

HUMAN TRANSLATION AND TRANSLATION BY MACHINE - 1

by

SILVIO CECCATO and BRUNA ZONTA

(Centro di Cibernetica e di Attivita Linguistiche,  
University of Milan)

I

**THE COMPREHENSION OF A TEXT, BY THE HUMAN TRANSLATOR AND BY THE MACHINE**

1. THE SUBJECT OF THE CONTRIBUTION

TIME, at congresses, is always too short, and the contribution of a team, even if split up into several papers, cannot possibly cover the entire ground of a subject as large and as complex as Mechanical Translation. Inevitably, we shall have to make some references to the material published in recent years by the members of the Centro di Cibernetica e di Attivita Linguistiche of Milan University.

Here we shall treat some of the more general problems of MT and, in particular, a question which, so far, has been given less attention than others by the students interested in this field. This particular question is: What operations does a man carry out when he translates, and how did one plan to reproduce them in a machine?

It is, for instance, perfectly obvious that whatever a man may want to do with a given text - unless he is exclusively concerned with its typographical aspect - he must first comprehend it, he must first understand it; that is to say, stimulated by the text he develops a strain of thought. The graphic material employed in the words refers him to something else. Only after this comprehension, after this reference to something else, can he go on to translate the text, or to summarise, to comment, or to expand it.

But in what does the comprehension of a text consist, What is the thought which it calls forth?

And what, of all this, must and can, remain in the operations of the machine if it is to translate?

2. THE AIM OF THE CONTRIBUTION

Although this paper stresses the similarities and differences between human and mechanical translation, its purpose is, of course, to contribute to the study of MT.

It is only our particular approach that leads us to look rather more towards man than towards the machine and, where man is concerned,

rather more towards thought than towards language (at least with regard to those aspects of language that usually are considered formal).

The results of this research, therefore, also belong to a general linguistics of which MT is an application and a test. As you will soon notice, ours is a novel kind of linguistics compared to traditional linguistics. For while the traditional studies start more or less directly from language and its formal aspects, ours start from an analysis of thought. In fact, we ask what language is, how it works, and how it matches thought, after an analysis of thought and its contents. The analysis is also of a novel kind, because the contents of thought have hitherto been conceived as static units, while we conceive them in terms of operations.

(b)

The individuation and description of the operations by means of which man translates also serve as a warning against certain dangers for those working in this field. The recent history of the first attempts to mechanise man's mental activities, such as perception, thought, translation, summarising, etc., clearly shows where these dangers lie.

Since the philosophers and psychologists have refrained from supplying an analysis of these superior activities in terms that might be of use for mechanical construction, the engineers set themselves up as philosophers and began to improvise. They did not hesitate to consider procedures identical that have little or nothing in common. A fictitious example may serve as an illustration of what has happened. Within certain limits it is possible to produce the results of arithmetical operations, either by actually carrying out these operations or by using results that have previously been obtained and memorised, that is, by simple substitutions; but would it be justified, in the second case, to say that the man or the machine "calculates"? Similarly, if machines or men do no more than substitute the words or groups of words of a prefabricated translation for the words or groups of words of a text, would it be justified to say that they "translate"?

Confusions of this kind are detrimental on various levels. On the theoretical level, for instance, they lead to the neglect of essential studies (in our case the most important branches of research concerning language and thought would no longer be furthered). On the psychological level they create an excessive initial optimism, because the problem would not only seem easier than in fact it is, but it would appear already solved; this optimism will then give way to a no less excessive pessimism, when it proves impossible to get beyond the limitations of the adapted solution, and one will then conclude that the mechanisation of this particular activity is impossible (in our case the conclusion would be that MT is impossible). On the practical level, finally, it might be that what seemed the quickest and most economical method turns out the slowest and most costly; because one might be led to constructing machines which yield only the one kind of result, instead of constructing machines which, since they reproduce the much richer human activity that produces these results among others, can be used - with small modifications - to obtain many other results as well (in our case, as we shall try to explain, the same analysis of an input

text can be used for output in diverse languages, for summarising, for documentation, etc.).

### 3. THE TYPES OF STUDY

Our approach to MT, as we have said, is but one of the possible applications of a series of studies aimed at an analysis of thought and its contents on the one hand and of language on the other.

The purpose of these studies is primarily that of constituting a psychology and a linguistics. Linguistics is particularly important to us, because, for the time being, thought has been approached above all through language. But our analyses have been carried out also with a view to the immediate application of the findings. These analyses are intended, as we have mentioned, to present results in terms of operations, and this makes it possible for them to be used both as working hypotheses by physiologists and anatomists who are trying to individuate the organs which carry out these operations and as indications by the engineers who intend to construct artificial organs which carry out these operations as their function. We have tried above all to assure that the analyses break up the studied activities in operations that are constituted by changes of state or changes of place which, on the one hand, can be supposed to be observable in the nervous system once adequate techniques of observation have been evolved, and, on the other hand, can be reproduced by an engineer given the actual possibilities of construction.

### 4. THE OBSTACLE OVERCOME

A dynamic conception of this kind made it necessary to overcome some difficulties inherent in the way in which thought and language have been considered in traditional philosophy and in the psychology deriving from this. According to this tradition we see in the brain, not operations, but a passive mirror which reflects all that surrounds us. The brain, that is to say, is supposed to double the physical objects of our environment by means of as many entities equal to the physical objects, yet lacking their physicality. (If for no other reason, because the brain is already a physical object and would thus have to give up its place and its matter to the other object.) Given this conception, however, the physiologist and anatomist as well as the engineer are put out of action, because these entities which are necessarily present in a negative form are neither observable nor reproducible; and thought, whose contents they are supposed to be, thus becomes equally unobservable and irreproducible.

It may seem strange, indeed, that such a conception should have become traditional; but there is one explanation that makes it plausible.

For the normal requirements of living it is important to know above all in what relation the observational objects are to one another; for instance, that fire heats water and that water quenches

fire, that salt can be found in the sea, and that certain mushrooms nourish and others poison our body; and so on.

Man has undoubtedly worked in this way for thousands of years, acquiring a particular ability and making a habit of it. In this research however, he has proceeded by searching for relations between objects that are always already present, and no attention is given to the activity of observation from which the objects result. Thus, when curiosity or some practical interest led man to investigate the very activity of observation, he did not, as would have been necessary, leave aside the already present objects in order to study the activity by means of which they are constituted, but tried, instead, to keep them present by devising an observational activity that might provide a double of them inside the head.

The split between outside and inside that became applied to all contents of thought, although not directly interfering with the studies concerning physical objects, created difficulties of every kind for research on non-physical things, such as figures and mental categories, as well as for research on any mental activity.

In language it had its repercussion inasmuch as it led to the belief that only those words which indicate physical objects had a corresponding nominatum. The remaining words were considered to be either *flatus vocis*, empty words, or elements of connection, not between nominata, but between the words themselves, etc. In this way language, whose constitutive function is designation, not only was understood contradictorily, but it was also lost as a way towards thought to which, in fact, it still is the most fertile and controllable way of access.

For our research on thought and language it was, therefore, necessary as a first move to get rid of that tradition which is linked to the doubling of observational objects.

##### 5. THE OPERATIONS UP TO THOSE OF THOUGHT

An analysis of thought and its contents that accounts for every different word and every different expression by isolating as corresponding to each a different operation or combination of operations, shows that four kinds of operation are required: Differentiation, Figuration, Categorisation, and Correlation.

**Differentiation** consists in changes of state. It gives rise to *Differentiata*, each of which results, of course, from a change of state and not from a single state; it is the function of two states and the direction of the shift from one to the other.

By differentiation we obtain the nominata of words like "dark", "light", "hot", "cold", "resistant", "yielding", "green", "red", "yellow", "silence", "noise", etc. The nominata of these words, however, often contain already more than the result of a single differentiation, because as a rule we use them to designate also the results of other operations which determine their place as content of a thought (cf. below).

Since differentiation is here taken as an elementary component operation, an analysis of the differentiation is excluded for this very reason. When we speak of states and of process we refer, in fact, to

a possible investigation to be carried out with regard to the functioning organ. It does, however, not exclude that we delimit a differentiatum by naming, for instance, its opposite, that is the differentiatum one obtains from the same states but by a shift in the opposite direction; or by naming the conditions under which we are inclined to effect the shift, that is, by indicating the dependences of the functioning. In this way we can say that we have noise when we make bodies vibrate, and silence when we make their vibration cease; that we have a certain colour when we put a certain salt into the flame of a Bunsen burner, and so on. But it would be a mistake to identify the two things with one another and to say that the differentiatum, as activity, is these other things which, in fact, are observata having their own figure, their own place, etc.

Differentiation, by itself, does not produce anything figured or localised (i.e. having a place in space or time). Nor, I should like to stress, does it correspond to sensation, which is, in fact, obtained by adding to differentiation the mental category of subject. Without this distinction we could no longer discriminate "green" from "green spot" nor "hot" from "sensation of heat".

**Figuration** consists in changes of place. It gives rise to *Figures* or *shapes*, each of which results, as in the case of the differentiatum, from a change of place and not from a single place; and, therefore, it is the function of two places and the direction of the shift from one to the other.

By figurating we obtain things which, as a rule, are not designated isolatedly but together with differentiatum (above all in the activities of perception and representation, as we shall see shortly). Although in some languages we find examples such as "lance" and "lanceolate" or (Italian) "uovo" (egg) and "ovale", there are usually no words to designate only the shape of the common objects of observation such as, for instance, apple, pear, tree, dog, horse, house, etc. Most of the shapes that are recognised and designated isolatedly belong to the technical realm of geometry, as, for instance, the circle, the ellipse, etc.

Most of the shapes which we name, either isolatedly or together with differentiatum, are not constituted by a single change of place. Mostly they result from several of these changes which thus constitute the elements of the shape. Among these elements there are the simple traces (lines of the shifts) constituted by two places, that is by shifting from the one to the other; there are the composite traces which have one place in common; there are the regions constituted by a trace and a place outside the trace, that is by shifting from a trace to a place or *vice versa*; they, too, can be composed if they have a place or a trace in common; and, finally, there are the volumes, constituted by a region and a place. The configuration of volume cannot be overstepped, because a shift from it to another place must necessarily lead through a region which thus becomes a region common to both volumes and as such gives rise simply to a composite volume.

**Categorisation** consists in combinations of a particular differentiatum, namely the differentiatum of attention, of consciousness, of presence; it is the differentiatum that corresponds to words such as "watch!", "look!", "listen!", and the like. This differentiatum can be combined with others of its type, because having effected one of them, this one can either be maintained or let go while a second one is effected. If the first one has been maintained, the second one

will be temporally superimposed on it and that gives rise to the simplest categorial combination. This corresponds to an attention that becomes focussed, as for instance when the word "watch!" is followed up with the word "there!". If this first, simplest combination is taken isolatedly it is designated by the word "something" or "thing" (corresponding to the Italian "cosa" in the question "che cosa?", or to the German "etwas", etc.). The further combinations are obtained by following up an isolated differentiatum of attention with a 'something' or a 'something' with an isolated differentiatum of attention, and so on.

By categorising we obtain things which are designated by words such as "or", "and", "not", "cause", "effect", "singular", "plural", "being", "can", "must", "want", "time", "space", "free", "necessary", "probable", "number", "point", "line", "surface", "substance", "accident", "subject", "object", "state", "process", etc.

Also mental categories are very often designated together with results of other operations. As an example, it is sufficient to think of the singular and the plural which occur in conjunction with the nouns of almost all languages.

Every differentiation, figure, or category can be combined with other elements that may be of the same or of the other two kinds; and the resulting combination derives its individuality from the particular elements combined in it, to the particular order in which they have been combined, to the time taken to combine them, etc.

Among the most usual combinations are perception and representation.

Perception consists in the following operations carried out in the following order:

- a) a succession of two differentiatum, and
- b) categorisation of the second differentiatum as object;
- b') the object-differentiatum may be given a shape (by figuration which, in every case, is guided by the separation between the two differentiatum).

In representation we have:

- a) a categorisation of something as object, and
- a') possibly, a figuration of this (in this case the figuration is free), and,
- b) differentiation of the object and the figure, that is, addition of a differentiatum to the object-figure.

This break-up into operations explains why perception is always felt to be constrained, or obligatory, in comparison to the sense of freedom, that is characteristic of representation. In perception, in fact, the object, being the result of the succession of two differentiatum, arises always coupled with something else, that is, together with its background, with another object, or with a determinate spatial or temporal relation; in representation, instead, the object arises without any link whatever. And this analysis in operations explains also why representation has always been felt to be poorer than perception even if it is always possible to effect a comparison between a representational and a perceptual result.

Other very usual combinations are the physical things and the psychical things. The first are the result of a spatial categorisation, the second the result of a temporal categorisation of the differentiatum;

hence a physical thing must always be in a certain place and distinct from at least one other thing in another place; whereas a psychical thing must always be at a certain moment and distinct from at least one other thing at another moment.

## 6. THOUGHT AND ITS VERBALISATION

If we ask ourselves which word designates the flow of the three kinds of operation we have so far discussed, there arises a terminological question: should the word "thought" be used to designate this flow already before any fragmentation into operational elements, or only after this fragmentation? And if after fragmentation, already with a discourse matching it or without any verbal accompaniment?

We believe that our current dictionary reserves the word "thought" for a certain fragmentation, or even a certain verbalisation of the operational flow (which, further on, we shall define with greater precision).

Moreover, a presentation of the flow as single operations, sequences, and groupings, already stems from a particular fragmentation effected with a view to articulation into linguistic units.

Some fragmentation, in any case, is necessary in order to achieve the conditions under which the flow can be accompanied by words. If one had to assign an individual word to each flow, or train of thought comprised between two pauses or stops, one would have to fix an unlimited number of semantic relations with, consequently, an unlimited number of words. In fact, we should have to spend our lives preparing this linguistic material without ever getting round to using it. In order to serve their actual purpose the semantic conventions must be of a relatively small number - small enough to be passed on and to be learnt during a short first period of our lives.

The criterion adopted for the articulation of the operational flow for the purpose of designation had to be one of economy: to isolate as units those single operations and combinations of operations that occur most frequently, and to leave the less frequent ones to be composed by combining the frequent ones.

For instance, "violet" and "light" certainly recur much more frequently in many combinations than does the particular situation "light violet"; and thus they are designated individually while the designation of the rarer situation is obtained by their combination. This is still more obvious, for instance, with the "singular" and the "plural"; sometimes they are found in conjunction with "horse", at other times with "tree", then again with "chair", etc. Thus they have been taken as units of designation and "horse in the singular" or "horse-" and "horse in the plural" or "horse-s" are obtained by combining them.

Among the operational elements that are to be taken individually we have single as well as composite differentiatata (composite above all if their localisations coincide giving rise to the materials), observata and their changes, mental categories and their applications, etc.

The most frequent of all are perhaps the operational elements that represent the modalities with which one passes from certain designated things on to others. If, for instance, one separates two things which initially occurred together, the designation is "with"; if one

unites two things which initially occurred separately, the designation is "of"; if, in a succession of things, attention remains focussed on all of them, the designation is "and"; if the focus of attention is shifted during a succession of things, the designation is "or"; etc.

The criterion in the examination of how, in a particular language, the operations have been grouped into units of designation, is the classical criterion of considering as a unit all that can be separated from one combination and used in another while preserving unchanged its phonic or graphic designatory material and its signification (or at least the latter).

At this point we have to consider the temporal relations subsisting between the operations that constitute an actual operational flow.

Once one has applied the analytical criterion we indicated when we spoke of three kinds of operation we have, of course, the possibility of diverse rhythmical figures within the single designated units. We can at once individuate the four possible rhythmical figures:

- 1) the operations are carried out simultaneously; that is to say, they begin and end together;
- 2) the operations begin one after the other, but end at the same time;
- 3) the operations begin at the same time, but end one after the other;
- 4) one of the operations begins before and ends after the other.

If two operations are carried out in such a way that the first ends before the second begins, it brings about a halt in the operational flow, an interval of non-operating such as occurs, for instance, when we "switch to another thought" or "stop one train of thought and immediately, or some time afterwards, embark upon another". The more complex rhythmical structures, however, result from combination of the four possibilities given above.

Also the single designated units are found to be in these temporal relations. To examine them becomes necessary if we want to account for the way in which a language designates an actual flow of operations. In fact, it is not enough if, in a discourse, there figure only the particular units derived from the fragmentation of the flow. The discourse must also contain the designations of their temporal order, that is to say, designations in order to constitute rhythmical figures consisting of the four possible ones and their combinations.

Before examining the possible designatory solutions, I should like to return once more to the terminological question concerning thought. We believe that the word "thought" is applicable to the operational flow only where this flow is articulated into operational elements that are grouped in a unitary structure by means of diverse rhythmical figuration. In particular the rhythmical composition corresponding to correlation is constitutive of thought; this rhythmical composition results from two operational elements succeeding one upon the other as correlata, or modalistata, while a third - the correlator, or modality of transition - persists; it



is the structure characteristic of all relations.

In fact, it is this structure which determines, more than any other, the unmistakable dynamism of thought, for every time we are presented with a correlatum and a modality of transition we necessarily have to wait for a second correlatum. Our thinking thus proceeds by a continual opening and closing of correlations. Expressions such as "fish and ...", "either red or ...", "a piece of ..." show the opening of a correlation and the consequent state of expectation and suspense that ceases as soon as the correlation can be closed ("fish and fowl", "either red or black", etc.).

However, there is some dynamism also in other types of modality, for instance in those of construction. "To want" and "to be able", for instance, - the first indicating that two equal developments attributed to the same subject must temporally succeed one upon the other, the second indicating that two different developments attributed to the same subject must be temporally superimposed one upon the other - if applied to another development such as "to go", distribute the going and its subject in the particular temporal order indicated by them. But the modalities of construction and their modalista are coincident and they do not create a void that has to be filled, as do the modalities of transition.

Stressing the parallel with music, one might say that the various other modalities correspond to the establishing of temporal relations between the single notes, whereas the modality of transition, the correlation, corresponds to the very bar itself.

A correlation can, of course, figure as a correlandum in a larger correlation; and, as a rule, our thoughts are constituted by a network of correlations, or correlational net. For instance "the meat and the fish are in the refrigerator" already represents a correlational net.

With regard to the discursive accompaniment, or verbalisation, of the operational flow, we can say that it is not absolutely necessary, in order to have "thought". Nevertheless discourse was presumably the prime motive for its articulation; by now, in any case, discourse, either precedent or subsequent, accompanies thought so universally that thought without discourse would be exceptional rather than the rule.

## 7. THE DESIGNATIONS

Different languages, in their presentation of single designated units and their temporal order in thought, display considerable similarities owing to similar exigencies; but they also display certain differences, especially when they are examined in their written rather than in their spoken forms. In some languages, for instance, there is a tendency to maintain in isolated words the single units of designation, and for this reason we call them isolating languages. In others, the inflecting or agglutinative ones, there is a tendency to group several units of designation together in one word; hence we speak of a root with suffixes, prefixes etc., of compound words, and soon. In spoken language, of course, such differences more or less disappear.

Usually, in any language which is to be spoken, the various units of designation have to be presented in linear succession; and

all languages use this sequence to impose an order on the flow of operating. But in this connection a complication arises: while the single units of designation can only succeed one another, each giving place to the next, we have seen that this is the only rhythmical figuration which cannot occur in thought. The figurations which, in thought, determine the units are always such that their elements present themselves simultaneously in one way or another; in other words, the units of designation appear in one single line and, thus, have monodic character; the designated units, on the other hand, appear on several lines and have polyphonic character. This means, in fact, that we carry out several operations concurrently, but designate them successively, and that we receive the indications one after the other, but carry out the corresponding things simultaneously.

It may be of interest to remember that polyphonic music began when a system of written notation had been devised which indicates temporal superposition by vertical alignment. But since in spoken language this is not possible, it is necessary to have conventions which will permit us to make a connection between the two sequences. And thus the listener or the reader always has to wait until a certain number of units of designation has been presented before he can effect the temporal arrangement of these units.

#### 8. COMPREHENSION IN KAN AND COMPREHENSION IN THE MACHINE

So far we have presented the operational flow, its fragmentation or articulation into units of designation, and the subsequent semantic conventions which give the units of designation their place in the flow, as something which could and must take place if an actual operational flow is to be accompanied exhaustively and univocally by its verbalisation.

In fact, however, the historical development has not been like that. On the one hand there was certainly a lack of operational awareness and also of a general plan of action; on the other hand presumably the same criterion of economy which, as we have seen, was applied in the choice of those operational elements that are to be taken isolatedly, has also exercised a certain influence. In our discourse we necessarily have to designate all those parts of a given operational stream, which cannot be known in another way, or from another source; but why - if the economy of designation is to be considered - should we designate things that everyone knows already?

Thus it has come about that texts sometimes lack those designations which would assure a univocal transition to the designated operations. Hence there are words that are linked to two or even more different operational elements; and, above all, it happens that the designations supplied by the words' places in their succession are equivocal. If the text is being understood by a man, he eliminates these ambiguities, because in the light of his knowledge only one alternative makes sense (see the examples in PART II).

Now, if, instead of a man, a machine is to understand the text - and in order to translate, to summarise, etc., it has to understand it univocally - it will be necessary to supply this machine not only

with all the semantic relations but also with a fund of integrating knowledge equivalent to that of the man. There is no reason why this should not be feasible, but at least for the time being, there are in practice two kinds of difficulties.

The first arises from the way in which men learn and remember things, which, at our present stage of technological development, is not reproducible. For man, to learn things is to carry out the operations that constitute these things; and to remember them is to set up an organic situation, that will always function in the same way. The memory of machines, on the other hand, has so far always been devised by means of registration, that is, it works on rigid non-intersecting lines and this makes it anything but suitable to function as a fund of encyclopedic knowledge.

But the second, greater difficulty arises from the fact that man, when he learns to think and to speak, actually carries out all the operations that he verbalises, whereas no machine could at present carry out all these operations, because neither engineering technology nor psychological analysis have reached the required stage of development. (Our research group is working on a project for a model capable of carrying out some of the human operations of observation, mental categorisation, thought and language, and will begin its construction before the end of this year on behalf of Euratom; but it is a very limited model and, even so, it has required extremely long preliminary analyses.) In any case, the machines employed at present for the immediate practical purposes of MT etc., are of the computer type which replaces the actual carrying out of operations by code numbers; and these code numbers in the machine represent only the results of our analyses of the operations designated by a discourse.

This imposes certain limitations. What unit of discourse is to be taken for analysis and codification before input into the machine? Obviously not entire texts, for this could be done only with texts which have already been composed at the time of the machine's construction, and in this case the machine would serve to do no other work except the work already done by us. Moreover no registration memory at present in existence could possibly contain as input units the variety of existing texts and the results of their analysis.

Hence there is a general tendency to restrict analyses to the manageable number of units represented by single words. Given this restriction we can ask ourselves what operations are designated by each single word; and this analysis must show both what the word designates as an operational fragment apart from the operational flow, that is, apart from its place in the rhythmical figuration in which it occurs, and its function in the building up of the rhythmical figurations characteristic of the flow of operations.

#### 9. THE ANALYSIS

In our work aimed at mechanical translation of Russian, English, Italian, and German we have chosen the word as the unit of discourse to be analysed, and we have left to the machine the task of reconstituting the operational flow from the data concerning the operational elements which are designated by the words of the four

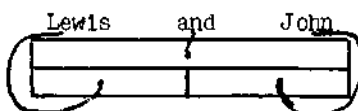
languages and which we supply to the machine.

It was, however, also necessary to decide the limits of this analysis and thus also the limits up to which the flow of operations is to be reconstituted.

By way of an experimental solution we have come to the following decisions:

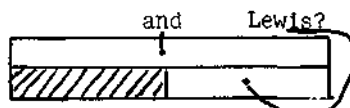
- a) never to give a rhythmical figuration to something which in one language appears as designatum of one unit of designation; that is to say, this designatum will always be taken as a component of a rhythmical figure;
- b) to arrange the designated things in two kinds of structure:
  - 1a) the explicit correlational structure in which the three correlanda (i.e. the two correlata and the modality of transition, or correlator) occur with the respective explicit designation.

e.g.



- 1b) the implicit correlational structure in which one of the correlata is not designated, either because it has been or will be designated in a context outside this correlation, or because it concerns the speaker and not what he says, and so forth.

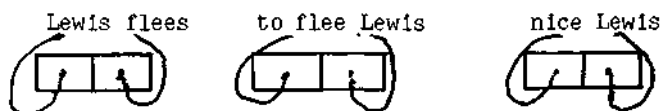
e.g.



In these structures the particular modality of transition is the determining factor.

- 2) the binary summative structure in which the resulting unit depends on the characteristics and on the order of the addenda.

e.g.



With regard to these summative structures, however, it has to be stated clearly to what extent one intends to keep apart, as separate addenda, the things designated by words. Here too, we have decided never to go beyond the single designated units and, further, to consider as units all the things that are given on a special list.

- c) to consider the operational flow broken off whenever no modality of transition, or correlator, has been designated - even if the things before and after the break suggest a specific rhythmical figure that might link them in our thought, but which, in order

to understand the text, can be gathered only from the two correlata (as happens, for instance, in the case of two sentences separated by a full stop).

## II TYPES OF ANALYSIS AND CLASSIFICATION FOR INPUT

AFTER the summary remarks on the verbalisation of thought (see PART I, section 6), we now have to consider what kind of and how many semantic connections are necessary in order to establish a univocal relation between thought and language.

We will only examine here those structures which are of correlational type.

As we have seen, every correlation is constituted by three elements which are characterised in two ways: a) as what they are isolatedly, b) according to the function they have in the formation of the correlation, that is to say, whether they are being used as first or second correlatum, or as correlator. From this it follows that a correlation requires at least five distinct indications: three concerning the particular things that are being correlated, and two to indicate the function of at least two of the three things (since the function of the third can be inferred from the other two). The indications must be at least five, since not even the one concerning the modality of combination or construction can be left out, for the kind of thing that usually functions as correlator can sometimes occur also without that particular function; although this is comparatively rare. It does happen, for instance, in the thought structure one refers to when one says: "And and or are modalities of construction".

The various languages have chosen different ways for supplying these necessary indications. For obvious reasons of economy the preference has been given to two basic methods of indication: a) the particular form given to the word; that is to say, part of the word is used to indicate the particular correlational function it is to have; b) the place given to a word relative to the others, i.e. its position in the propositional sequence of words. The situation becomes more complicated whenever a designation points to more than one correlation. In such a case the alternatives are no longer those admitted by two or three words that designate the three things with which the three places of a single correlation have to be filled; instead the alternatives are those admitted by many words placed in a linear sequence, for they must designate the many things which are to fill the many places of the correlational net. With regard to this it is important to realise that the formal characteristics of words, if used to designate the particular things put in correlation as well as their function in constituting the single correlation, can no longer be employed for the groups of words designating the whole correlation, which, nevertheless, must be an element of another one. Two solutions have been evolved for this problem: a) stress and timing; in other words, the play of accentuation and pauses in speech, and a system of punctuation marks in writing, and b) - at least in a great number of languages - formal grammatical agreement, that is to say, discrimination according to gender, number, person, and finally, according to case (wherever the case has not a direct designatory function). Thus, where the position in the sequence of words gives no indication, it is owing to these other kinds of indication that we can decide of which correlation the designated things are a part.

Although the system of classification is still under revision we will give here an outline of how analysis is at present carried out.

The input vocabulary for our current programme includes about 50,000 inflected Russian forms, corresponding to about 2,500 headwords and punctuation marks. At the first level these were divided according as they can or cannot designate contents of thought. Four classes have been distinguished in this respect:

a) constructive words, i.e. words corresponding to correlanda. These in turn are divided, according to the number of correlanda which they represent into: monoconstructive, if only one correlandum corresponds to them, polyconstructive, if more than one correlandum corresponds to them, in the same or in different correlations (see *figs 3 and 4*).

b) directive words, i.e. words which only indicate operations to be performed on correlanda. Purely directive words include quotation marks, whose function is to indicate that what is between them is to be treated as a correlandum:

e.g. "A" is the first letter of the alphabet.

Another function of punctuation marks is to indicate that a correlandum is exercising a correlational function which is not its characteristic one:

e.g. "a" and "the" are articles.

An example of a purely directive word is, in Italian, the accent introduced to eliminate a dictionary polysematicity. There are in this language cases in which an identical spelling has more than one meaning:

e.g. àncora (anchor)  
ancòra (still, yet)

In the same way a change of typeface or letter size can be considered as directive.

In our figurative representation purely directive words do not occupy any place in the rectangles.

c) there are other words which, according to the particular context, may have constructive or directive functions. These include the comma, constructive when it acts as a correlator, directive when it only marks a pause of grouping

e.g. Hadrian, (constructive) Emperor of the Romans, (directive) where the first is the correlator of the apposition correlation, while the second simply indicates that the appositive group is finished.

d) There are also words which always exercise both functions. The full stop, for example, on one hand indicates that what has gone before is to be taken as a unit of thought, and on the other represents the correlator between one sentence and the next. If after the full stop the text does not go on, the mental category "end" will figure as second correlatum.

This kind of classification is only made on the word matrices, it will not appear on the product matrices nor on the construction-control matrices, since it belongs to the preconstructive classification of input.

After the words had been divided into directive words and constructive words, analysis was next applied, of course, to constructive words, in respect of the operations which they represent.

Since the Russian language is an inflected one, we first analysed the suffixes of nominal flexion, or declension, and of verbal

flexion, or conjugation. Other suffixes, apart from their designation of correlational function, were analysed in respect of agreement, that is, when the words designate a certain correlation by the use of identity of case, person, number, or gender. For the systematisation of agreement it was necessary to classify not only the nominata but also the words. It is well known that in many languages the gender of the word corresponds only rarely with the sex of the nominatum. For example, some names of animals in Italian have only one form to designate both male and female, and this form is sometimes masculine and sometimes feminine ("il leopardo", "la pantera", etc.). The gender of the names of objects is always conventional; some names masculine in the singular become feminine in the plural ("il lenzuolo" and "le lenzuola"); and others *vice versa*. Classifications of this type appear on all matrices, and they appear in the columns under the items "Agreement" and "Individuation".

There is, however, another solution which arises by itself and which, at least within certain limits, makes it possible not only to do without any rule of the kind mentioned on page 16 but also to go against such rules as have been established. To understand how this is possible we have to remember what has been said about the unitary flow of thought and the way in which this is broken up into correlational structures. Certain things in it not only have arisen together, but usually arise together, at least inasmuch as regards that particular relation that constitutes the modality of their construction, even if this is indicated separately. If they are then broken up, every one still knows or notices their common origin, their reciprocal appurtenance. For instance, a quantity can be large, or small; but that could not apply to material, which never appears in pieces. Thus, if one says in Italian "quantita di acqua grande" no one could help understanding that "grande" has arisen together with "quantita", and not together with "acqua", and that it refers to the first, even if its position in the sequence of words, according to the normal rules, would imply that it should refer to the second. And similarly, if one says "The trees and the fruit which hung from the branches...", no one could not understand that the relative refers only to "fruit", and not to the "trees"; but the situation would be different in a sentence such as this: "The trees and the fruit which the land produces...", where the relative can correctly refer to both the "trees" and the "fruit". This situation is very common, and always occurs when an expression containing "and" is referred to by a relative or followed by "from", "of", etc. See, for instance:

"The boat and the fish which he has filled ...", but

"The boat and the fish which he has bought ...".

It.: "Una pozza di acqua circolare", but

"Una pozza di acqua sporca".

It.: "Un cassetto del tavolo aperto", but

"Un cassetto del tavolo apparecchiato".

Such expressions produce, at the formal level of analysis a double correlational net, and, if the output language requires a particular agreement, they produce a double output.

Our representation of things already places them in a certain way and suggests certain relations, while others are excluded; this is so at least wherever there is a choice of alternatives.



One of the things which most helps a human translator in understanding a text is the whole representational world which words continually evoke. It is on the rich representational material that the human translator really begins to work, guided only secondarily by the formal suggestions of the words. Someone who reads, for instance, a newspaper headline sets up a whole mental and representational network; if at a certain point in the text a word is found that the dictionary defines polyvocal, it will nevertheless be understood immediately in one single sense (the sense best adapted to the general setting). In fact the human reader, or translator, more often than not does not notice polyvocality in a text at all. But the machine, only to eliminate polysematicity, has recourse to a "notional sphere".

In the case of a machine, if we want the input to be limited in number, the largest pieces acceptable will have to be words, or even smaller units; and of the things designated by these units one will have to indicate from the beginning all the correlational possibilities of the two relevant kinds (i.e. which things are put in correlation, and what correlational functions they carry out). Then, as the words are put in one by one, one will check what happens when the things indicated by them meet one another. The makers of our languages to some extent relied on the relations between *nominata*, which relations are known independently of their linguistic expression in a discourse.

Thus they did not always provide all the linguistic indications to assure the correct connections between language and thought.

In order to overcome this difficulty it will be necessary to carry out another kind of analysis which has to concern just these relations between the *nominata*; and this can be done in two ways: a) by breaking things up into the smallest possible component operations, and assigning to the machine as its program the task of finding the possible relations between things by examining their compositions; b) by analysing the relations between certain things, and things in certain relations, as notional sphere, and, that is to say, as a unit of knowledge.

In the last two years, we have isolated about 100 salient types of relations, in a field of 500 headwords (see *figure 9 and Appendix I*). It was not always easy to find the salient relations and the formulation of them was not always simple. For one reason in fact a long chain of indirect relations was sometimes involved, at another time a more or less sophisticated cultural reference is called upon.

An interesting example of this arose when, in the course of making a figurative representation of the correlational net of the Lord's Prayer (Latin input) for a Euratom Report, we realised that the words "fiat voluntas tua sicut in caelo et in terra" give rise to as many as three different correlational nets.

- I. fiat voluntas tua sic- (-ut in caelo et in terra),  
where the whole expression "ut in caelo et in terra" is taken as a term of comparison (the indicative "facta est" is to be interpolated in the expression).
- II. fiat voluntas tua sic- in caelo -ut in terra,  
which gives exact parity to the two terms (sive ... sive), and interprets "et" as referring back to the earlier "ut".
- III. fiat voluntas tua sic- (-ut in caelo) et in terra,

where "ut in caelo" is taken as a term of comparison, and "et" equal to "etiam" (the indicative "facta est" is to be interpolated in the expression "ut in caelo").

In such a case the human translator might be in doubt as to which of the three structures should have the preference, given that we become aware of all the three possibilities, but he would certainly give a single final translation, choosing only one of them. But at this level it still proves difficult to make a satisfactory analysis, apply classifications, and formulate rules.

In any case, given the actual situation of the relations between language and thought, the usual discrimination into syntactic analysis and semantic analysis is not the most convenient; even if there is no other reason but that inherent to the conventional way in which these analyses have been understood, and which contains the supposition that language consists of two parts: meaningful, or semantic words, and logical or syntactical words. This division is misleading, because if something enters to make part of language, and does not remain mere phonic or graphic material, it is always because one considers its designatory function.

A satisfactory program of analysis aimed at the mechanisation of understanding a text - even if this is limited to its substitution by the correlational net that corresponds to the text when a man understands it, - would, in my opinion, have to comprise, like the following, an analysis in four directions:

A) Examine:

- 1) all that a word, taken singly, designates by means of its form and all it designates taken in conjunction with others, by means of its place in the sequence;
- 2) completion of this examination by an examination of the relations arising between the nominata, as a result of their operational contents.

B) Discriminate neatly the designations:

- 3) into designations with the purpose of indicating which particular things are to be put in correlation; and
- 4) into designations with the purpose of indicating the correlational function these things have in constituting the correlation; that is, whether they function as correlator, as first correlatum, or as second correlatum.

SOME EXAMPLES OF

1) Monoconstructive Words

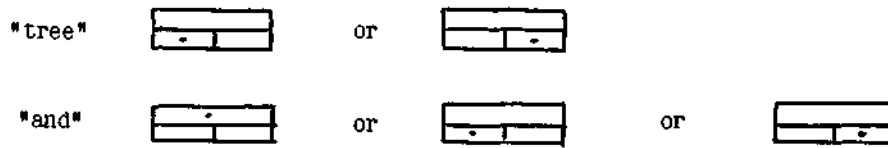


fig. 1

2) Polyconstructive Words

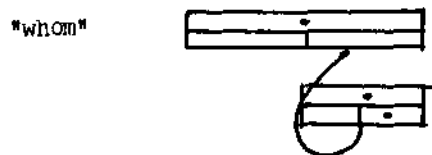
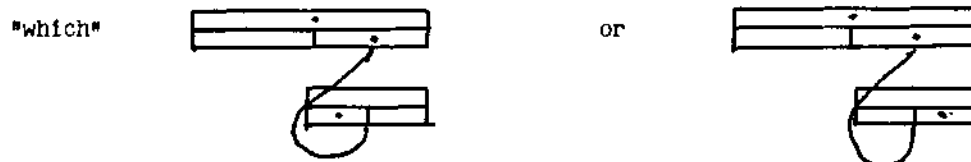
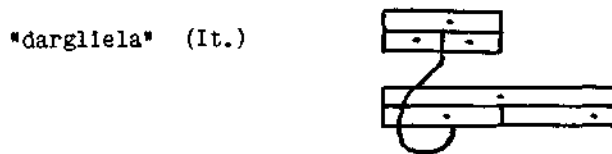
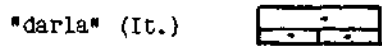


fig. 2

(see the examples of relative pronouns: figs. 3 - 8.)

1) LUX QUAE SOLE EMANAT

1 2 3 4

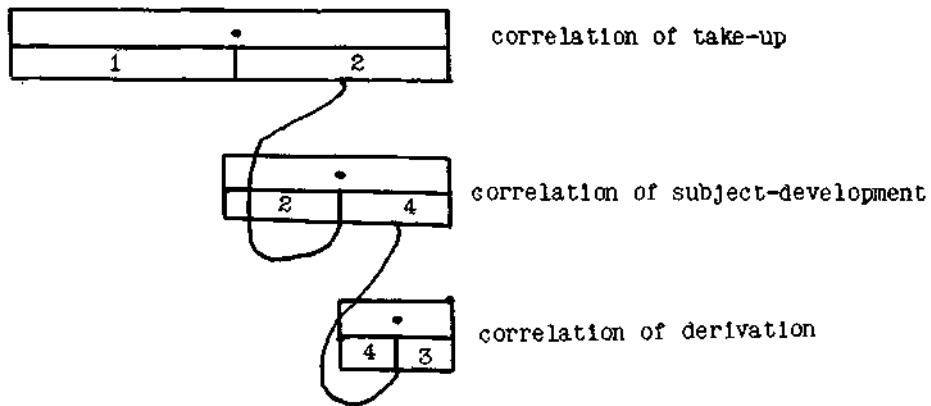


fig. 3

2) LUX QUAM SOL DIFFUNDIT

1 2 3 4

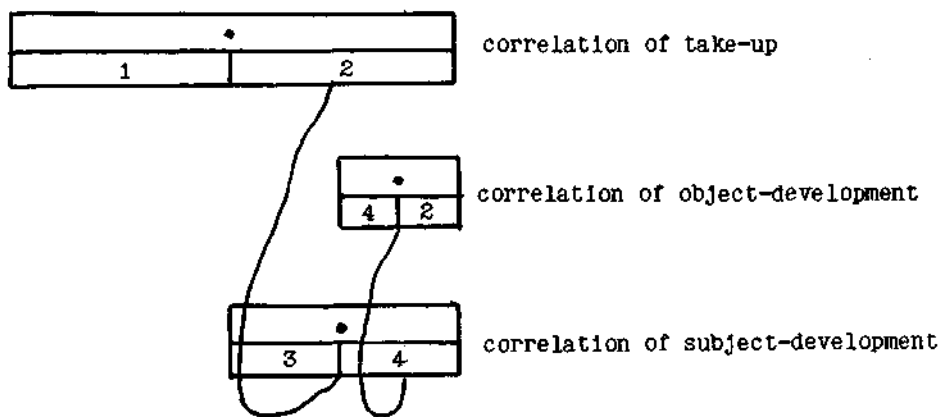


fig. 4

3) I AM WHERE HE IS  
 1 2 3 4 5

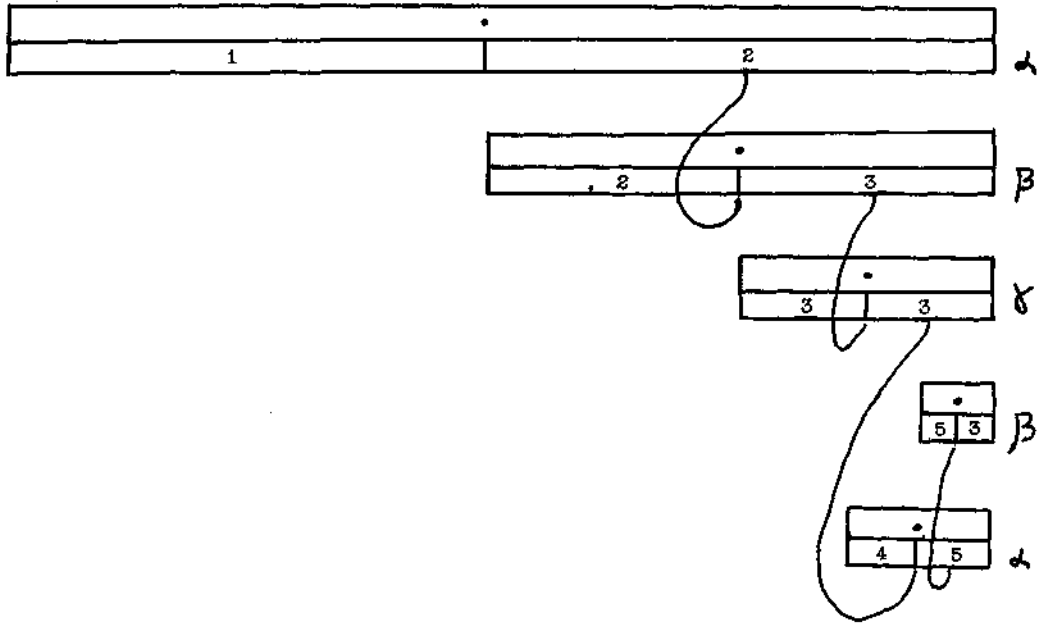


fig. 5

4) I AM WHERE HE GOES  
 1 2 3 4 5

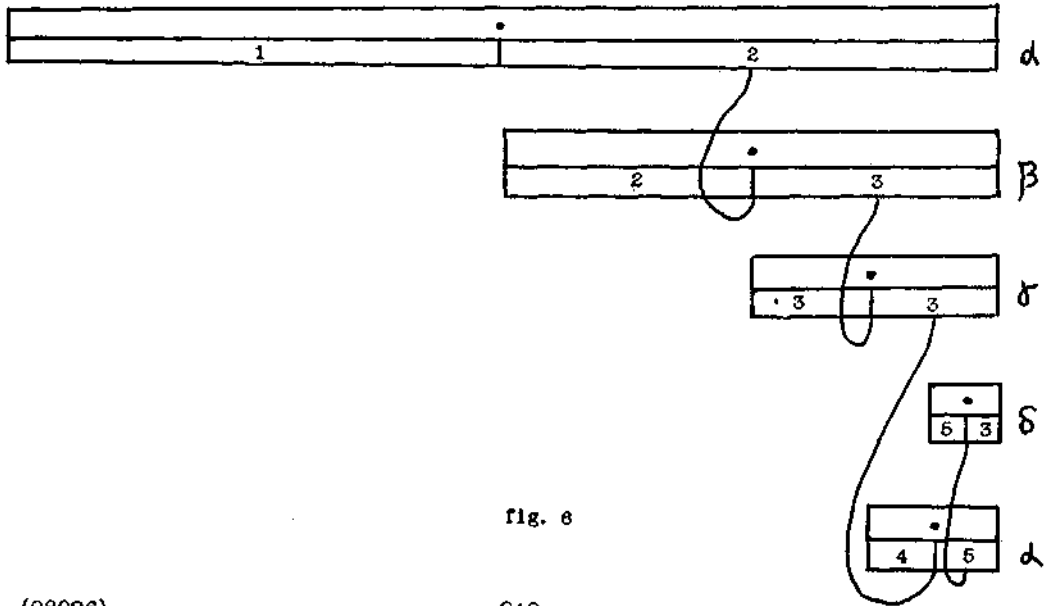


fig. 6

6) I GO WHERE HE IS  
 1 2 3 4 5

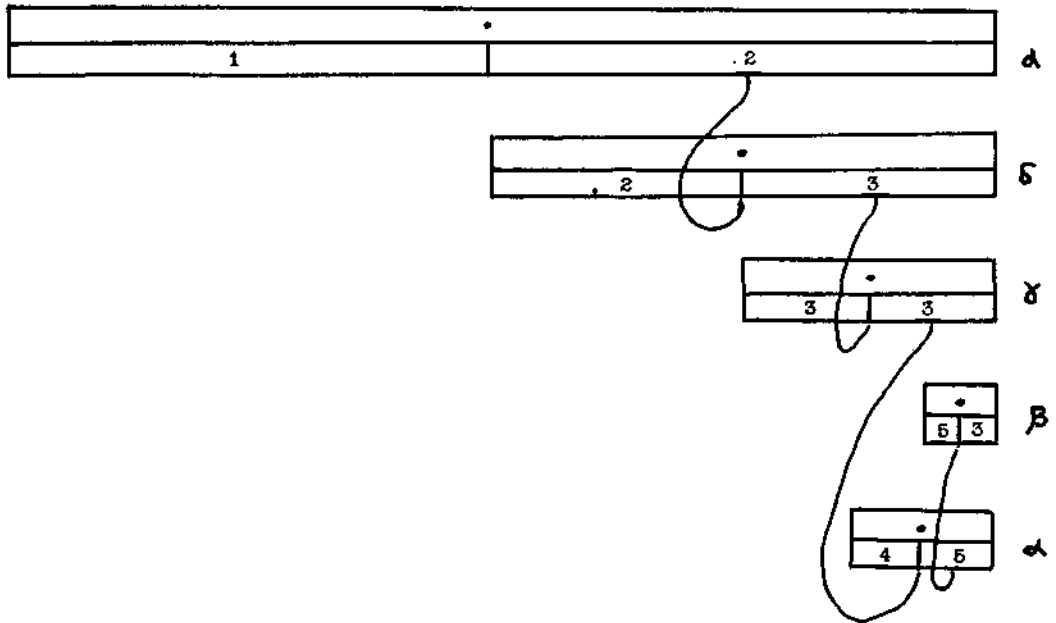


fig. 7

6) I GO WHERE HE IS  
 1 2 3 4 5

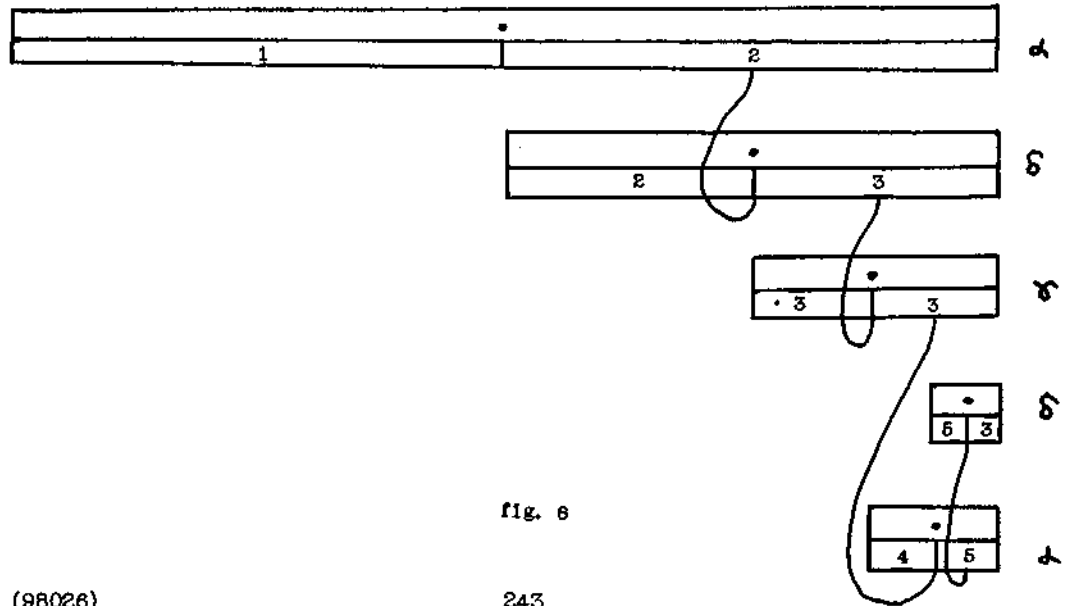


fig. 6



## APPENDIX I.

## NOTIONAL SPHERE - LIST

NUMBER OF THE RELATION	1st TERM	2nd TERM
1	element	collection
2	specimen	class
3	species	genus
4	part	whole
5	container	contents
6	support	supported
7	thing pulling	thing pulled
8	thing looking after	thing being looked after
9	thing directing	thing being directed
10	component	composite
12	ornament	thing adorned
13	thing provenient	provenience
15	consecutive element	thing composed
16	thing preceding	thing preceded
18	object	accessory
19	contiguity	contiguity
20	opposition	opposition
21	ascending relation	descending relation
	(3rd TERM: collateral relation)	
23	object	cover
24	means of protection	thing from which they protect
25	historical appurtenance	historical appurtenance
26	thing generated	generator
27	thing generated	ambience of generation
28	first stage	second stage of genetic
33	object	physical state
34	material	form
35	constitutive element	thing constituted
38	economic relation	economic relation
39	semantic relation	semantic relation
40	subject	activity
41	subject	material
42	subject	result
43	subject	instrument
44	activity	object
45	activity	result
46	activity	instrument
47	activity	modality
48	activity	ambience
49	activity	habitual place
50	activity	time
51	material	product
52	object	instrument
53	object	ambience
54	object	habitual place
55	object	time



NUMBER OF THE RELATION	1st TERM	2nd TERM
58	organ	function
61	object	material
62	complementary instrument	complementary instrument
63	instrument	result
65	subject	place
66	activity	material
67	complementary object	complementary object
68	subject	object