[From: *Sprache und Datenverarbeitung* 5 (1/2), 1981, 5-8]

MARGARET KING

Semantics and Artificial Intelligence in Machine Translation

ISSCO, University of Geneva

Abstract

Die Autorin grenzt zunächst die Probleme ein, deren Lösung man von der Einbeziehung semantischer oder AI-Methoden erwarten könnte: Disambiguierung von Wörtern, Strukturen und Relationen. Unter diesem Aspekt untersucht sie die Arbeiten von Schank, Riesbeck, Minsky, Charniak und Wilks und kommt zu dem Schluß, daß die beschriebenen Systeme bei der Entwicklung operationaler MÜ-Systeme kaum hilfreich sein werden, es sei denn, diese beschränken sich auf eine wohldefinierte Welt. Das Letztgenannte illustriert die Autorin mit einer Beschreibung des Edinburgher Mecho-Projekts. Da jedoch die Mehrzahl der MÜ-fähigen Texte sich nicht in der beschriebenen Art und Weise eingrenzen läßt, sind solche Verfahren insgesamt gesehen für die praktische Anwendung unbrauchbar. Es liegt jedoch bei MÜ-Systemen wie EUROTRA, die Ideen der AI geschickt für ihre Zwecke zu nutzen.

The author exemplifies three types of ambiguity that the introduction of semantics or of AI methods might be expected to solve: word sense, structural, and referential ambiguity. From this point of view she examines the works of Schank, Riesbeck, Minsky, Charniak, and Wilks, and she comes to the conclusion that the systems described will not be of much help for the development of operational MT-systems, except within a well-defined, constrained world. The latter aspect is illustrated by the author by means of a description of the Edinburgh Mecho-project. But, as the vast majority of texts destined for MT does not come from a constrained world, such systems will hardly be used as MT production systems. Still, MT-systems like EUROTRA give the chance of making intelligent use of AI ideas.

Introduction

Let us start by looking at the sort of problems that the introduction of semantics or artificial intelligence methods into a machine translation system might be expected to solve. In each of the following three sentences, the word 'stocks' has a completely different sense

Stocks of coal are falling at an alarming rate.
Stocks are a flower commonly found in English gardens.
The use of the stocks as a punishment has long fallen into disuse.

Examination of the syntactic structure of the sentence gives no clue as to the particular sense involved in each case: it is only the surrounding semantic context which allows us to distinguish them.

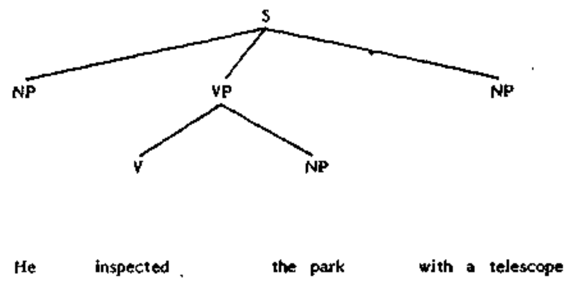
A special case of this type of word-sense ambiguity is the treatment of prepositions, a notoriously difficult problem for machine translation.

Je le sais par ma fille → I know it from my daughter.
Il viendra par un jour d'hiver → He will come on a winter's day.
Il l'a fait par paresse → He did it through laziness.

Here again, the syntactic structure of the sentence is of no help in determining the correct translation: indeed, each of the three sentences given has the same syntactic structure.

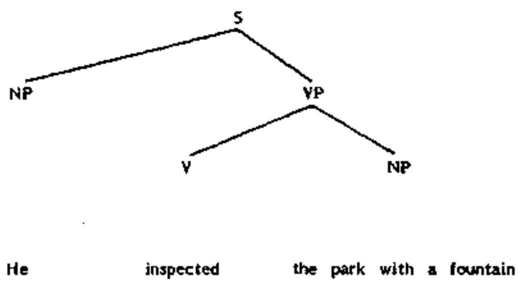
It is not only the interpretation of single words which can require semantic or pragmatic knowledge. Sometimes determining the correct syntactic analysis of the sentence can only be done by use of non-syntactic tools:

He inspected the park with a telescope →



i.e. He used a telescope to inspect the park

vs. He inspected the park with a fountain →



i.e. He inspected the park in which there was a fountain.

So far we have exemplified two types of ambiguity: word-sense ambiguity and structural ambiguity. Referential ambiguity also poses severe problems, as exemplified in the two sets of sentences below, based on Winograd and Wilks respectively:

(i) The town councillors refused a permit to the women
- because they feared violence (they = councillors)
- because they advocated revolution (they = women)

(ii) The soldiers shot at the women and
- several died (several = women)
- several missed (several = soldiers)

Here, in fact, the problem is even more severe, since it is only our knowledge of the world around us which allows us to determine the correct reference. (It should be noted, too, that knowledge of the world implies also knowledge of *a* particular culture. A different culture from ours might well reverse the references in (i)).

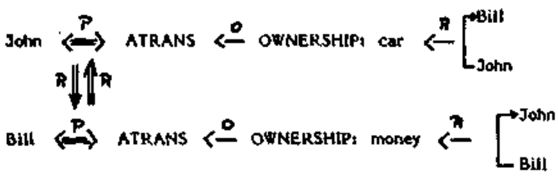
This, then, is the problem area. In what follows we shall examine some proposals coming mainly from work in artificial intelligence intended to help in the resolution of these problems. The systems discussed are not ail machine translation systems. Relatively few semantics or world-knowledge based machine translation systems exist, of which the best known are those of the Yale Group and of Wilks. However, the problems described above are not limited to machine translation. Any system depending on an analysis of language will encounter the same problems. We shall therefore also look at some other systems within the general framework of artificial intelligence to see what suggestions they offer.

The Yale project

The Yale project, developed by Schank and his collaborators is concerned with a set of general theses about language comprehension, rather than specifically concerned with machine translation. Nonetheless, experimental machine translation systems have been developed based on the Yale philosophy.

Within the project, two separate aspects are relevant to our present concerns, and should be distinguished. The first is a belief that language can be mapped on to a universal language independent representation, on which 'conceptual dependency' in Yale terminology forms a central part. Conceptual dependency is too familiar to need more than a very brief recapitulation here. Essentially, the representation consists of a small set of primitive acts, each with an associated set of case slots. The example below, taken from Schank (1973), can serve as an intuitively clear example.

'John sold his car to Bill'



As can be seen, two instantiations of the primitive act ATRANS (non-physical transfer of something) are involved, each of which is seen as the result of the other. In one of these primitive acts, John transfers ownership of a car to Bill: in the second Bill transfers ownership of money to John.

More recent versions of conceptual dependency differ in details from that described in Schank (1975), but the basic principles remain unchanged: the representation aimed at is very abstract, can be used to carry information required for inference making based on real world-knowledge, and can itself trigger such inference making.

Some of the difficulties inherent in using such a representation as a pivot-language in a machine translation system are fairly obvious. First there is the (perhaps carping) doubt about how abstract the conceptual dependency representation really is. It is rather evident, in the examples given in the literature, that while predicates are reduced to primitive acts, substantives, qualifiers and adverbial modifiers retain essentially their surface form. This being so, one is then tempted to ask how much more advantageous the mapping into and out of conceptual dependency representation really is, as compared to, say, a fairly standard case grammar based system, where the cases are regarded as case slots looking for fillers. (See, for example, Waltz, 1978).

This point lead into the second difficulty, that of setting up a universal representation in the first place. The Yale project has worked with very small subsets of language, and has even so finished up with some very language dependent elements still present in the representation. It is certainly not self-evident that with persistence and ingenuity a universal representation of a convincingly large subset of languages) could be found.

The final difficulty is more specifically concerned with the utilisation of such a representation for machine translation. It seems legitimate to suppose that the conceptual dependency representations of 'John sold his car to Bill' and of 'Bill bought a car from John' would be essentially the same. If this is the case, there can be no guarantee that a translation based on generating the target text directly from the conceptual dependency representation would not translate the first as 'Bill a achete une voiture de Bill' and the second as 'John a vendu sa voiture a Bill'. Yet no translator would regard these as accurate translation. In other words, whilst such a system could, in principle, guarantee the production of a truth-preserving paraphrase, it could not guarantee the production of a meaning-preserving translation.

Riesbeck's Parser

Let us now turn our attention to the second aspect of the Yale project, that of parsing procedures designed to map the source text onto the conceptual dependency structure. The first such, that of Riesbeck, is again too well known to need much description here.

Riesbeck's parser is based on 'expectations'. Essentially, each lexical item known to the system is coded with a set of expectations to be set up when that item is encountered in the text. The table below is the standard example and is self-explanatory. It shows what would happen if the sentence being analysed were 'John gave Mary a beating'.

Step	Word read	Requests waiting	Requests triggered	Actions taken
0	(none)	1-is there an NP?	none	none
1	John	1-is there an NP?	1	Assume John is subject of the verb to follow
2	gave	2-is the current NP a human? 3-is the current NP an object? 4-is the current NP an action? 5-true (i.e. default)	5	Assume the word 'to' if it appears, introduces the recipient of the 'giving'
3	Mary	2-is the current NP a human? 3-is the current NP an object? 4-is the current NP an action?		Assume Mary is the recipient of the 'giving'
4	a	3-is the current NP an object? 4-is the current NP an action 6-true (as before, default)	6	save the current list of requests and replace it with 7: does the current word end an NP?
5	beating	7-does the current word end an NP?	none	none
6	.	7-does the current word end an NP?	7	Build the NP 'a beating' and reset the list of requests
7	none	3-is the current NP an object? 4-is the current NP an action?	4	Assume the NP action is the main action of the clause, the subject (John) is the actor and the recipient (Mary) is the object

The most obvious remark to be made about this parsing procedure is that it uses syntactic methods intermingled with semantic tests. For the purposes of this paper, however, that is not a major issue, especially since any claims that it was a purely semantics based parser were probably mainly a product of prevailing fashion at the time the parser was written. It is more interesting to remark that the semantic tests used rely on the use of single binary semantic features, a la Katz and Fodor. The following set of examples is intended to show that reliance on single binary features will frequently lead to incorrect analysis, or rather to failure of the analysis process.

1. The proposal broke the previous agreement.
   ('Break' expects a non-abstract subject)
2. Darkness fell before emergency supplies could be flown in.
   ('Fall' expects a subject which is a physical object)
3. The document gives the reasoning behind the proposal.
   ('Give' expects a human subject)
4. Prices froze during the latter half of the year.
   ('Freeze' expects a liquid subject)
5. Opposition melted in the face of this argument.
   ('Melt' expects a solid subject)

The point here is that single features are too crude, if they are used as yes/no criteria in establishing the correct analysis of the sentence.

(It is worth noting here that Wilks' preference semantics system, which is not discussed here, includes both the possibility of structuring the semantic markers, thus achieving rather more subtle codings whilst maintaining reasonable economy in the number of markers used, and a technique for using the semantic codings which relies on selecting the most acceptable analysis from a range of possible analyses, rather than on assuming that there is a 'right' behaviour for each linguistic item which can be predicated in advance and used to eliminate all interpretations which are in conflict with it. The interested reader is referred to Wilks' own accounts and to Ritchie (1983) for further discussion).

One further remark is worth making before we look at a further parser associated with the Yale group. Riesbeck's parser is essentially top-down and depth-first: that is it assumes that it will always get the right answer first. This is a dangerous assumption, in that it relies again on the assumption that there is one and only one correct interpretation, which can be known to be the correct one without even inspecting any alternatives. Furthermore, of course, it excludes completely the possibility of there being two or more genuinely correct interpretations and being able to produce them all. (For more detailed discussion on these points, see Wilks 1983). The moral to be drawn here is that the system should provide for back-up or for parallel exploration of alternative interpretations.

Frame based parsers

Later Yale parsers are based on 'frames', an idea first introduced by Minsky (1975). Although different frame-based systems differ quite widely in the kind of frame employed, the underlying idea remains unchanged. The essential hypothesis is that a great deal of our knowledge of the world can be expressed via structures describing stereo-typic situations (frequent examples are birthday parties, visiting a restaurant, painting a piece of furniture). The description is in terms of elementary units, linked by chronological relations, causal relations and so on. Normally, a parser using such structures not only uses the frame as a parsing device but also parses into a representation which consists itself of frames, filled out by specific information drawn from the text.

Thus, to take a very simple example based on Charniak (1975), there might be a frame describing the normal sequence of events when shopping in a supermarket. The frame itself would mention, among other things, that objects are bought. When a text dealing with shopping is being analyzed, an attempt is made to map the sequence of events described in the frame onto the text, simultaneously filling out detail in the frame instantiation; for example, specifying that, in this particular case the objects bought are milk and bread. The end result is a version of the frame, with the particular information drawn from the text included. (Once again, the description given here is unfairly brief; and the reader is referred to the literature for more detail).

The problems with frame-based systems are by now fairly well known. It is difficult, in the general case, to know what frames will be needed for the analysis of a particular text, just as it is difficult to know what frame out of the sets of frames available should be activated when, or to de-activate it at the appropriate moment.

I have deliberately refrained from going into much detailed discussion of frame-based systems, despite their one time popularity in the world of artificial intelligence, simply because, in what is, admittedly, an entirely personal view, I see very few really advantageous ideas to be picked up from such systems when the aim is to produce an operational machine translation system. The exception to this rather sweeping condemnation is that such systems might prove very efficient if one were working in a constrained and definable world. In other words, if it were possible to build a model of the world about which the text to be translated is talking, then a description of that world via a set of inter-related frames might prove a very powerful parsing device.

The Mecho project

To illustrate this point, let us briefly look at a system which works in a very tightly constrained world, and where some very elegant methods for dealing with reference disambiguation have been developed. (The closed world of the system is in this case not described by frames, but this is irrelevant to our main point here, which is the use of specific properties of a constrained world to resolve linguistic problems).

The program is described by Mellish in the papers listed in the bibliography. It acts as a front end to a large program which reads in and solves elementary mechanics problems of a sort found in a school text book (Bundy et. al., 1979). The constrained world is therefore the world of simple mechanics, with known and easily describable physical laws. Furthermore the language used in talking about the world is also constrained: most sentences are declarative sentences, on the basis of which the system builds up a small world model of physical objects and relations between them. The following example sentences illustrate this:

1. A bridge sixty feet long is supported by a pier at both ends.
2. Small blocks, each of mass m, are clamped at the ends and at the centre of a light rod.
3. One painter stands on the scaffold four feet from the end.

Because of the type of text being dealt with, the number of objects in the text which could potentially be referents for a noun phrase requiring reference disambiguation is usually small, sometimes even directly listable.

Referents are computed using any information available. This may be linguistic information - the noun phrase itself, the enclosing sentence, the general context - or information coming from general inferences about the world of mechanics. All this information contributes to defining a set of constraints, which limit what the noun phrase refers to. Constraints may apply not only to single items in isolation, but may also consist in expressing relations between two or more items - that one item supports a second item, for example. Because this is so, elimination of an item as a possible referent for one noun phrase may, via the relations, result in cutting down the possible referents for a different referring expression.

The following over-simplified example, taken from Ritchie (1983), may illustrate the process of elimination. In the sentence

A uniform rod is supported by a string attached to its ends

the word 'its' could, syntactically, refer to either the rod or the string. The procedure whereby the system decides, as does a human reader, that the referent is the rod goes as follows.

Assume that there are two objects in the world, the rod and the string, which are possible referents for three symbolic expressions R1 (it), R2 and R3 (the ends). Certain constraints can be stated

(1) left-end (R1, R2): R2 is defined to be the left end of R1.
(2) right-end (R1, R3): R3 is defined to be the right end of R1.
(3) attached (R2, STRING): from the phrase 'a string attached to its ends.
(4) attached (R3, STRING): from the same phrase.

Amongst the general inference rules available to the system is one which states that an object cannot be attached to its own ends. Applying this rule to constraints (3) and (4) above we can deduce the following further constraints:

(5) SEPARATE (R2, STRING)
(6) SEPARATE (R3, STRING)

By a fairly simple inference, given these two constraints, the string cannot be the referent of R1. Since only two possible referents were available, the rod and the string, the rod must therefore be the reference of R1, 'its', and the pronoun reference is resolved.

The problem, of course, with a method of this type, where pronoun problems are resolved via strict inferential chains based on known properties of the world, is that they can only work within a very tightly structured world. It is perhaps worth saying that it is not the size of the world that matters, it is its tight structure. In other words, if one were to construct a machine translation system working on this type of text, it would probably be possible to produce perfect translation, even if the world were many times larger than that of the Edinburgh project. Sadly, however, the vast majority of texts destined for machine translation are not of this type, and, indeed, could not be. The logic underlying discussion of the agricultural policy is of a quite different nature to the logic of mechanics, and not susceptible to description in terms of stable properties and inexorable inferences based upon them.

Conclusion

Despite perhaps the rather critical tone taken in this paper towards some suggestions put forward by workers in artificial intelligence, it is not the author's intention to argue either that the problems described in the introduction are Insoluble or that artificial intelligence work has not contributed enormously to possible solutions. On the contrary, a system making use of semi-proved techniques in an intelligent way could, I believe, resolve most of the problems there described, with the sole exception of cases where a great deal of highly specific world knowledge is required, like the Winograd example quoted. What follows, very briefly, describes some characteristics of a system currently being developed under the auspices of the EEC designed precisely to make use as much as possible of the available techniques (King & Perschke, 1982).

The text representation to be achieved by the analysis module includes the explicit mention, for each constituent in the text, of its semantic relation to other constituents. The relations used strongly resemble an extended set of case relations of the type found in, for example, Fillmore (1968). The chief difference is that the set of relations is considerably larger than the standard set.

Each of these relations in the interface structure carries a property describing the strength of its (valency) binding to the predicate with which it is associated. It is the value of this property which distinguishes central from peripheral arguments.

In order to reach the interface representation, use may be made of semantic information coded with lexical units. Any lexical unit may have associated with it a case frame, which describes the type of context in which the lexical unit prefers to find itself. Also, any lexical unit may have a semantic formula associated with it. The semantic formula may be as heavily structured an object as the linguist wishes, providing he can express the structure in terms of a limited set of semantic features and two basic relations between them, subclassification and implication. There is no restriction on when and how these two types of information may be used by the rules.

The semantic formulae may be used either in analysis or in transfer to assist in building the interface structure and/or in disambiguation, but they are not intended as 'universal' or even as 'euroversals'. The system is constructed with mono-lingual analysis and generation modules, and bilingual transfer modules in such a way that there is no necessity for language A to use the same formula for the 'same' (= corresponding) lexical unit as language B. Thus, the semantic formulae, whilst offering more possibilities for subtle or complex description than are achievable via simple union of a set of binary features, make no pretension to being the correct universal descriptions of basic concepts.

The lexical unit may also have interferencing information attached to it, which may thenbe manipulated by inference making rules. For example, it is possible tostate that a particular predicate normally has a particular consequence or is done for a particular reason.

This facility is not intended, however, to be used for heavy inference making in the immediate future. It is not yet clear that the price to be paid, in terms of development time and even running time, is acceptable in a practical application, where the end user might well prefer to accept a certain (small) number *of* errors rather than wait a long time for his output. On the other hand, it would be foolish to exclude the use of heavy inference mechanisms by designing the system in such a way as to make them impossible. A fundamental underlying principle of the system design is to make maximum use of known techniques whilst leaving the door open for future extension towards what are at the moment speculative methods once they have proved their worth.

References

Bundy et al. (1979): Solving mechanics problems using meta-level inference. Proceedings of the Sixth International Conference on Artificial Intelligence, Tokyo

Charniak (1975): Inference and Knowledge in Charniak & Wilks (Eds) Computational Semantics, North Holland

Fillmore (1968): The case for case In Bach & Harms (Eds) Universals in Linguistic Theory, Holt, Rinehart and Winston

Katz and Fodor (1963): The structure of a semantic theory. Language 39, no. 2

King & Perschke (1982): EUROTRA and its objectives. Multilingua, 1,1

Lehnert et. al. (1981): BORIS, an experiment in in-depth understanding of narratives. Computer Science Department Memo No. 188, Yale University

Mellish (1980): Some problems in early noun-phrase interpretation. Proceedings of AISB conference, Amsterdam

- (1981): Coping with uncertainty - noun phrase interpretation and early semantic analysis. Ph. D. Thesis, Department of Artificial Intelligence, University of Edinburgh

Minsky (1975): A framework for representing knowledge in Winston (Ed) The Psychology of Computer Vision, McGraw-Hill

Riesbeck (1974): Computational understanding: analysis of sentences and context. Ph. D. Thesis, Department of Computer Science, Stanford University

- (1975): Conceptual analysis in Schank (Ed) Conceptual Information Processing, North Holland

Ritchie (1983): Semantics in parsing in King (Ed) Parsing Natural Language, Academic Press

Schank (1975): Conceptual Information Processing, North Holland

Waltz (1978): An English language question answering system for a large relational data base CACM 21/7

Wilks (1975): A preferential pattern-making semantics for natural language inference. Artificial Intelligence, vol. 6

- (1983): Deep and Superficial Parsing in King (Ed) Parsing Natural Language, Academic Press