Appendix E: Language and knowledge considerations

Concept of semantic fields: Appendix E1
Use of several languages and translation problems: Appendix E2
A discipline's model as a "language": Appendix E3
Knowledge, information, and documentation: Appendix E4
Knowledge dynamics: Appendix E5
C. TRAITEMENT PRÉLIMINAIRE DES DONNÉES

Before studying the various movements of time-series it is essential to ensure the compatibility of the observations included by making necessary adjustments for, inter alia:

(a) variations in the interval of time (caused by fluctuations, resulting partly from the incidence of holidays, in the numbers of whatever days or working days per month, per quarter, etc.);

(b) population changes (the eliminate their influence time-series are often calculated on a per capita basis);

(c) price changes (to eliminate their influence value data forming time-series are divided (related) by the price index for the corresponding periods and thus quantity indexes are established);

D. ANALYSE DE L'ÉVOLUTION (SÉCULAI) TENDANCE

The first step to be taken in the analysis of time-series is to plot them on ordinary paper or logarithmically plotted paper (Ch. IV, A.2. (2) and (3)). The secular trend may be drawn on such a chart-freehand or with a transparent ruler—on a curve or as a straight line thus enabling us to study it as well as the fluctuations from it due to cyclic, seasonal or random factors (p. B. 1. above). This method of trend-drawing consists, however, of certain subjective elements and is very approximate.

If we want to make formulae based on the

This Appendix consists of extracts from S. Ullman, *Dictionnaire des notions de temps chronologiques*, p. 243-253.

The concept of semantic fields (**):

The semantic field theory (*) was first put forward by José Luis Trías and is based on the conception of fields as closely-knit sections of the vocabulary, in which a particular sphere is divided up, classified and organized in such a way that each element helps to delimit its neighbors and is delimited by them. Their contours fit into each other like pieces of different shapes in a mosaic. In each field, the raw material of experience is analyzed and elaborated in a unique way, differing from one language to another and often from one period to another in the history of the same idiom. In this way, the structure of semantic fields embodies a specific philosophy and a scale of values.

The field theory has been strongly criticized from various quarters and some of the claims put forward by its champions are no doubt extravagant and unconvincing. The neatness with which words delimit each other and build up a kind of mosaic, without any gaps or overlaps, has been greatly exaggerated. This is true only of specialized and rigidly-defined systems; in ordinary language, vagueness, synonymy, ambiguity and similar factors will produce a much less tidy picture... Quite apart from overlaps between the various conceptual spheres, it is clear that many of these have no systematic organization of any kind.

These limitations must not, however, be allowed to obscure the outstanding importance of the field theory.

Ullman cites: its introduction of a truly structural method into a branch of linguistics, the possibility of formulating problems which would otherwise pass unobserved, and a method of approach to the problem of the influence of language on thinking.

A semantic field does not merely reflect the ideas, values and outlook of contemporary society, but it crystallizes and perpetuates them, it hands down to the oncoming generation a ready-made analysis of experience through which the world will be viewed until the analysis becomes so palpably inadequate and out-of-date that the whole field has to be recast.

At this point, the field theory links up with another recent development in linguistics, the so-called Sapir-Whorf hypothesis.


thesis on the influence of language upon thought. (see Appendix E2) Modern philosophers, "linguistic analysts" and others, are deeply concerned about the possibility that some philosophical problems are pseudo-problems generated by the structure of our languages. Benjamin Lee Whorf approached the question in a novel and fruitful way: by comparing our own European languages -- "Standard Average European," as he called them -- with the totally different structure of American Indian idioms. His researches convinced him that each language contains a "hidden metaphysics," that it embodies a unique way of viewing the world and imposes this outlook on its speakers. "The linguistic system of each language," he argued, "is not merely a reproducing system for voicing ideas, but rather is itself the shaper of ideas, the program and guide for the individual's mental activity, for his analysis of impressions, for his synthesis of his mental stock-in-trade. We dissect nature along lines laid down by our native languages." During the last few years, a new concept of semantic fields has been evolved by the French linguist George Matroix: c'est en partant de l'études du vocabulaire que nous essayerons d'expliquer une société. Aussi pourrons-nous définir la lexicologie comme une discipline sociologique utilisant le matériel linguistique qui sont les mots." His teaching and example have also stimulated a number of enquiries into specialized vocabularies, ranging from feudalism to railways and from fashion to medicine. These investigations have found a focal point in the recently-established centre of lexicological studies at Besançon, where large-scale research projects are being organized with the aid of modern mechanical devices. (*)

Use of several languages and translation problems.

A. Absence of a lingua franca.
It would be optimistic to expect wide acceptance of the system if it was based on one language only. The UNIST Study notes (pp. 72-73) that:
(i) English now accounts for about 40% of the world literature, regularly yielding (as are French and German) to the rising group of "Eastern" languages e.g. Slavic, Chinese, and Japanese.
(ii) No one can predict what the situation will be twenty or fifty years ahead, nor does anyone possess reliable data on the present use of foreign language material in the scientific community (but see Appendix E2, section E).
(iii) The position of English as a lingua franca of science is contested by some governments either to consolidate a new country via a national language or in the belief that language can be artificially maintained as a vehicle of a culture.
(iv) The chances of securing international acceptance of English as the standard language of science are, in present circumstances, very poor.

B. Language preferences.
Apart from these aspects, there is the extremely serious problem that social scientists in one language group tend to either ignore foreign language materials or find it "very relevant" to their particular concerns. This is particularly significant across the English, French, German divide. Concepts given in foreign languages may be difficult to comprehend if one is less than completely at home with the language in question. An unconscious hostility to concepts expressed in foreign languages may even build up (see section E).

C. Language group incompatibilities.
There is also the possibility that a concept may first be expressed or may only be expressible in a given foreign language. It would be an advantage to be able to file it as such and worry about the translation afterwards. The author who fails to do much to emphasize the difficult-to-comprehend contrasts between meanings in the standard Indo-European languages and those in other language groups is Benjamin Lee Whorf. These contrasts are well-illustrated in the following extract from one of his papers (*):

"The growth of the Indo-European language-culture complex dates from ancient times. Much of its metaphysical reference to the nonspatial by the spatial

(*) The centre, headed by S. Quezada, publishes a series entitled "Cahiers de Lexicologie," and a "Bulletin d'Information du Laboratoire d'Analyse Lexicologique".

was already fixed in the ancient tongues, and more especially in Latin. It is indeed a marked trait of Latin. If we compare, say Hebrew and Yiddish, while Hebrew has some allusion to not-space as space, Latin has more. Latin terms for nonspatial, like educe, move, principle, comprehends, are usually metaphorized physical references: lead out, tying back, etc. This is not true of all languages -- it is quite untrue of Hopi. The fact that in Latin the direction of development happened to be from spatial to nonspatial (partly because of secondary stimulation to abstract thinking when the intellectually crude Romans encountered the ancient Greek culture) and that later tongues were strongly stimulated to mimic Latin, seems a likely reason for a change, which still lingers on among linguists, that this is the natural direction of semantic change in all languages, and for the persistent notion in Western learned circles (in strong contrast to Eastern ones) that objective experience is prior to subjective. Philosophies make a strong case for the reverse, and certainly the direction of development is sometimes the reverse. Thus the Hopi word for "heart" can be shown to be a late formation within Hopi from a root meaning "thing we remember." Or consider what has happened to the word "radio" in such a sentence as "he bought a new radio" as compared to its prior meaning "science of wireless telephony."

"To sum up the matter, concepts of "time" and "matter" are not given in substantially the same form by experience to all men but depend upon the nature of the language or languages through the use of which they have been developed. They do not depend so much upon any one system (e.g., tense, or nouns) within the grammar as upon the ways of analyzing and reporting experience which have become fixed in the language as integrated "fashions of speaking" and which cut across the typical grammatical classifications, so that such a "fashion" may include lexical, morphological, syntactic, and otherwise systematically diverse means coordinated in a certain framework of consistency. Our own "time" concept is markedly from Hopi "duration." It is conceived as like a space of strictly limited dimensions, or sometimes as like a motion upon such a space, and employed as an intellectual tool accordingly. Hopi "duration" seems to be inseparable in terms of space or motion, being the mode in which life differs from form, and consciousness in turn from the spatial elements of consciousness. Certain ideas born of our own time-concept, such as that of absolute simultaneity, would be either very difficult to express or impossible and avoided by means under the Hopi conception, and would be replaced by spatial conceptual categories. Our "matter" is in the linguistic sense a subtype of "substance" or "stuff" which is conceived as the formless extensional item that must be joined with form before there can be real existence. In Hopi there seems to be nothing corresponding to it: there are no "empty" extensional items: existence may or may not have form, but what it also has, with or without form, is intensity and duration, these being nonextensional and at bottom the same."

The differences are not restricted to high level abstractions such as "time" and "matter" but may permeate the whole perspective. The famous hypothesis associated with the work of von Humboldt, Sapir and formalized by Whorf suggests:

"that the commonly held belief that the cognitive processes of all human beings possess a common logical structure which operated prior to and independently of communication through language, is erroneous. It is Whorf's view that the linguistic patterns themselves determine what the individual perceives in this world and how he thinks about it. Since those patterns vary widely, the modes of thinking and perceiving in groups utilizing different linguistic systems will result in basically different world views." (1)

"We are thus introduced to a new principle of relativism which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar... We cut up and organize the spread and flow of events we do largely because, through our mother tongue, we are parties of an agreement to do so, not because nature itself is segmented in exactly that way for all to see." (2)

Each language becomes a classification and organization of experience in its own right. As such each may be significantly different from the other and may structure the forms and categories by which the individual not only communicates but also analyzes nature, perceives or neglects particular phenomena or relationships, and constructs his model of the world (3).

A striking example of the possible differences is given by Marshall Walker in discussing the social factors which affect scientific models:

"The language of the Wintu Indians of California seems to indicate a way of thinking quite different from our own. Imagine the surface of a table with a

2 B.L. Whorf, Collected Papers on Metalinguistics, Washington, Foreign Service Institute, Department of State, 1952.
3 See Whorf, op cited.
book lying on it. The remainder of the surface is bare. In English one describes the situation by saying, "The book is on the table." In Winne you say, "The table bums." The English phrase has already committed the speaker to an entirely analytical philosophy of the situation: (1) there are two objects; (2) there is a polarity such that one object is above the other; (3) there is an implication that the book is supported by the table. None of this analysis is present in the Winne sentence, which is purely topological... The scientist who wishes to be as objective as possible in his study of the external world will try to free himself from the possible constraints of his own language." (1)

Such languages may not have parts of speech or separate subject and predicate. In Indian Languages such as Nootka and Hopi events as a whole are signified. Instead of "a light flashed" or "it flashed", Hopi uses a single term, "flash" to signify that a happening has occurred. There is thus no distinction between tense, for the Hopi has no general notion or intuition of time as a smooth flowing continuum in which everything in the universe proceeds at an equal rate, out of a future, through a present, into a past. The focus is rather on the totality accessible to the senses at any moment, with no distinction between past, present, or even the future or physically distant where the latter are accessible or represented in memory. Nauko is different again with little development of tenses by type of activity or aspects of action. The finest concern of Indo-European languages can be defined as time of Hopi, the validity a statement has (in terms of fact, memory, expectation, or custom); and of Nauko, the type of activity. (2)

Von Bertalanffy suggests that the Whorfian hypothesis may be extended. He argues that the categories of knowledge depend on biological and cultural factors. In particular, he argues that Aristotelian logic actually covers only the extremely small field of subject-predicate relations. The all-or-none concepts of traditional logic fall short of continuity concepts basic for mathematical analysis. He is with Whorf in hoping that other languages may permit basically different kinds of "science" which would represent other aspects of reality as well or even better than does the current scientific world picture.

The suggestion has been made, for example, that a language like Hopi might be better suited to verbalizing the concepts of modern physics than English. But some of the non-Indo-European languages may also have important and hitherto unknown concepts concerning the functioning of social processes -- an area in which continuity is even more vital to understanding than is the natural sciences.

Some languages may in fact constitute rich sources of concepts which could prove useful to the understanding of organized social complexity. Little work seems to have been done on this possibility -- most of the examples refer to contrasts of interest, to the natural science perspective. In fact the field of comparative linguistics seems to be made up of "one shot" studies with very little comparison. Where comparisons are made it is at the formal rather than the conceptual level (1), so that with the exception of a few startling examples which augur for a fascinating variety of thinking styles, little information is available. It may be that few linguists are competent to write on the concepts of more than one or two non-Indo-European languages, so that no wide-ranging study or classification is possible, and no "handbook" is available. The absence of such a study only helps to conceal the many differences from the Indo-European perspective -- the existence of such differences is certainly not widely recognized (2).

The whole argument raises the possibility that the computer record design envisaged (see Appendix A6) would not be sufficiently general and flexible to be able to "contain" the concepts of some other language groups. The either-or distinction between "entities" and "relationships" may only amount to a magnificent exercise in handling Aristotelian "substance" and "attribute" as represented in Indo-European nouns and predicate adjectives. In the concept of distinct, 1

1 One reason is that a major school of linguists denies the need to consider "semantics" and "concepts" claiming that all understanding relevant to the discipline can be gained from analysis of syntax. A second reason may be, as Sapir has argued, that many linguists consider such languages "primitive" and therefore unlikely to constitute a source of concepts unknown to the Indo-European culture.

2 Marshall Walker (The Nature of Scientific Thought) notes (p.103-4): "The student of science also has a vital need for comparative linguistics, especially his experiences in the recreation of concepts from their language matrix. The usual language departments of a university are not much help for this type of study... There is need for a course for undergraduates (not language majors) which is designed to illustrate the expression of concepts by different language families. Pending the arrival of such courses the student of science will have to do it himself as best he can."
persistent "entities" common to all languages and can all concepts of "relationships" be adequately represented by graph-theoretic type areas?

David Bohm, a theoretical physicist interested in Piaget's and Gibson's work on the problems of perception, gives detailed arguments against permanence of "entities" and concludes (1):

"It is clear that both in common experience and in scientific investigations, the objects, entities, substances, etc., that we actually experience, perceive, or observe, have always (thus far) shown themselves to be only relatively invariant in their properties, this relative invariance having often been mistaken for absolute permanence" (p.14)

"It is evident then that by considering entities and structures as relatively invariant, with an as-yet-unknown domain of invariance, we avoid making unnecessary and unprovable assumptions concerning their absolute invariance. Such a procedure has enormous advantages in research, because one of the main sources of difficulty in the development of new concepts -- not only in physics but also in the whole of science -- has been the tendency to hold onto old concepts beyond their domain of validity." (p.121-2)

Colin Cherry, a telecommunications engineer interested in the psychology of communication with developing countries, considers that relationships may not be meaningfully represented by graph-theoretic links and that other forms of representation might be preferable.

One response is in the work on linguistic universals. It is suggested that terms exist in all languages to designate objects which meet a condition of spatio-temporal contiguity. And, in general, that all languages are cut to the same pattern without there necessarily being any point by point correspondence between particular languages (2). It is recognized that work in this area is only at the early stages (3).

A close look at the logical assumptions built into the computer record design seems to be necessary.

It would seem important to avoid losing the richness of alternative perspectives by confining this project to one or two languages in one language group -- particularly as the concepts inventoried are supposed to be in some way relevant to the cultures using such languages. That this is significant is indicated by the fact that ~1% of the world's population currently uses non-Indo-European languages (4). This includes the Chinese, who are unlikely to remain a minor influence on world society. The argument that many learn an Indo-European "second" language is weak in that being present in classes at which such a second language is taught or used is no evidence that the language and its perspective "take" in the individual -- as most school leavers know. Even if they do take, it is questionable whether it is satisfactory to ignore the individual's problems of translating between the two conceptual systems (1).

D. Problems of translation.

It may astonish many people to know that contemporary linguistics has concluded that translation between languages is theoretically impossible. Chomsky notes (p.202):

"In fact, although there is much reason to believe that languages are to a significant extent cast in the same mold, there is little reason to suppose that reasonable procedures (not involving extralinguistic information) of translation are in general possible"

Georges Mounin, who notes the same conclusion, has summarized the theoretical difficulties prior to considering why, how, and within what limits the practical operation of translations is relatively possible (7).

Some of the difficulties he notes argue against any attempt to force this project into a unilingual mode.

--certain languages have highly developed terminologies in areas where there are few Indo-European equivalents, e.g. the Pyollum Indians and "salmon"; the Eskimos and "snow" (30 terms), some African languages and "palm trees", the Argentine gaucho's and "horse colouring" (200). There is little value in attempting a definitive translation when no exact equivalent exists.

--the situation becomes more complex when dealing with socio-cultural terms, e.g. how can "brother" and "sister" be translated into Maya when that language only has terms for "younger brother" or "older brother" (2). Much closer to the concerns of this project is the simple problem of translating "people's capitalism" into French (3).

5. A special issue of the ETC (Institute of General Semantics), 15, 2, March 1958 is entirely devoted to interpretation and intercultural communication. It gives many examples of this sort of problem.
Another excellent example, noted by Culkin Cherry (XVII), is that whilst there is no difficulty in translating the colour "red" into and from Russian, the association in the two languages are very different. In English, blood red, red in tooth and claw, red with anger, red light district, etc. In Russian the translation of "red" is synonymous with "beautiful" and has associations equivalent to the English "golden" -- hence "Red Square" and the "Red Army" should be meaningfully translated as the "Golden Square" and the "Golden Army". How much has international tension been aggravated and reinforced by this simple error? Similarly, in Chinese, "red" is primarily associated with "joy", "prosperity", "luck", and "happiness" (*). Thus greeting cards, invitations, decorations, etc., are usually in red. (To what extent have the positive associations of the colour in the two cultures influenced the marked success of socialism there, compared to that in Anglo-Saxon culture, where it has more negative associations?)

E. Use of foreign language material by social scientists

A recent study of 1000 social science research information users in Great Britain has just been completed (**). It shows that 16% of the sample read English only, 72% read French, and 27% read German. Of those who said they were able to read a foreign language, only one-third regularly scan literature in that language. There is even a reluctance to follow up articles in another language.

It was also noted that 22% make no use of abstracts or indexes, 34% never use bibliographies, 22% do not use library catalogues, and 45% do not consult the librarian.

F. Administrative delays

If the attempt is made to translate every theoretical formulation into English, before filing, there will be a hold-up similar to that associated with the modelling activity. There is also bound to be disagreement as to the adequacy of translations. It may be preferable therefore to conceive of a Translation Phase in parallel with the filling, modelling, and term allocation phases, and to give priorities to the translation of given terms according to need.

(*) I am grateful to Mr. Thai Wo Tsean for this information.


A discipline's model as a "language"

For such entity or class of entities within a discipline we can attempt to indicate which other entities, relationships, attributes may be associated with it (i.e., in a political science parlance, what entities or classes of entities can the attribute "democratic" possibly be meaningfully applied to?). In a given discipline the number of such permissible relationships should be quite limited for a given concept -- and even if the number is large it can be reduced by redefining the entities in question as a class or classes.

This approach permits us to define the possible meaningful subunits of propositions which may be generated from the discipline vocabulary. In effect extremely stringent "grammatical" rules are thus set up for the creation of valid elements of sentences in the discipline jargon.

By this approach, extra conditions on sentence formulation are imposed, based on the knowledge obtained by use of the discipline. In natural language such conditions cannot be imposed because many meaningless sentences in natural language are grammatically correct. Because of the richness and redundancies, size of vocabulary and indifference to truth values in natural language, any natural language project would lead to very large numbers of permissible relations which would be of little use, besides being impracticable. (Basically, in natural language, any adjective may be used to qualify any entity in the class of "noun"; any adverb may be used with any entity in the class of "verbs", etc. -- and it is difficult to introduce restrictions at a more detailed level.)

Once the sub-propositional units have been determined (perhaps as "concept pairs") these in effect amount to new compound concepts (e.g., "voting procedure"). These may either be registered as new concepts or left as permissible "generatable" concepts (merely indicating the relationship between the component units, rather than showing the component units as components of a new compound unit). The choice would depend on the frequency of usage of the compound unit.

The procedure may then be repeated for the compound units, where this is considered useful, in order to build up the more complex permissible units commonly encountered. Clearly at a certain point in a particular domain:

- it becomes difficult to determine whether higher level concepts are permissible because there meaningfulness is as yet untested, i.e., they are "new"
- it is increasingly useful to create new units because of the quantity

In certain cases, however, the build-up can continue to the level of defining permissible propositions, i.e., a propositional inventory is built up from the units. Other sub-units are...
held in such a form that many probable propositions may be generated automatically for inspection and possible coding as requiring investigation, false, etc. This procedure introduces further rules restricting the manner in which the units may be combined. Modification and additions may of course be made as new insights and data are obtained.

Once the concepts of a discipline are held in this structured form, some interesting investigations of levels of analysis and degree of equivalence may be made. Where a set of propositions exists employing a given entity (e.g., "nation") which is itself made up of sub-units (e.g., "provinces"), or is itself a sub-unit of a larger entity (e.g., "continental region") (**). "New" propositions may be systematically generated for the higher or lower level by treating the terms as equivalent. These propositions may then be inspected as before, to eliminate the obviously meaningless and inapplicable at the new level. The remaining propositions may be added to the inventory if required.

A similar approach may be adopted between disciplines. In some case new insights may be suggested by treating key entities in different disciplines as equivalent and substituting the entity from the second discipline into the propositions of the first containing the proposed counterpart. (e.g., "individual" in psychology may be substituted for "nation" in political science or vice versa; "cell" from biology for "organization" in organization theory(**)). In the case of a given set of propositions containing a limited number of concepts, equivalents for many of the concepts may be selected from the second discipline, so that only the formal structure of the first discipline proposition is retained. This amounts to a general system investigation of propositional invariance or isomorphy across discipline boundaries -- without the need to define any questionable "meta language" in which the isomorphy is established (**).

"In fact, similar concepts, models and laws have often appeared in widely different fields, independently and based upon totally different facts. There are many instances where identical principles were discovered several times because the workers in one field were unaware that the theoretical structure required was already well developed in some other field." (*)

"Thus, there exist models, principles, and laws that apply to generalized systems or their subclasses, irrespective of their particular kind, the nature of their component elements, and the relations or "forces" between them... we can ask for principles applying to systems in general, irrespective of whether they are of physical, biological or sociological nature... To take a simple example, an exponential law of growth applies to certain bacterial cells, to populations of bacteria, of animals or humans, and to the progress of scientific research measured by the number of publications in genetics or science in general. The entities in question, such as bacteria, animals, men, books, etc., are completely different, and so are the causal mechanisms involved. Nevertheless, the mathematical law is the same." (p. 32-33)

Clearly, investigation of propositions generated by this technique would facilitate the elimination of meaningless and false propositions, and the identification of isomorphisms as distinct from superficial analogies. (**)