THIS QUESTION OF LATTICE THEORY

by
R.M. Needham & A.F. Parker-Rhodes

The Cambridge Language Research Unit have, for some time, been using lattice algebra as a semantic theory in its studies on Mechanical Translation and Information Retrieval. Its doing so has been criticised on the grounds that no semantic theory is possible; on the grounds that lattice theory is too strong to function as a semantic theory, in the sense that the algebra of partly ordered sets is more suitable; on the grounds that lattice theory is too weak to function as a semantic theory since it is so general that it will fit any state of affairs.

In order to say what a semantic theory is, as misconceptions on this are probably at the root of statements that it is impossible to have one, it is probably easiest to give, by discussion of the M.T. process, an outline of what semantic theory is for. An entry in a mechanical dictionary has typically consisted of the following:

a. An input language word, or part word.
b. Grammatical facts about it.
c. One or more output language equivalents for the input word, with, in the case where there is more than one equivalent, marks of some kind for effecting a choice between them,
d. Grammatical facts about the output language equivalents,

I wish to concentrate here on the nature of the marks for choosing in (c). Usually some of the choices correspond to different grammatical uses of the input word. In this case, there will be a kind of correspondence between (b) and parts of (c). For example, (b) might contain "noun or verb", and (c) might contain "if used as a noun, render as P or Q; if used as a verb, render as R." However, it is notorious that not all choices can be made on such grammatical grounds as these, and the difficulties involved in making the choices have been a major obstacle to M.T. research for years, under the names, (in order of increasing pomposity) of multiple meaning, semantic ambiguity, and polysemy. In M.T. contexts, the purpose of a semantic theory is to enable the program for M.T. to make these residual choices correctly, or to avoid the necessity for such choices.

I shall now discuss the problem that has been separated off above in greater detail. It is agreed on all sides that if the remaining choices can be made at all, they must be made on the basis of context, that is to say, on the basis of the surrounding discourse to some undetermined extent. Assuming, therefore, that this is the policy that will be adopted, it is necessary to find something that will do for the discriminating marks. It is here that some mathematically sufficiently strong notation (i.e., a theory) is needed. If no such notation is admitted, the only possible solution is explicit anticipation and listing by the dictionary maker of the precise verbal contexts in which the word in the input will be used, which is not only in practice, but, in principle, impossible. What is required is a closer classification of contexts and a notation for the classes.* The first attempt that was made involved the use of subject classifications of contexts (Re. e.g. Booth[(1)]). It was assumed that for M.T. purposes all choices between alternatives could be made by specifying the context matter of the discourse, as, for example, physics, biology, etc. Thus section (c) of the dictionary entry would perhaps contain renderings "P (physics) Q (biology) R (psychology)". Supposing that it were possible to decide which of the classes was "on top" at any stage, it would then be possible to select unambiguously from the list of renderings. Usually the same class would be "on top" throughout a piece of discourse. It is worthy of note that this method involves giving a context class even to words which are quite unambiguous in order that it may be possible to compute as the text runs on which class is "on top" at a given stage. There is a strong case for saying that classifications of this kind are not good enough to achieve the desired result. Many words have far too many different uses for this method to be feasible,

---

* That such a classification is possible follows mathematically (i.e. for practical purposes in a very weak sense) from the fact that there is at most a finite number of renderings for an input word, viz, the total output vocabulary, while there are an infinite number of possible and actual contexts. It is thus impossible for each context to affect the rendering in a quite individual way, which is the same as saying that there are classes of contexts.

and they differ far too subtly from one another. Now it familiar from studies in library classification that the generalisation of such a scheme as this is to use, instead of an exclusive list of classes, a lattice of classes. This follows straightforwardly from the fact that given a pair of classes defined by properties A and B, it is possible to conceive the class consisting of things with both the property A and the property B, and the class of things with at least one of the properties A and B. This system, for any given finite set of properties, generates a lattice in the sense of Birkhoff [(2)]. For the justification of this statement, see Birkhoff and Maclane [(3)], Mooers [(4)], Fairthorne [(5)], and many others. It is thus possible to consider a lattice of classes, of any desired largeness, which will provide the class-marks attached to the separate alternatives. Now these classes are by definition classes of contexts. They are a generalisation of such contexts as physics contexts, human relations contexts, musical contexts, and so on. It is not my purpose here to argue in favour of the possibility of making any classification of contexts which affect the use of words of a language. The reader is referred to the companion paper by Margaret Masterman on "What is a Thesaurus?" where this question is discussed at length. The point that is to be made here is that, if such a classification of word-affecting contexts is possible, the scope and subtlety of the tags for semantic choices can be greatly increased.

With this increase in subtlety of classification must go an increase in the subtlety of the means by which the correct choice is made. It is no longer possible simply to decide for example "This document is about biochemistry", and expect all the choices to be determined by this one decision. It is necessary to compute for each word something called a <u>specification</u> that will correspond exactly or roughly to the context tag for (we hope) one only of the given equivalents.

The step that is next taken, therefore, which is also the step which distinguishes the Cambridge language Research Unit approach from (as far as we know) all others, is this.

The context tag attached to an equivalent E for a word W is a description of a context (an extralinguistic context, that is) in which W is used, and therefore a description of a similar context in which E is used. Because both of these contexts are extralinguistic, it is thus possible to regard this tag as a characteristic of E quite independent of W, and to assemble an Output Dictionary consisting simply of a list of output language words together with extralinguistic context tags which would have accompanied with them when they occurred in lists of equivalents. The Cambridge Language Research Unit asserts that it is possible to use the specification mentioned in the previous paragraph to select, not from a particular list, but from the whole output dictionary, thus avoiding the process of attaching a short list Of equivalents to an input language word, and making possible the use, in effect, of very much larger ranges of renderings. This assertion amounts to a very high claim for the selecting power of the specification; it also makes immediately plain the extent to which the notation of the specification is, and is not, an "interlingua". That is, it is an interlingua in that it is the notation which is used to pass from a word in one language to a word in another; it is not at all an interlingua in the sense of an intermediate language which could be used for independent communication. It is on the basis of this assertion that the Unit has proceeded for the past two years, and a number of experimental results are available which will shortly be described or referred to.

It will be obvious from the nature of the classification of contexts that the selecting power of a specification, in terms of contexts, will be limited in various directions. I shall use the terms "Resolution" or "Resolving power" to refer to the extent to which a scheme of contextual classification distinguishes the words of a language one from another. If there is a set of words which can occur in virtually the same contexts, as can for instance the names of various vegetables, they will not be distinguished by the context classification. The distinction between such words rests upon an entirely different principle, to wit, different (contingent) facts about the objects to which they refer. The experiments performed to date accept this lack of resolution, though methods are now available to deal with it to some extent (see below).

Our experiments have made use of Roget's Thesaurus as a classification of contexts. This document is in the form of 1,000 heads, or type-contexts, words being placed in as many as can be justified by their uses. It only contains English words, but it is possible to classify words in other languages according to the same scheme (e.g. Dornsieff ( )). I shall henceforth speak of the class of contexts of a word as its head-set. that is to say, the list of heads in which it can be put. An example of its use is as follows. In an experiment on the semantic translation of a Latin sentence, the following procedure was adopted[6]. Each Latin chunk was given a context classification by a skilled linguist in the form of a head-set. Of the heads in the head-sets for the chunks, those were rejected which did not occur elsewhere in the sentence. The remaining "purged" head-sets of the chunks were then used as specifications, and the cross-reference index of Roget's Thesaurus was used as an output dictionary. Those words from it were selected which contained in their head-sets as many as possible of the heads in the specifications. Thus in terms of the previous discussion:

Each Latin chunk has attached a tag which is notionally the union of all the tags attached to its list of equivalents were it the list of equivalents to be explicitly found.

The algorithm for finding the contexts consists of rejecting heads which do not occur at least twice in the sentence. The extent to which a head-set of an entry in the output dictionary corresponds to a specification head-set is measured by the number of heads that they have in common.

The procedure just described is a very simple one, and is intended as an example of the type of algorithm upon which the Cambridge Language Research Unit experiments. It clearly cannot be asserted in general that only those heads should be retained for the output specification which occur more than once in the sentence; for example, a sentence might have one word, and only one word, which conveyed sarcasm, which nevertheless converted the whole sentence into a sarcastic utterance. The heads to do with sarcasm would only come once. It is part of present research to examine further the conditions for retention of heads in this and other similar cases.

I have now given an example of the possible and actual use of a scheme of context classification which is at least potentially a lattice. Before going on to discuss whether it is really a lattice (since clearly not all possible context classes will be used), I shall consider whether the limitation of resolution mentioned above is fatal to the scheme*.

The application of lattice theory, and indeed the possibility of M.T. at all, is frequently questioned on grounds that amount to saying that the lack of resolution is fatal. It is not the purpose of the present paper to argue about the possibility of M.T. and the validity of the objections to it which have been raised, for example, by Bar-Hillel. These points are discussed in another paper of the present set. I shall, however, now describe a method of improving the resolution within the general lattice framework.

The information that is lacking from the context-class specification of a word can be summed up as "factual information about the empirical thing to which the word refers". This must not be taken as asserting that there is usually a one thing to which the word refers. It would perhaps be truer to say, "...about <u>one</u> of the empirical things...". This information can usually be put in a very straightforward way, provided that it is not required explicitly. (For the case where a fact is explicitly required, see other paper.)

It can be truly said that this process (i.e., that of lumping together all the uses of a word) can in some circumstances import noise into the system. It is necessary to distinguish carefully the aggregation of uses of a word which is simply treating language as it really is from the aggregation of two or more quite distinct sets of uses. It is the latter that can import noise, which may in some cases be too much for the context algorithm to remove. An example of this occurs in the Latin sentence translated, as between those uses of the chunk, "Terr-" associated with "earth" and those associated with "fear". A procedure is given for resolving this case without importing noise under the same of "semantic pun removal".

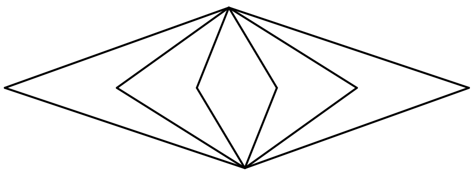
Suppose that there is a list of facts about elm trees and oak trees which adequately distinguish them. We may simply adopt an arbitrary sign, say 5, as an abbreviation for all the elm-facts, and, say, 6, as an abbreviation for all the oak-facts. Then the input language word will have tagged on to its context-class specification the appropriate sign for the facts, (called by the C.L.R.U. the list number) a direct match for which may be sought in the output dictionary. In lattice terms, this is equivalent to putting below (i.e. as included by) the elements (i.e. combination of context-classes) corresponding to trees a spindle* , the middle elements of which are the list numbers.

If it is necessary to put in factual information that must be capable of explicit use, it can be fitted into the lattice scheme provided that it can be interpreted as classificatory and unstructured. In this case, we simply have factual classes as well as the context classes, which generate a lattice for the same reason. It is a matter of present research to discover to what extent lattice theory requires supplementation (as for example, by some combinatory scheme), to deal with language as it is, and particularly with structured information. The work that has been done so far has all been on simple lattice lines, with a view to discovering how far these methods would suffice+.

It is now necessary to expand the earlier discussion of the lattice of classes and its generation. Clearly, not all the possible combinations of classes will occur. This

---

* A spindle is a lattice of any order with only three levels. It is thus a rather trivial case of a lattice - a way of accommodating an unordered list in the lattice framework.



+ As to M.T.: Agricola incurvo terram dimovit aratro; Gallia est omnis divisa in partis tres.
Detailed semantic programs have been gone through on these rather difficult sentences, for the results of which see the C.L.R.U. workpapers.
As to Library Retrieval: C.L.R.U. scheme, as in Joyce & Needham, Amer. Doc. 58, where in a system where there was prior reason to suppose that unstructured methods would be reasonably adequate, the lattice method has been applied with great thoroughness.

being so, it may be objected that it remains to justify the use of lattice theory since that theory makes very stringent demands of the presence of unique meets and joins which there is no reason to suppose will occur.

The set of classifications given by combinations of a set of _n_ basic classes is, under the set-inclusion relation, the Boolean lattice of order $2^{\underline{n}}$. If not all the elements of this lattice are present, the resulting system can fail to be a lattice through the existence of pairs of elements with no meet, or through the existence of pairs of elements with more than one meet.

I am going to show this means of an example. The general argument, however, can be stated as follows. While the absence of a meet can in a purely formal sense be supplied be inserting an 0-element which functions as all the missing meets, it is possible to deal with this point in a much more theoretically interesting way. The system of classes formed by a set of non-exclusive heads, is as said above, potentially a Boolean lattice of degree equal to the number of heads involved. While we know which of the elements are used for naming the contexts of words, which elements may very well fail to form a lattice, we do NOT know which combinations of heads, i.e. elements of the Boolean lattice, will be used for specifications. We are, therefore, not in a position, _a priori_, to rule any of them out[1], so we must consider the whole Boolean lattice as possibly occurring. Whenever a "retrieval" from the system takes place, a meet of certain heads is specified. The retrieval consists of taking that element from the dictionary which is in a defined sense[2] nearest to the specification. It is irrelevant that the list of elements in the dictionary does not of itself constitute a lattice.[3]

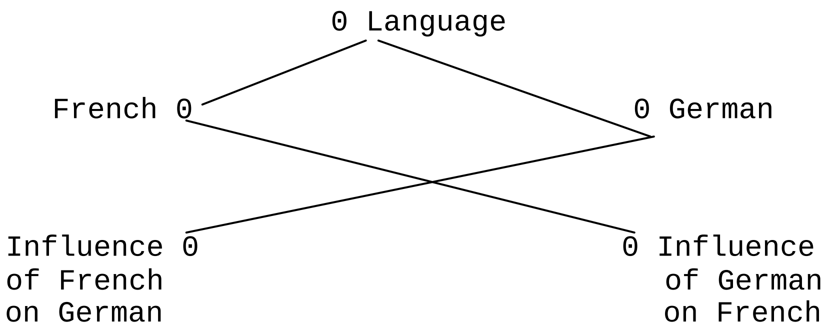
---

1. Except for some which appear to be semantically self-contradictory such as "inanimate life".
2. For the existence of a metric in a Boolean lattice, see Birkhoff. Once such a metric is established, it may be transformed into more suitable forms for any particular purpose.
3. Nonetheless, the word-points may, for coding purposes, be embedded in some lattice (Parker-Rhodes & Needham, passim) provided that the original encoding is in principle recoverable.

I shall now give an example to illustrate this argument, which was worked out by A. F. Parker-Rhodes and C. Wordley in answer to an objection to the Cambridge Language Research Unit's use of lattice theory. Though the example is specifically directed to library retrieval, it is equally apposite here.

The C.L.R.U, library retrieval system is described by Joyce and Needham ( ), and was criticised by a worker in that field on the grounds that the absence of certain elements from the system made it only an application of partially ordered sets rather than of lattices, and further that to include the extra points required to satisfy the lattice axioms, would contribute nothing to the flexibility of the system, and only cause a vexatious increase in bulk.

The example proposed was a sub-system including the subjects "language", "French", "German", "Influence of French on German", and "Influence of German on French", which was asserted to be correctly represented by the following diagram:

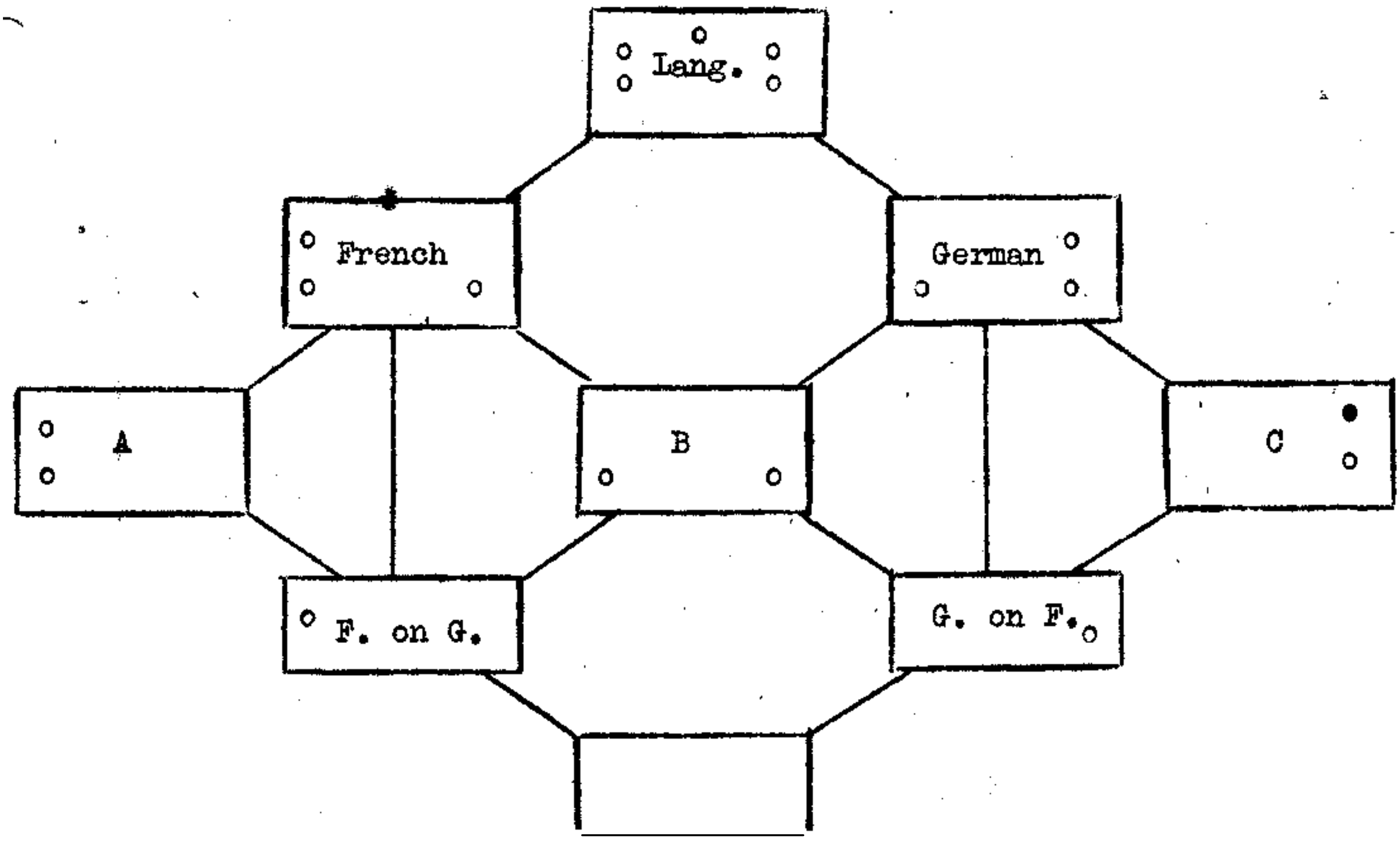


This diagram does not represent a lattice, because two of the elements, those in the bottom row, are both greatest lower bounds of the two elements representing French and German; whereas by definition lower bound, and dually. The justification for the use of such a diagram would be, if it were the case, that any document on the influence of French on German would be classified under both French and German, and so would a document on the influence of German an French, but that no other descriptors would be available to distinguish them.

Now consider what would happen if we were to seek to retrieve a document on this system. Let us say that we wished to retrieve one on the influence of French on German. The C.L.R.U. procedure is as follows: the catalogue will consist of a set of cards, each of which will represent one of the subjects represented, i.e. one of the points of the above diagram. On these cards will be punched holes, each place corresponding to one document, a hole in a given place indicating that this document is classified under the subject which the card stands for. Since the purport of the diagram is that a subject represented as including another will contain all documents attributed to the latter, we have:

The hole for a document classified under	will be punched on the cards representing
Language	Language
French	Language French
German	Language German
French-on-German	Language French German F.on G.
German-on-French	Language French German G.on F.

A request for a document on the influence of French on German would be dealt with by taking out the cards representing French and German, offering all the available documents represented by the holes which remain when these two cards are superimposed. The patterns on the cards can be represented thus:



From this it appears that the pattern we shall get in response to the suggested query is that represented at B above. This will induce us to offer both the documents classified as French-on-German and those under German-on-French. This is all right, but we obviously could do better. To do better, we should obviously need to introduce at least one additional descriptor to the two, French and German, we have used up to now. Suppose we introduce one for the influence of French on other languages. This would include documents under French-on-German, but not those on German-on-French. It would therefore have the pattern of holes on the corresponding card represented at A above. The given request would then be represented by superimposing card A and the cards for French and German, and obtaining their "meet". The result would evidently be the one-hole pattern corresponding to French-on-German. This is correct. Symmetry suggests we should also introduce a descriptor having a card punched like C.

For actually, our method (which is among other things a coding method), presupposes the possible existence of a subject and therefore, of a document, represented by any combination of holes on a card; and this allows for new subjects to be created. Of course, most of these "subjects" may not be needed; but in fact the system allows them to be used if they do turn up (and experience shows that they do very frequently turn up) implies that the "diagram" is not merely a <u>lattice</u>, but, potentially a <u>Boolean lattice</u>. This means that the critic is also wrong in saying that to take into account the extra points needed to make it a lattice would increase its size. It would not increase it at all, because, in principle, <u>the system presupposes the existence of not merely a lattice but of all the points that are needed to make it into a Boolean Lattice</u>. The describing of a system as a Boolean lattice gives the system the greatest possible flexibility, as any of the points may be used whenever necessary. It is not the case that we have one card for each of the points in the system, but we need have cards only for those points corresponding to which we have at least one document. Therefore, the number of points that we imagine to exist will not be reflected in the number of any material object, and will affect neither the cost nor the speed with which the system operates.

There remains a question which has to do with the justification of lattice theory, but has not been included in the present paper, both because of its length, and complexity, and because it did not fit naturally into the expository scheme. This is the question of inexact matching for M.T. purposes; the notion of selecting the nearest correspondent to a word in a given context. It would have been possible to discuss this whole matter starting from a theoretical demonstration of the necessity of inexact matching, and proceeding to the nature which a semantic system must have if such inexact matching is to be possible. One then reaches lattices as being a set of systems with a satisfactory metric for semantic purposes. This approach to the question of using lattice theory, however, would have to form the subject of a whole other paper.

Conclusion

The general argument of this paper has been that the list of possible translation equivalents for a source language word must, for realistic translation, be very long, and therefore that more refined methods than, say, that of making subject-classes must be employed to choose between them. Subject-classes are envisaged as context classes and then generalised to form a lattice of context classes; it is then asserted that, with the aid of the "list number" technique, the context classification is sufficiently powerful to select the output rendering not merely from a list of equivalents, but from the whole output vocabulary. The initial procedure of having a list of translation equivalents may therefore be dropped and the context class specification retained as the "interlingual notation". Finally, the technical question of the satisfaction by such a system of the lattice axioms as against those of a partial-ordering is considered and resolved by way of an example.

Roger Needham
University of Cambridge
Mathematical Laboratory
Cambridge Language Research Unit

A. F. Parker-Rhodes
Cambridge Language Research Unit

References:

1. Booth, A.D.; Cleave, J.; Brandwood, L. - Mechanical Resolution of Linguistic Problems. London, 1958.

2. Birkhoff, G. Lattice Theory, 2nd Ed. Amer. Math. Soc. Colloquium Publications.

3. Birkhoff, G; Maclane, S. - Survey of Modern Algebra.

4. Mooers, C. N. - Zatocoding & Information Retrieval. Aslib, 1956.

5. Fairthorne, R. A. - Patterns of Retrieval, Amer. Doc. 1956.

6. "Agricola Incurvo ...... ". Appendix to Bastin & Needham (ML92), this volume.