[*Machine Translation Today: Translating and the Computer 15*.
Papers at a conference…18-19 November 1993 (London: Aslib)]

MT – MYTH, MUDDLE OR REALITY?

Seamus Derrington
Nissan European Technology Centre, Language Services

Summary: the trend away from development of fully automatic machine translation (FAMT) is the result of failure to develop the foundation level of machine translation (MT) systems design theory. In order to create this level and establish reliably whether FAMT is achievable or not we will have to revise our view of the inter-disciplinary approach. The paper ends with a personal critique of the interdisciplinary approach as applied to date.

The aim of MT research is to create computer systems which will simulate human translation activity and stand in for it when required. Until quite recently, it was hoped that it would not be long before it was possible to develop these systems in such a way that the work would be done automatically and without human assistance — FAMT, in other words.

This aim has always been well received by the media and the public because it seemed like a hefty step towards the dream of the independent computer with a mind of its own and able to communicate with humans like HAL in the film 2001. But the current trend in MT systems design is away from FAMT towards systems with more limited aims:-

a) Machine assisted human translation: "work station" systems in which the translation work is done by the human translator aided by a range of software tools such as dictionaries, parsers, editors, modem access to technical databases, etc.

b) Human assisted machine translation: these are the surviving general purpose systems, eg. SYSTRAN, which rely on pre-editing of source texts or post-editing of target texts to make up for the system's limitations.

c) Restricted language systems: these systems operate only in a particular application area, eg. weather reports, production of technical manuals, stock markets reports, etc. (the advantage of this approach to system design is it results in an automatic system which is relatively error-free because it restricts the use of language to a fixed vocabulary used in a limited set of sentence structures in order to exclude ambiguity).

The trend towards a more piecemeal approach to MT is based primarily on the poor success rate of research into FAMT systems over the forty odd years since this research started. Such projects are increasingly being seen as black holes in funding terms so it is hardly surprising that people are starting to look elsewhere for more positive results.

What is disturbing however is the fact that no-one appears anxious to call a post-mortem. No-one seems to be willing to ask what has gone wrong. Many authorities have pointed out a number of the difficulties involved but no-one has provided substantive reasons why these difficulties should be insoluble. Instead there seems to be a rising swell of intuitive judgement that the task of developing FAMT is simply not achievable.

The need to resort to intuitive judgement arises from the fact that there is no widely accepted foundation level of systems oriented research into language phenomena. This

level is essential if we are ever to make a reliable decision on whether or not FAMT can one day be achieved. What we have here, after all, is an attempt to computerise an activity which has a long tradition of paper-based human handling. If this had been a commercial company, the systems builders would have been straight in there doing a full systems analysis of every aspect of the business before considering a single line of code.

However, although some money can be made from translation, the language skills on which it is based are definitely not a commercial business. And, of course, there has always been available, from the very first attempts at MT systems building, a large corpus of non-systems oriented expertise in language phenomena in the form of scientific research. So it has seemed only natural that MT research should rely on the output of other disciplines to provide it with descriptions of language phenomena on which to base judgements regarding what is achievable and what is not. But this, I feel, has been MT's biggest mistake.

To provide MT systems design theory with an effective foundation level, reliable systems oriented information about the following topics is essential:-
- the evolutionary origins of language
- the nature of the physiological mechanisms controlling language
- the nature of the relationship between language, society and culture
- the nature of the relationship between language, semantics and psychology
- the nature of the relationship between language, logic and problem-solving
- the nature of the learning process which permits more than one language to be learnt by a single individual.

Unfortunately, no one scientific discipline covers this list of topics and little of the information currently available regarding them is in a systems oriented form. In fact, the biggest current barrier to a rigorous systems approach to building a foundation level for MT development is the multi-disciplinary nature of the knowledge required (together with the fact that each of the disciplines involved has a traditional structure centring round a set of issues which evolved long before the advent of computers).

THE LIMITS OF THE INTERDISCIPLINARY APPROACH

a) The Kuhnian nature of the problem.

Kuhn, in his "Structure of Scientific Revolutions", said that there are always limits to the questions which can be answered within the structure of knowledge at a given point of human history. These limits are imposed by the basic model of the universe possessed at that point He called this basic model the "paradigm".

The traditional view of the evolution of human knowledge is that it has been a steady piling up of experience. In Kuhn's view, this is not the case. Knowledge doesn't just accumulate in a steady linear progression. It has also, more significantly, sometimes gone through periods of catastrophic change, eg. as with the acceptance of Copernicus' view of the solar system. In such periods, the underlying paradigm changes or shifts. Large amounts of knowledge are virtually dumped and previously inconceivable avenues of insight are opened up.

With hindsight we can look back on the history of the interdisciplinary approach to MT and detect several Kuhnian features.

b) Kuhnian features of MT.

Modem scientific disciplines have only recently modified the severe territoriality which has customarily characterised the maintenance of interdisciplinary boundaries. In fact, the development of interdisciplinary teams working on a range of artificial intelligence (AI) problems (of which MT is only one) has always itself been acknowledged as something of a minor paradigm shift

However, research has been forced down this avenue because (as already mentioned in the case of MT) no one discipline seems to provide a sound basis for tackling AI problems. The limited success of the new interdisciplinary research raises the suspicion that what we are, in fact, observing in it is a cosmetic process of papering over the cracks in a paradigm crumbling under the stress imposed on its inner limits.

If it is true that the research aims of the individual disciplines involved are pointed away from the areas which need to be investigated to solve AI problems, we need to ask ourselves why it should have been imagined that this situation would change substantially when the disciplines combine. The only way an interdisciplinary approach to these problems might have worked is if the issues involved had been re-opened, unpacked and re-examined from the new viewpoint. Unfortunately, no one has attempted such an enterprise.

One outstanding example of the failure of the interdisciplinary approach can be seen in its inability to redefine the traditional paradigm regarding the relationship between perception and movement in the organisation of knowledge. Researchers have been aware for some time now that human knowledge and consciousness are dependent on the constant feedback relationship between movement and perception.

Unfortunately, the traditional disciplines are based on a paradigm which separates perception and cognition from movement and gives them a higher priority. The result is:-

i. We still do not have the kind of grasp of how the vertebrate system coordinates movement which would permit us to develop a computer simulation of vertebrate movement.
ii. We are left with the impossible situation in AI where we are attempting to simulate highly flexible phenomena like language with computer architectures which are condemned to a primitive rigidity from which there appears to be little likelihood of escape.

c) Consequences for MT.

In this situation, it is almost inevitable that no progress will be made towards the solution of fundamental problems like deciding on the achievability of FAMT. The only factor which makes the artificial separation of MT from the other areas of AI seem acceptable is the way it appears so reminiscent of the interdisciplinary boundaries of traditional sciences.

This, however, has the negative effect of ensuring that, as topics, machine translation, intelligent knowledge based systems, robotics and pattern analysis will start to develop the same type of mutual exclusivity which will increasingly block the cross-fertilisation essential to further progress. If the current trend towards concentration on short-term gains in MT and AI in general takes hold there is a danger that the situation will ossify in this position.

The long-term issues will only retain their vitality and significance if we are prepared to face up to and act on the following sobering possibilities:-

i. That we have finally reached a point when a purely empirical science approach with its objective analysis of observable data is no longer enough and we have to ask ourselves: is it, perhaps, the assumptions we have been working on at the basic paradigm level that are blocking our path?
ii. That we are witnessing a critical turning-point in the development of our culture where we have to acknowledge that we have come up against its inner limits.

Whichever route we decide to take out of the current situation we will only keep a grip on the realities of understanding the nature of the main medium of communication for our species, language, if we are willing to recognise the true scale of the problems involved. This a cultural Everest we are confronted with and we translators are its Sherpas, conscious of the deep mystery still lying within its slopes and watching with bemused curiosity the fluctuating energies, now flagging, now in full flow, of the technically minded climbers exploring its surface. MT research may one day provide us with the oxygen bottles to reach its summit but it is unlikely ever to deprive us of a job.

Roger Penrose recently went to great theoretical lengths to prove that computers can never be made to simulate natural cognition. I feel there is some irony in the fact there is a much simpler reason that computers cannot be made to simulate natural language activity. The reason is simply that we still do not know what language is and are unlikely to find out what it is as long as we cling as fiercely as at present to our illusions about what it should be.

This is good news for translators and, maybe, for the human race. But it must make disappointing knowledge for anyone who feels that computerised language technology should have a sound theoretical base. It is also very frustrating for those like myself who would like to see revealed more clearly what language is and what it can still do for us.