MARGARET MASTERMAN

THE SEMANTIC BASIS OF HUMAN COMMUNICATION

The purpose of this study is to examine the foundations of a new hard science, the raw material of which is at present proceeding out of my mouth.

This fact, — that the primary data for any study of language are fleeting and auditory, not visual and permanent, — constitutes the first difficulty in establishing the study of language as a scientific field. To confront this difficulty, the special science of phonetics has, as you know, been founded and developed. This study, the object of which is to identify and examine the actual sounds made by any language speakers, has achieved some remarkable initial results, though it is now running into theoretic difficulties. Strictly speaking, moreover, all examination of language and of languages, without exception, should be founded upon the data furnished by the science of phonetics ; and we shall see later what new light has been thrown on the study of the foundations of human communication by the fact that one man who was investigating in this field is a Professor of Phonetics who takes phonetic data seriously.

Traditionally, however, the human race has, up to now, largely by-passed the challenge presented to it by phonetics

by substituting, for the study of the primary phonetic data of language, the study of differing languages, — or, speaking more exactly, of texts, — presented visually in some of the world's many language scripts.

This substitution, however, of script for sound destroys at once the generality of the subject. For the first thing to realise, in approaching the general study of scripted language, is how very many of these scripts there are, and from what diverse scripting-principles they have been built up. Thus, at the beginning of the English-language-teaching illustrated primer, *English Through Pictures*, there is a short preface written in 40-odd different languages. These require, between them, nearly 20-odd different scripts. Moreover, these scripts embody, between them, four vitally differing scripting principles. Thus, there is the Chinese script, which is often incorrectly described as an ideography, which has a different visual sign for each Chinese « word ». Then there are the Japanese and Korean scripts, which are syllabaries, — one sign for each syllable ; then there are the Hebrew and Tamil scripts, usually described as consonantial, in them, the syllabary principle has begun to be phonetically systematised, in that they distinguish the consonants within the syllable. And finally, there is the much larger group of alphabetic scripts, Greek, Cyrillic and Roman, in some scriptings of which the different letters of the script represent fairly well the fundamental sounds of the language being scripted, and in other scriptings of which, — as, for instance, in the Roman-letter scripting of English, — they don't. Thus, in beginning any comparative studies of scripted languages, and starting with the scripts, the student always has to contend with, and

allow for, the differently-distorting media of the various scripting principles, which make the different languages which they variously portray look even more unlike one another than they really are.

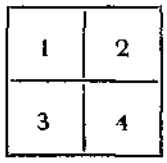
Nevertheless, it has got to be faced that the scripted text, rather than the phonetic transcription, is still now the accepted point of start, even for those who are attempting a scientific study of the nature of languages.

Here are examples of the ideographic, syllabic, consonantal and alphabetic scripting principles. The sentence which is being scripted in the various languages, when translated into English, reads as follows :

ENGLISH THROUGH PICTURES

The book will teach you the first steps of English. It gives you about 500 important words in sentences ; with the meaning shown through pictures.

Read each page liks this :



CHINESE

英語圖解

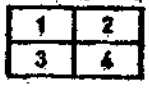
這本是學英國話的入門書,教給你話句子裏常用的重要字,大約有五百,一看圖,意思就明白了。

每面都這麼看下去:

a. IDEOGRAPHIC

KOREAN

그림을통하야
영어공부하는 방법

당신이 이책을 갖이고 영어쇼보를 공부할수있읍니다。 중요한 단어오백을 갖이고 문구를 만들고 그림으로 그뜻을 설명하게 되어있읍니다。 각페지를 공부할때에 다음순서로 하십시요。 

문구와 그림을 비교하면

b. SYLLABIC

HINDI

चित्रों द्वारा अंगरेजी

यह किताब आपको अंगरेजी की पहली शिक्षा देगी। इसमें लगभग ५०० मुख्य शब्दों का वाक्यों म प्रयोग है जिनके अर्थ चित्रों द्वारा बताए गए हैं।

हर पृष्ठ को इस तरह पढ़िए :

१	२
३	४

c. CONSONANTAL

RUSSIAN

АНГЛИЙСКИЙ ЯЗЫК В КАРТИНКАХ

Эта книга научит Вас основам английского языка. В ней Вы найдете около 500 важных слов во фразах, значения которых объясняются в сопровождающих текст картинках.

Читайте каждую страницу в следующем порядке:

1	2
3	4

d. ALPHABETIC

Accepting now the current script-based situation as it is, if we ask ourselves what scientific procedure is suggested by the comparative study of the pieces of script just shown, the obvious first answer is : « The study of their patterning ». Not only can we discern individual recurrent signs in the pattern of any fairly long piece of script ; we can also discern even with the naked eye, recurrences of combinations of such signs. And if we code up the signs of a text into binary code, which there are accepted principles for doing, and then examine the resulting sequence using a digital computer, we find that we can discover, and define in terms of sign-concomitance, and of sign recurrence, many initially non-discernible very interesting features of the text, and of its style. Moreover, there is hope that, by pushing this method further, we shall be able, in the course of time, to discover many more. And of course, there is nothing to stop us comparing the results reached from such study of scripted texts with results reached from a phonetic transcription of the same text, provided the language which is being studied is one which is still spoken ; thus putting the findings of the machine-aided science of phonetics at the service of this second machine-aided science of scientific, or descriptive, linguistics. And it goes without saying that all the findings of descriptive linguistics, being as they are exact findings, findings objectively arrived at from the examination of agreed material, have got to be taken, like the findings of phonetics, extremely seriously.

Nevertheless, these findings are still not what I want to talk about ; and for two reasons. The first reason is that any study of patterning in any scientific field, will always serve to throw into primary relief the complexities of the field rather than its homogeneities ; to show how unlike any one piece of the raw material of the field is to any other ; and how very little emerges as being really in common between one pattern and another once the procedure for examining pattern is made really exact. So it is in language.

Once the procedure for examining the patterning of any text is made really strict, — as it has to be if a digital computer is to be used as the tool for doing the examination, — not only do we find that the patterning of any text in any natural language differs fundamentally from the patterning of any text in any other. We also discover that natural language patterning changes drastically from place to place, — thus defining dialects, — and with the passage of time, — thus defining diachronic variation. Finally we find that even within one dialect, recorded at one time, there occur different types of text, — that is, texts which have considerably different patternings, as defined by the procedure ; and, to cap all, that even within one type of text different patternings are characteristic of the work of different authors.

Now, if we are going to have a new scientific *field*, — a field, as contrasted with a quantity of initially-similar-looking material which can be so prepared as to be examinable by a single scientific procedure, — you have got to find not only what is complicated in your material, but also what is simple : not only what diversifies, but also what stays the same. In other words, if you are going to create and establish a new scientific field you have not only got to describe and classify your data ; you have also got to interpret it, in the light of some picture of it which is simpler than itself. And immediately we try to interpret, as well as to compare, descriptive linguistic findings of texts, it becomes clear that we have got to find some new and deeper principle of interpretation of them, — that is, if we are to establish that there is a general scientific field of language at all, — which, actually, we all intuitively feel there is.

To find this new principle of interpretation we have to place the texts back in their settings ; which means that we have to reconsider them more generally as *communications*, or *messages*, and not just as meaningless patterns. And this is just what, — science apart, — we have always known that they were. People speak and write to communicate with one

another ; not just to create, by the written record of the sets of recurrences of the sounds which they utter, beautiful and variegated two-dimensional mosaics. Spoken language is a human social institution for communicating, not just a very unmusical and erratically rhythmic tune.

Our first step, then, having decided to widen the basis for data finding for our field, is to reexamine language as a social institution. As soon as we do this, we find that indeed, alongside phonetics and linguistics, there have grown up some very interesting and suggestive studies of the sociology and the psycho-biology of language. Unfortunately, however, these do not give us our principle of interpretation ; both because most of the social studies of language are still at the natural history, anecdotal stage ; and also because the procedures for investigation which are used in them, — especially by comparison with the beautiful and exact procedures used in machine-aided phonetics and in machine-aided linguistics, — are very vague indeed.

So we ask ourselves : « Is there not any exact and, if possible, mechanisable procedure, with the aid of which we could examine communication itself, in such a way that, having once found out and cracked open the principles of its essential nature, we could then use these principles to interpret the linguistic-cum-phonetic data, and so create, from the study of language material, the new hard scientific field of which we are in search ? »

There is yet one more exact scientific procedure with which we are not now concerned but on which we must now comment ; and that is, the set of procedures used in communication-theory. These have produced a new scientific discipline; and their raw material is the sets of signals, or codes, which human beings use to convey communication. But it is always assumed, in communication-theory, that before we can use these codes, we already know what human communication is. The procedures are most elegant ; the results, in such applications as the simplification of telephone systems, are

often startling. Moreover, communication-theory, unlike linguistics, is a simplifying procedure, not a diversifying one ; but it simplifies too much ; for it examines communicating sets of signals on the assumption that any communication, conveyed by means of any such set, is, for scientific purposes, the same as any other ; whereas we all know very well that the vital fact about human communications is that they are not all the same as one another ; they differ. So communication-theory, lastly, is not what we are looking for.

What shall we do then, and how shall we proceed, to find out what human communication really is ?

For centuries, a traditional answer has been given to this question from the province of logic, which, in the West, dates from Aristotle :

A human communication essentially consists of a subject and a predicate, these two together forming a sentence. Any sentence can be connected up with other sentences by using connectors such as « and », « or », « if », « then », « all » or « some ».

However, take a look at this machine-derived output, generated on an I.B.M. 7090 computer at the Massachusetts Institute of Technology :

THREE SELECTIONS FROM LOGICALLY CORRECT SETS OF SENTENCES MECHANICALLY GENERATED AT I.B.M.

(a) (223) * AFTER HE IS COVERED, IT NEVER ADMIRES HIS STEAM.

(224) * A SHORT SMOKESTACK IS OILED, AND IT ISN'T BRIGHT.

(225) * IN WHAT WAY ISN'T A RUG OLD, AND WHY IS HE WIRY AND COOL.

(226) * SHE MAKES SIX BOILERS AND HIS TRAGIC AND NEW LEG.

(227) * WHAT IS HE COOLED FOR. Q

(228) * BESIDE NO TRUCKS DOES HE OPERATE THE LEGS BEHIND THE FOUR BLACK WHEELS AND COTTON AND THREE FLOORS.

(229) * IT NO LONGER RUNS IT, NOR DOES SHE PUT IT UNDER FOUR SMOOTH FLOORS.

(230) * NEVER IS MACPHERSON APPRECIATIVE OF IT.

(231) * HE DOESN'T MAKE A THIN HOUSE, NOR IS IT UPHOLSTERED.

(232) * ISN'T SHE POLISHED UNDER MISS MACPHERSON AND PLANTS. Q

(233) * WHEN ISN'T IT CLOSED. Q

(091) * IS HE BLUE, PROUD AND COLD, IS SHE HEATED, ISN'T HE UPHOLSTERED, OR DOES SHE SEEK HER CHAIR AND WILLITS. Q

(092) * ISN'T HE BLACK BEHIND GENTLE AND HEATED RATS. Q

(094) * WHAT IS IT PAINTED FOR. Q

(095) * IS SHE COVERED OFTEN. Q

(096) * WHERE IS HE HOT. Q

(097) * WHO DOES IT SEEK. Q

(098) * WHAT IS HE FULL OF THREE LIDS AND A SAND-DOME FOR. Q

(099) * WHY ISN'T GENERAL SMALL GOOD AND COOL. Q

(100) * ISN'T HE PAINTED. Q

(101) * IT IS SELDOM PAINTED, AND HE ISN'T PAINTED EITHER.

(c) (173) * HE DOES HEAR THE WOOD AND THE FUNNY NEWSPAPERS BESIDE THE WATER.

(174) * IS SHE OILED, AND IS HE COVERED. Q

(175) * HOW IS SHE WIDE AND SHINY ? WHY ISN'T HE CLOTHED, AND HOW DOESN'T HE SEE THE RAT. Q

(176) * THEN ISN'T SHE HEATED. Q

(177) * WHEN ISN'T SHE FULL OF IT. Q

(178) * ONLY UNDER THE RUG ABOVE SIX DRIVING WHEELS IS HER SANDDOME AWARE OF THE COLD SHOES.

(179) * SHE IS NEVER GENTLE AND OLD, NOR IS STEAM FOND OF SIX LONG PENCILS.

(180) * IT KEEPS IT, AND IT DOES HEAR THE PLANTS BELOW THE TWO TRAINS AND FIVE ROOFS.

(181) * NOT EVEN IN FRONT OF MISS SMALL AND THE STEAM IS HE HEATED.

(182) * HE DOES KEEP TWO BRIGHT ENGINES, AND HE IS TAME.

(183) * IT DOES MAKE IT.

In spite of the fact that, on the logical definition of human communication which I gave just now, all of these sentences, in this output, are logically faultless, I think we would all agree they have a screw loose somewhere. In fact, as you can guess, this set of sentences was produced by allowing the

machine to choose between sets of words of the correct syntactic sort by using a randomiser. In one sense, the machine production of them by M.I.T. marks a scientific milestone in fundamental language research, in that it provides a scientifically objective definition of nonsense. For just as any sequence of numbers produced by an accredited randomiser counts by definition as a random sequence, even if that sequence should in fact consist of the number-sequence 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, so any sentence produced by the M.I.T. semantic-randomising procedure counts by definition as nonsense, or non-communication, even if the sentence so generated should turn out to be

SHE IS NOT CLOTHED : HE DOES NOT NOTICE : HE IS OILED.

Our problem, then, in trying to discover the semantic basis of human communication turns out be that of reprogramming the machine which produced this output so as to improve it, and, as we say, « make the machine write sense ». We have got to impose upon the possibilities of combination of these words such semantic restrictions, — semantic restrictions, remember, not grammatical ones — that everybody shall spontaneously say, « How well-connected, and how sensibly-expressed, this new machine-output is ! »

Now, without going in general into the matter of the extreme difficulty, for a human being, of distinguishing sense from nonsense when you have only got isolated sentences to judge from (for almost any isolated sentence can be made, by ingenious interpretation, into some sort of sense) I will now exemplify, a simple procedure which was used to analyse mechanically one sentence of this output, in order to make the machine give a judgement of what, in the sentence, had semantically gone wrong.

The sentence to be analysed was No. 226 in the M.I.T output, which runs :

SHE MAKES SIX BOILERS AND HIS TRAGIC AND NEW LEG.

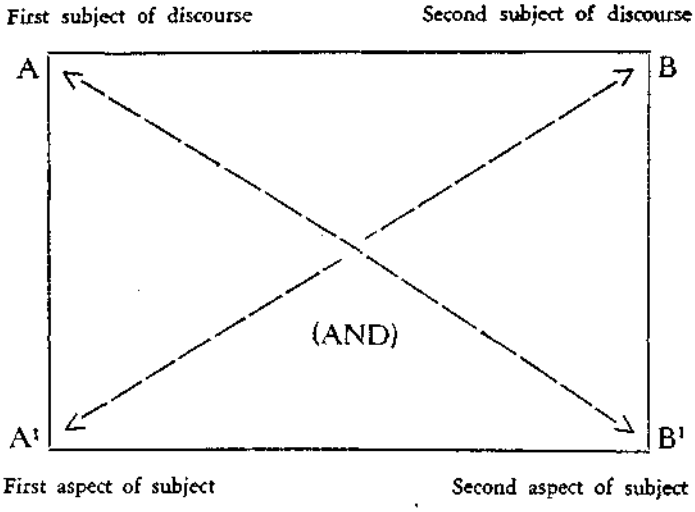
This sentence, like the majority of the sentences in this output, is not plain gibberish ; but, on the other hand, it is also not plain sense. Two contexts can be imagined to justify it. One : she is a female toymaker, and he is a broken animated toy, with whom she has fallen in love. The boilers which she makes are the wooden boilers of toy engines ; so is his leg, which she has just made to replace the older leg which he had lost. But there is something wrong with the new leg ; perhaps it is made of wood when it should have been made of metal, and so the magic or mechanical force of life which animates him does not work on it. This fact makes it a « tragic and new » leg. However, even this context does not fully explain this sentence : for even if this whole surround were the right one, the sentence should properly say, « SHE MAKES BOILERS AND *SHE MADE* HIS TRAGIC AND NEW LEG.

The other context in which this sentence could make sense causes the reader squarely to swallow the initial implausibility that she is now the female managing director of a boiler factory, but suggests that she is now holding down this job merely because she is replacing her lover, who has been recently incapacitated by a boiler-room accident. This accident cut off his leg. She is now using the plant of the factory, therefore, to make him a new leg, which is presumably in shape like a very thin boiler, or, alternatively, one of the metal legs used in boiler-rooms for boiler support purposes. This leg is tragic, because it has had to be made as a result of an accident. It is new, evidently ; all too novel, judging from its general appearance. But even this context, like the last, does not really explain the sentence we have given ; for even if all this story of her devotion, resourcefulness and fidelity were true, the sentence should still run : « SHE MAKES BOILERS, AND SHE IS ALSO NOW MAKING HIS TRAGIC AND NEW LEG. »

Nevertheless, I have said enough to show the extrem difficulty of devising a mechanizable procedure for sense finding in a single human communication. Provided alway that the sentence in question is not grammatically and synt actically, as well as semantically, gibberish, somebody wil always be able to make up some context which will mak sense of it. And yet we all of us know quite well that it i not a really sensible sentence ; something has gone wrong.

That I am able to exhibit a mechanizable procedure fo semantically assessing this sentence, as well as for semanticall assessing at least some other sentences of the M.I.T. output is not due primarily to my own work, not to that of an other member of my group, but to an insight which was ha by a Professor of Phonetics, Professor Pietar Gubérina, of th Institute of Phonetics, Zagreb, Yugoslavia. At the beginnin of this essay I stressed the distortions and limitation imposed upon fundamental scientific language research b the current habit of starting it only from conventionall scripted text. Professor Gubérina has escaped from thi limitation because his present primary occupation, — he i eminent in three fields, — is that of teaching the deaf t hear and to speak. In the course of doing this he was provoke to ask himself the question, « In a communication, what is really vital for the deaf man to get hold of ? » In order t help himself answer this question, he drew on some muc earlier work which he had done when working for hi doctorate under Charles Bally, the linguist, in Paris. In hi thesis Guberina, extending and developing Bally's theory c stylistics, had maintained that the basic semantic unit c human communication was not the sentence, but the paire set of two semantically matching sentences which huma conversation, even at its simplest, always requires. This matche pair of sentences has constituents ; basic *subjects-of-discours* and *aspects* of them, (usually predicative aspects of them) which semantically interact in the squared pattern give below :

THE BASIC SEMANTIC SQUARE



Ordinary script, says Gubérina, obscures the existence of this form ; but its existence can be divined as soon as you work from heard speech, from the question-and-answer lilt of the overall intonational form, which, in one way or another, can be detected in every language. This lilting pattern covers the assertion of, at most, two semantically cognate subjects of conversation ; and of, at most, two semantically cognate remarks which are made about them, or aspects which are asserted of them. Moreover, the semantic cognateness of the two subjects-of-discourse is so tight that the first remark, semantically speaking, could have been made about the second, and the second remark could have been made about the first ; semantically speaking, the sentence would still make sense.

Take an example. « Was it Mary who milked the cows

this morning ? » « No, it was John. » In this exchange, the twice referred-to subject of conversation, — the piece of information which the deaf man must, at all costs, catch, — is that it is the milking of a cow which is being talked about. In the first half of the « square », this milking activity is ascribed to one human being, Mary ; in the second half of the square it is reascribed to a second human being, John. Both Mary and John are semantically cognate entities, being both human beings ; and both are concerned with a single human activity, namely, the milking of a cow. Once the deaf man, already primed to pick up such a « square », can get hold of the general subject being talked about and the kind of thing that is being said, he can reconstruct, says Guberina, quite a lot of the details of the conversation for himself ; partly because he knows, semantically speaking, what kind of thing to expect, and partly because he knows what happens when people milk cows.

Of course, it is not always as easy as that : in the sentence, « Mary milks the cows, but John deals with the goats, » the semantic squared form is still quite easy to discern ; but in many cases, particularly in scientific prose, it is distorted, elaborated, inverted, truncated in endless ways, and sometimes half of it is left out altogether. Nevertheless, although at first, most people are extremely unwilling to admit that spoken human communication could have this single, simple, squared-up semantic basis, after they have got used to the idea they usually have to confess that they are being haunted by finding Guberina semantic squares everywhere. As a Polish logician at a recent Colloquium said, « Yes, I am converted. Like Molière's Bourgeois Gentilhomme, who all his life had been talking prose without knowing it, I lay awake last night realising that all my life I have been making semantic squares without knowing it. » Take, for instance, the paragraph which I have just written, which was composed without the least intention of exhibiting squares. « In some sentences, the semantic square form is quite easy to discern ; in others, it is

liable to be distorted in various ways. » SQUARE 1. « It can get expanded, inverted and truncated, in some cases, half of it can get left out altogether. » SQUARE 2. « Nevertheless, although at first people are unprepared to admit that the semantic basis of human communication is this simple semantic square ; when they have considered the matter for a little, they usually begin seeing semantic squares everywhere. » SQUARE 3. And so on ; and so on. (That last sentence of mine, incidentally, is a degenerate semantic square.)

Now, this sort of fundamental speculation is all very well ; but, unless some determinateness and accuracy is injected into it, it is not the sort of thing which you can put into a machine. It is one thing to run phonetically riot, intonating squares ; it is quite another thing to make a digital computer pick them up from texts. (That last sentence, by the way, was quite a nice semantic square.) And I think that Gubérina's phonetically-based vision of human communication as a series of interlocking semantic squares might have remained indefinitely without scientific application, but for the possibility which has opened up of combining it with other, more detailed work which is currently being done, and in which digital computers are used for semantic analysis. This second line of work is aimed at producing semantic results in two fields ; firstly that of programming computers to do interlingual mechanical translation ; secondly that of devising mechanical methods to extract wanted scientific information from large libraries. In the first field, my own group leads ; in the second, by far the most massive effort at present is that being put out at Western Reserve University, Cleveland, Ohio. For each of these two enterprises, a mathematical model of a semantic system is required which enables different uses of the same word to be separated and classified by reference to the general subject-matter to which the uses refer. Thus, if the set of uses of the English word « *plant* » be taken as an example, « plant » meaning « vegetable », might be classified under *Gardening*, « plant » meaning, « to insert into the

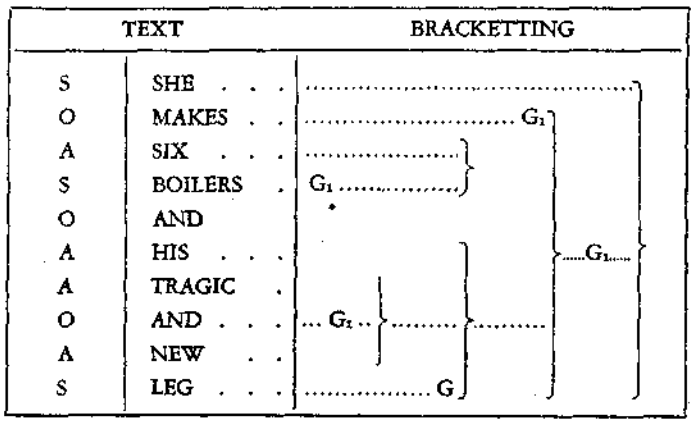
ground », under *Agriculture,* « plant » meaning « engineering plant »; under *Engineering,* and « plant » meaning « trick » (as in the sentence, « *It was all a plant* ») under *Deceit,* or *Subterfuge.* If this method of classifying word-uses is applied to a whole language the result is a very large and variegated semantic subject-dictionary, which looks, — in so far as it looks like anything, — like a distorted and developed form of Roget's famous *Thesaurus.* We have one of these, at Cambridge, punched on punched cards ; Western Reserve have another, of a slightly different kind. Such a dictionary is laborious indeed to make ; but it has the great advantage over ordinary dictionaries, that in it the notions of « semantic overlap » and of « semantic cognateness » can be exactly, though crudely, defined. Consider the two-word phrase, « flowering plant ». The word « flower » has, a use which could be classified under *Gardening ;* so has the word « plant ». In the phrase « flowering plant », therefore, once the constituent words are coded up into the system constituted by a semantic dictionary, the *Gardening* reference will occur twice ; all other references, used to classify uses of « flower » and of « plant » which do not occur in this phrase, will, — we hope — occur only once. It is a simple matter, therefore, to program a machine to reject all the uses of words in a sentence which do not occur twice ; and it is simple also, having done this, to define the semantic overlap of any two words as that subset of subject-references which their two sets of uses have in common. In such a system, also, « semantic cognateness » can be defined as numerical nearness of any two subject-references in the list of subject-references. For instance, in the example already taken, *Gardening* and *Agriculture* will be found nearer together, in any semantic dictionary, or thesaurus, than, for instance, *Engineering* and *Deceit.* We can therefore say that *Gardening* and *Agriculture* are semantically more cognate to one another than *Engineering* and *Deceit,* and set up a scale whereby semantic cognateness can be measured.

Armed now with a determinate, if crude, hypothesis as to the basic nature of human communication, and with a determinate, if crude, semantic system for applying it, let us return to our original none-too-sensible sentence, M.I.T. output No. 226,

SHE MAKES SIX BOILERS AND HIS TRAGIC AND NEW LEG.

The series of diagrams which I am about to present, as a conclusion to this study, represent the outputs, at successive main stages, of a fully mechanized syntactic-cum-semantic analysis of this sentence. The first two of these stages, which are syntactic only, have been fully programmed, and the program debugged, on the Cambridge computer E.D.S.A.C. The other four stages, which are semantic, were worked through by hand, using 5 by 3 cards ; but we shall soon be starting on the project of programming these stages also on the machine.

INITIAL STATE

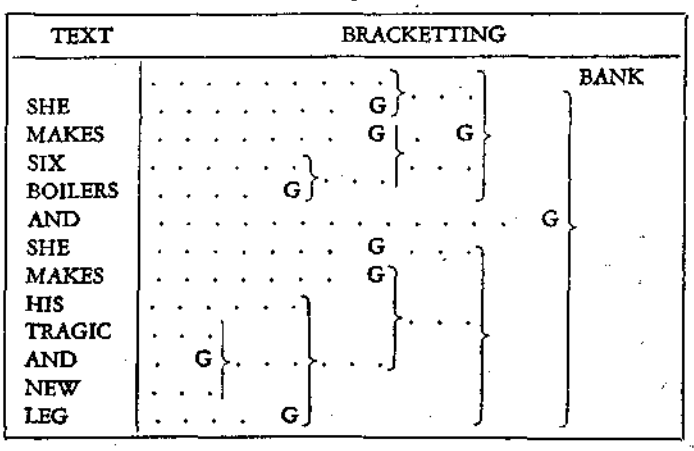


In this diagram, a mechanized syntactic analysis is made of the sentence, and the result of this analysis is shown in simplified form. The sentence itself is shown in the second column from the left ; it is assumed here that the two hyphen- « -S » at the ends of MAKE-S and BOILER-S, the first of which, in English, is a singular present tense verb ending, and the second a plural noun ending, have already been correctly interpreted and sorted out by the machine. In the left-hand column you have the set of grammatical markers which the machine finally picks out as being right for the words of this sentence ; LEG, for instance, here has the grammatical marker « S », for Substantive, in spite of the fact that, in other circumstances, you can « leg it away », and so it can be a verb. Similarly, MAKE, SIX, and NEW are here correctly labelled respectively as Operative and Adjuncts ; even though, in other circumstances, they can all be abstract nouns. On the right, you have the actual syntactic analysis ; this shows two things explicitly, and one more thing which this diagram does not indicate. The syntactic groupings into which the words of this sentence fall are here shown by brackets, which, going from left to right, are of progressively increasing size ; and the chief, or head word of each syntactic group, which is here called its Governor and marked G, the function of the whole group, within that bracket, being defined by reference to the finally selected grammatical category of the word marked G. For instance, in the extreme left-hand bracket-group, which groups together the three words TRAGIC AND NEW, the word marked G. is AND ; its function is marked C, « Conjunction », and so the whole bracketgroup is called a Conjunctive Group. However, the third thing which the program very ingeniously does and which the diagram does not indicate is to divide the bracket-groups which it finds into two classes. These are firstly endocentric syntactic groups, — that is, word-groups the total function of which is the same as that of one of its members ; and exocentric syntactic groups, such as prepositional phrases used adverbially, or

subject-predicate groups made up of substantives, adjuncts and operatives, where the total function of the group is different from that of any of its constituents. This information is used later in the program.

For the benefit of linguists, the technical description of the program is that it mathematicises a touched-up version of the Immediate Constituent model of Syntax, which is familiar from the linguistic literature, by representing the set of possible functions of the simple substituents (or « single » words) of any language as points on a finite self-dual product lattice. These points are so arranged that the functions of endocentric syntactic groups, in the language, come out as lattice meets of the points of the constituent functions ; and the functions of exocentric groups as meet-points which fall within the principal ideal of the positive central of the lattice. This central represents the subject-predicate function ; and if a meet-point falls within its principal ideal, the polar-point of the lattice is substituted for the original meet, and a new meet found, which represents the new function of the exocentric group.

AFTER SQUARING



This diagram shows the second syntactic analysis, which is made after identifying the sentence as a semantic square. This is done, in this case, by the machine picking up the first AND in the sentence as a « squarer » (which it does easily) and determining, in the end, that the second AND of the sentence, is not such a squarer, (which it does with great difficulty). Since there is an exocentric substantive-operative group just before the squarer, as defined by the system, and none after, the machine repeats the words of the first group after the squarer, to make the square, thus producing the sentence, SHE MAKE-S SIX BOILER-S AND SHE MAKES HIS TRAGIC AND NEW LEG. The machine then reanalyses the sentence, marks the Governors, and numbers the syntactic bracket-groups in five ranks, as is shown by the Roman numerals along the top of the diagram.

EXAMPLES OF STAGE I DICTIONARY

MAKE	C	CAUSE /
NEW	F	FLOW :
TRAGIC	E	FEEL :
SHE	H	MAN :
SIX	D	COUNT :

The machine then matches the words of the sentence with the first semantic dictionary, five sample entries from which are shown here. This dictionary is an extremely simple dictionary, each entry in which consists of two and only two parts. The right-hand part consists of a very general classifier ; « CAUSE », « FLOW » (of time), « FEEL », (Something felt), « MAN », (human being). The upper-case letters on the left stand for some special subject-reference ; this reference is substituted for the letters at the match with the No. 2 dictionary. The peculiarity of this dictionary consists in the fact that, simple as it is, the entries in it have their own

internal syntax. This is shown by the two punctuation-signs which occur after the classifiers. These signs distinguish two parts of speech only ; the slash indicates an operative, or verb, and the colon a descriptor, (roughly, an adjective or noun). It follows from this fact that each entry represents a syntactic group, of which the right-hand part, in each entry, is the Governor, and the left-hand part, — specified here only by the upper-case letters, — the dependent.

SEMANTIC ANALYSIS BEFORE SQUARING		SEMANTIC ANALYSIS AFTER SQUARING	
H	MAN :	H_1	Man :
C	CAUSE /	H_1	Cause /
K	THING :	K_1	Thing :
(Squarer)	AND :	(Squared)	And :
K	THING :	H_2	Man :
		C_2	Cause /
		K_2	Thing :

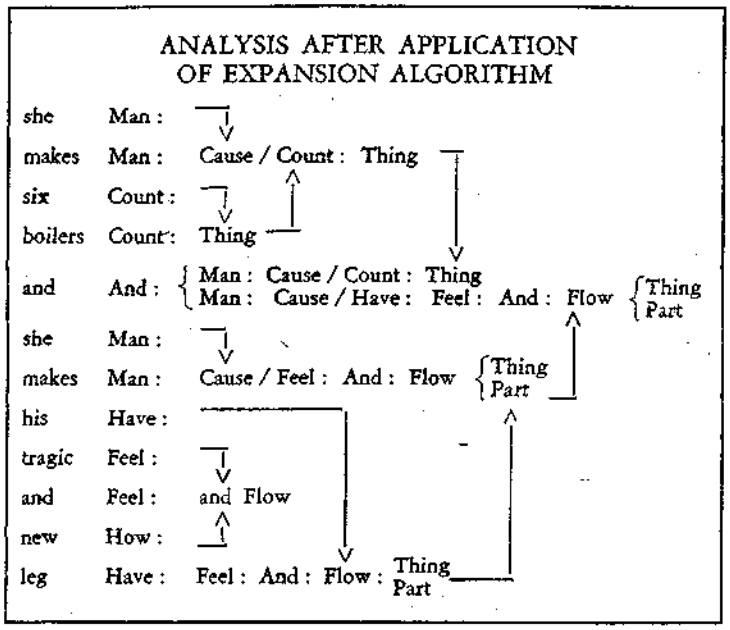
This shows the semantic analysis of the sentence after the first-stage dictionary-entries have been substituted for the words ; for comparison, the semantic analysis is shown both before and after squaring, though in fact, the first dictionary match occurs after making the square. The analysis has been simplified, for diagram-making purposes, by compiling it of the Governors only of Bracket-ranks II-V. The rank of the bracket of which the word, in the original sentence, was the Governor can be shown by the number of lines drawn under the Governor of its dictionary-entry. Thus CAUSE/would have four lines under it, because it was the dictionary-entry which matched with MAKE-S, which was the Governor of a bracket-group of Rank IV.

The machine also adds contrasting subscripts to the upper-case letters which represent the special subject references, thus adding semantic contrast to the sentence. Notice now, how, though building on linguistic foundations, we are already beginning to interpret the linguistic data in accordance with a hypothesis which was inspired by phonetics. The sentence is no longer flat text ; it has contours. Moreover, the contours are semantically tinted in contrasting colours which mirror (crudely) the corresponding features which the contours refer to in the extra-linguistic world.

This first semantic analysis, made after squaring, represents the machine's first attempt at determining the sentence's semantic message. « A man causes a thing », we might interpret it as saying, « and another man, (or the same man) causes the same thing (or another thing). »

This analytic procedure has been criticised by linguists as unscientific, because it begins with an intuitive first step, namely that of making the semantic dictionary entry. But so does any analysis from script begin with an intuitive first step ; for every language was scripted by human intuition. Scripted language is reliable material only because scripting can be repeatably taught. If this new scripting principle, exemplified here, can also be repeatably taught, it is no more and no less intuitive than any other. And why should we not re-script language for our own purposes, especially when we consider how many diverse kinds of language-scripts there already are in the world, and how many different analytic principles they embody ?

The machine now proceeds to fill out the analysis of the squared sentence. It does this by transferring the governors of the dictionary-entries of the original dependents of the sentence to the governors of the original sentence. In other words, it makes the contours of its analysis yet steeper, and tints them more strongly. This it does in order to be able to



match them with the stage-two dictionary-entries of the constituent words of the sentence in such a way as to choose correctly between two alternative uses of a word. Thus, MAKE-S is here used very concretely, for a man actually constructing something ; this fact is now shown in its stage-one dictionary-entry, which will enable stage-two match to be successfully made. The matching complexities of MAKE-S have been deleted ; but two stage-one entries for LEG are here given, to indicate the existence of the problem. One of these entries has the semantic governor THING, and one has the semantic governor PART. When we come to the stage two match, the stage two dictionary gives no less than three entries for LEG, not counting the verbal use, which the

machine is supposed to have thrown out at the syntactic stage. The next diagram shows the machine choosing between the three stage-two dictionary-entries for LEG.

SAMPLE INTERSECTIONS

Boiler	List 23, Engineering	[Cause/Heat] Thing :	No Intersection
Leg 1	List 40, Body, Human	[Man : Limb] Part :	
Boiler	List 23, Engineering	[Cause/Heat] Thing :	Intersection Thing
Leg 2	List 23, Body, Human	[Prob : Part] Thing :	
Boiler	List 23, Engineering	[Cause/Heat] Thing :	Intersection Thing
Leg 3	List 42, Medical	[Man : Do : Cure] Thing :	

At the final stage of the analysis, the machine reconstructs the square, using the stage-two dictionary-entries instead of the stage-one dictionary entries to replace the upper-case letters, and their subscripts. The upper-case letters are replaced by the interesting part — that is, the semantically overlapping part, — of the stage-two dictionary entries ; the subscripts are replaced by the non-intersecting part, in order to give the sentence semantic variegation. And here we at last see what is semantically wrong with this sentence, put in a form which a computer can assess. For half the constituent words of the square intersect completely, — in fact, they repeat ; and the other half do not intersect at all. And this is unacceptable ; both intuitively, and also to the machine. For if a semantic square has a repeating subject and predicate, but a non-repeating complement, the two variants of the complement must semantically be very like one another indeed. THEY MAKE BOILER-S, AND THEY ALSO MAKE SPARE

PART-S. If, on the other hand, the two complements are semantically unlike one another, then the subject and verb must also vary. HE MAKE-S BOILERS, AND SHE MAKE-S WOODEN LEGS. And thus, as soon as a semantic square can be located in a piece of discourse, and exact criteria established in terms of intersection for assessing semantic likeness and unlikeness, its limits of semantic tolerance, as it were, can also be assessed.

The last diagram summarises the intersecting unbalance which makes the sentence from the M.I.T. output an unacceptable square.

FINAL STAGE

(Letters replaced by intersecting parts subscripts replaced by non-intersecting parts.)

Boilers K_1 —> Thing : cause/heat. List 23 engineering

Leg 2 K_2 —> Thing : prep. part. List 30. Furniture

Leg 3 K_2 —> Thing : —> Man : do : cure. List 42 medical

whereas

She $\begin{cases} H_1 \\ H_2 \end{cases}$ —> Man : one : she (no-subscript)

Makes $\begin{cases} C_1 \\ C_2 \end{cases}$ —> Cause/do. List 50 : constructing (no subscript)

It is clear that the semantic analysis which I have given is only a crude analogue of what adequate mechanised semantic analysis will have to be. It is clear also, however, that it is a prototype. Such analysis can be developed further in any number of ways.

Looking at this new semantic research more generally ,if we ask ourselves what is happening in this new scientific field of fundamental semantic research, the only possible answer seems to me to be that computer-using children are pushing on where humanistic wise men fear to tread. Academic semantics, as Ullmann says, is in an ambiguous position ; for its practitioners hesitate before using the precise classifying methods of linguistics to explore what they assume to be « the loosely organised and virtually unlimited world of meanings ». But it sometimes happens, in science, that situations arise in which wise men, who see the complexities of the problem, are outpaced by fools, who do not. In this matter, of finding foundations for semantics, I myself am schizoid as between wisdom and folly ; for I started my academic life as a French specialist, admiring Ferdinand Brunot ; and I now write programs, and devise crude semantic dictionaries and systems to go on machines. I am therefore in a good position to say that while linguists consider and reconsider whether they dare speak of « meaning », mechanical translation researchers are combining meanings to form semantic squares. While linguists classify with more and more refinement, information-retrievers, using the crudest possible systems, are making fundamental predictions. As a result, our whole picture of language is gradually changing but not because of the main trend of work being done by linguists and a new hard science is being founded, not, as it should be, from an academic discipline, as it often has been, but from a confluence of minor technologies.