri-word members in some instances have synonyms consisting of a single word, cf. lymph, lymphnodes (section 3.3) and in some articles are treated as single units by means of abbreviations, e.g., the use of "pl.c." for plasma cells, of "SRBC" for sheep red blood cells. More importantly, in respect to the transformations, these members are transposed or otherwise altered as a unit.

An objective of the analysis, not mentioned above, is that the formulas be informationally additive, i.e., that material which would otherwise alter the information stated in a formula is represented in the formula. This objective requires that various operators which indicate whether the sentence is, e.g., asserted, negated, questioned, are somehow represented in the formula for that sentence. In

only some cases has this objective been reached (principally for operators of negation). In other cases, the indications of "assertion-status" are contained in segments marked 'M' and have not been organized further (e.g., We therefore consider it improbable above).

Two aspects of the procedures deserve further consideration -- the formation of word-classes and the use of transformations:

Classification. Classification into word-classes is made, as noted before, on occurrences of words within the articles. One consequence of this, addressed in section 3.3, is that various occurrences of what in an operator grammar is considered the same word are classified in some instances into different word-classes (e.g., produce in different occurrences is identified as belonging to V, W, and the colon word-class). Word-occurrences are grouped into classes on the basis of their recurrence in particular word combinations. Certain word-sequences, particularly adjuncts which often comment on methods, do not recur, and are therefore not assigned to any word-class, e.g., bled to death in ...tissues rich in plasma cells from highly immunized animals bled to death (from 3,122.4.1). Within a particular article there may not be sufficient recurrence of particular combinations of words to establish clear-cut word-classes. However, over the course of the articles, these word-classes become "entrenched" (the term is

Goodman's) with additional repetitions of these combinations. In other cases, the assignment of word-occurrences into sublanguage word-classes and subclasses may be revised as other combinations of words are encountered. This means that, overall, classification of word-occurrences is made on the basis of the entirety of word-combinations within the articles. It may be of some interest to compare this classification with one made either for each individual article considered separately, i.e., a discourse analysis of each article, or one made for the articles taken cumulatively, i.e., in chronological order (the analysis made here is closest to the latter but for the revisions). The different classifications and "rates of entrenchment" might themselves prove instructive as to the course of change in a research area. Within particular articles, there are differences in word-combinations which would distinguish, for example, different antigens as subclasses of the word-class G. However, in respect to all the combinations in the various articles, such subclasses cannot be established. Finally, it should again be noted that even within the procedure sketched above, alternative classifications are possible, with differing "costs".

Transformations. Transformations (reductions and their inverses) enter at several points into the course of the analysis. Initially they are employed in identifying within each text-sentence the main operator and its argu-

ments. Given a tentative classification of word-occurrences into some class, transformations are used to test this classification by noting whether in terms of these transformations one can alter word-occurrences into specified grammatical relations with one another. Finally, given a statement of the word-classes and sentence-types, some text-sentences (or segments of them) are further transformed to set members of the word-classes contained in them in alignment with those in other sentences or segments.

Two final notes on transformations. The first concerns the fact that a good number of the text-sentences have two or more readings, and thus, different reconstructions can be offered for these sentences. Some of these reconstructions, however, can be eliminated from consideration in line with the objectives of obtaining frequent recurrence of sentence types, each of which encases as much of the text-sentence as possible. For example, the sentence -- This experiment served to demonstrate the early appearance of agglutinins in the regional lymph nodes and serum (from 1,789.4.1) -- is ambiguous. Reconstruction of the sentence into segments with lymph serum or regional serum would yield word combinations which do not otherwise occur in the articles. In order that as much of the text-sentence as possible is included within a formula, the sentence, under the meta-scientific operator This experiment served to demonstrate, is reconstructed to This experiment served to demonstrate the early appearance of agglutinins

in the regional lymphnodes and the early appearance of agglutinins in serum. That the source of ambiguous sentences is often that which yields the largest repeating sentence types is confirmed by the judgments of "speakers" of the sublanguage who were consulted and itself confirms the adequacy of these methods as a means of representing information in the subscience (these methods, as opposed to informal -- and thus non-controllable -- judgments of author's intent). In other cases, the reconstruction is readily discerned by examination of neighboring sentences, as well as consideration of the reconstruction's similarities with other sentences over the bulk of the articles. For instance, article 6,164.3.2 contains the segment thereafter mature plasma cells diminished rapidly which is ambiguous considered in isolation: the diminishment may be in the size or the number of the cells. The preceding sentences indicate the latter as the source: Mature plasma cells began to appear in large numbers only on the 4th day; they were the predominating cells on the 5th and 6th days. Examination of other articles may yield instances of ambiguity which cannot be resolved in terms of the criteria noted above. In such cases, interpretation of the sentences may require, e.g., the elicitation of sentences which are assumed in the science.

Secondly, the segmentation of text-sentences into components and word-class members may be considered the

resultant of a battery of paraphrastic transformations in the sense of Hiż (1961, 1964: 101-102).[40] That is, the transformations yielding the regularized text-sentences impose as well a structure on the sentences which conforms to (i.e., corroborates) the segmentation based on word occurrences. Within the domain of a sublanguage, there is the opportunity to test the conjecture (of Hiż, see 1961:49) that each (sublanguage) word is distinguishable by a unique battery of transformations.[41]

3. Results and Discussion. The results of the investigation whose procedures were reported in section 2 are presented in the tables of Appendix I in FIS and are discussed in chapters 3 and 4 of that work. Results of a parallel effort in French for research articles on the same problem occupy the second appendix and are discussed in chapter 7 of FIS.[42] The tables in these appendices give, for each of the articles, analyses of those sentences carrying the main argument of the article: together with the actual text-sentence, the regularized transform and its formulas are displayed. In presenting these results here, a selection obviously had to be made -- I present the major word-classes and sentence types of the sublanguage along with results which will prove pertinent to a description of cross-reference in the "Influenzal Antibodies" article. Following presentations in various subsections, I discuss connections

to the present essay on cross-reference. Other implications of these results, of general linguistic import, are discussed in section 4.

3.1 Word-classes. Listed below are, in order, the major argument and operator word-classes of the sublanguages. Next to each word-class, some of its members are given together with various subclass designations where applicable.

Argument Word-Classes:

G antigen, diphtheria toxin, influenzal virus, disease (G_f)

A antibody, agglutinin, protein (A_p)

T blood (T_b), lymphnodes (T_n), lymph (T_l), spleen (T_s)

C lymphocytes (C_y), plasma cells (C_z), macrophage (C_m)

S cytoplasm (S_c), ribosomes (S_b)

B rabbit, mice[43]

Operator Word-Classes:

J inject, administer

U stimulate, uptake by (U_s^t), found in (U_i)

V formed by (V_p), appear in (V_i), secretion by (V_s), absorbed to (V_u^t)

W change (W_c), large (W_g), mature (W_m), flows (W_u), inflamed (W_f)

Y were typically, derived from (Y_c^f)

Some members of these word-classes consist of more than one word: this, as noted before, is the result of the structure

imposed on the sentences by the regularizing operations. In the present essay it is hypothesized that at least each of the major argument-classes in the "Influenzal" article has a referential classifier. More precisely, the hypothesis states that there are cross-referential relations (as defined in chapter 1) in which the referential classifies its referend as well as all other members of the word-class to which the referend belongs, e.g., that an occurrence of lymphocytes is cross-referred to by the cells, and that other members of C (plasma cells, reticulum cells) are classified by cell: Plasma cells are cells (chapter 3, sections 3.21-22 provides a detailed discussion).[44]

3.2 Sentence Types. Regularization over the set of articles yield sets of partially similar sentence types. In the listing below, parentheses around a particular word-class symbol indicate that there are text-segments in which a member of that class does not appear. Each of the sentence types is followed by a sample sentence of that form.[45]

GJ(B)	Schick test toxin was injected at the same time intradermally.
GU(^{f}T)T(^{y}T)	Pathogenic bacteria are often arrested in the glands.
AV(C)	Antibody is concentrated chiefly within lymphocytes.
AV(T)	Antibodies had been formed in the right lymphnodes.

CW(T) Plasma cell proliferation becomes much more intensive in the tissues.

TW(T)(B) The lymphnodes were found inflamed.

Sentence types which have a V, W, or Y_c operator are termed "response sentences" (chapters 2.6, 4.3, FIS). In many occurrences these response sentences are conjoined with sentences of the types GJB or GUT. These conjunctions form a word-class designated by colon (':'), e.g., following, after, produce. Such occurrences can be represented by a macro-sentence type, i.e., GJB: response sentence:

> On immunization with several antigens simultaneously, the concentration of antibody in the blood becomes much higher.

Occurrences of these sentence types may have referential classifiers as well, e.g., these findings/observations (cf. chapter 5, sections 3-6). One can substantially reduce the number of sentence types of the sublanguage if one considers, for example, an occurrence of antibodies are present in large numbers (AV) as an incomplete AVT or AVC sentence type. Such a reduction is proposed in the present essay by expanding incomplete sentences through tacit referentials, e.g., the sentence above could be expanded to antibodies are present in large numbers in the cells, with the cells a referential. The possibility of this reduction depends on (1) the regular occurrence of instances

of the full sentence types and (2) the hypothesis above that referential classifiers are available for the major word-classes (see chapter 3, section 2 for a thorough discussion).

3.3 Subclassification/Synonymy. A large number of the subclasses only have a single member, e.g., liver (T_v). Several subclasses have a few or many members which are in this sublanguage synonyms of one another. A sample follows.

T_b blood, serum, vascular[46]

T_n lymphnodes, glands

T_ℓ lymph stream, lymph[47]

V_i present in, contained in, appears in

V_p produced by, synthesized by, formed by

V_u seep through, drain from

Y_c^f descend from, originate from, arise from

U_i is arrested by, is held by

W_p formation, -poetic

Synonymy relations also obtain among members of local operator modifier classes:

\- few, little, low

b begin, start off, induction

l primary, single, sensitizing

c changing, developing, differentiating

r have a role in, participate in, is concerned with, is responsible for

(for a more thorough listing, see FIS, pp. 55-66). The consequences of this result are many-fold: section 4.5 dis-

cusses some of these in detail. Here it can be noted that examination of cross-references provides additional evidence for some of these synonymy relations, cf. 7, 1.5.1-2.

> McMaster and Hudack gave unequivocal evidence of the direct importance of lymphoid tissue in antibody formation. After subcutaneous injection of the antigen it was mainly the lymphnodes that were responsible for this production.

where this production refers anaphorically to the occurrence of antibody formation.[48]

Harris notes another relation -- "dependent synonymy" -- where a particular subclass member occurs given the occurrence of a particular operator or argument subclass; for example, coated occurs in V_i with protein (A_p).

The restricted selection of words in a sublanguage also permits consideration of factorizing single words into synonymous word-sequences. This has been attempted to a limited extent in the study of this subscience: is free of can be factored into does not contain ($W_i^{\sim}$), proliferated into are produced in large numbers (W_p^{+}), cf. FIS, chapter 5.9.

There are as well a fair number of homonymities in the sublanguage -- that is, instances in which the same English word (in respect to the operator grammar of the language)[49] appears in two or more classes. As noted (section 2.2), this situation is a consequence of classification in respect to word-occurrences in a regularized text. For example, immune and morphologically-related

forms (immunize, immunization) appear in both J and A; produce sometimes occurs in V_p, in other occurrences it is a member of W_p, Y_c, and the colon. Other instances of homonymity are presented in FIS, chapters 3 and 4.5 [50] (see section 4.5 of the present chapter for more discussion).

3.4 Open Questions. Noted below are a few areas which require further research.

(1) Beyond those word (-sequences) grouped into the conjunctional colon word class, e.g., after, following, produce, cause, other occurrences of conjunctions and conjunctional verbs (here including points to, suggests, indicates) have not been, and perhaps cannot be, organized into word-classes. Pending examination of a larger corpus, they may be considered operators of English on sentence-segments represented by the formulas, and on meta-scientific segments (section 3.5, 4.1). In mathematics, sequences of sentences are subject to well-formedness conditions as is seen in proofs. If certain of the conjunctions can be organized into word-classes and subclasses, this will likely prove useful in characterizing argumentation in the articles.

(2) As with the conjunctions, so too word-sequences expressing quantity, e.g., cell-counts, measurements of volume, of time, could not be grouped into closed word-classes. Such terms may be considered as part of an assumed (prior) science -- arithmetic.

(3) In work on this sublanguage, we excluded from further consideration those sentences which were found to contain only one (and in some instances none) of the sublanguage word-classes (this class was generally an argument word-class) insofar as these sentences did not yield sentential relations among word-classes. While such sentences included the majority of those in "Materials and Methods" (or "Procedure") sections of the articles, this is not an arbitrary exclusion: stray occurrences of sublanguage sentence-types are found in these sections: conversely, other sections of the articles such as "Results" contain sentences of this sort. There are apparently connections between sublanguage sentences and those containing a member of only one sublanguage word-class. For instance, the 'procedural' operator were extracted is evidently related to the (superscripted) modifier of various members of T, e.g., lymphnode- extracts. Other 'procedural' operators may be related to measurements, cf. (2). A description of these connections would prove instructive as to relations of laboratory techniques and results in the immunology subscience.

3.5 Meta-scientific Segments.[51] Determination of the sublanguage sentence types yields, as a residue of the segmentation imposed by the regularizing operations, portions of sentences in the text which are here identified as "meta-scientific". These portions, marked "M" in the

tables of FIS, may be loosely characterized as appertaining to the investigators' epistemic relation to results, etc. reported by the sublanguage sentences (occasionally referred to as "science sentences"). The M-segments can be delimited in many cases as being operators on the science sentences (which are their first or second arguments).

Included in M are operators which have as their object a science sentence whose subject is not identical with that of the operator. These include: assume, note, state, find, contend that, demonstrate that, hold that, expect, etc. Subjects of these operators include particular investigators' names. These subjects also occur with another group of operators, whose object is a member of an argument word-class of the sublanguage, e.g., excise, use, analyze, examine (these verbs can be noted as M').[52] Finally, there are operators which have a science sentence as their only argument: these (M'') include is likely, is probable, is a fact (theory, problem), emerges, results. Certain verbs, termed conjunctional above (section 3.4 (1)) might also be considered meta-scientific.[53]

"Meta-scientific" as used here should be clearly distinguished from metalanguage and from such metalinguistic referentials as the latter, the second sentence above. The operators noted above do not belong to statements about word combinations in the sublanguage, as would be the case

for "Antibody" does not occur as the subject of the operator "is enlarged". Further remarks relevant to this distinction are in Hiż, 1982.

While some preliminary distinctions concerning M can be made, at present these segments cannot be collected into sublanguage word-classes. It may be that they are part of a meta-science common to several science sublanguages. One might expect to find different M segments in the harder sciences, e.g., where particular results are assigned probabilities.

It is of some interest to see whether, and if so, to what extent M and the science-sentences can be separated. In a number of text-sentences, operators of a meta-scientific character have been retained in the science-sentences, e.g., demonstrable as in A significant amount of antibody protein was demonstrable in the nodes. These operators may be extracted -- It was demonstrated that a significant amount of antibody protein is present in the nodes -- and It was demonstrated that is then assigned to M (that was demonstratable can be considered a variant of was demonstrated requires eliciting judgments of paraphrase from "speakers" of the sublanguage). The science-sentence fact is put into the present tense.[54] Other sentences may contain segments for which there are no precise criteria for extraction; judgments of scientists would have to be used, at least provisionally.[55]

In chapter 5, section 6.3, I have attempted to 'translate' a number of text-sentences of the article into a form where meta-science segments are separated from the science sentences and the science segments can be read independently of these segments. This aim has been reached only in part: the section supplies the "rules of translation" (some of these are transformations) and comments on some of the difficulties confronted. Some of these can be circumvented if the translation is made after resolution of particular cross-references in the article.

The examination of cross-reference in the "Influenzal" article tests this division of texts into meta-science and science sentences. It might be expected that referentials in meta-science segments have referends both in other segments and in science sentences (the latter inasmuch as these meta-science segments operate on science-sentences). In this work, it was supposed that referentials in science sentences have their referends only in other science-sentences of the article (cf. section 4.1). The results, reported in chapter 5, section 6.2, support this conjecture only in part.

4. Specifying the Immunology Sublanguage: Issues and Implications. In section 1.2, a scientific sublanguage was "initially" and externally, i.e., extra-grammatically, characterized as a set of articles concerned with a particular research problem (or: theory); in the present case,

the cellular site of antibody formation. This section attends to issues associated with a grammatical characterization of the sublanguage of immunology. The previous section sketched a partial grammar of the sublanguage -- a set of word-classes, subclasses, sentence types. At the present stage of investigation, the grammar is but a partial one -- as noted in section 3.4, there remain questions concerning the description of meta-scientific material, conjunctions, and expressions of quantity. Pending the resolution of these questions, there are options in the specification of the sublanguage; these are surveyed and discussed in section 4.1. Section 4.2 examines in what sense the sublanguage can be considered to extend beyond particular portions of the articles and their regularized, i.e., transformed, counterparts. The present essay on cross-reference also involves an extension of the sublanguage by way of implicit sentences and incorporates a hypothesis as to its closure in respect to resolution of cross-references (section 4.3). Section 4.4 addresses the issue of what is ungrammatical in regard to the sublanguage grammar. A discussion of synonymy and ambiguity in the immunology sublanguage is presented in section 4.5.

4.1 Delimiting the Sublanguage. The grammar sketched in section 3 does not describe the entirety of the research articles which form the corpus. Notably it separates out those text-sentences which were found to contain only one

member (in some instances, no member) of the word-classes; such sentences obviously do not exhibit sentential relations among the word-class members. These sentences generally occur in sections of the articles entitled "Materials and Methods" or "Procedures". It should again be noted that this is not an arbitrary exclusion: stray sentences conforming to sentence types of the grammar are found in these sections; conversely, "methods sentences" are encountered in, e.g., sections describing or discussing results. This situation is not unusual in descriptive linguistics -- having established regularities in a corpus of linguistic material, particular fragments can be identified which do not exhibit these patterns, e.g., quoted material from another language (cf. Harris, 1951:375). In FIS the text-sentences found in the "Methods" section of the articles were not described; they may be supposed to comprise a sublanguage of laboratory procedures. There are clearly connections between this "laboratory sublanguage" and the immunology sublanguage. For instance, in the "Influenzal" article described in chapter 4, the sublanguage modifier 'x' for extracts is evidently related to the 'procedural' operator were extracted; modifiers relating to quantity in the sublanguage may be related to measurement procedures; the occurrence of the term lymph-plasma (and its synonyms) is preceded, it appears, by a sentence describing centrifugation of the lymph.[56]

It is important to note that the basis for excluding "methods" sentences from the sublanguage is the absence of regularities characteristic of the remainder of the articles. Otherwise, one might, informally, specify the sublanguage as occurrence in the texts with closure under the regularizing operations noted in sections 1 and 2. As mentioned above, the regularized texts are described in terms of sentence-types (more precisely, formulas), e.g., AVC, GUT, CWT, occurring under meta-scientific operators as well as conjunctional and quantificational operators which have not, and perhaps cannot be established as particular word-classes. This situation presents an option in respect to demarcating the sublanguage -- one may include the meta-science operators, the conjunctions, and quantificational operators in the sublanguage or exclude them.

The latter possibility would restrict the sublanguage to instances of the sentence-types ("science-language") sentences). These sentences could then be taken as closed under particular operators, e.g., and, or, not. Meta-scientific operators and conjunctions are considered to form higher order languages; arguments of these operators would be various referential pro-forms (usually pro-sentential) to the science-language sentences.[57] Given a sentence such as The authors demonstrated that antibody is found in the lymphnodes, one would form a meta-science sentence -- The authors demonstrated this, with this

referential to the science-language sentence, Antibody is found in the lymphnodes. To delimit the sublanguage along these lines involves the following considerations.

(1) Under various meta-scientific and conjunctional operators, a number of science-language sentences may be deformed, i.e., nominalized. To establish these sentences as self-standing (grammatically independent) will thus in many cases entail further regularization. For instance, the underlined portion of the following sentence would have to be denominalized:

> Ehrich and Harris have demonstrated high titers of antibody in the regional lymphnodes after injection of antigen. (3,121.7.1)

One could then establish a sentence of the sublanguage -- e.g., High titers of antibody are present in the regional lymphnodes after injection of antigen. (are present is obtained as a reconstruction of an appropriate zeroing) and a higher-order meta-scientific one -- e.g., Ehrich and Harris have demonstrated this. (cf. item 3 below). Another issue which arises here is the need to establish criteria for meta-science" material so that various words, e.g., demonstrable (cf. section 3.4) can be extracted from the science-language sentences.

(2) Even if science-language sentences can be established as grammatically independent of M-operators and conjunctions, these sentences would not in all cases be

"informationally independent" of them. That is, under various M-operators and conjunctions, the science-language sentence is in various ways negated or stated to be likely, improbable, etc., e.g., the negation nothing has emerged which speaks directly in favour of in On the other hand, nothing has emerged which speaks directly in favour of the participation of the lymphocytes in the formation of antibodies (4,121.4.1). The science-language sentence alone would in such cases inaccurately render what is stated in the text-sentence.

(3) This option presumes either that there are no cross-references from the science-language sentences to material in M, or that such cross-references, if any, have been resolved. The present work tests the hypothesis that in the "Influenzal" article there are no cross-references of this kind. As is noted in chapter 5, section 6.2, this hypothesis does not quite hold; this entails that demarcating the sublanguage along the lines suggested requires the prior resolution of cross-references.

(4) Finally, it is not clear in what way quantificational operators can be separated out from the science-language sentences (see section 4.2 for some further considerations).[58]

The other possibility is to include these various operators -- meta-scientific, conjunctional, and quantificational -- in the sublanguage. As noted above, the

grammar presented in FIS does not characterize these; a larger body of articles is needed to establish whether any of these forms can be organized into word-classes. At the present stage of research, the most cautious course would be not to definitively decide between these possibilities. The hypothesis noted in the third item above obviously is a consideration in deciding among them.

4.2 Extending the Sublanguage. The grammar of the sublanguage describes not only the regularized texts but also the original text-sentences (minus those of the Methods sections) in that the former are obtained by (paraphrastic) regularizing operations from the latter. Alternatively, one can state that the sublanguage, i.e., the regularized texts, is closed under these operations -- now, taken in an inverse order. The scientific sublanguage so defined conforms to the definition presented before: "proper subsets of a language...closed under some or all of the operations defined in the language".[59]

The regularized texts and their original text counterparts may be referred to as "the language of the articles". In these articles, authors forwarded differing positions as to the site of antibody formation. Some researchers claimed the lymphocyte as the site; others claimed the plasma cell; in early articles in particular, it was often assumed that these positions were mutually exclusive. Part

of the resolution of this problem consisted in establishing that lymphocytes and plasma cells are in fact different stages of a single cell line.[60] Thus, this language cannot be taken as a 'truth set' in the sense of Hiż (cf. chapter 1, section 5, and Hiż, 1969a).[61] In respect to the entry of new word-classes, subclasses, and sentence-types over the successive articles, one might speak of the language of the articles as one which is changing. However, if the corpus is viewed as a sample or a selection from a 'larger' sublanguage, these changes in sentence-types, etc. can be said to reflect differing opinions and results within that common sublanguage.[62]

To speak of the language of the articles as a 'selection' from a 'larger' sublanguage is to pose the question in what sense(s) does the grammar describe more than the language of the articles. If the description were only of this "language", the situation would be comparable to that of the philologist describing a closed text. However, the grammar of the immunology sublanguage incorporates a prediction that articles within this subfield which are added to the corpus will contain sentences analyzable with respect to the established word-classes and their combinations (with explicable extensions).[63] This is one obvious sense in which the grammar is to describe more than the language of the articles.

The language of the articles can also be augmented (i.e., extended) by specifying various closure operations on its sentences. For instance, the science-language sentences, perhaps under various meta-science operators, can be taken as closed under such operators as and, or, and not. These sentences could also be said to be closed under operators such as know, suspect, find dubious, deny, confirm, etc. -- given a science-language sentence, e.g., Antibody is concentrated chiefly within lymphocytes, the sentence The authors suspect that antibody is concentrated chiefly within lymphocytes is also within the sublanguage. Extending the sublanguage in this manner would be in line with the claim that opinions and results in the research reports can be discussed and opposed within a sublanguage common to both the reports and the discussion of them.[64]

One can also consider closure of the sublanguage under various rules of inference. For instance, closure with respect to a rule of inference which "drops" modifiers derived from appositive relative clauses (cf. chapter 1, section 8)[65] would yield, when applied to, e.g., A few scattered plasma cells were found in the retriperitoneal fat in 3 animals (from 3,125.8.1), such sublanguage sentences as -- Plasma cells were found in the retriperitoneal fat., Plasma cells were found in the retriperitoneal fat in animals. (cf. section 4.3 for related discussion.)

Specifying other closure operations requires further consideration of the word-classes and subclasses of the

grammar. The word-classes are equivalence classes of words and word-sequences which occupy grammatically similar positions in the regularized texts. Less obviously perhaps, the subclasses are equivalence classes of elements which share a particular limitation in distribution, i.e., which are restricted to occurring with particular members of another stated word-class. Thus, it appears that the sublanguage may also be augmented by closure under all substitutions of one word-class member for another in the science-language sentences[66] (see qualifications below). For example, given a sentence of the GUT sentence-type Pathogenic bacteria are carried on the lymph stream, one can substitute another member of G, e.g., S. typhi for pathogenic bacteria and obtain a sentence of the sublanguage.[67] Intersubstitutability of all members of a given word-class, however, would extend the sublanguage to a point where subclasses could no longer be discriminable distributionally. To preserve distinctions among the various subclasses would require restricting substitution in instances where a word (-sequence) is a subclass member to other members of that subclass.

Substitution of one member of a word-class (or subclass) for another such member is at present limited by two considerations. Firstly, as noted above, in the regularized texts of FIS, a "reading" arrow is used in the stead of various transformations, e.g., antibody/ produce/ plasma cells is regarded as an instance of the AVC sentence

type although it is read in the order of CVA. In such a sentence, i.e., employing an arrow as a reading device, produce is an "inverse member" of the word-class V; generally, "inverse members" of a word-class are not substitutable for their non-inverse counterparts, e.g., is produced by. Secondly, certain word-classes of the sublanguage have as members words or word-sequences which are not of the same grammatical category in English, e.g., vascular (an adjectival operator) and serum are both members of 'T_b', which again limits possible substitutions. Both of these restrictions on substitution are eliminable given further regularization of the texts. In regard to the first consideration, sentences regularized by means of the "reading" arrow can alternatively be, e.g., passivized or linearized, thus dropping the distinction between "inverse" and non-inverse members of a word-class or subclass. In regard to the second -- a regularized text-sentence such as Antibody is absorbed vascularly can be transformed to Antibody is absorbed by (or: along) the vascular system; vascular system and serum as nominal phrases are substitutable for each other.

4.3 Extensions in the Present Work. In the present investigation of cross-reference, the sublanguage, at least with respect to the "Influenzal" article, is augmented by implicit sentences needed to resolve particular referen-

tials (cf. chapter 1, section 2.4, and chapter 3, section 3.2). In many instances a hypothesis can be made as to the implicit sentence required. This is then confirmed or amended by consultation with the authors of the article. In some cases it may be more appropriate to characterize the implicit sentence not as part of the immunology sublanguage but as part of some assumed prior science. For example, to obtain a referend for the referential The enlargement of the node in the passage --

> Microscopically there was marked diffuse hyperplasia of lymphoid tissue reaching a maximum two days after the injection. The enlargement of the node was seen to be due to swelling of the cortex.... (200.3.1-2)

requires a number of implicit sentences -- among them, hyperplasia is of cells., cells are in a node. Such sentences might be described as part of histology.

The present work also examines the hypothesis that the sublanguage is closed under resolution of referentials in the "Influenzal" article. That is, the consequences or paraphrases of various science-language sentences with replacement of a referential by its referend are supposed to be analyzable in terms of the established word-class combinations of FIS. In this supposition of closure under particular rules of consequence, the present work invites comparison with Carnap's definition of sublanguage in his Logical Syntax of Language: S_2 is a sublanguage of S_1 (S_1 and S_2 are formally constructed language systems) if (1) every sentence of S_2 is a sentence of S_1, and if (2) R_2 is

a consequence-class of R_1 in S_2 (R_1, R_2 are sentential classes), it is likewise a consequence-class of R_1 in S_1 (Carnap, 1937:179). R_2 is a consequence-class of R_1 if every sentence in the former is a consequence of the latter (Ibid., 172). The first condition for a sublanguage is close to the requirement that a sublanguage be a "proper subset" of the language -- in the present case, the languages are not formal systems. The second condition posed by Carnap corresponds, albeit more loosely -- there are no ready counterparts to Carnap's sentential classes here -- to the requirement that the proper subset be closed under some or all of the operations defined in the language (cf. section 1).

4.4 Sublanguage Ungrammaticality. In English a sequence of words is ungrammatical if it is not analyzable in terms of the operator-argument relations and reductions of the grammar, e.g., Happily your down fragrant between a. Sentences which satisfy operator-argument requirements can be considered more or less acceptable: compare John fell, A chair fell, A report fell, A void fell. In Mathematical Structure of Language (1968:52), Harris suggests that the characterization of sentences as more or less acceptable is "replaceable" by noting in what discourse contexts the sentence can occur.[68] Discourse-context may be equated with membership in particular subject-matter sublanguages. Sentences which are questionably acceptable in everyday

English, e.g., _-I went to the hospital_, may occur "naturally" in some story like those of George Gamow. Thus, the constraints on word-combination stated in an operator grammar of English do not preclude the admission as grammatical sentences of word-sequences which are regarded as "nonsensical" -- these sentences can be considered to have low likelihood.[69]

Characterizing certain word-sequences as ungrammatical in the immunology sublanguage is a difficult matter. Some sentences, acceptable in English, are excluded from the sublanguage as they are composed of a different vocabulary, e.g., _No two species with identical niche requirements can continue to exist_. Other sentences, composed of the sublanguage vocabulary, e.g., _Antibody contains lymphocytes._, can perhaps be excluded inasmuch as they, taken together with other sentences of the language, e.g., _Lymphocytes contain antibody._, and general rules of inference, lead to a "collapse" of the word-categories A and C.[70] The body of sentences which comprise texts in the science of immunology, as opposed to sentences in English as a whole, must satisfy the demands of the science for coherence and warranted assertion. A patent falsehood such as _Rabbits are immortal_ is rejected by an immunologist (the sentences presented here have been checked with informants, i.e., immunologists). The negations of these sentences, however, may be acceptable within the immunology sublanguages or

some prior science and used implicitly in the course of argumentation.

Other sentences are acceptable, though only with allowances for "loose usage" or metaphor. In the immunology articles analyzed in FIS, metaphor is conspicuously absent[71] -- this points to the limited possibilities for a word to extend its selection within the sublanguage. Nonetheless, certain extensions of a word's selection do appear to lead to acceptable sentences, albeit ones with low likelihood. For example, tissue proliferates is acceptable as an instance of "loose usage" as is cells are inflamed. Actually, it is a collection of cells that proliferates and organized groups of cells that are said to be inflamed (thus, a cell is inflamed is rejected by my informants). The sentence Antibody is produced by reticulum cells is regarded as questionable -- if acceptable, it is only as elliptic for Antibody is produced by lymphoid cells of which reticulum cells are an earlier stage. If the occurrences of these words in these sentences are classed with other of their occurrences, i.e., if class-cleavage is avoided (cf. section 2.2), one obtains instances of established sentence-types:

Tissue proliferates is an instance of TW

Cells are inflamed is an instance of CW

Antibody is produced by reticulum cells is an instance of AVC

Proliferates and are inflamed can thus take as arguments words which in other of their occurrences accept other W-operators. Other members of the C word-class, e.g., plasma cells, lymphocytes, occur as second arguments of is produced by (with first argument antibody).[72]

On the other hand, sentences such as Antibody proliferates, Antibody is inflamed are rejected by immunologist-informants. If class-cleavage is to be prevented, antibody would be classed as A, proliferates and is inflamed as W. These sentences would then be instances of an AW sentence-type. This sentence-type is not encountered in the corpus and could only be accommodated within the sublanguage by revision of the definitions of well-established word-classes. Such sentences, i.e., which do not conform to the well-entrenched sentence-types of the sublanguage, may be considered ungrammatical. A more definitive answer to the question of sublanguage ungrammaticality presumes the study of a larger corpus, e.g., one amplified by elicitation from informants.

Whether sentences in a scientific sublanguage exhibit a grading of acceptability, whether certain sentences -- formed out of the sublanguage vocabulary and grammatical in English -- can be described as "sublanguage ungrammatical" -- are clearly questions which are important to the topic of the relation of sublanguages to the system of which they are part. Further investigation of these mat-

ters is pertinent as well to questions posed in other disciplines. One is the various attempts by various of the logical empiricists, e.g., Carnap, Hempel, Reichenbach, to establish a 'criterion of cognitive significance' whereby certain sentences would be excluded from scientific discussion.[73] Another is the discussion of 'appropriateness' in "pragmatics". 'Appropriateness' -- insofar as this notion does not refer to the social constraints posed by etiquette and the like -- may be characterizable as grammaticality in respect to a sublanguage.

4.5 Implications. In addition to the general result reported in FIS -- the statement of a (partial) grammar of a scientific sublanguage, there are specific results which are of interest in regard to the use of language in science, and, more broadly, to central topics in grammar. This section addresses two of these topics -- synonymy and ambiguity. Chapters 3 and 4 of FIS discuss other issues of relevance.

Synonymy. In section 3.3, instances of synonymy-relations among members of particular subclasses of the grammar were cited -- for instance, between lymphnodes and glands in T_n, between is produced by and is synthesized by in V_p. This result has a number of significant consesequences. As Harris remarks:

> The importance of this synonymy lies in the fact that the openendedness of the English vocabulary in science is only apparent and not real: an author can draw for the V_p position upon any word that

> even remotely means "to make", but in so doing he is not using the particular meaning of the word, but merely using different phoneme sequences for the one entity V_p.[74]

An earlier informal observation of Bloomfield locates the "source" of this result in the character of scientific inquiry:

> As, by convention and training, the participants in scientific discourse learn consistently to ignore all private factors of meaning, the lexical, grammatical, and stylistic features of their informal discourse become indifferent.... We say that scientific discourse is translatable, and mean by this that not only the differences between languages but, within each language, the difference between operationally equivalent wordings has no scientific effect. (1939:47)[75]

In the sublanguage analysis, synonymy among word-occurrences is a resultant of paraphrastic regularizing operations performed on sentences in a discourse. This result stands as a confirmation of transformational analysis: whatever the priority of word-meaning in learning a language is (and this bears more scrutiny than the position has generally received), transformational analysis is a description of relations between sentences (cf. MSL, sections 4.0-4.1) in respect to which meaning-relations among words is derivative.[76] One effect of these synonymy-relations is an extension of the paraphrase relation beyond sentential forms which share the same co-occurrents.[77] That is, substitution of a synonym in a sentence of the sublanguage (here taken to be what was referred to above as the "language of the articles") yields a paraphrase of

that sentence. A corollary of this in regards to ambiguity is noted below.

It may also prove instructive to compare these results with the distributionalist theory of synonymy as articulated by Hoenigswald in a number of papers, in particular in his review of John Lyons' Structural Semantics.[78] In accord with this position, two words are synonymous if in all environments they are interchangeable (Hoenigswald, 1960:18). Two familiar difficulties with this position are: (1) environments cannot be exhaustively listed over an entire language, and (2) the existence of forms, e.g., color-words and numerals, which appear to be intersubstitutable, though do not conform to what we would want to call synonymy. To the first 'difficulty' Hoenigswald responds by first noting a parallel 'difficulty' in phonemic analysis where an exhaustive listing of environments is not to be had but where domains are tentatively set up which are later justified as 'optimal'. That is, consideration of wider environments does not yield an alteration in the phoneme (combinations) established. Moreover, the 'unboundedness' of language does not prevent an in principle assessment of interchangeability given the transformation of sentences in a discourse into the kernelized form.[79] To the second 'difficulty' Hoenigswald counsels consideration of the discourse environments of sentences in which the troublesome forms (e.g., numerals)

occur: the supposition is that the wider environments of, e.g., Four plus two is eight, will differ from that of, e.g., Four plus two is six., i.e., that the environments of 'true' and 'paradoxical' sentences will differ.[80]

The status of the distributionalist theory in respect to the immunology sublanguage depends on how the sublanguage is specified (sections 4.1-4.3). If the sublanguage is restricted to the "language of the articles", then the environments of particular word-occurrences can be exhaustively listed. In this case, the first supposed difficulty does not even arise. However, few, if any, of these occurrences share exactly the same environment so that in the "language of the articles" there would presumably be no synonyms as defined by the distributionalist theory. If the sublanguage is extended to allow for all substitutions within the same word-class, then the basis upon which those subclasses which are synonym-sets are distinguished is lost (section 4.2). If substitution in the case of subclasses is restricted and closure in respect to various consequence operations is admitted, synonymy as defined by the theory may obtain. The question being raised is: under what extensions of the language of the articles can one test the distributionalist theory to see whether sufficient repetition, confirming the theory, obtains. This requires further study.

In any event, the results noted in section 3.3 give point to the concern with the discourse-environments of

words as regularized by means of transformational analysis. In a closed corpus, Hoenigswald's supposition that 'truth' is a distributionally discriminable environment (Hoenigswald 1965:192) is perhaps not an issue. It may be said that within each article in the language of the articles, the sentential environments of particular words and sentences are, if not 'true' statements, those which are warranted assertions (cf. section 4.4).

Ambiguity. Within the immunology sublanguage, and likely in all scientific sublanguages, the presence of ambiguity is considerably reduced. The 'source' forms, i.e., the regularized reconstructions, of many text-sentences which in isolation would be adjudged ambiguous appears to be readily determinable, either by examination of neighboring sentences, e.g., mature plasma cells diminished rapidly (discussed in section 2.22), or by the establishment of recurring sentence types, each of which encases as much of the text-sentence as possible (cf. section 2.2). One should note the possibility in some instances of differing analyses of text-sentences in terms of the sentence-types established in the grammar (cf. p. 141-43): this may be termed 'sublanguage ambiguity'. That the appropriate reconstructions are typically those which yield the 'largest repeating sentence-types' is confirmed by immunologist-informants and is itself evidence for the dependencies stated in the grammar of the sublanguage.

Whether all such cases of ambiguity can be settled in this manner requires detailed study (cf. a hypothesis to this effect in MSL:201-2, 212).

'Dictionary' ambiguity, i.e., homonymy, also plays a reduced role in the immunology sublanguage. Certain words which are 'ambiguous' in English as a whole (see below) have within this sublanguage only one of their senses. For instance, while a cell is interpreted differently -- though in each case definitely -- in discourses on, e.g., prison conditions, political movements, and cytology, in the immunology sublanguage -- only the 'cytological' sense is used.[81]

There is another sense in which 'dictionary' ambiguity can be reduced in this sublanguage. Hiż, in "The Role of Paraphrase in Grammar" notes the "open problem" of "whether dictionary ambiguities are reducible to grammatical ambiguities" (1964:99); "The distinction between a dictionary ambiguity and a grammatical ambiguity is made mainly by the fact that in the case of a dictionary ambiguity we give paraphrases that contain different words (dictionary definitions) from those present in the ambiguous sentence, whereas in the case of a grammatical ambiguity we come up with a paraphrase which contains only words appearing in the ambiguous sentence" (Ibid.). With the above-mentioned extension of paraphrase relations by

synonym-sets, the homonymity of, e.g., produce (cf. section 3.3) can be eliminated by choosing appropriately different synonyms for it in each of its occurrences in different subclasses of the sublanguage. For instance, produce in V_p can be taken as synthesize, in Y_c as develops into, and in colon as cause.

Harris notes that "if the symbols (rather than the words) are looked upon as the real vocabulary of the science, sufficient for its reports, then both homonymity and synonymy disappear" (FIS, chapter 3:85). The sufficiency of the symbols for the reports is supported in part by the results of the previous investigation, though problems in regards to quantifiers and "indicators of assertion" (section 3.4) require resolution before this claim can be fully substantiated. As is, the situation in the immunology sublanguage is close to that conjectured in "Report and Paraphrase": there (pp. 688-90), it is supposed that, in a language of science, one may define syntactic elements positionally, rather than as morphemes which require representation for the entire range of their occurrences.

5. Summary. This chapter provides, in an admittedly abridged form, an introduction to the study of scientific sublanguages. Following a brief description of the notions of 'discourse' and 'sublanguage' as areas of linguistic

investigation (section 1), the methods of analysis employed in studying a corpus of articles in immunology are presented and some aspects discussed in detail (section 2). Section 3 presents a sketch of the grammar of this sublanguage and notes some of the ways in which the present examination of cross-reference in the "Influenzal" article builds upon, tests, and extends results obtained in the prior investigation reported in FIS. In section 4, the important though difficult question as to grammatical specification of the immunology sublanguage is discussed. Possible extensions of the "language of the articles" are presented in sections 4.2-3. Section 4.4, addresses the questions of what is ungrammatical in the sublanguage; in section 4.5, the topics of synonymy and ambiguity are considered.

FOOTNOTES CHAPTER 2

1. A discussion of these procedures -- whereby various units of analysis are established -- may be found in Harris, Structural Linguistics (Chicago: University of Chicago Press, 1951) (hereafter cited as "SL"). In like fashion the procedures in historical descriptions are founded upon a comparison of discourses. In a comparison of earlier and later stages of a language, it is supposed that each stage has been synchronically described (cf. Hoenigswald, Language Change and Linguistic Reconstruction, especially chapter 3).

2. Harris, SL, p. 7, fn. 4, and chapter 4; Mathematical Structures of Language (MSL), chapter 3.1.

3. For this formulation, see Harris, Notes du cours de syntaxe, pp. 13-16 (Paris: Editions du Seuil); MSL, chapter 2.1.

4. Harris, SL, p. 367, and MSL, chapter 2.1.

5. Chapter 10 and 13 of SL provide examples and discussions of simultaneous components and morpheme alternants, respectively; they are discussed as cases of regularization in MSL, pp. 158-160.

6. As Harris noted in the paper "Discourse Analysis" (Paper XXIX of PSTL), this is due to the fact that "descriptive linguistics generally stops at sentence boundaries. This is not due to any prior decision. The techniques of linguistics were constructed to study any stretch of speech...but in every language it turns out that almost all the results lie within a relatively short stretch, which we may call a sentence. That is, when we can state a restriction on the occurrence of element A in respect to the occurrence of element B, it will almost always be the case that A and B are regarded as occurring within the same sentence." (p. 314, PSTL). This remark should be considered along with the stochastic procedure for marking 'sentence boundaries' noted in chapter 1, section 1.1, and fn. 1 of that chapter.

7. For the property of word-repetition under various conjunctions, see chapter 1, section 2.4.1 above. MSL, chapter 5.8, provides a more detailed discussion of word-repetition (cf. the work of Hoenigswald cited in fn. 1, p. 1 fn. 2 for a related point). Earlier discussions of discourse analysis are presented in papers XXIX-XXI of PSTL and in Harris, Discourse Analysis Reprints (The Hague: Mouton and Co., 1963).

8. Harris, "Discourse Analysis", pp. 340-42.

9. These results are reported in Harris, et al, The Form of Information in Science (to be published by D. Reidel in the series Boston Studies in the Philosophy of Science); page numbers are to the typescript version of this work, hereafter referred to as FIS. Chapter 3 and chapter 4, "Introduction", of FIS provide some preliminary discussion of the results' connection with the notion of information.

10. This formulation was suggested by the discussion of 'slants of meaning' in Hiż, "Aletheic Semantic Theory", pp. 448-50.

11. The phrase is due to Putnam in "The Meaning of Meaning", pp. 215-71, in Mind, Language, and Reality (Cambridge: Cambridge University Press, 1975). Some of the questions raised by Putnam in this paper can be approached by the study of sublanguages; in other respects, the position adopted there concerning 'word meaning' diverges sharply from the conclusions reached in research on the immunology sublanguage, cf. sections 3.3 and 4.5 below.

12. See the sample definitions of 'dialect' collected and discussed in John Reinicke, Language and Dialect in Hawaii (Honolulu: University of Hawaii Press, 1969), pp. 7-13, and the article "Dialect" of Edward Sapir, reprinted in Selected Writings of Edward Sapir, David G. Mandelbaum, ed., (Berkeley: University of California Press, 1949), pp. 83-88.

13. MSL, chapter 2.6.

14. In particular I have used examples from the Sager and Lehrberger articles in Kittredge and Lehrberger (eds.) Sublanguage (New York: DeGruyter, 1982).

15. This point is made in Harris, "New Views on Language", p. 248, in Senta Plötz (ed.), Transformationelle Analyse: die Transformations-theorie von Zellig Harris und ihre Entwicklung (Frankfurt am Main: Athenaum Verlag, 1972).

16. Harris, "Discourse and Sublanguage" in Kittredge and Lehrberger (eds.) Sublanguage.

17. Carnap, Logical Syntax of Language, especially Part VB. (cf. section 4.3 below.) Ajdukiewicz, in his article "Language and Meaning", does not speak of sublanguages and his conception of semantics differs markedly from the point of view adopted here. Still, his suggestion that for precise semantic investigations one should not speak of "one English language", but "a medical English language, a physical English language, etc." (1978:64) is in the same vein as the study of scientific sublanguages.

18. Cf. Gumperz on "linguistic range" and "degrees of compartmentalization" (p. 230) in his article "The Speech Community", in Giglioli (ed.) Language and Social Context (Harmondsworth: Penguin, 1972), pp. 219-231. For the suggestion on borrowing, see the Kittredge article in the volume noted in fn. 14.

19. The term 'language-like' is used in MSL, chapter 7.4. Language-like systems are akin to what Sapir called "transfers". Carnap (cf. fn. 17 above) provides an example of a card-index system (the card being an 'object-name' and the rider, e.g., "Lent", being a predicate), p. 6 in Logical Syntax of Language in which the symbols are movable items. Other language-like systems are noted in Goodman, Languages of Art, chapter IV, especially section 10, and chapter V, and in the Clerk Maxwell article on diagrams, Encyclopedia Britannica, vol. 8 (1910), pp. 146-49, and the Boltzmann article on models in vol. 18 (1911), pp. 638-40, cited by Goodman on p. 171.

20. For a discussion of information along the lines of this investigation, see chapters 6 and 7 of Thomas Ryckman Grammar and Information (Columbia University dissertation). A related concept of information, drawing upon results of Carnap and Tarski, is presented in several papers by Hiż, cf. Hiż, 1979, 1984.

21. "Removed" in some situations would be more accurately rendered as "transferred to domains of reductions", see below. The formulation offered is presented in MSL, p. 189. There is no suggestion here of any unique way by which redundancies are eliminable.

22. GEMP, chapter 1.

23. See, for instance, those presented in the section on Operator Grammar in Harris, Papers in Syntax (Hiż, ed.) (Dordrecht: D. Reidel, 1981), papers 12-16.

24. For a discussion of operator-grammar of French, see Maurice Gross, "Les bases empiriques de la notion de prédicat semantique", Langages, vol. 63 (1981), pp. 7-52.

25. A note on the word "requirement": certain of the elementary arguments mentioned below, e.g., dog, table, lamp, occur in the singular only with a preceding a (occurrences without a are derived as reductions of forms containing a). This requirement in GEMP is stated in terms of an automatic addition of a to these words and not as an independent entry. Thus, for instance, not as part of the argument requirement of lamp (GEMP, pp. 64, 251-52). Other

morphological requirements on words -- such as case endings, are taken as argument-indicators which are not entries (cf. below in text, and GEMP, pp. 40, 167, 264). "Argument-requirement" is thus to be distinguished from these morphological requirements -- some considerations on argument indicators are presented below.

26. More precisely, the 'third person present' affix -s is analyzed in GEMP (2.03) as an indication of operator status and not as an operator, e.g., call operating on Rudi, Mary as its first and second arguments yields the sentence Rudi calls Mary (the indicator -s has a zero variant when the first argument of the operator is I, you, or is of the form N and N or plural). In the case of prepositional, nominal, and adjectival operators, the affix -s is not attached to the operator itself but to a "carrier" be. Under various higher operators, the indicator -s is replaced by tense morphemes, e.g., -ed, will; this does not alter the status of be as a carrier, e.g., was, were, in I was late, You were peevish.

27. For specification of parallel-position, see GEMP, 2.52, 3.41.

28. Pages 191-98 in Ryckman and Gottfried, "Some Informational Properties of Prepositions", Lingvisticae Investigationes, vol. V (1981), pp. 169-214.

29. In GEMP, the discussion is phrased in terms of likelihood-inequalities among operators and arguments (section 1.2). The reformulation of this in terms of likelihood of reconstruction was suggested by Henry Hiz.

30. The course of these reductions is quite complex and cannot be fully described here, see GEMP, sections 3.65-67, and passim.

31. GEMP, 3.1 and 3.21.

32. The base sentences are thus analyzed by a restricted categorial grammar: the operators are functors with various argument-requirements. The grammar is restricted in that there are no functor-forming functors. It is possible that the grammar should be extended to such functors. It is possible such functors can be established for sublanguage grammars.

33. For the conjecture, see Harris, "Mathematical Characterization of the Structure and Informational Power of Language", pp. 7-9, 21.

34. The following encapsulates points addressed more fully in MSL, chapter 1.

35. For further discussion, see section 5 of the paper cited in fn. 33 and references given in fn. 20 of that paper.

36. MSL, chapter 2.3.

37. From article 1,796.2.2.

38. It may be possible to devise some "discovery procedure" for the establishment of word-classes despite this arbitrary starting point. Alternative procedures could be provided as well leading to alternative classifications; for some relevant considerations, see Hiż, 1957, "Types and Environments", Philosophy of Science, vol. 24, pp. 215-220.

39. the side injected with that antigen and the other side can be wh- clauses attached to the respective sides: these secondary sentences are instances of the sentence-type GJB.

40. These two papers differ: in the former, transformations were not required to be rules of "constant semantic difference".

41. A sublanguage word might be taken as a set of synonyms cf. section 3.3 and 4.5 below.

42. Interestingly, it was found that the word-classes and sentence-types constructed for the English corpus sufficed for representation of the French articles. This result, i.e., of largely similar grammars of sublanguages in different languages, is corroborated by those of Kittredge and his associates for weather-reports in French and English (see the Kittredge paper in the volume cited in fn. 14). Whether sublanguages of particular sciences in different languages tend to share the same or similar grammars would ideally be tested by the choice of languages less closely related than are English and French. In respect to such a test, the adoption of particular languages as generally accepted "vehicles" of scientific communication, e.g., Russian and English, presents a complication.

43. A revision of this word-class, based upon the examination of referential-classifiers, is noted in chapter 3, section 3.22.

44. Such classifier words are akin to variables in logical systems, see chapter 1, section 10.

45. The sentences given below are taken from the articles. Other sample sentences, given in FIS, presume an extension of the sublanguage, such as is discussed in section 4.2-3.

46. That vascular is of a different grammatical category than the other members of T_b presents a complication, addressed in section 4.2.

47. Alternatively, one might consider stream as a classifier word.

48. The passage presents other synonymy relations: the first sentence may be considered epiphoric (though not referential) to the second. This points to a connection between lymphoid tissue in the first and the lymphnodes in the second; similarly the direct importance of is linked to were responsible for. Some remarks a propos such links are presented in the discussion of epiphora in chapter 5.

49. An operator-grammar attempts to reduce instances of "class-cleavage" by analyzing occurrences of words in different positions as the result of reductions, GEMP, 2.07.

50. FIS, pp. 84-85, discusses synonymy and homonymy relations among the symbols for the operator subclasses.

51. A thorough discussion of these segments from which the following presentation is excerpted is found in FIS, chapter 2.1.

52. M' might alternatively be considered part of the "procedural sublanguage", cf. section 3.4 and 4.1.

53. For some instructive examples of conjunctions which are in some instances part of the object-language and in others metalinguistic, see Danuta Hiż, "Some English Connectives lead a Double Life: in and outside the object-language", unpublished manuscript, 1973.

54. This in line with the position that facts are tenseless. For a discussion of tensing of a lower operator in respect to a higher one, see GEMP 6.12.

55. A complication arises in respect to such extraction of meta-scientific operators: in the case of operators such as supposed in (3,128.3.4) If the plasma cell is supposed to be a highly active cell type,... extraction to, e.g., If it is supposed that... removes from the science-sentence the indication of its assertion-status

(section 2.22), contra the objective that the formulas be informationally additive. As the entire question of "assertion status" is at this time poorly understood, extraction of M-operators might be argued to be a necessary first step in the identification of operators affecting the status of a sentence as asserted, hypothesized, etc. Referentials can be seen to be relevant here as well. Cf. article 2, 1.1.3-4 On the other hand, Ehrich, Harris, et. al. emphasize the importance of lymphocytes in this respect. [this respect is referential to the process of antibody formation in a sentence not given here] Dougherty and White seem to have produced further evidence in favour of the last-mentioned opinion. The referential -- the last-mentioned opinion -- in the second sentence "picks up" the assertion-status of its referend, the importance of lymphocytes in the process of antibody formation, under the M-operator emphasize.

56. There is also a close relation between agglutination and the antibody (agglutination is evidence for the presence of antibody). The Methods and Materials section of the "Influenzal" article is also described here in regard to cross-reference. It is further discussed in chapter 5, sections 5 and 6.

57. This formulation of the relation between science-sentences and meta-science material was suggested by Zellig Harris.

58. These operators might be considered as part of some assumed science such as arithmetic or as part of the "methods sublanguage" which operates on science-language sentences.

59. One should note here a question raised by Lehrberger in an examination of aircraft-maintenance manuals, namely, the status of such forms as Check indicator rod extension (see fn. 14). The form (or: others like it) is doubtfully grammatical and thus, not part of a proper subset of the language. An alternative suggested above (section 1) is to consider such forms as "stylistic reductions" of grammatical sentences, e.g., Check the indicator of the rod's extension (Lehrberger presents a list of these "reductions"), which do comprise the sublanguage in the sense given.

60. An overview by two immunologists of the course of research is given in chapter 8 of FIS.

61. Given a consensus among immunologists as to the truth of a substantial number of sentences in the articles, it

may be possible to isolate a truth-set within this language. If so, some extensions of the sublanguage suggested below, e.g., by negation, would have to be re-evaluated.

62. Nonetheless, it may be expected that the rate of change in scientific languages may in some sense be greater than that in the language as a whole owing to the pace of scientific inquiry and the "life-situations" in a science. For some apposite remarks, see the Harris paper cited in fn. 15, and Hoenigswald, 1960, section 1.2.

63. This prediction is borne out by work done by Janis Vieland on articles which investigated the effect of hormones on the immune response.

64. By way of contrast, if the sentences in "Methods" sections are considered to comprise a sublanguage, such closure operations would be questionable: these sentences are for the most part assertions.

65. Cf. Estival, et al. Information in Comparatives, pp. 29-32 for discussion of this rule.

66. As presented here, this closure operation is a hypothesis. If borne out, i.e., if the sublanguage can be so extended, then in respect to the augmented sublanguage, grammatical categories (word-classes) can be taken as "semantic categories" in the sense of Husserl's Logical Investigations, volume II. For a discussion of Husserl's definition of semantic categories, see Hiż, "Intuitions of Grammatical Categories", Methodos 12 (1960), pp. 311-19, and TDAP number 21, Syntactic Completion Analysis and Theories of Grammatical Categories, especially section 2.

67. Cf. fn. 45.

68. The notion of 'acceptability' in MSL is distinguished as inequalities of likelihood in GEMP, 1.2.

69. The distinction between sentences of low acceptability (or: low likelihood) and sentences which are ungrammatical is close to, if not identical with, Husserl's distinction between Unsinn and Widersinn in his Logical Investigations.

70. More precisely, inclusion of such sentences would be an exception to the regularities which lead us to establish A and C as word-classes. In respect to immunology, it would lead to an "absurdity".

71. There does occur one case of a rather elaborate metaphor, referring to a "biological minuet" (pp. 67-68 in

Elizabeth Leduc, et al., "Studies on Antibody Production II. The Primary & Secondary Responses in the Popliteal Lymph Node of the Rabbit", Journal of Experimental Medicine, volume 102 (1955), pp. 61-72). The sentence occurs in a Discussion section of the article where the authors are speculating as to the mechanism responsible for a particular result.

72. Cf. Hoenigswald, 1960, chapter 4, for a discussion of extension of a morph's selection.

73. For the comparison, see Harris, "On a Theory of Language", Journal of Philosophy, vol. 73 (1976), pp. 253-76 (p. 276). The literature on a 'criterion of cognitive significance' is enormous; for some insightful discussion (with references), see Hempel, "Empiricist Criteria of Cognitive Significance: Problems and Changes", pp. 101-122 in Aspects of Scientific Explanation (New York: The Free Press, 1965).

74. Harris, "On Grammars of Science", appearing in a festschrift for Rulon Wells.

75. The use of "operational" in this passage does not appear to be subject to the usual objections voiced against operationalism, see "A Logical Appraisal of Operationalism" for these objections, pp. 123-33, in the Hempel book cited in fn. 73.

76. This result also suggests comparison with the Fregean dictum that words have meaning only in the context of a sentence (p. 71 in J. L. Austin (trans.) The Foundations of Arithmetic (Evanston: Northwestern University Press, 1980)). Such a comparison is made difficult by Frege's use of 'Satz' for what is usually rendered as "proposition". There is also the question as to how "meaning" is to be interpreted, i.e., as 'Sinn' or Bedeutung'. Since the sentences regularized in the analysis are in a discourse, the result might more properly be said to confirm the "integrity" of discourses as a domain of investigation.

77. See Hiż, 1964, and the section on ambiguity below.

78. The review of Lyons appears in Journal of Linguistics, vol. 1 (1965), pp. 191-96; also Hoenigswald's paper in the American Journal of Philology, vol. 79 (1958), pp. 290-93, and section 3.2 of Hoenigswald, 1960.

79. The status of Hoenigswald's response here is unclear inasmuch as the hypothesis of kernelization, presented in

earlier work on transformational analysis (papers XXIII, XXVI, XXVII in PSTL, give the essential development), has been superceded in operator-grammar.

80. Hoenigswald, 1965:192.

81. In the Methods and Materials section of the "Influenzal" article, there are occurrences of the word cell in a non-cytological sense, namely as photoelectrical cell in sentences 196.1.3 and 196.1.12.