

SEMANTIC MESSAGE DETECTION FOR MACHINE TRANSLATION, USING AN INTERLINGUA*

by

MARGARET MASTERMAN,

(Cambridge Language Research Unit, England)

This paper is dedicated to R. H. Richens

1. THE NEED FOR THE ESTABLISHMENT OF A SEMANTIC DISCIPLINE OF "MESSAGE-DETECTION" FOR MACHINE TRANSLATION (M.T.)

1.1. Comment on the Theoretically Unsatisfactory Nature of the Present Situation

The purpose of this paper is to draw attention to a semantic feature of language which must be well understood before it is even imaginable that machines should be programmed to do genuine M.T.- M.T., that is, of which the input is both unrestricted, and also composed of heterogeneous and randomly-chosen texts. This feature is the argument, or *message*, of any piece of discourse. It has been shown that failure to detect this can produce dangerously misleading "garbage"¹⁾, even when the machine, though it does not detect message, is programmed to detect many features of the input language's grammar and syntax, and when it is provided, as well, with a considerable bilingual dictionary.

In my view, the present "critical situation" in M.T.,²⁾ is not due to the fact that genuine Mechanical Translation is inherently impossible, as Bar-Hillel thinks,³⁾ but to the fact that the mechanizable techniques at present being used to analyse language are not powerful enough to detect the message, or argument, of any particular text. Linguistics deals only with syntactic groupings of words picked up unilingually from text, and with the interrelations of these. Information Theory deals with codes, not with languages; it is assumed throughout that the message itself, before it is coded and after it is decoded, is intuitively comprehensible and that it need not be further analysed. Metamathematics deals only with the interrelations of sentences, assuming to start with that these are

* This paper was written with the support of the United States Office of Naval Research Washington D.C., and the research on which it is based was supported by the National Science Foundation, Washington D.C.

intuitively detectable as sentences, which is not, in fact, the case. What is needed is a discipline which will study semantic *message-connection* in a way analogous to that in which metamathematics studies *mathematical connection*, and to that in which mathematical linguistics now studies *syntactic connection*. Until we have this discipline, in however crude a form, there can be no real understanding of the nature of M.T., even between pairs of languages. For, as R.H. Richens has repeatedly pointed out ^{4,5,6)} M.T., by its very nature, has got to be an application of this so far non-existent technique.

1.ii. Message-detection is the Common Goal of Both M.T. Research and of Documentation-retrieval

As a matter of fact, the new discipline required is not entirely non-existent. A first version of it is coming into existence via the techniques currently being used by documentalists for information retrieval from libraries, ^{7,8,9,10)} and a first attempt has been made to apply this technique to M.T. ¹¹⁾ Moreover, an analogous technique is now being developed by an experimental psychologist of perception, ¹²⁾ after years of prophesy by experimental psychologists, to the effect that, in the end, it would be required. The documentalists' technique, which is the most clear-cut of all these, consists, in fact, of semantic model-making; they actually create and encode semantic classification-systems, and then try them out on sets of documents, or simulated sets, in order to see what their retrieval powers are ^{13,14,15,16)}. Mathematically speaking, nearly all these systems are definite, being specifiable either as lattices or as trees. The trouble with them, as they are at present constructed, is that even when they are successful, they will retrieve subject-matter, but not detect message; and, for purposes of M.T., this is not enough. A more complex semantic model, though still of the information-retrieval sort, is therefore submitted in this paper.

This model is semantically stronger than the ordinary retrieval lattice in that (1) within it formulae can be syntactically bracketted, though they need not be, (2) it contains a device for converting any general classifying element, a, into either of two basic "parts of speech", a/, a verb form, and a :, a noun form. (3) It is very highly faceted. It contains 100 very general basic classifiers, which can be used in formulae both singly and in combination, and to these can be added, (though they are not added in this paper) 2 sets of numbered classifiers, the membership of which can be extended as required. Thus the system is both powerful and flexible. It pays for these advantages by being of very considerable size. For this reason, no attempt has as yet been made to code it for a machine, though preliminary work is being done on this.

1.iii. Research Used as Data for the Construction of T

(a) Conceptual Dictionary for English

The uses of the main words and phrases of English are mapped on to a classificatory system of about 750 descriptors, or *heads*, these heads being streamlined from *Roget's Thesaurus*. (This dictionary was designed by E.B.May.)

Format The dictionary is multiply punched on to Hollerith punched cards, with the X and Y rows, and the 79th and 80th columns being left blank to receive further information.

Size 15,000-20,000 cards.

Status This conceptual dictionary is finished up to Q. It is estimated that it will be completed in six months.

Remark A single card in this dictionary covers a root, in English, not just a word. For instance, a single card covers Disappoint, Disappointed, Disappointing, Disappointment. The estimated coverage is 100,000 English words.

Purposes (a) to test the resolving power of an unfacetted conceptual dictionary with about 1,000 unordered descriptors. It was found that this dictionary will sufficiently distinguish the synonyms of *Roget's Thesaurus*, but that a set of unordered descriptors is not informative enough to use as an interlingua for M.T.

(b) to produce, in punched-card form, a realistic dictionary to be used for semantic research in English, such that, (by using the X and Y rows) samples of it could be recorded in a system such as T, and (by using, for a reference-number, the 79th and 80th columns) separate cards could be punched, without loss of information, both for every word, and also for every use of a word, in the case of any word selected for special semantic study.

(b) Italian - Interlingual and Interlingual - English Conceptual Dictionaries, using the interlingua "Nude", designed by R.H. Richens

50 interlingual elements and a negator were used to make this dictionary, the subject-matter of which is oriented to Plant Genetics. The two connectives, / ("slash") and : ("colon") and a word-order rule are used as in T to replace R.H. Richens' three subscripts, and every two pairs of elements are bracketted together, two bracketted pairs of elements counting as a single pair for the purpose of forming 2nd order brackets. Two special-subject list numbers are allowed for every dictionary-entry.

Format The dictionary is coded on to Hollerith punched cards, with a special code which economically handles the brackets, but which prevents the system from being used as a classificatory system. The dictionary also exists on 5 x 3 cards.

Size About 1,000 entries, the unit of entry being a chunk, or sub-word, not a word. Estimated coverage, in Italian, 35,000 words.

Purposes (a) to try out an interlingua using a very small number of elements (so that the elements would frequently intersect with one another when multiple-meaning choices were being made).

(b) to construct an interlingua which could itself function as a pidgin-language.

Remarks When tests were made using 5x3 cards, it became clear that this interlingua could be an M.T. instrument of great power, both for the resolution of semantic multiple-meaning problems, and for the detection of semantic message. In fact, it was from tests done with this interlingua that the first notion of semantic message was obtained, since it was noticed that, once they were coded in the interlingua, sentences which were semantically similar but which differed from one another grammatically and syntactically displayed identical, though crude, semantic messages, and this although the constituent dictionary-entries had by no means been made with this end in view.

In the punched-card form, however, the dictionary proved mechanically and semantically untractable, for the following reasons:

- (a) total absence of commutativity and associativity
- (b) ambiguity of negator
- (c) excess of brackets.

Since the first of these research dictionaries had proved inadequate for M.T. because it could only be used as a classificatory system, and the second had proved mechanically intractable because it could only be used as a pidgin-language, the notion gradually grew up that the next system to try would be one which, in principle, could be used either as a classificatory system or as a pidgin-language.

This third system is the system presented here, the system T', the major constituent of which is the Thesaurus T.

2. *MATHEMATICAL SPECIFICATION OF T

2.i. The Free Lattice, L

Let there first be a semantic lattice, L, generated from a finite number of classifiers, A .. N.

* This section has been written jointly with Mr. R.M. Needham.

These classifiers are used in the following way:

If to a word-use or set of word-uses W there apply classifiers, A, B, C , W is entered as " $A \cap B \cap C$ ".

If there would apply to W a composite classifier " A and/or B ", it is entered " $A \cup B$ ".

Any combination of these is permitted, for Instance, " $(A \cap B) \cup C$ ".

Unfortunately, this system gives rise to the Free Lattice, (Birkhoff¹⁷), p.29) generated by the set of N classifiers, which is infinite for $N > 2$.^ø In order to avoid this infinity, which arises because the laws obeyed by " \cap " and " \cup " do not cause long and complicated expressions to collapse on to simple ones, it is necessary to amend the system so that the interpretation given to the two connectives is to some extent separate.

This can be done in the following way:

2.ii. The Semantic Net and Semantic Tree

When connected solely by the connective " \cap ", a finite set of N classifiers generates the Boolean lattice of degree N . (The same result would of course be reached by using solely the single connective " \cup "; by choosing " \cap " we make the initial assumption that the fundamental operation which we perform when using language is that of specifying or subdividing. Let us call this lattice L' , the *Semantic Net*. Within this lattice, which is of course finite, we may operate according to the rules of lattice-algebra, taking meets and joins as needed. However, within this lattice, it is not possible to express such composite classifiers as " A and/or B "; so that we cannot, for example, specify an element to which - given that we have classifiers for, say, "man" and "woman", say, "HE" or "SHE", within our budget of classifiers - we can apply the composite classifier, "HE and/or SHE".

One might think that this specification of composite classifiers could be made by a simple addition to a Boolean lattice. However, a partially-ordered set formed from a Boolean lattice by the addition of elements between one of its bounds and the neighbouring minimal elements can be a lattice only on condition that all the added elements form a single chain, which is not sufficiently general for our requirements.

^ø This fact, that they are generating Free Lattices, does not seem to have been noticed either by the lattice-using documentalists, nor by their critics, e.g. by Bar-Hillel.

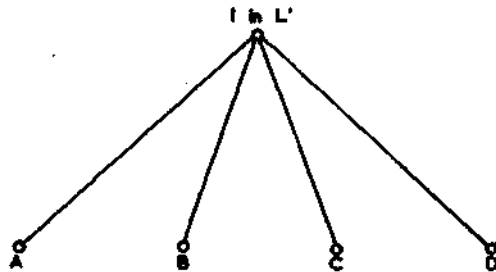
Proof: Suppose that there are four minimals a, b, c, d , at least three of them distinct. Let us add to the Boolean lattice two new elements $a \cup b$ and $c \cup d$. Now since it is a Boolean lattice both $a \cap c$ and $b \cap d$ exist and are distinct. But both of these are evidently contained in both $a \cup b$ and $c \cup d$. Both these extra elements are therefore upper bounds of $a \cap c$ and $b \cap d$. Before we added them, the least upper bound was I , therefore the new elements are now least upper bounds, unless one of them includes the other. If not, the system is no longer a lattice; if so, by repeating the construction the theorem is proved.

We therefore express the set of possible composite classifiers in a quite different system, namely, the system of the *Semantic Tree, K*.

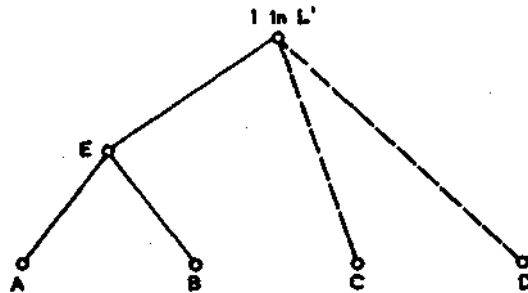
We construct the Semantic Tree, K , as follows:

Suppose we have the set of classifiers, $A \dots N$, as before, these now being the minimals of the *Semantic Net, L'*, and therefore placed directly under the I -element of the lattice, I in L' :

The case is shown below for $N = 4$:



We now create, as below, the element E , forming, of the subset of elements I in L' , E , A and B , the tree k_1 . We shall now say that A and B are the *base points* of k_1 , and that I in L' is the *tree-point* of k_1 :



(The system connected by continuous lines forms the tree k_1 .)

Here the element E, in k_1 , corresponds to the composite classifier " $A \cup B$ ". When the total system is interpreted, the use of the tree k_1 indicates that the classifying elements, A and B, are semantically so closely related in the system, that a means has been provided within it whereby the semantic distinction between them can be dropped if need be.

Suppose now it is discovered, by using the system, that the elements C and D are also closely semantically related; that, in fact, in the system, the pair of elements A and B, and the pair of elements C and D, are more closely related to one another than any of the other 4 pairs of elements which can be formed within the system. It is clear that, by creating the element F, so that we form the tree k_2 , we can create the composite classifier " $C \cup D$ ". We can then say that the total semantic tree for this system, namely the semantic tree K, is made up of the tree k_1 and the tree k_2 , which have the common tree-point I in L' .

Extensions of K If we are to be faithful to our empirical material, however, three types of extension of K will be required.

(1) *Rankings of k's* The possibility has got to be allowed for that the new elements E and F may themselves turn out to be sufficiently close in semantic relationship for it to be necessary, on occasion, to blur the distinction between them.

To provide for this, we create the element G, forming a k with base points E and F; that is, a k with base points $A \cup B$ and $C \cup D$. It will be noticed that, in the case exemplified, G will fall on I in L' , so that no more k 's can be meaningfully created in this K.

If we employ this extension in other cases, however, we shall form rankings of k 's; the set of k 's of which the base-points are minimals of the semantic net we shall call the set of k 's of rank 1; the set of k 's of which the base-points are of the form $A \cup B$, A and B being any minimals of the semantic net, we shall call the set of k 's of rank 2; and so on, up to N orders of k 's, N being the number at which the number of ranks is equal to the number of minimals, and so at which the new element of the final k falls on I in L' , after which no new k 's may be meaningfully created in any K.

It will be noticed here that, in making this extension, we have provided a constructivist method of extending a tree beginning at the bottom, rather than as is customarily done, at the top; and that in order to provide ourselves with a cut-off for the tree-extending process, we have used the essential datum behind Koenig's Stop-Rule Theorem, in the form that, in any binary tree which is extended upwards from its base by a step-by-step k -creating procedure, the number of possible ranks of upward extension is equal to the number of elements at the base. (Koenig¹⁸⁾)

(ii) *Composite rankings of k's* The possibility has got to be allowed for that a close semantic connection may be observed between pairs of elements, other than the minimals, of the Semantic Net, thus making it convenient, on occasion, to obliterate the distinction between them.

Let us take two such pairs of elements, the first, $A \cap B$ and $C \cap D$, and the second $A \cap B$ and $B \cap C$. In the Semantic Net, L' , the join of $A \cap B$ and $C \cap D$ is I in L' , since there is no common element between them. Let us call any pair of elements in L' , which have no common element between them a *disparate pair*. The join of $A \cap B$ and $B \cap C$, on the other hand, is the minimal B , B being the common element between them. Let us call any pair of elements in L' which have one or more common elements an *overlapping pair*.

We can now say that if we create a k with new element G and base points a disparate pair of points in L' , G will not fall on any existing point in L' ; since our upward-reaching tree-growing method is such that I in L' can be the tree-point, but not the new element, of any k . If, however, we create a k with overlapping base-points and with new element, H , H will fall on the point of overlap between the base points, namely, on their join in L' ; and the number of new k -ranks with overlapping base-points will be $N - 1$, N being the degree of L' . If we now add to this number of ranks 1 k -rank to hold the k 's with disparate base-points, the number of new composite k -ranks will be $N - 1 + 1 = N$.

Let the composite k -rank whose elements fall on the minimals of L' be rank 0; the composite k -rank whose elements have been created from pairs of disparate base-points rank - 1; the composite k -rank whose elements fall on the rank of L' with points of the form $M \cap N$ rank - 2; and so down.

(iii) *Reorganization by inconsistent extension, of K* The possibility has got to be allowed for that, a close semantic connection having already been observed between a pair of elements in L' , say A and B , and a k having therefore been generated from them with base-points A and B , a new and closer semantic connection is subsequently observed between one of these base-points and some other element in L' , say A and C . Now if this second semantic connection is to be held as a reason for creating a second k , k' , the existence in it of both k and k' will prevent K , the total set of k 's, from being a tree, and will make it, on the contrary, start to "Boole out" so as to generate a piece of the Free Lattice.

To prevent this, we make the following rule: The creation of any k' , with new element F , which is inconsistent with the creation of a k with new element E , (in the sense that a K containing as parts both k and k' will not be a tree,) shall be deemed to obliterate the creation of k and of any other antecedently created k which is inconsistent with k' (in the sense of "inconsistent" given above) so that K shall always remain a tree.

The semantic principle behind this rule is that when, in a text, one semantic distinction is blurred, other, cognate semantic distinctions will have to be kept, or the whole text will become unintelligible.

A mechanizable method for creating new k's from a text is given later in this paper, in section 3.

Minimals of the System

The following, in alphabetical order, are the minimals of the system:

AIR	COLD	GIVE	LAUGH	ONE	SIGN	USE
AND	COME	GO	LAW	PAIN	SMALL	WANT
ANSWER	COUNT	GOOD	LESS	PAIR	SMELL	WET
ART	COVER	GUESS	LIFE	PART	SOFT	WHEN
ASK	DO	HARD	LIKE	PLANT	SPREAD	WHERE
BAD	DONE	HAVE	LINE	PLEASE	STUFF	WHOLE
BANG	DOWN	HE	MAN	PRAY	TALK	WILL
BE	DREAM	HEAR	MANY	POINT	TASTE	WORLD
BEAST	EAT	HOT	MATE	RE-	THAT	YES
BODY	FEEL	HOW	MORE	ROUND	THING	YOU
BUT	FIGHT	I	MUCH	SAME	THINK	
BUY	FOLK	IF	MUST	SEE	THIS	
CAN	FOR	IN	NAME	SELF	TO	
CAUSE	FORM	KIND	NO	SELL	TRUE	
CHANGE	FROM	KNOW	NOT	SHE	UP	

Illustrative sections of the initial semantic tree are shown in *Appendix I*.

The table in *Appendix II* shows part of a table of equivalences between the K-numbers (that is, the numbers of the nodes of the initial Semantic Tree, numbering from the top) and the k-specifications, (that is, the semantic specifications of the same nodes, starting from the minimals of the Semantic Net, i.e. from the bottom). Certain abbreviative mnemonics are also given, which have attached themselves to certain nodes in the course of tests. These mnemonics are of no logical consequence, and are only inserted for convenience sake.

It must be emphasized that this is only the *initial* Semantic Tree, which is liable, at any time, to be changed, either by the operations of the system itself, or by fiat. For instance, at the time when it was copied for this paper, the branch K3222, (mnemonic: "Date") happened to be part of K3, (mnemonic: "The Introspected World"). At an earlier point this same branch was part of K 1, (mnemonic: "Human-Beings-and-their-ways") where it just as reasonably belongs. Similarly, during one test the nodes WHOLE \cup MANY and ONE \cup PART were created, instead of the nodes WHOLE \cup PART (K21211 and K21212) and ONE \cup MANY (K21221 and K21222) as on the initial tree.

Thus this initial tree must be taken as altogether fugitive and provisional. It is useful, however, to have a Semantic Tree constructed and at hand, when experiments are made with full and ramifying dictionary entries, or with pieces or randomly chosen texts. These often require the creation of new minimalis; and it is useful to be able to fit these into an already constructed tree, so as to see how much their creation is liable to disturb the whole system, and to require changing other dictionary-entries already made. For instance, the system will not easily deal with farming and agriculture. It might well be convenient, therefore, to add 101, GROW and 102, TOOL, as minimalis to the system; or even to insert these instead of e.g. LAW and PRAY. So long as it is clear, however, that they belong in the tree under the branch K3222 (mnemonic: "Date") it is clear also that the system is not seriously disturbed by their being added to it.

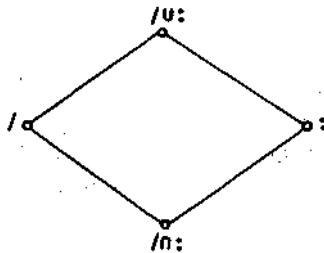
In spite of the mnemonics being individually only abbreviative conveniences, and in spite of the evident crudity of the model, I believe that the overall impression given by it, of language as an anthropomorphically oriented tool, is in substance correct. Note, however, that the minimalis of this system, though man-centred, are not concrete. On the contrary, as the next section will show, these minimalis are not only indeterminate in their possibilities of application; they are also highly abstract.

2.iii. The M factor: the Connectives of the System

(a) The system contains two connectives, / ("slash"), and : ("colon"). The slash is to be interpreted as a verbalizing connective, and the colon as a nominalizing connective.

These two connectives occur as minimalis on a Boolean lattice of 4 elements, M, which occurs as a multiplicative factor in the system, thus making of L a direct product L x M lattice. Let us call this 4-element lattice the *M-factor* of the lattice L.

The M-factor is to be interpreted as below:



/ ∪ : means "Can have either slash or colon"

/ ∩ : means "The property of being a connective; i.e. "both slashlike and colon-like"".

From the fact that $L \times M$ has factor M , it follows that every minimal, a , in L will have 4 forms: $a/$, $a:$, $a(/ \cup :)$, $a(/ \cap :)$. It further follows that every combination of minimal elements in the system will have 4 forms of each constituent element. It also follows that there will be no formula constructible in the system which will not have at least 1 M -element as one of its elements.

(b) It will be noticed that the interpretation given above of every I -element of any m (m being any 4-element constituent sublattice of the M -factor) is suggestive for the interpretation given earlier to any node of the Semantic Tree, K ; for the attachment of the element I -in- m to any minimal, a , in L , indicates that the distinction between "having a slash attached" and "having a colon attached" is to be disregarded. Thus the I -element of any 4-element sublattice m of the M -factor can be taken as having been invisibly attached to any minimal of L , a , when there is no indication given as to which connective is to be attached to a ; as is the case in both the forms of the initial Semantic Tree, as given above.

It follows from this that every terminal node, a , in K can now have a binary sub-twig attached, forming two more nodes for every a in K , these two new nodes being, respectively, $a /$ and $a :$. A diagram showing this extension made to two twigs of K is given below. Alternatively, if we wish to keep the rule that extensions to K can only be made by creation of k 's, these two new terminal nodes, $a /$ and $a :$, can be taken as base-points from which we now form a k with e -element e :

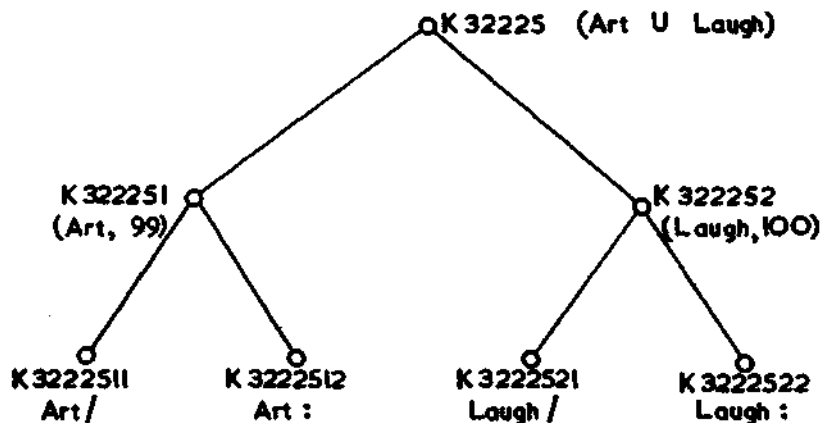


Diagram showing one spray of the Initial Semantic Tree, showing the binary extensions produced at each terminal node, or twig, by the insertion into the system of the factor M .

2.iv. The Thesaurus T

Let the total semantic system consist of the following:

- (a) the Semantic Net, $L \times M$. Let this be fixed and unalterable in the system,
- (b) the Semantic Tree, K . Let this be alterable within the system by use of the procedures given in Section 3.

Let this total semantic system be called the *Thesaurus T*. The following rules govern the construction of formulae in the Thesaurus T.:

- (a) $(a :) \cap (b :)$ converts to $(b :) \cap (a :)$ and conversely
 $(a :) \cap (b /)$ converts to $(b /) \cap (a :)$ but not conversely
 $(a /) \cap (b /)$ does not convert to $(b /) \cap (a /)$ nor conversely.

(b) When constructing any formula in which the minimalis of the system can be interpreted as words in a pidgin-language, the commutativity of the system is to be initially still further restricted by constructing the formulae with the elements ordered according to the word-order rules of that language, in terms of which the minimalis of the system have been interpreted. As soon as these "words" have been bracketted or grouped (as in the next section), and so stored, this restriction can be lifted, and the system T can be allowed to resume its restricted commutativity.

Basically, the minimalis of the thesaurus T should not be regarded either as "words" or as "classifiers" (or descriptors) in any language, but as interlingual *aspect-indicators*, referring to recurrent *aspects* of basic situations which occur in real life. The philosophy of doing this is given in the paper *Translation*,¹⁹⁾ in which the aspect-indicators of the system, which are there called *tags*, are interpreted not by using any terms in any language, but by reference to a pack of cards bearing interlingually recognisable stick, picture ideographs (see later for further discussion of this in section 4).

In the specification given above, however, the minimalis of the thesaurus T are interpreted as monosyllabic words in English, on the assumption that, in any M.T. experiments for which the thesaurus as here specified will be used in the near future, the target-language of the experiment will be English.

3. THE CONSTRUCTIONS AND DETECTION OF MESSAGE IN T

3.i. The Construction of Semantic Shells

(a) We start *by constructing a text* in T; that is, we construct a formula which can then be interpreted as a sentence in a pidgin-language. Any formula interpretable as a pidgin-sentence will do, since the text in question is only to be used for illustrative purposes:

THIS ∩ : ∩ MAN ∩ : ∩ HE ∩ : ∩ CAN ∩ / ∩ DO ∩ / ∩ MUCH ∩ :
 ∩ EAT ∩ : ∩ BUT ∩ : ∩ NOT ∩ : ∩ HE ∩ : ∩ CAN ∩ / ∩ DO ∩ /
 ∩ MUCH ∩ : ∩ FIGHT ∩ :

Since the only connective used in this text is "∩" (which means that the text itself is a point on L x M) we will replace "∩" by a space, which, will have the effect of making the total formula more like a sentence in pidgin-English, without it ceasing to be a formula in T. We will not, however, shorten it by applying the associative or commutative principles to it - and thus lose the semantic information given by the word-order - because, at a later stage, we intend to bracket it.

With spaces instead of ∩'s, the sentence now runs:

THIS : MAN : HE : CAN / DO / MUCH : EAT : BUT : NOT :
 HE : CAN / DO / MUCH : FIGHT:

(b) We now proceed to *construct a pattern of semantic replacement* for the text.

To do this we must first define a *pair of elements* in T:

A *pair of elements* in T is an element in L followed by an element in M.

We now proceed to ask ourselves, just as the linguist does, by which other pairs of elements in T the pairs of elements which constitute the text could be replaced. This question, as linguists know, is highly contentious, since it is the characteristic of semantic replacement, in linguistics, that, as opposed to grammatical and syntactic replacement, there is no end to it. That is to say, as soon as a natural language is being used, the class of possible semantic replacements for any position in any sentence is an open set. The thesaurus T, however, is not a natural language, but a mathematical system with a finite number of minimals. If for the moment we make the rule that we will deal in minimals only, for replacement purposes, (thus cutting out the use of combinations of minimals) the replacement operation thus becomes closed and finite.

It might be objected that the minimals of T are so uncouth and so vague that it is impossible to decide intuitively, in the case of any text, which will replace which. Experience shows, however, that this objection for the most part does not hold, *provided that the text given above is initially treated as a separate text from all other cognate texts* which "mean more or less the same thing and could perfectly well be taken as translations of it". To make this point clearer, let us construct such cognate "texts", coupling each up with a different kind of sentence in full English which might be taken to be a "translation" of it. In the series of sentences given below, the full English sentence is put first, and the formula in T which most nearly corresponds to it immediately below:

- (a) "This man can eat, all right; but he can't, for the life of him, fight."
- (aa) THIS: MAN: HE: CAN/ DO/ MUCH: EAT: BUT: NOT: HE: CAN/ DO/ MUCH: FIGHT:
- (b) "This man is an ace at eating; but he's an exceedingly poor hand at fighting."
- (bb) THIS: MAN: HE: BE/ WHOLE: GOOD: EAT: MAN: BUT: HE: BE/ WHOLE: BAD: FIGHT: MAN:
- (c) "This man makes a beeline for his food, but he shies off at the first sign of a scrap."
- (cc) THIS: MAN: HE: COME/ MUCH: BANG: HOW: TO/ EAT: POINT: BUT: HE: GO/ MORE: BANG: HOW: FROM/ FIGHT: POINT:*
- (d) "This man is greedy, but pusillanimous."
- (dd) THIS: MAN: HE: MUCH: WANT/ EAT/ BUT: HE: SMALL: WANT/ FIGHT/

or,

- (dd') THIS: MAN: HE: SELF: MUCH: PLEASE/ EAT: HOW: BUT: HE: SELF: MUCH: PAIN/ FIGHT: HOW:

Although all the above texts in T begin "THIS: MAN: HE:", and all contain, at later positions, the elements EAT and FIGHT; moreover, although, in addition, all contain the pair of elements BUT: in a central though not invariant position, experience shows that excessive complication and subjectivity is generated by any direct attempt to make a combined pattern of replacement for them. Moreover, if the elements EAT and FIGHT are themselves considered to be replaceable by all the other elements in T which can be used to denote human activity - a step which is obviously necessary if any generality is to be obtained - then the resultant replacement operation becomes totally impossible.

As long as the originally given text is strictly adhered to, however, the replacement operation, with one or two doubtful cases, is easy to perform. Below is a set of intuitively-given replacements in T of the given text, obtained by two people performing the replacement operation independently and then comparing results:

*Notice how, when T is interpreted as pidgin-English, the grammatical paucity makes postverbs of all full English prepositions.

TABLE GIVING PATTERN OF SEMANTIC REPLACEMENT FOR THE GIVEN TEXT

POSITION IN TEXT	1	2	3	4	5	6	7	8	9	10	11	12	13	14
AIR OF ELEMENTS	THIS:	MAN:	HE:	CAN/	DO/	MUCH:	EAT:	BUT:	NOT:	HE:	CAN/	DO/	MUCH:	FIGHT:
SET OF POSSIBLE REPLACEMENTS IN TEXT	THAT: ONE: MANY:	BEAST: FOLK:	SHE:	WILL/ WANT/ MUST/			ASK:* ANSWER: DREAM: SEE: HEAR: TASTE: SMELL: FEEL: FORM: KNOW: GUESS: THINK: USE: BUY: SELL: LAW: PRAY: MATE: FIGHT: TALK: ART: LAUGH:			(As for 3, but same replacement must be used)	As for 4, but same replacement must be used)			(As for 7, but same replacement must not be used)

*When the system T is itself used as a pidgin-language - which is not its normal use for M.T., - it is not possible at the same time to use it as a semantic classificatory system, which it was primarily designed to be. Thus it is impossible, by creating patterns of replacement in T, adequately to model the fact that all the set of full English sentences (a), (b), (c) and (d) are pejorative sentences. They remark of a living being, of either sex, that he, or she, or it, makes at once for, excels at and revels in, some human activity which the speaker of the sentence, in present circumstances, thinks is BAD; whereas this same living being evades, will not tackle; shies off from, some other contrasting human activity which the speaker of the sentence, in the same circumstances, thinks is GOOD. Contrast the set of sentences (a), (b), (c), (d), with the sentence

(e) "This man likes his food, but, you know, he's not greedy",

(ee) THIS: MAN: YES: HE: SELF: PLEASE/ EAT/ BUT: NOT: YES: HE: MUCH: RE: WANT/ MORE: EAT/

If we want to point the pejorative-approbative contrast between the T-formulae (aa) and (ee), we can enlarge the (aa) text to run: THIS: MAN: BAD: BE/ IF: HE: CAN/ DO/ etc.; and then enlarge the (ee) text to run: THIS: MAN: GOOD: BE/ IF: etc. But if T itself is used as a pidgin-language, the actual replacement patterns of these two sentences will never themselves tell us whether an approbative or a pejorative remark is being made.

Whereas if the constituent words of a full English sentence have their dictionary-entries coded up in T, and the sentence is then mechanically "translated" into T, (take as an example the full English sentence, "He is greedy but pusillanimous") it may well be that the dictionary-entries themselves will supply the pejorative indications that an unfavourable and not a favourable judgment is being made.

(c) We now proceed to reorganize K to make it conform to the pattern of replacement given above. We will call this pattern of replacement, *p*.

The general method of reorganizing K consists of taking the total set of pairs of elements given in the table above for each position in *p*, and of then either locating, or creating, a K-point to represent this set.

Thus, for Position I in *p*, the set of mutually replaceable pairs of elements in THIS:, THAT:, ONE:, MANY:. Since there is no existent K-point which immediately connects these four, we grow a new binary branch in K, (see Appendix III). The point of origin of this new branch is then K2113 (with new mnemonic "Specify"), which we shall say is the K-point representing the range of replaceability of Position I in *p*.

Using the same method, we can obtain K-points giving the ranges of replaceability for all the positions in *p*. These K-points are given in the table below:

TABLE GIVING K-POINTS FOR THE RANGES OF REPLACEABILITY FOR ALL POSITIONS IN *p*, TOGETHER WITH THE PROCEDURES WHEREBY THESE K-POINTS WERE OBTAINED

POSITION in <i>p</i>	SET OF PAIRS OF REPLACEABLE ELEMENTS	PROCEDURE	K- POINT
1	THIS:, THAT:, ONE, MANY:	Grow a new binary branch in K	K2113
2	MAN:, BEAST:, FOLK:	Grow a new ternary spray in K	K1111
3	HE:, SHE:	(K-point already existent)	K1121
4	CAN/. WILL/, WANT/, MUST/	Exchange two twigs in K	K311
5	DO/	Create a binary sub-twig in K	K311312
6	MUCH:	" " " " "	K212411
7	ASK:, ANSWER:, DREAM:, SEE:, HEAR:, TASTE:, SMELL:, FEEL:, FORM:, KNOW:, GUESS:, THINK:, USE:, BUY:, SELL:, LAW:, PRAY:, MATE:, FIGHT:, TALK:, ART:, LAUGH:	Create a new ternary fork in K, transplant an existing fork and exchange two branches	K33
8	BUT;	Create a binary sub-twig in K	K21123211
9	NOT:	" " " " "	K2112121
10	HE:, SHE:	(K-point already existent)	K1121
11	CAN/WILL/WANT/MUST/	(K-point created under 4)	K311
12	DO/	(K-point created under 5)	K311312
13	MUCH:	(K-point created under 6)	K212411
14	ASK:, ANSWER:, DREAM:, SEE:, HEAR:, TASTE:, SMELL:, FEEL:, FORM:, KNOW:, GUESS:, THINK:, USE: BUY:, SELL:, LAW:, PRAY:, MATE:, FIGHT:, TALK:, ART:, LAUGH:	(K-point created under 7)	K33

(d) We have now accounted for all the possibilities of semantic replacement in p , with the exception of the special replacement restrictions noted in the replacement table opposite positions 10, 11 and 14.

We formalise these supplementary replacement restrictions as follows:

Creation of Replacement-Variables in T

(i) Let the restriction, "The pattern of replacement is as for the earlier position, X , in p , except that the same replacement must be used", be formalised as follows:

We insert the symbol X_1 after the K-point for position X , and X_2 of the K-point for the position which, in the replacement-specification, carries the instruction. If more than one position carries this instruction, the symbols W and V can be used, in addition to the symbol X .

Let the replacement-restriction just specified above be called Restriction A.

(ii) Let the restriction, "The pattern of replacement is as for the earlier position, Y , in p except that the same replacement must *not* be used", be formalised as follows:

We insert the symbol Y after the K-point for position Y , and the symbol $-Y$ ("not- Y ") after the K-point for the position which, in the replacement pattern, carries the instruction. If more than one position carries the instruction, the symbols, Z , YY and ZZ can be used.

Let the replacement restriction just specified above be called Restriction B.

3.ii. Definitions of Semantic Shell and of Semantic Message in T

We are now in a position to convert the replacement-pattern of p into a formula in T:

$$p = K2113 \quad K1111 \quad K1121X_1 \quad K311W_1 \quad K311312 \quad K212411 \quad K33Y \\ K21123211 \quad K2112121 \quad K1121X_2 \quad K311W_2 \quad K311312 \quad K212411 \quad K33-Y$$

If now we redefine the K-points as lattice-joins, (using the table of equivalences in *Appendix II* and if we replace the spaces by lattice-meet signs, we have an ordinary formula in lattice-algebra. Since this formula will be associative and commutative, however, we shall lose, by creating it, all the information which the word-order, as opposed to the replacement-pattern of p , can give us.

To avoid this loss, we bracket the K-formula which we have created, using the replacement-pattern in order to create the brackets. I here use the Parker-Rhodes syntactic bracketting method (the actual bracketting-procedure is given in *Appendix IV*), but it is probable that any mechanizable syntactic bracketting method could be adapted to apply to T. Let us say merely that we bracket *p* "Rhodewise".

The theoretic advantage, however, of using the Parker-Rhodes' method of syntactic bracketting is that his secondary syntactic lattice for T, (shown in *Appendix IV*) can be incorporated in T as a multiplicative factor, producing a direct product lattice, T', of 1 order higher than T itself. Thus T' = T x S (S being the secondary syntax lattice shown in *Appendix IV*).

This gives:

$$p = (K2113 K1111) (K1121X_1 (K311W_1 K311312 (K212411 K33Y))) \\ (K21123211 K2112121) (K1121X_2 (K311W_2 K311312 (K212411 K33-Y)))$$

This formula can now be stored in a machine memory without loss of semantic information, since the information used for the bracketting is defined in S, and the information used for the K-points will be defined in T, so that the total information used, for syntax and semantics, is defined in T'.

We will define a semantic replacement-pattern in T, so converted and so bracketted, a *Semantic Shell* in T'.

We will define a *value* of such a shell - that is, any single pidgin-formula which can be constructed from the Semantic Shell by following the replacement-pattern given in the Shell - as a *Semantic Message* in T'.

Thus,

(THIS: MAN:) ((HE: (CAN/ DO/(MUCH: EAT:))) (BUT: NOT:) (HE: (CAN/ DO/ (MUCH: FIGHT:)))

is a Semantic Message in T', because it is a value of a Semantic Shell.

3.iii. The Semantic K-Region of *p*, and the Total K-region of P

It will be recalled that, in order to construct *p*, we had at every stage either to reorientate or to make use of some determinate existing part of K.

We can obtain a "semantic K-region" for *p* in the following way: First we take all the parts of K which have been employed, whether reoriented or not (i.e. those parts of K which have been shown red in *Appendix III*, had the set of diagrams of which only two are there, been shown in full); second, we put these parts together, and, by analogy with the procedure given earlier for reorienting the tree, exclude any parts inconsistent with the resulting structure being a lattice.

Now consider the set of semantic messages (aa), (bb), (cc), (dd).
Call this set *P*.

By extracting semantic K-regions for every *p* in *P*, and then conflating these, we can obtain a *total K-region* for *P*.

This total K-region will no longer be in tree-form. We do know about it however.

- (a) that it will be a lattice (since the method of k-creation ensures this).
- (b) that it will be finite, (since it is made up of a finite set of operations, each of which is itself finite).
- (c) that the semantic K-region of any *p* can be mapped upon it. It will thus be possible to speak of the *semantic overlap* between the K-regions of any subset, *p1*, *p2* in *P*.

3.iv. The Bracket-repetition-pattern of *P*

Consider now comparatively the bracket-patterns of the members of *P*.

All of these bracket-patterns are different. The bracket-repetition-pattern, however, is always the same, namely SZCZ, S being a Substantive Group, Z being a Predicative Group, and C being a Conjunctive Group. Moreover, the sequence of K-numbers enclosed in each Z-group, in any *p*, is identical in spite of the fact that this repeating K-sequence in Z differs for each *p* in *P*.

3.v. Shell-Clusters in *T'*

We thus have three criteria for semantic cognateness in *T'*. They are the following:

- (a) Semantic overlap (the measure of this to be empirically determined).
- (b) Repetition of bracket-pattern.
- (c) Identity (or, partial identity within a measure to be empirically determined) of K-sequence within an already repeating bracket-pattern.

We can now say, in the case of any two semantic shells *p₁* and *p₂*, for which the criteria of semantic cognateness given above are satisfied, that they form a *semantic shell-cluster* in *T'*.

In the examples above given, the members of P form a semantic shell-cluster in T'.

4. EXPERIMENTAL APPLICATION OF T'

4.i. The Application of T' to M.T.; Blue Sky Paragraph

It is easy enough, in principle, to see how the Semantic Shells of a system such as T' should be used in programs for interlingual Mechanical Translation. Antecedently, the Master-List of Semantic Shell-clusters has been stored in the machine; the program then proceeds, in broad outline, as follows:

Stage I: The incoming text is converted, by paragraphs, not by sentences, into sequences of dictionary-entries in T'. Each sequence, (one sequence here representing one paragraph) is then syntactically bracketted, and bracket-repetitions looked for, which are then matched with the bracket-repetition-patterns of the Master-List. Where a match is found, repetitions of pairs of semantic elements within the repeating brackets are looked for, multiple-meaning choices, within the dictionary-entries, being made as required. Where sufficient repetition is obtained, the machine decides that it has detected a matching shell-cluster, and searches for unique values of the K-numbers and upper-case variables of the constituent shells of the cluster in the subsequence of dictionary-entries covered by the previous successful match. If values are found, a semantic message will then be obtained. If values are not found, the extent of semantic overlap between the sub-sequence and the shell-cluster must be computed in order to decide which constituent shells, if any, of the cluster, are semantically cognate to that piece of text. Thus the detection of interlingual semantic message consists in successively establishing, for any piece of text, the three criteria of semantic cognateness, as given above, as between the piece of text, and the Semantic Shell most cognate to it in the Master-List.

Stage II: Once the input text has been converted into a sequence of Semantic Messages, the conversion of these, (using the methods currently being developed by Chomsky and Yngve) into pleasantly-running sentences in any output language, should be, by comparison with Stage I, a straightforward task.

4.ii. The Actual Situation

When we turn back from blue-sky speculation to consideration of facts, the size of the semantic universe of discourse hits us like a blow. Not only, in practise, is there no existent computer capable of scanning and syntactically analysing adequately a whole paragraph at a time (let alone of handling a realistic semantic dictionary); it is also the case, quite apart from the hardware problem, that the semantic universe itself is built on the astro-physical, rather than the normal scale. Consider, for instance, the pattern of semantic replacement determining the single Semantic Shell, *p*. From this single shell - taking no account of any other members of the semantically overlapping set *P*, nearly 50,000 *p*-values, that is, semantic messages, can be generated.* Then hand-to-mouth detection of semantic Shell-Clusters, though it gives faint indications that ultimately it might level off, will by no means be small. Finally, and most fundamentally, there has to be considered the amount of semantic information which must be taken account of in any realistic semantic dictionary compiled for M.T. The compilation of realistic samples entries, using T', (a back-breaking task) and the diagramming of their ramifications when they were encoded in T' showed that, when it came to compiling a realistic dictionary for M.T., the scale of a bi-lingual large Oxford English Dictionary (18 volumes) would be far too small; a 200-volume dictionary would be more like what would be required, and, again, the mechanical aids do not yet exist which would be needed to make this.

4.iii. Limitation of the Field in Projected Experiment, using T'

On the other hand, the situation of finding a field, as defined, to be too large to handle, is not new in science; and whenever determinate theoretic analysis comes up against such a "size-barrier", there are always two experimental procedures which can be tried. The first of these is the experimental search for more structure in the field; in this field such a search must take the form of research into the semantic nature and build-up of dictionaries. An initial essay into this research is now being undertaken at the Cambridge Language Research Unit, both by encoding sample dictionary-entries in T', and then analysing the result, and also by programming a computer to make a more direct semantic analysis

* This number is obtained by the following procedure: count the replacement sets of pairs of elements for each position in *p*. Reduce any set carrying the Restriction A to 1 member, and subtract 1 from any set carrying the Restriction B. Ignoring now all 1-member sets of pairs, of elements, multiply the sequence consisting of the remaining sets. Thus, $4 \times 3 \times 2 \times 4 \times 23 \times 22 = 48,576$.

of "natural" dictionary material ²⁰⁾. One solid piece of progress has been made, using T', which will be reported in due course; but the problem is immense.

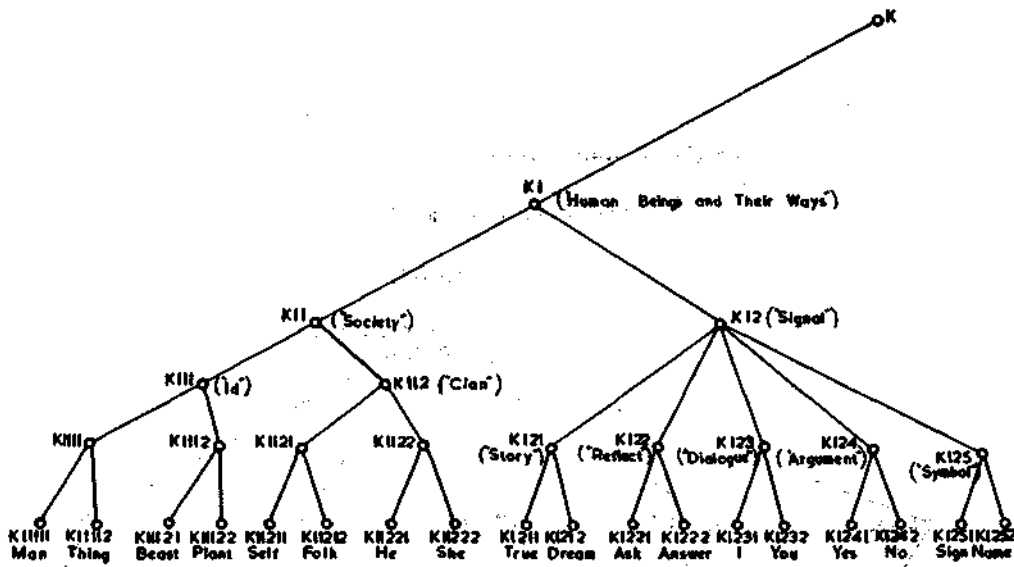
The other experimental procedure for controlling size of field consists in examining, by means of a model, the structure of an artificially limited form of the field, and of then taking the results of such an examination as a rough-and-ready guide to the structure of the whole. Here an unexpected experimental opening has been discovered for the application of T' (T' here being considered as the model) in the material provided by the six *Language Through Pictures* books ²¹⁾. If the semantically similar ideographic schemata round which these books have been built up are correlated and completed to produce a single scheme, the successive sections into which this scheme naturally subdivides are, quite clearly, each intended to define a single semantic shell.* Moreover, the single pictures in the section each suggest a single value for the shell (see Appendix V), and therefore the set of descriptions of them in T' constitute an ascertainable, though not a complete, replacement-pattern for the shell. By using the *Language Through Pictures* ideographic scheme, (or part of it) as though it were a total semantic "world", a limited and determinate, (but yet non-trivial) semantic universe of discourse can be obtained.

Work on mechanizing the analysis of this, though it is still in the hand-test and concordance-making stage, is being put in hand, and will be reported on in due course. It is already evident that the experiment will be a considerable undertaking; but, unlike most semantic experiments, this one is feasible, both in terms of man-hours and in terms of hardware, and it should be interesting to see what it gives.

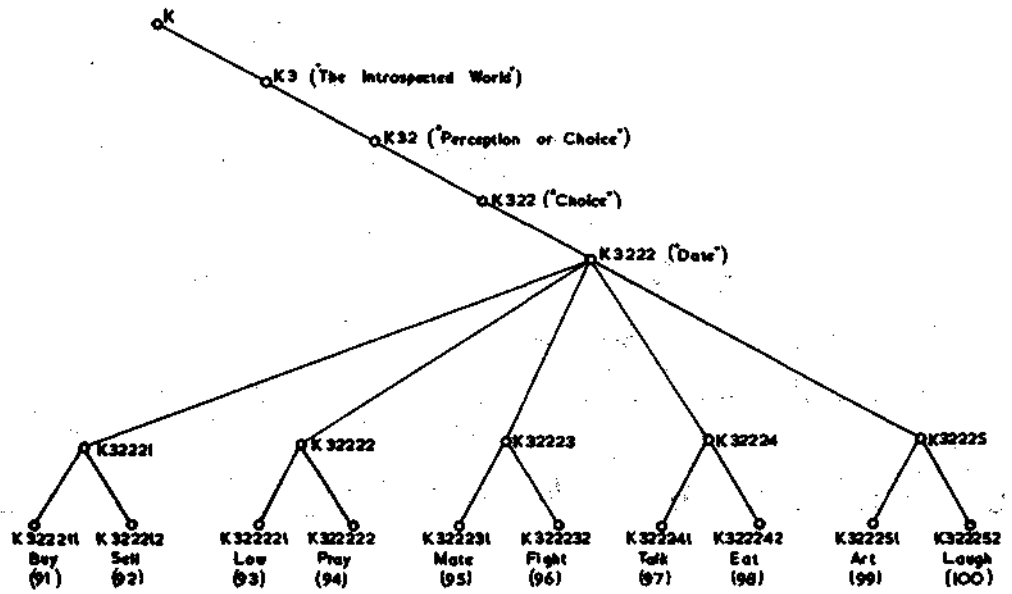
*Such shells are called "key-patterns", or "meaning-patterns" in the various prefaces of the books.

APPENDIX I

(a) EXTREME LEFT-HAND SUB-TREE
OF THE "INITIAL SEMANTIC TREE", K.



(b) RIGHT-HAND BRANCH OF THE
"INITIAL SEMANTIC TREE" K



APPENDIX II

TABLES OF K-NUMBERS AND K-SPECIFICATIONS
OF THE INITIAL SEMANTIC TREE

List No. of Minimal	Tree-No.	Semantic k-specification	Mnemonic (if any)
(as in diagram)	(as in diagram) K i.e. the point of origin of the tree, Which is the join of all the 100 minimals of the lattice)	MAN ∪ THING ∪ BEAST ∪ PLANT ∪ SELF ∪ FOLK ∪ HE ∪ SHE ∪ TRUE ∪ DREAM ∪ ASK ∪ ANSWER ∪ I ∪ YOU ∪ YES ∪ NO ∪ SIGN ∪ NAME ∪ KIND ∪ HOW ∪ HAVE ∪ FOR ∪ RE ∪ NOT ∪ IF ∪ MUST ∪ AND ∪ BUT ∪ PAIR ∪ COUNT ∪ THIS ∪ THAT ∪ WHOLE ∪ PART ∪ ONE ∪ MANY ∪ POINT ∪ SPREAD ∪ MUCH ∪ SMALL ∪ MORE ∪ LESS ∪ SAME ∪ LIKE ∪ IN ∪ COVER ∪ TO ∪ FROM ∪ LINE ∪ ROUND ∪ COME ∪ GO ∪ UP ∪ DOWN ∪ WHERE, ∪ WHEN ∪ WORLD ∪ LIFE ∪ BODY ∪ STUFF ∪ CAUSE ∪ CHANGE ∪ BE ∪ BANG ∪ DO ∪ DONE ∪ CAN ∪ WILL ∪ WANT ∪ GIVE ∪ PLEASE ∪ PAIN ∪ GOOD ∪ BAD ∪ SEE ∪ HEAR ∪ TASTE ∪ SMELL ∪ WET ∪ AIR ∪ HARD ∪ SOFT ∪ HOT ∪ COLD ∪ FEEL ∪ FORM ∪ KNOW ∪ GUESS ∪ THINK ∪ USE ∪ BUY ∪ SELL ∪ LAW ∪ PRAY ∪ MATE ∪ FIGHT ∪ TALK ∪ EAT ∪ ART ∪ LAUGH	"Limit" (from the quotation "The Limits of my Language mean the limits of my World" Wittgen- stein)
	K1	MAN ∪ THING ∪ BEAST ∪ PLANT ∪ SELF ∪ FOLK ∪ HE ∪ SHE ∪ TRUE ∪ DREAM ∪ ASK ∪ ANSWER ∪ I ∪ YOU ∪ YES ∪ NO ∪ SIGN ∪ NAME	"Humans" i.e. "Human- Beings and their ways"
	K11	MAN ∪ THING ∪ BEAST ∪ PLANT ∪ SELF ∪ FOLK ∪ HE ∪ SHE	"Society"
	K111	MAN ∪ THING ∪ BEAST ∪ PLANT	"Id"
	K1111	MAN ∪ THING	
	K1112	BEAST ∪ PLANT	
1	K11111	MAN	
2	K11112	THING	
3	K11121	BEAST	
4	K11122	PLANT	
	K112	SELF ∪ FOLK ∪ HE ∪ SHE	"Clan"
	K1121	SELF ∪ FOLK	

List No. of Minimal	Tree-No.	Semantic k-specification	Mnemonic (if any)
	K1122	HE ∪ SHE	
5	K11211	SELF	
6	K11212	FOLK	
7	K11221	HE	
	K12	TRUE ∪ DREAM ∪ ASK ∪ ANSWER ∪ I ∪ YOU ∪ YES ∪ NO ∪ SIGN ∪ NAME	"Signal"
	K121	TRUE ∪ DREAM	"Story"
9	K1211	TRUE	
10	K1212	DREAM	
	K122	ASK ∪ ANSWER	"Reflect"
11	K1221	ASK	
12	K1222	ANSWER	
	K123	I ∪ YOU	"Dialogue"
13	K1231	I	
14	K1232	YOU	
	K124	YES ∪ NO	"Argument"
15	K1241	YES	
16	K1242	NO	
	K125	SIGN ∪ NAME	"Symbol"
17	K1251	SIGN	
18	K1252	NAME	
	K322	FEEL ∪ FORM ∪ KNOW ∪ GUESS ∪ THINK ∪ USE ∪ BUY ∪ SELL ∪ LAW ∪ PRAY ∪ MATE ∪ FIGHT ∪ TALK ∪ EAT ∪ ART ∪ LAUGH	"Choice"
	K3221	FEEL ∪ FORM ∪ KNOW ∪ GUESS ∪ THINK ∪ USE	"Brood"
	K32211	FEEL ∪ FORM	
	K32212	KNOW ∪ GUESS	
	K32213	THINK ∪ USE	
85	K322111	FEEL	
86	K322112	FORM	
87	K322121	KNOW	
88	K322122	GUESS	
89	K322131	THINK	
90	K322132	USE	
	K3222	BUY ∪ SELL ∪ LAW ∪ PRAY ∪ MATE ∪ FIGHT ∪ TALK ∪ EAT ∪ ART ∪ LAUGH	"Date"
	K32221	BUY ∪ SELL	

List No. of Minimal	Tree-No.	Semantic k-specification	Mnemonic (if any)
	K32222	LAW ∪ PRAY	
	K32223	MATE ∪ FIGHT	
	K32224	TALK ∪ EAT	
	K32225	ART ∪ LAUGH	
91	K322211	BUY	
92	K322212	SELL	
93	K322221	LAW	
94	K322222	PRAY	
95	K322231	MATE	
96	K322232	FIGHT	
97	K322241	TALK	
98	K322242	EAT	
99	K322251	ART	
100	K322252	LAUGH	

APPENDIX III

PROCEDURE FOR REORGANIZING THE SEMANTIC TREE, K, TO MAKE IT
CONFORM TO THE PATTERN OF REPLACEMENT, R

We first specify the parts of the Semantic Tree, K, as follows:

- (a) Let every K-point with a 1-digit number, KM, be called a *sub-tree* in K
- (b) 2-..... KMN,*fork*
- (c) 3-..... KMNO,*bough*
- (d) 4-..... KMNOP*branch*
- (e) 5-..... KMNOPQ,*spray*
- (f) 6-..... KMNOPQR,*tune*
- (g) 7-..... KMNOPQRS,*sub-twig* ...

(N.B. A sub-twig will consist of a *pair of elements* in T, and will be formed under the procedure described in the section on *the factor M.*)

We can now say that the procedure for reorganizing K to conform to P requires that the following operations be performable in K:

- (i) Growing a new binary branch.
- (ii) Growing a new ternary spray.
- (iii) Exchanging two twigs.
- (iv) Creating a new ternary fork.
- (v) Exchanging two branches,
- (vi) Transplanting a fork.

Of these operations, we give the first two, with illustrative examples, and an order-code, below, from which the other four can easily be worked out.

(1) *Growing a new binary branch in K*

We are given two points in K, KMNOP 1 and KMNOP 2, (MNOP here being any number which is at, or below, branch-level in K).

We create a *k*, with base-points KMNOP 1 and KMNOP 2 and with e-element KMNOP (A), KMNOP (A) being the point of origin of the new binary branch.

We renumber the base-points of the new branch KMNOP (A) as KMNOP (A) 1 and KMNOP (A) 2 respectively.

We renumber the branch or branches in K from which KMNOP 1 and KMNOP 2 have been removed, so that the K-numbering in these branches runs continuously.

We redraw K in accordance with the new numbering.

Example:

By the set of replacements in Position 1 of P, we are given K2122, (THIS \cup THAT) and K21125, (ONE \cup MANY).

We create a k with e-element K2113, (and with new mnemonic "Specify"), and with base-points K21125 and K2122.

We renumber K21125 as K21131, and K2122 as K21132.

We renumber all K-points in branch K212 and with 4th digit > 2 by subtracting 1 from 4th digit.

We redraw the relevant parts of the fork K21 (mnemonic "Meta-") in accordance with the new numbering (for diagram see below).

We now create the following order-code:

Renumbering any new entity in K = RENUMBER (A).

Renumbering existing entities in K from which a K-point or K-points have been removed = RESHIFT.

Redrawing the diagram in accordance with the new numbering = REDRAW.

Growing a k with 2 given base-points, 1 and 2 = k(1,2).

Growing a new ternary spray in K

We are given 3 K-points, KMNO PQ, 1, KMNO PQ 2, KMNO PQ 3, (MNO PQ being any number in K which is at, or below, spray-level in K).

k (1,2)

k (2,3)

RENUMBER (A)

RESHIFT

REDRAW.

(N.B. Under RESHIFT, 1, 2 or 3 further ternary sprays may have to be created, to provide for the K-points left unattached by k (1,2) and k (2,3).)

Example:

By the set of replacements in Position 2 of P we are given K11111 (MAN), K11121 (BEAST), K11212 (FOLK).

We create a k with e-element K1111 and base-points K11111 and K11121. We renumber K11121 as K11112.

We create a k with e-element K1111 and base-points K11112 and K11212. We renumber K11212 as K11113.

We create a k with e-element K1112 and base-points K11112 (THING) and K11122 (PLANT). We renumber K11122 as K11122.

We redraw the relevant parts of the fork K111 (mnemonic: "Id") in accordance with the new numbering (for diagram see below).

DIAGRAM ILLUSTRATING PROCEDURE FOR GROWING A NEW BINARY BRANCH IN K

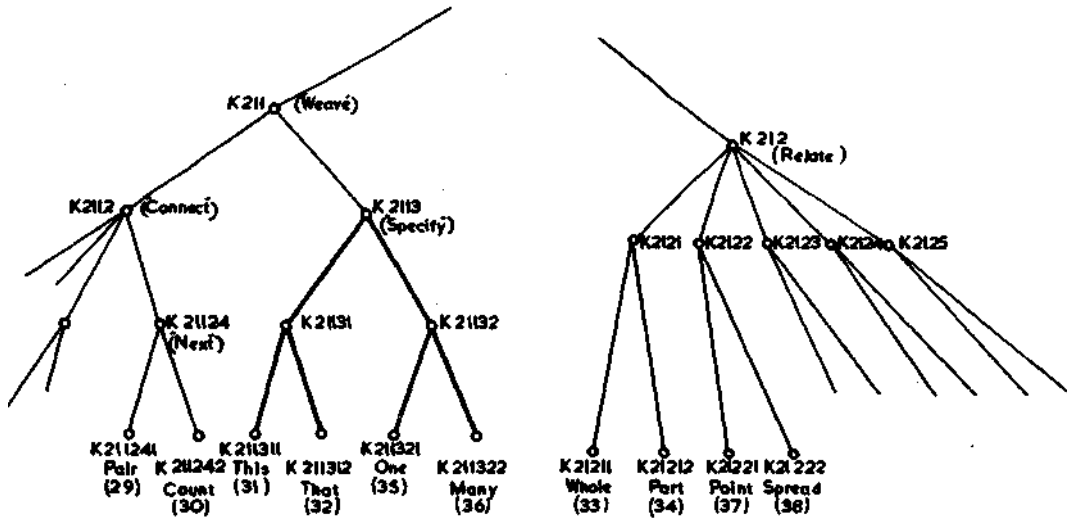
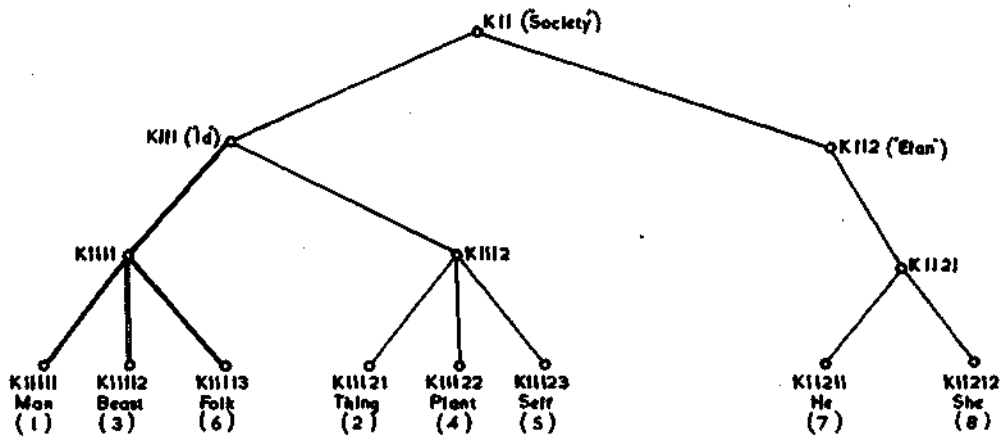


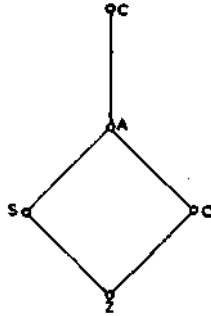
DIAGRAM ILLUSTRATING PROCEDURE FOR GROWING A NEW TERNARY SPRAY IN K



APPENDIX IV

This operation may actually be done with the same simplified primary lattice which is used to illustrate the ideas of principal ideal, and of taking a polar, in Parker-Rhodes' theoretic paper ²²⁾ submitted to this conference. It thus provides quite a good exercise in applying the theory.

The simplified primary lattice is the one below, which gives Lattice Position Indicators A, S, O and Z. These may be assigned to the positions of *p* straight from the replacement-table, also given below:



Primary Simplified Lattice, giving L.P.I.'s for *p*

TABLE OF SYNTACTICALLY MUTUALLY EXCLUSIVE
REPLACEMENT-CLASSES FOR *p*

Replacement-Class	Position in <i>p</i>	Substituent-Type	L.P.I.
THIS:, THAT:, ONE:, MANY:, MUCH:	1, 6, 13	Adjunct	A
MAN:, BEAST:, FOLK:, HE:, SHE:, ASK:, ANSWER:, DREAM:, SEE:. HEAR:, TASTE:, SMELL:, FEEL:, FORM:, KNOW:, GUESS:, THINK:, USE:, BUY:, SELL:, LAW:, PRAY:, MATE:, FIGHT:, TALK:, ART:, LAUGH:.	2, 3, 7, 14	Substantive	S
CAN/, DO/.	4, 5, 12, 13	Operative	O
BUT:, NOT:.	8, 9	Conjunction	C

The product-lattice made by "multiplying" this with its dual, and the place of the primary syntax lattice given above as the dual of the principal ideal on the centre ZZ are all given in the theoretic paper. Since, except for the subject-predicate group itself, Z, *p* consists entirely of

endocentric groups, which can be made by applying the meet algorithm to the primary lattice, m , the polar algorithm is not applied, except vacuously, to explain why the bracket-group SO converts to Z , (Z being the I -element of the principal ideal, and so also a centre of the product lattice). The conjunctive device is used to convert XCX to C .

We thus get the following bracket-groups:

(1) *Conjunctive groups*

$CC \rightarrow C$
 $XCX \rightarrow C$

(11) *Endocentric groups*

$AO \rightarrow O$
 $AS \rightarrow S$
 $OO \rightarrow O$
 $SS \rightarrow S$

(iii) *Exocentric groups*

$SO \rightarrow Z$

We then form bracket-groups according to the following program, finding endocentric groups first, exocentric second, and conjunctive third.

- (i) Form bracket-groups $AO \rightarrow O$, $AS \rightarrow S$, $OO \rightarrow O$,
- (ii) Form bracket-groups $SO \rightarrow Z$.
- (iii) (a) Form bracket-groups $CC \rightarrow C$.
- (b) Search text before and after C for longest repeating sequence.
 Call this sequence X .
- (c) Form bracket-group XCX

In the bracketting schema below, which proceeds from left to right, the penultimate column represents the bracketting analysis of p which is used in the text of the paper.

BRACKETTING-SCHEMA APPLYING BRACKETTING PROGRAM TO P

1	K2113	A	}	S				
2	K1111	S	}					
3	K1121X ₁	S	}					
4	K311W ₁	O	}	O				
5	K311312	O	}					
6	K212411	A	}					
7	K33Y	O	}					
8	K21123211	C	}					
9	K2112121	C	}					
10	K1121X ₂	S	}					
11	K311W ₂	O	}	O				
12	K311312	O	}	O				
13	K212411	A	}	O				
14	K33-Y		}					

APPENDIX V

SPECIFICATION OF A SAMPLE SEQUENCE OF PICTURES FROM THE INTERLINGUALISED
 SCHEMA OF PICTURES MADE BY A.PARKER-RHODES, L.BRAITHWAITE AND
 R.BOSTON BY CORRELATING THE FIRST 5 BOOKS OF THE
LANGUAGE THROUGH PICTURES SERIES

N.B. The picture-descriptions in square brackets do not occur in any
Language Through Pictures book yet published, but have been put in
 to complete the sequence.

Title of sequence: "People speak of place"

<i>No.</i>	<i>Description of Picture</i>	<i>Book-reference</i>
1	A man describes his own position	English 13 French 13 German 7 Spanish 73 Hebrew 13
2	[A woman describes her own position]	
3	[Description of the position of a nearby man]	
4	Description of the position of a nearby woman	German 18
5	Description of the position of a nearby boy	French 27 Spanish 91
6	Description of the position of a nearby girl	French 29 Spanish 89
7	Description of the position of some nearby men	German 27
8	[Description of the position of some nearby women]	
9	Description of the position of some nearby boys	French 31
10	[Description of the position of some nearby girls]	
11	Description of the position of a far-off man	German 20
12	[Description of the position of a far-off woman]	

<i>No.</i>	<i>Description of Picture</i>	<i>Book-reference</i>
13	Description of the position of a far-off boy	French 29 Spanish 92
14	Description of the position of a far-off girl	French 30 Spanish 90
15	[Description of the position of some far-off men]	
16	Description of the position of some far-off women	German 28
17	[Description of the position of some far-off boys]	
18	Description of the position of some far-off girls	French 32
19	*Statement of the relative positions of a nearby man and woman, and of a far-off man and woman	German 53 Spanish 104

etc.

*No.19 is the first member of a new sequence, "Two groups of people discuss their relative positions".

REFERENCES

1. MASTERMAN, M. and KAY, M.: "Mechanical Pidgin Translation", C.L.R.U. Workpaper, M.L. 133. (Distributed at the Princeton Conference on Machine Translation, July, 1960.)
2. BAR-HILLEL, Y.: "The Present Critical Situation in Machine Translation" Colloquium held in Washington D. C. under the auspices of the United States Office of Naval Research, February, 1961.
3. BAR-HILLEL, Y.: "Report on the State of Machine Translation in the United States and Great Britain" (Report to the United States Office of Naval Research), Hebrew University, Jerusalem, 1959.
4. RICHENS, R. H.: "A General Program for Machine Translation between any two languages via an Algebraic Interlingua", *M. T.* 1956, **3**, No.2.
5. RICHENS, R. H.: "Tigris and Euphrates" - a comparison between human and machine translation", Symposium on the Mechanization of Thought Processes, National Physical Laboratory, Teddington, 1958.
6. RICHENS, R. H.: "Interlingual Machine Translation", *The Computer Journal*, 1958, **1**, No.3.
7. MEREDITH, G. Patrick,: "Semantic Matrices", International Conference on Scientific Information, Washington D.C., 1958.
8. MOOERS, C.N.: "Information Retrieval on Structured Content", 3rd London Symposium on Information Theory, 1955.
9. LUHN, H. P.: "A Statistical Approach to Mechanized Literature Searching", I.B.M. Research Centre, Yorktown Heights, New York, 1957.
10. NEEDHAM, R. M. and JOYCE, T.: "The Thesaurus Approach to Information Retrieval", *American Documentation*, 1958, **9**, No.3.
11. MASTERMAN, M., NEEDHAM, R.M. and SPARCK JONES, K.: "The Analogy between Machine Translation and Library Retrieval", International Conference on Scientific Information, Washington D.C., 1958.
12. QUILLIAN, R.: "The Elements of Human Meaning; A Design for an Understanding Machine", University of Chicago, 1961.
13. "Universal Decimal Classification" Trilingual (Abridged) Edition, British Standards Institution, 1958.
14. TAUBE, M.: "The Comac", International Conference on Scientific Information, Washington D.C., 1958.
15. NEEDHAM, R.M., MILLER, A. H. J. and SPARCK JONES, K.: "The Information Retrieval System of the Cambridge Language Research Unit", C.L.R.U. Report, M.L. 109.
16. STEVENS, M. E.: "A Machine Model of Recall", International Conference on Information Processing, Paris, 1959.
17. BIRKHOFF, G.: "Lattice Theory", American Mathematical Society Colloquium Publications, 2nd Ed. 1948.
18. KOENIG, D.: "Theorie der endlichen und unendlichen Graphen", Leipzig, 1936.

19. MASTERMAN, M.: "Translation" to be read at the Joint Session of the Aristotelian Society and the Mind Association, July, 1961.
20. SPARCK JONES, K.: "Mechanised Semantic Classification" to be read at the International Conference on Machine Translation of Languages and Applied Language Analysis, National Physical Laboratory, Teddington, September, 1961. ,
21. RICHARDS, I. A. and GIBSON, C. M.: "English through Pictures" Pocket Books Cardinal Ed. New York, 1952.
See also: Italian through Pictures
 French through Pictures
 Spanish through Pictures
 German through Pictures
 Hebrew through Pictures
22. PARKER-RHODES, A. F.: "A New Model of Syntactic Description", to be read at the International Conference on Machine Translation of Languages and Applied Language Analysis, National Physical Laboratory, Teddington, September, 1961.