
Machine Translation **16:** 1–19, 2001.
© 2001 *Kluwer Academic Publishers. Printed in the Netherlands.*

Has There Been a Revolution in Machine Translation?

PIUS TEN HACKEN
Universität Basel, Abt. Geisteswiss. Informatik, Petersgaben 51, 4051 Basel, Switzerland

Abstract. When we compare the contributions on MT in the proceedings of Coling 1988 and Coling-ACL 1998, it seems obvious that in the period between them a revolution has taken place. Often this intuition is formulated as the replacement of linguistic approaches by statistical approaches. On closer inspection, however, this position cannot be defended. An analysis of Rosetta, concentrating on the different levels of discussion and of underlying assumptions, shows that the choice of knowledge from linguistic theories or information theory and corpora is by itself not a decisive issue. More important is the question of how the problem to be solved by an MT system is defined. An analysis of the decisions underlying Verbmobil, resulting in a list corresponding point by point to the one for Rosetta, shows how far-reaching the new approach to defining the problem of MT is. As it is shown that these systems are representative of the work in MT as it was done ten years ago and today, it can reasonably be argued that a revolution in MT has taken place, though not in exactly the way it is often believed.

Key words: evaluation criteria, linguistic vs. probabilistic approaches, MT as applied science, paradigms, Rosetta, scientific revolution, Verbmobil.

In the community of researchers concerned, there is a widespread feeling that in the past ten years or so computational linguistics (CL) in general and MT in particular have undergone a process of rapid change. Moreover, the change in the field, as seen in the contrast between, for instance, the proceedings of Coling 1988 and Coling-ACL 1998, was so radical that many researchers have the impression that they have experienced a revolution. The aim of this paper is to analyse the nature of this development: has there been a revolution and, if so, in what sense? MT is a specialization of CL which is particularly suited to illustrating the development in question.

1. Introduction

In order to assess the validity of the claim that a revolution has taken place, we need a more precise concept of "revolution". Since Kuhn's (1970) foundational work, revolutions in the context of science have generally been understood as a change of paradigm. Although the term "paradigm" has often been at the basis of confusion, we will provisionally continue to use it here until a more precise

term ("research programme") can be introduced in Section 2.2 below. In the case of CL, the two paradigms are usually characterized as probabilistic, stochastic, or statistical approaches on the one hand and non-probabilistic, symbolic, or linguistic approaches on the other, e.g., by Sampson (1987), Kay et al. (1994), and Klavans and Resnik (1996). These two types of approaches are both represented in MT.

If we take this perspective as a basis for the comparison of the sessions on MT at Coling 1988 and Coling-ACL 1998, the contrast is indeed striking. In 1988, the MT sessions encompassed 22 papers. Six of them were devoted to or based on the Eurotra project, which was therefore the most prominently represented single project. The linguistic orientation of this project is shown not only by some of the presentations at this conference, e.g., Steiner and Winter-Thielen (1988) and van Eynde (1988), but also by the general overview of Durand et al. (1991). They typically take linguistic phenomena as discussed in theoretical linguistics as a basis for the identification of topics in MT. Other projects presented at Coling 1988 did not challenge this orientation, with one exception. The common view as expressed in an overview which appeared in the same year was that "The obstacles to translating by means of the computer are primarily linguistic" (Lehrberger and Bourbeau, 1988:1). The one exception to this generalization of opinions at Coling is Brown et al. (1988), probably the first presentation of the ground-breaking IBM project to a large audience.[1]

Turning to Coling-ACL 1998, the first striking observation is the decrease in the relative importance of MT in the overall field of CL. Whereas in 1988 the 22 papers on MT constituted about 19% of the 117 papers at the conference, in 1998 the proportion of papers on MT had dropped to 10% (around 26 out of 261).[2] Next, papers with titles on a pattern such as "The treatment/analysis of [linguistic phenomenon] in [MT-system]", quite frequent in 1988, are largely absent in 1998. Finally, the majority of papers are on statistical approaches or related issues such as bilingual text alignment.

It seems then that this comparison of two conference proceedings with a ten-year time interval shows quite convincingly that the statistical approach to MT has gained prominent status at the cost of the previously dominant linguistic approach. In order to answer the question of whether this development should be termed "a revolution", it is necessary to establish whether the two approaches qualify as paradigms. Ironically, doubts in this matter are raised by a paper which assumes a positive answer to the latter question. Gazdar (1996) discusses the competition between "Two 1980s Paradigms" and describes the most recent developments as a "paradigm merger". If the two approaches to MT are considered as paradigms in the sense of Kuhn (1970), however, a merger is inherently excluded. Let us therefore turn to the question of what characterizes a paradigm in CL and in MT.

2. Paradigms in MT

In discussing Kuhn's philosophy of science and applying it to CL, there are two potential sources of confusion and misunderstanding. First, in introducing terms such as "revolution" and "paradigm", Kuhn (1970) appealed to a commonsense interpretation, but he elaborated them in a highly specific sense. Many scholars criticizing or using his terminology projected their own ideas onto Kuhn's terms. As a consequence, it is useful to summarize the aspects of Kuhn's theory assumed here before proceeding (Section 2.1). Furthermore, Kuhn explicitly intended his philosophy of science for empirical branches of natural science, e.g., chemistry and astronomy. CL deviates from such disciplines in a double sense, because of its concern with language and its applied nature. In Sections 2.2 and 2.3 each of these extensions will be discussed separately.

2.1. KUHNIAN PARADIGMS AND REVOLUTIONS

A paradigm is, according to Kuhn, what a group of scientists share so that their communication in the relevant scientific field is successful. Given that theories are highly underdetermined by observations and that the selection and interpretation of observations is theory-dependent, successful exchanges and cooperation are only possible if scientists share particular orientations in these questions. Margolis (1993) calls them "habits of mind", which are in most cases highly productive ways of getting to see the relevant facts, although they can sometimes become barriers to gaining certain specific insights.

If two scientists work in different paradigms, the differences between their views are much more radical than if they only have different theories. As long as scientists share a paradigm they have a basis for discussing and evaluating their theories which is accepted by all of them. It is probable that, if they are skilled and honest enough, they will more or less agree on the relative evaluation of their theories. Often, there will not be conclusive evidence in favour of one of the theories, so that the assessment of the strength of the case remains in part a matter of personal judgement. What will not occur, however, is a situation where what one side considers a strong argument or even conclusive evidence in favour of their own theory is rejected as utterly irrelevant by the other side. The common paradigm gives a basis for deciding which data are relevant and can serve as evidence in the evaluation of a theory.

Theories developed within different paradigms, however, cannot be compared as easily. They are incommensurable in the sense that there is no encompassing set of criteria which is accepted by proponents of both sides of the discussion and at the same time powerful enough to bring about a ranking of the success of the two theories. Incommensurability is not a matter of bad will on the part of the scientists concerned but an inevitable consequence of the existence and the nature of paradigms. Because of incommensurability the replacement of one paradigm by

another is a revolution, a change of perspective which cannot be explained in terms of rigidly logical reasoning.

The far-reaching nature of such a change makes it dependent on a preceding crisis. Only if part of the scientific community perceives the state of the field they are working in as a crisis will they start to scrutinize the basic assumptions they have always entertained and consider alternatives, calling into question some of the habits of mind that have served them well enough so far.

2.2. THE EXTENSION TO LINGUISTICS

Despite Kuhn's own reservations in this respect, cf. Kuhn (1970:209f), his concepts of paradigms and revolution were soon applied in a variety of areas other than natural science. Only three years after the appearance of Kuhn's first edition, Thorne (1965) refers to "the Chomskyan revolution in linguistics". The discussion on whether there has actually been a Chomskyan revolution has been marked by ideological arguments. Opponents, such as Percival (1976), often attack Kuhn's or Chomsky's theories or both when they oppose the use of the term "Chomskyan revolution". Proponents of this use, such as Newmeyer (1986), often tend to define the concept of revolution in the context of linguistics so as to fit what happened in the rise of Chomsky's theoretical framework. One of the reasons why the discussion took on such a vehement aspect was the widely held opinion that a paradigm should, according to Kuhn, be generally accepted in the relevant field of science. While it is doubtful that he ever held this view strongly in the first place, it is clear that when Kuhn (1970) refers to "the relative scarcity of competing schools in the developed sciences" in his postscript, he does not make uniform consensus a condition for the existence of a paradigm.[3]

In order to get away from the emotional load based on doubtful extensions and reinterpretations of Kuhn's theory, a new term "research programme" is proposed, which, though related to Kuhn's "paradigm", concentrates exclusively on the intellectual rather than the social components (ten Hacken, 1997, 1998a). A research programme can be taken as a solution to the old problem of how to validate the empirical cycle, which links data and theory. Science formulates theories to explain observations. A theory can be tested by confronting its predictions with further data (experiments). Explanation goes beyond mere generalization and assumes that certain aspects of reality are singled out as questions with respect to which the data gathered are to be explained. One of the main functions of the research programme is the choice of such a question. At the same time, we need assumptions which are not questioned in order to find a basis for explanation. The need for such a basis arises as explanation is never absolute, because at any point further *why*-questions can be formulated. The only way to have such a sequence bottom out is to postulate certain assumptions as "beyond doubt". These assumptions are justified by the potential of the entire research programme to find and explain new data.

Research programmes in linguistics thus establish assumptions which can serve as a basis for explanations and determine which aspect of language should be explained as a priority. An example of an opposition between research programmes is found when, for instance, Chomskyan linguistics aims first of all to explain language acquisition, whereas Bresnan's (1982) Lexical-Functional Grammar (LFG) takes language processing as the primary question.

A revolution in this model is more closely linked to the end of an old research programme than to the emergence of a new one. It is possible for different research programmes to continue to exist alongside each other. When a new research programme replaces an existing one, however, there is a revolution. The rise of Chomskyan linguistics and its replacing of post-Bloomfieldian linguistics can be called a revolution because the latter is now obsolete. By contrast, the rise of LFG does not constitute a revolution because both it and Chomskyan linguistics continue to exist alongside each other.

2.3. THE EXTENSION TO APPLIED SCIENCE

In the case of the extension of the scope of application of Kuhn's concept of revolution to fields other than natural science, the barrier seems first of all an ideological one. The application itself seems straightforward and Kuhn's reservation about doing so seems rather a form of resistance to abandoning the special status of science as regards the role rationality plays in it (cf. Kuhn, 1971). By contrast, Kuhn considered the extension of his framework to applied science, though less straightforward, as a worthwhile enterprise (personal communication, 1995).

The difference between empirical science and applied science is one of goals. Whereas empirical science aims at knowledge, applied science aims at working solutions to practical problems. Medicine is an applied science because it aims at cures for diseases or effective procedures to prevent diseases from afflicting people rather than being content with formulating theories as explanations of why and how people suffer from them. Not all problem-solving is applied science, of course. Practical problems are solved routinely without any thought of science coming up. Only if a non-trivial problem is solved in a way which can be explained against the background of the relevant field of science can we speak of applied *science*.

A good illustration is the comparison of the first and second green revolutions in agriculture, as described by Reijnders (1997). The first green revolution involved dramatic increases in agricultural yields through the use of town refuse as fertilizer, starting in the Netherlands in the late Middle Ages. The solution to the problem of producing enough food for a growing population was highly effective. In the absence of sufficient relevant scientific knowledge at the time, the success of the method could not be explained, however. In fact, it was often described as a miracle. The second green revolution started in the 19th century with the application of scientific theories of plant growth to the production of artificial fertilizers. The success of this method of increasing the yield can be explained in terms of the

theories used. Therefore, the second green revolution involves applied science but the first does not.

Summing up, we can distinguish applied science from empirical science by the practical aspect of its goal and from commonsense problem-solving by the epistemological aspect of its goal. The goal of applied science is to develop working solutions to practical problems and to explain how and to what extent they work.

3. Research Programmes and Theories in CL and MT

We will assume here that CL is a branch of problem-solving which, at least in favourable contexts, can be an applied science. Focusing on MT constitutes a narrowing down of the range of problems considered. We will furthermore assume that a revolution is a change of research programme by the scientific community at large, making a previously widely adhered to research programme obsolete. Given these assumptions, a first step in answering the question of whether the radical change in the field of MT described in Section 1 constitutes a revolution is the identification of research programmes in an applied science.

3.1. A CASE STUDY: ROSETTA

One approach to identifying the components of research programmes in CL is to start from a concrete example in the form of a system or project and progressively generalize from the specific decisions taken in the project to the more abstract decision issues. A system representative of the state of the art in MT around 1990 is Rosetta. The extensive, systematic documentation in Rosetta (1994) makes this MT project a particularly suitable example to start from.

Before entering the bottom-up study of decisions in a particular system, we have to take a few top-down steps to get started. In considering Rosetta as an embodiment of applied science, the possibility of explanation is the primary point to concentrate on. Explanation demands an object, the data to be explained, and a framework with respect to which they are explained. The object of explanation is the performance of the system. In the case of an MT system, the central point in performance is the quality of the translations produced by the system. The framework is the background such that if data are made to fit in with this background they count as explained data. It encompasses the assumptions postulated to make explanation possible (cf. Section 2.2).

The data to be explained in applied science are instances of the performance of the solution developed. In the case of Rosetta they are properties of the translations produced by the system. What is particularly worth explaining is the degree to which output produced by the system corresponds to what its designers intended as translation. Rosetta (1994) distinguishes between the best translation and a possible translation and takes the goal of the MT system to be to produce the set of possible translations into the target language (TL) of an input sentence in the source

language (SL). A TL sentence S'_i is a possible translation of the SL sentence S iff there exists a reading of S such that S'_i and S are built up compositionally in the same way out of corresponding rules. The performance data consist of answers to questions such as which readings are produced in the analysis of an SL sentence S, which TL translations are produced for each reading, and how the set of possible translations produced for S corresponds to the set of translations which are actually linguistically possible (i.e., whether all and only the possible translations of S are found).

The explanation of the data links them up with the background framework assumed. In the case of Rosetta, the presence or absence of a reading S_i in the set of analyses of an SL sentence is typically explained in terms of the linguistic coverage of the system: whether S_i contains linguistic phenomena not in the scope of the system, whether the linguistic theory used is adequate for the case at hand, and whether the rules in the system are a correct implementation of the linguistic theory. Similarly, the presence or absence of a translation can be explained in terms of the correct analysis of the underlying SL reading, the correct TL rules and the correct correspondence rules. Again, "correctness" is interpreted in terms of an underlying linguistic theory. The compositional nature of Rosetta is intended to guarantee that when the correspondences at the level of individual words and rules are correctly defined, the set of translations produced for an SL sentence is in fact the set of linguistically possible TL translations of this sentence.

3.2. GENERALIZATION FROM PROJECT TO THEORY AND RESEARCH PROGRAMME

Rosetta is not only an MT project, but also, and at the same time, an instantiation of a theory of MT which is in turn embedded in a research programme. Each of these constitutes a level of more general decisions or choices. In order to disentangle these three levels of choices, let us first list the main choices made in the Rosetta project, then consider the consequences of different types of alternative choices.

(a) The problem description in Rosetta assumes sentences as input, sets of possible translations as output, and compositional translation as the relation between an element of the input and an element of the output.
(b) Evaluation criteria justified by the approach in Rosetta are the formal correctness of the mapping and the correspondence to actual language.
(c) The general architecture of the system is interlingual rather than transfer-based. Compositionality is implemented by tuning the grammars of the different languages of the project to each other, and by the use of transformations for language-specific properties.
(d) The knowledge used in the system is taken eclectically from a variety of linguistic theories with an emphasis on Montague grammar.

In the interpretation of (a) and (b), the concept of grammaticality has a key role. Sentences in the input are assumed to be grammatical, so that if the system is

formally correct it can be guaranteed that each sentence which is recognized is also translated. The correspondence to actual language in (b) concerns the relationship between the sets of grammatical sentences in the system and in the language.

The order in the above list is not random. It reflects more or less how one type of decision evolves from or elaborates on the previous ones. The general architecture in (c) can be seen as a particular analysis of the problem specified in (a) and (b). The choice in (d) can almost be considered a consequence of the other choices listed. Its validity can in principle be verified experimentally.

The order of the issues involved in (a)–(d) is also reflected in various types of discussions on MT. As a typical example of a conference paper such as found in the proceedings of Coling 1988, van Munster (1988) discusses how, given the choices (a)–(c) for Rosetta, a particular phenomenon can best be treated. This involves the choice of knowledge adopted, i.e., issue (d), and the way this knowledge can be implemented.

In reviews of Rosetta (1994), such as Dorr (1995) and Van Eynde (1998), it is especially the choices in (c) which are scrutinized. Reviews, as opposed to conference papers, do not contribute to the development of a system, but provide a context for a general evaluation of the approach taken. The choices in (c) set off Rosetta from other MT projects such as Eurotra and Dorr's (1993) Unitran. It is less straightforward to evaluate a system than a particular treatment in a system. Elements of system performance should be related to the architecture in such a way that shortcomings are plausibly analysed as a consequence of the architecture independently of the particular solution chosen in terms of issue (d).

The practical evaluation of a system typically takes the form of an evaluation of the choices in (c) and (d) taken together. Melby (1988) proposes a procedure which evaluates the overall performance of an MT system on a previously unknown set of sentences. In order to avoid favouring the tuning of a system to a particular set of test sentences, only the vocabulary is given to system developers in advance. This evaluation procedure indicates the range of variation in which discussion of the design of MT systems takes place in this period, restricted basically to different choices in areas (c) and (d).

As far as areas (a) and (b) are formulated at all, they are either presented in introductions to MT or as a basis for introducing a new approach to areas (c) and (d). In the latter case, a new formulation may be proposed which highlights particular aspects, not previously emphasized, and relates them in a way that is slightly different from the usual one. An example is the following:

> The task of MT can be defined very simply: the computer must be able to obtain as input a text in one language (SL, for source language) and produce as output a text in another language (TL, for target language), so that the meaning of the TL text is the same as that of the SL text. (Nirenburg, 1987:2)

The degree of variation is demonstrated by the subtle differences between the view expressed by Nirenburg and the more explicit one in Rosetta (1994). It would be unthinkable, however, to find a conference paper specifically and exclusively

on, for instance, the correct problem definition in MT. This constitutes a strong indication that the Rosetta theory of MT is embodied essentially in (c) and (d), whereas (a) and (b) are part of the research programme shared with most of the other MT systems of the time.

4. The Nature of the Change

In the introduction it was observed that the changes which took place in the period from 1988 to 1998 were widely perceived as a revolution. Assuming that, as discussed in Section 2, a revolution is the result of a competition between two research programmes, and that the research programme described in Section 3 was widely adhered to in 1988, we have to look for a new research programme which made the earlier one obsolete by 1998 in order to justify an account of the change observed in terms of a revolution.

4.1. STATISTICAL VERSUS LINGUISTIC APPROACHES TO MT

The opposition between statistical and linguistic approaches to MT seems a promising place to start looking for a competition between research programmes. However, Gazdar's (1996) observation that the competing "paradigms" involved in this opposition were merging should make us extremely cautious. A central property of Kuhnian paradigms which is taken over in research programmes is their incommensurability. Two theories in different research programmes adopt incompatible world views, so that, even if they use the same vocabulary, a translation between the two is extremely difficult. In fact, this situation is well known in the object of MT as the existence of translational ambiguities. For general, epistemological reasons, a merger of statistical and linguistic approaches is therefore only possible either if they do not constitute competing research programmes or if the meaning of *merge* is relaxed so as to allow for the emergence of a new research programme loosely inspired by two earlier ones.

The intuitive description of the opposition between the two approaches in Section 1 involved the type of knowledge used in the system. It is for this reason that the different approaches were called statistical and linguistic. In terms of the analysis in Section 3.2, the selection of knowledge belongs to area (d), which does not even determine a distinction between different theories, let alone between different research programmes. This discrepancy requires an explanation.

First, the shallow influence of the type of knowledge chosen is attested by the facts. Even in Rosetta one could imagine the use of statistical techniques for specific, well-determined tasks without affecting the theory of MT adopted. In Eurotra, which explored a broader range of avenues because of its decentralized organization, there were actually experiments in the use of neural networks for the translation of prepositions. Therefore the use of statistical knowledge as such is by no means incompatible with the predominant research programme at Coling 1988.

Second, the IBM project which is usually taken as the prime example of statistical MT constitutes a more radical departure from the assumptions adopted by Rosetta than, for instance, Eurotra or Unitran. The latter three systems differed in taking Montague's compositional semantics (Rosetta), Jackendoff's conceptual structure (Unitran), and an eclectic definition of the interface structure (Eurotra) as the goal of linguistic analysis. Brown et al. (1988, 1990, 1993) do not assume linguistic analysis at all. Somers (1998:22) quotes Peter Brown as having stated that "Every time I fire a linguist, my system's performance improves". The shock this caused was due to the neglect of the commonly accepted boundaries of the area within which a solution for the problem of MT was sought.

In the case of the IBM system, the choice of knowledge from outside the area of linguistics is not a decision based on rational arguments from within a transfer-based architecture, but rather the consequence of a radical rethinking of the general architecture of an MT system. Weaver (1949) had already foreseen two approaches to MT, one taking it to be a linguistic problem, the other considering it as a problem of information theory. In the latter analysis, the correct translation of a word or phrase is found by statistical evaluation of a bilingual aligned corpus. The architecture of a system is then no longer in terms of morphological, syntactic, and semantic analysis, but rather in terms of parameter estimation (language modelling and translation modelling) and the search for the global maximum of the product of the two probabilities.

Nevertheless, the departure from earlier consensus concerns mainly issues (c) and (d) in Section 3.2. The problem to be solved is still to produce a TL sentence as a translation of the SL input sentence. Evaluation can still take place in terms of the procedure proposed by Melby (1988). It is claimed that the correspondence to actual language is improved by the use of a corpus of actual language. Only the concept of grammaticality, which was important in the interpretation of the decisions on issue (b) for Rosetta and similar systems, has no role in the IBM system. The explicit claim is that this improves translation as conceived of in the research programme to which Rosetta belongs.[4]

Therefore, while Brown et al. (1988, 1990, 1993) propose a more radical change in architecture and choice of knowledge than usual, they remain within the predominant research programme of Coling 1988 because they take sentences as input and assume the correspondence of meaning of these sentences and their translations as the standard for evaluation.

4.2. SIGNS OF A CRISIS

In Kuhn's (1970) theory, the notion of a crisis is essential to avoid relativism. It is not by mere fashion or the whim of an eloquent scholar that a revolution takes place, but only in reaction to a crisis. A crisis in a Kuhnian paradigm arises when scientists lose faith in the problem-solving potential of their paradigm and start considering more radical changes.

In a theory of research programmes in which a revolution is interpreted as the end of a research programme, the connection between crisis and revolution is less close. A crisis is a reason to develop a new research programme. As described in ten Hacken (1997), the sense of crisis may be limited to a part of the relevant section of the scientific community, in which case a revolution is unlikely to occur. Thus, the emergence of LFG, described by Bresnan and Kaplan (1982), was a response to a problem which for many linguists was explained satisfactorily by Chomsky (1980).

It is common for a well-functioning, undisputed research programme that its basic assumptions need hardly be stated and that cutting-edge research explores only a tiny part of the range of possibilities logically available within the research programme. In the research programme predominant at Coling 1988 a number of signs of a crisis can be recognized. In MT, one of the main problems was that despite large-scale investment in terms of time and money, projects considered as state-of-the-art failed to produce solutions which could be used in actual practice. As far as MT was available, the technology it used was outdated.[5]

The reactions to a crisis can initially be classified into three types. First, there is a group of scientists who refuse to consider the problem seriously. They continue their work in the usual way, trusting in the problem-solving potential of the methods and assumptions implicit in their research programme. This group is bound to shrink when problems grow more acute. Second, there is a group of scientists who attempt to formulate explicitly the underlying assumptions of the mainstream of the research programme in their most convincing form in order to defend their past work. Third, there is a group of scientists who explore the borderlines of the research programme in order to find out whether non-mainstream versions might be better.

It is probable that most of the MT researchers at Coling 1988 belonged to the first group. Crucial questions such as what counts as a translation were addressed rather reluctantly, if at all, as illustrated for instance by the remarks in Maegaard and Perschke's (1991) contribution to this journal's special issue on Eurotra. A prototypical example of the second type of reaction is found in Rosetta (1994), with its explicit statement of a theory of possible translations. The third type is represented by Brown et al. (1988).

By the mid 1990s the crisis had reached such proportions that we even find an explicit description of it in Melby (1995). The tone of this work is highly pessimistic in the sense that MT as it had been attempted for a long time was a hopeless enterprise and should be given up.[6]

4.3. A NEW RESEARCH PROGRAMME

A crisis can be resolved by the specification of a new direction for research within the same research programme. This means that the basic context of the research programme is preserved, but the possibilities are used in a way which used to

be non-mainstream. As has been argued elsewhere (ten Hacken, 1997, 2000), the fairly radical changes in the Chomskyan research programme, such as the transition from Chomsky's (1965) Standard Theory to Chomsky's (1981) Government and Binding Theory, can be interpreted in this way.

An alternative outcome of a crisis is the emergence of a new research programme. This happens when the positions taken in issues such as described in Section 3.2 are changed in a more radical way than we have seen in the approaches discussed so far. We will argue here that this is what occurred in the field of MT and that one of the key publications expressing the approach of the new research programme is Kay et al. (1994).

Kay et al. (1994) give an overview of the fields of research relevant to the research project Verbmobil. It was not meant as a feasibility study but rather as a set of recommendations on how to pursue the goal of developing a system for fully automatic MT of spoken dialogues. A remarkable feature of the overview of the field of MT is a new type of attack on the approach taken by Rosetta. Whereas all the criticism considered above concerns the issues of the general architecture and the choice of knowledge to be used (issues (c) and (d) in Section 3.2), Kay et al. launch a frontal attack against Rosetta's positions on issues (a) and (b), claiming that Rosetta "carries to an extreme that fallacy of thinking of translation as a function from a source to a target text" (1994:85). This remark refers to the lengthy discussion at the start of the book, whose conclusions are summarized as follows.

> It is clear that a translation is some transformation of the source sentence into the target language which preserves certain properties. What exactly must be preserved is difficult to say. Many of the examples we have discussed suggest that it is not the meaning, in any but the loosest sense of that word. What we will prefer to say is that a good translation is one that preserves to the extent possible, the *intention* of the original. In other words, it preserves the intended effect on the recipient." (Kay et al. 1994:27, emphasis original)

The point here is simple. It expresses the general opinion among translators and theorists of translation on the nature of a good translation as performed by human translators.[7] The reason why Kay et al. have to pile up so much evidence to support it is that their statement is in flat contradiction to the general opinion in the MT community as expressed by Nirenburg (1987), cited above.[8]

The new research programme incorporated in Kay et al.'s (1994) recommendations for Verbmobil can be summarized in a way similar to the one in Rosetta (1994) by a description of the four issues corresponding to those in Section 3.2.

(a) The problem of MT is the substitution of a TL text for a given SL text such that the effect on the recipient is as far as possible the same.
(b) Evaluation criteria should test the success of communication.
(c) The architecture of the system should allow for a negotiation process, comparing different TL translations and reconsidering the analysis of the SL input as necessary in order to choose between them.

(d) Knowledge should not be taken only from linguistic theories. One of the other promising sources of knowledge is the empirical study of human translation.

The problem description in (a) is a paraphrase of Kay et al. quoted above. The formulation of (b) is particularly appropriate for a dialogue system such as envisaged in Verbmobil, but then this setting is typical of the new research programme. The negotiation model as referred to in (c) is presented by Kay et al. (1994:93). On the types of knowledge, Kay et al. remain rather vague. They imply a fairly deep analysis along linguistic lines and are sceptical about statistical approaches (1994: 204). The empirical study of human translation is explicitly recommended (1994: 200), but no reference is made to the large body of translation theories available.

5. The Representativity of Rosetta and Verbmobil

Before drawing any more general conclusions, we have to address the question of how representative our example systems are for the classes they are meant to exemplify. In the case of Rosetta, this is not so controversial. As shown in Section 3 the discussion of aspects of Rosetta illustrates quite convincingly how choices relating to the theory of MT are differentiated from choices relating to the research programme. Whereas the architecture of the system and the choice of a linguistic theory as a source of knowledge to be applied are the subject of controversial discussion, the assumptions on the nature of translation and the proper evaluation of the MT system are not questioned in the late 1980s.

Taking Verbmobil as a representative example of a more modern approach is less straightforward. It should be kept in mind, however, which aspects of Verbmobil are meant to reflect more general tendencies and which aspects are individual choices. This is comparable to the way Rosetta is representative of the earlier approach: it is the issue of interlingua vs. transfer that is typical rather than the interlingua architecture. In the case of Verbmobil, the following cluster of connected properties is proposed as determining the general trend it is representative of. Translation is considered as a matter of communication of intentions rather than an operation on language. Evaluation is therefore based on the successful completion of a communicative task. As a consequence, greater flexibility as to the architecture of the systems and the choice of theories prevails.

One way of establishing to what extent the choices made by Verbmobil, as contrasted with the ones made by Rosetta, are representative of the general trend in MT is an analysis of how closely different conference papers are compatible with these choices rather than with the Rosetta ones. Carrying out such a comparison is more difficult than it seems. This is an immediate consequence of the fact that what is discussed at a technical conference such as Coling or ACL is not the foundations of the research programme, but results of research assuming them. It is common for a new research programme to take over as much as possible of the results of earlier ones, while reinterpreting them in terms of the new background assumptions.[9] As a consequence, the discussion of alignment techniques for translation modelling

is not by itself a reason for the classification in one or other research programme. In papers such as Ahrenberg et al. (1998) and Collier et al. (1998) the relevance of the alignment problem is simply assumed and the ultimate context of use not mentioned.

A number of interesting observations can be made concerning tendencies in the papers at Coling-ACL 1998. Insofar as the general use of MT is mentioned, the discussion focuses on real-life problems rather than on formal settings. This is to be expected for the papers on Verbmobil. In their presentation of the generation module, Becker et al. (1998) emphasize the need for real-time processing imposed by the context of use, and Emele and Dorna (1998) describe a treatment of ambiguity in line with the proposal of a negotiation module, such that resolution of ambiguity is postponed until it is inevitable. There are also various other papers with explicit references to practical applications and the constraints they impose on the MT process, including telephone communication taken on by Akbar and Caelen (1998), cross-linguistic information retrieval by Chen et al. (1998), web-based translation by Choi et al. (1998) and a multilingual web agent by Read and Barcena (1998). The general impression is that evaluation is not only much more important than at Coling 1988 but also more geared towards successful communication.

An objection to this type of argument by one of the reviewers concerns the fact that Coling and ACL are more general CL conferences, which might systematically skew the type of papers. Let us therefore consider in some detail the proceedings of a specialized MT conference in the same year, AMTA (Farwell et al., 1998). These proceedings contain 43 papers which can be divided into two categories of almost equal size. The first category contains papers discussing components and techniques at a micro-level, the second one papers considering MT systems at a higher level of generality, discussing their appropriateness for specific tasks or their general design.

Many papers in the first category are neutral as to the research programme assumed. An example is the discussion of methods for the development of lexicons, as in Fung (1998), Melamed (1998), and Miller and Zajic (1998). Even in this category, however, we find presentations which suggest an increased awareness of the practical use of MT systems. For all their difference in techniques, McCarley and Roukos (1998) and Dorr and Katsova (1998) both refer explicitly to information retrieval as the domain for which their MT system is intended.

In the second category we find a number of papers which illustrate the importance of the assumptions listed in Section 4.3, underlying the research programme of which Verbmobil was taken to be representative. Thus, Yang and Lange (1998) and Fourla and Yannoutsou (1998) present analyses of feedback by end users of generally available MT services. They underline the impact of assumptions (a) and (b), because in the earlier research programme such information was not taken to be important. Another example is Loehr (1998), who explores a new type of knowledge, not usually considered in MT, in line with assumption (d). Papers like these can only be found in such a specialized conference as AMTA and would be

misplaced at Coling or ACL. A flexible architecture as advocated by Kay et al. (1994) and listed as assumption (c), in which components are not in a rigid linear order, is also used by Woszczyna et al. (1998). Even Hong (1998), who presents work in CAT2, a system rooted in the Eurotra project and originally conceived in a context in which many of the assumptions underlying Rosetta were shared, suggests the concurrent use of multiple sources of knowledge for the solution of the same problem.

Therefore we may conclude that the assumptions listed in Section 4.3 for Verbmobil are representative of a large amount of work in MT in the late 1990s.

6. Conclusion

In order to answer the question in the title, we have to determine not whether a new research programme has emerged, but whether it has supplanted the old one. The two research programmes in question have been outlined in Sections 3.2 and 4.3, respectively. As argued in Section 5, papers at Coling-ACL 1998 and AMTA 1998 can often not be unequivocally attributed to one of these research programmes, but to the extent they can they are almost exclusively in the second one. The only examples of papers representing the older research programme are McCord and Bernth (1998) and Gdaniec (1998). They both present work on the LMT system of McCord (1989). Among the 43 papers at AMTA and the 26 papers on MT at Coling-ACL, they constitute a small minority.

These observations at least strongly suggest that the research community producing papers which are accepted at Coling, ACL, or AMTA has largely turned to the new research programme. Those researchers still clinging to the old values have either chosen to work in a way acceptable in the new perspective as well or did not produce accepted papers. Others have included at least a token reference to the new values in order to increase their chances of being accepted. Therefore, one can quite safely conclude that the revolution replacing the research programme as in Rosetta by the one exemplified by the recommendations of Kay et al. for Verbmobil is at least at an advanced stage.

Acknowledgements

I would like to thank Jan Landsbergen and Martin Emele for comments on earlier versions and Hilde Young and Margaret Gibbins for proofreading.

Notes

[1] Somers (1998:22) mentions a presentation at the Second TMI conference at Carnegie Mellon University in the same year, without referring to its proceedings.

[2] In 1998, topics for regular papers were much narrower, so that there were no sessions devoted to MT as such. Furthermore, project notes were not organized into thematic sessions. This leaves

more room for subjective interpretation. It is unlikely, however, that other counts would result in a significantly higher number of papers on MT than the 26 given here.

[3] Hoyningen-Huene (1989:143–145) gives an extensively documented overview of the development in Kuhn's thought in this respect. He shows that the loose formulations of the point in Kuhn's earlier work would be misinterpreted if taken to impose a strict condition of general consensus on paradigms.

[4] This is not to suggest that Melby (1988) was the final word on evaluation of MT systems in the research programme under discussion here. Arnold et al. (1993) distinguish evaluation procedures based on a test suite, for systems strongly governed by linguistic knowledge, and declarative evaluation procedures "particularly suited to systems which are robust and weakly rule governed" (1993: 10). In all cases, however, correct translation of sentences is the basis for evaluation. Operational evaluation, their third type, requires an operational system. One of the major problems of Rosetta and similar projects is their failure to produce an operational system for practical use (cf. the discussion of the crisis in Section 4.2).

[5] The most successful MT systems in terms of translated text are Systran and Météo. The former relies heavily on post-editing of low-quality translation and involves *ad hoc* adjustments to improve translation quality. The latter is bound to a limited domain, which proved to be remarkably resistant to any effort to extend it in an interesting way.

[6] Melby incorrectly describes this as a consequence of the Chomskyan view of language. The views he attributes to Chomsky are more typical of the formalist position defended by Montague (1970), Gazdar et al. (1985), and Katz and Postal (1991), among others. Cf. also ten Hacken (2000) for an elaboration of this opposition and its relevance to CL and ten Hacken (1998b) for a critical review from this perspective.

[7] Stolze (1994) gives an overview of different approaches to translation theory. The general trend to adopt intention and use of a text rather than meaning in the restricted sense as a basis of translation seems to be well established by the late 1970s.

[8] It is interesting to note that the object of translation is a sentence for Kay et al. and a text for Nirenburg. One would actually expect the opposite. An explanation of the usage might be that the intention of a sentence in Kay et al. can only be found in context, whereas the meaning of a text for Nirenburg is typically approached as compositionally arising from the sentences it consists of.

[9] Laudan (1977) describes this process in much detail and with ample exemplification for his concept of "research tradition". While not identical to research programmes as treated here, the general idea is similar enough to assume that Laudan's argument can be transferred to the present context.

References

Ahrenberg, Lars, Mikael Andersson and Magnus Merkel: 1998, 'A Simple Hybrid Aligner for Generating Lexical Correspondences in Parallel Texts', in Coling-ACL (1998), pp. 29–35.

Akbar, Mohammed and Jean Caelen: 1998, 'Parole et traduction automatique: le module de reconnaissance RAPHAEL' [Speech and Machine Translation: the Recognition Module RAPHAEL], in Coling-ACL (1998), pp. 36–40.

Arnold, Doug, Louisa Sadler and R. Lee Humphreys: 1993, 'Evaluation: An Assessment', *Machine Translation* **8**, 1–24.

Becker, Tilman, Wolfgang Finkler, Anne Kilger and Peter Poller, Peter: 1998, 'An Efficient Kernel for Multilingual Generation in Speech-to-Speech Dialogue Translation', in Coling-ACL (1998), pp. 110–116.

Bresnan, Joan (ed.): 1982, *The Mental Representation of Grammatical Relations*, Cambridge, MA: MIT Press.

Bresnan, Joan and Ronald M. Kaplan: 1982, 'Introduction: Grammars as Mental Representations of Language', in Bresnan (1982), pp. xvi–lii.

Brown, Peter, John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, Robert L. Mercer and Paul S. Roossin: 1988, 'A Statistical Approach to Language Translation', in Coling (1988), pp. 71–76.

Brown, Peter, John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer and Paul S. Roossin: 1990, 'A Statistical Approach to Machine Translation', *Computational Linguistics* **16**, 79–85.

Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra and Robert L. Mercer: 1993, 'The Mathematics of Statistical Machine Translation: Parameter Estimation', *Computational Linguistics* **19**, 263–311.

Chen, Hsin-His, Sheng-Jie Huang, Yung-Wei Ding and Shih-Chung Tsai: 1998, 'Proper Name Translation in Cross-Language Information Retrieval', in Coling-ACL (1998), pp. 232–236.

Choi, Sung-Kwon, Han-Min Jung, Chul-Min Sim, Taewan Kim, Dong-In Park, Jun-Sik Park and Key-Sun Choi: 1998, 'Hybrid Approaches to Improvement of Translation Quality in Web-based English-Korean Machine Translation', in Coling-ACL (1998), pp. 251–255.

Chomsky, Noam: 1965, *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press.

Chomsky, Noam: 1980, 'Rules and Representations', *The Behavioral and Brain Sciences* **3**, 1–61.

Chomsky, Noam: 1981, *Lectures on Government and Binding*, Dordrecht: Foris.

Coling: 1988, *Coling Budapest: Proceedings of the 12th International Conference on Computational Linguistics*, Budapest: John von Neumann Society for Computing Sciences.

Coling-ACL: 1998, *Coling-ACL '98: 36th Annual Meeting of the Association of Computational Linguistics and 17th International Conference on Computational Linguistics: Proceedings of the Conference*, Université de Montréal, Montréal.

Collier, Nigel, Kenji Ono and Hideki Hirakawa: 1998, 'An Experiment in Hybrid Dictionary and Statistical Sentence Alignment', in Coling-ACL (1998), pp. 268–274.

Dorr, Bonnie J.: 1993, *Machine Translation: A View from the Lexicon*, Cambridge, MA: MIT Press.

Dorr, Bonnie J.: 1995, 'Review of Rosetta, M.T. (1994), *Machine Translation*, Dordrecht: Kluwer', *Computational Linguistics* **21**, 582–589.

Dorr, Bonnie J. and Maria Katsova: 1998, 'Lexical Selection for Cross-Language Applications: Combining LCS with WordNet', in Farwell et al. (1998), pp. 438–447.

Durand, Jacques, Paul Bennett, Valerio Allegranza, Frank Van Eynde, Lee Humphreys, Paul Schmidt and Erich Steiner: 1991, 'The Eurotra Linguistic Specifications: An Overview', *Machine Translation* **6**, 103–147.

Emele, Martin C. and Michael Dorna: 1998, 'Ambiguity Preserving Machine Translation using Packed Representations', in Coling-ACL (1998), pp. 365–371.

Farwell, David, Laurie Gerber, and Eduard Hovy (eds): 1998, *Machine Translation and the Information Soup: Third Conference of the Association for Machine Translation in the Americas, AMTA'98, Proceedings*, Berlin: Springer.

Fourla, Athanassia and Olga, Yannoutsou: 1998, 'Implementing MT in the Greek Public Sector: A Users' Survey', in Farwell et al. (1998), pp. 300–307.

Fung, Pascale: 1998, 'A Statistical View on Bilingual Lexicon Extraction: From Parallel Corpora to Non-parallel Corpora', in Farwell et al. (1995), pp. 1–17.

Gazdar, Gerald: 1996, 'Paradigm Merger in Natural Language Processing', in Ian Wand and Robin Milner (eds), *Computing Tomorrow: Future Research Directions in Computer Science*, Cambridge: Cambridge University Press, pp. 88–109.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag: 1985, *Generalized Phrase Structure Grammar*, Oxford: Blackwell.

Gdaniec, Claudia: 1998, 'Lexical Choice and Syntactic Generation in a Transfer System: Transformations in the New LMT English-German System', in Farwell et al. (1998), pp. 408–420.

Hong, Munpyo: 1998, 'Multiple-Subject Constructions in the Multilingual MT-System CAT2', in Farwell et al. (1998), pp. 174–186.

Hoyningen-Huene, Paul: 1989, *Die Wissenschaftsphilosophie Thomas S. Kuhns* [Thomas S. Kuhn's Philosophy of Science], Braunschweig: Vieweg.

Katz, Jerrold J. and Paul M. Postal: 1991, 'Realism vs. Conceptualism in Linguistics', *Linguistics and Philosophy* **14**, 515–554.

Kay, Martin, Jean Mark Gawron and Peter Norvig: 1994, *Verbmobil: A Translation System for Face-to-Face Dialog*, Stanford, CA: CSLI.

Klavans, Judith and Philip Resnik (eds): 1996, *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, Cambridge, MA: MIT Press.

Kuhn, Thomas S.: 1970, *The Structure of Scientific Revolutions, Second Edition, Enlarged*, Chicago: University of Chicago Press.

Kuhn, Thomas S.: 1971, 'Notes on Lakatos', in Roger C. Buck and Robert S. Cohen (eds), *Boston Studies in the Philosophy of Science VIII*, Dordrecht: Reidel, pp. 137–146.

Laudan, Larry: 1977, *Progress and Its Problems: Towards a Theory of Scientific Growth*, Berkeley: University of California Press.

Lehrberger, John and Laurent Bourbeau: 1988, *Machine Translation: Linguistic Characteristics of MT Systems and General Methodology of Evaluation*, Amsterdam: Benjamins.

Loehr, Dan: 1998, 'Can Simultaneous Interpretation Help Machine Translation?', in Farwell et al. (1998), pp. 213–224.

Maegaard, Bente and Sergei Perschke: 1991, 'Eurotra: General System Design', *Machine Translation* **6**, 73–82.

Margolis, Howard: 1993, *Paradigms and Barriers: How Habits of Mind Govern Scientific Beliefs*, Chicago: University of Chicago Press.

McCarley, J. Scott and Salim Roukos: 1998, 'Fast Document Translation for Cross-Language Information Retrieval', in Farwell et al. (1998), pp. 150–157.

McCord, Michael C.: 1989, 'Design of LMT: A Prolog-Based Machine Translation System', *Computational Linguistics* **15**, 33–52.

McCord, Michael C. and Bernth Arendse: 1998, 'The LMT Transformational System', in Farwell et al. (1998), pp. 344–355.

Melamed, I. Dan: 1998, 'Empirical Methods for MT Lexicon Development', in Farwell et al. (1998), pp. 18–30.

Melby, Alan K.: 1988, 'Lexical Transfer: Between a Source Rock and a Hard Target', in Coling (1988), pp. 411–413.

Melby, Alan K.: 1995, *The Possibility of Language: A Discussion of the Nature of Language, with Implications for Human and Machine Translation*, Amsterdam: Benjamins.

Miller, Keith J. and David M. Zajic: 1998, 'Lexicons as Gold: Mining, Embellishment, and Reuse', in Farwell et al. (1998), pp. 484–493.

Montague, Richard: 1970, 'Universal Grammar', *Theoria* **36**, 373–398.

van Munster, Elly: 1988, 'The Treatment of Scope and Negation in Rosetta' in (Coling 1988), pp. 442–447.

Newmeyer, Frederick J.: 1986, 'Has there been a 'Chomskyan Revolution' in Linguistics?', *Language* **62**, 1–18.

Nirenburg, Sergei: 1987, 'Knowledge and Choices in Machine Translation', in Sergei Nirenburg (ed.), *Machine Translation: Theoretical and Methodological Issues*, Cambridge: Cambridge University Press, pp. 1–21.

Percival, W. Keith: 1976, 'The Applicability of Kuhn's Paradigms to the History of Linguistics', *Language* **52**, 285–294.

Read, Tim and Elena Barcena: 1998, 'JaBot: A Multilingual Java-Based Intelligent Agent for Web Sites', in Coling-ACL (1998), pp. 1086–1090.

Reijnders, Lucas: 1997, *Het boerenbedrijf in de Lage Landen: Geschiedenis en toekomst* [Farming in the Low Countries: History and Future], Amsterdam: Van Gennep.

Rosetta, M. T.: 1994, *Compositional Translation*, Dordrecht: Kluwer Academic.

Sampson, Geoffrey: 1987, 'Probabilistic Models of Analysis', in Roger Garside, Geoffrey Leech and Geoffrey Sampson (eds), *The Computational Analysis of English: A Corpus-Based Approach*, London: Longman, pp. 16–29.

Somers, Harold L.: 1998, ' "New Paradigms" in MT: The State of the Art Now that the Dust Has Settled', in Frank Van Eynde (ed.), *Proceedings of the Workshop on Machine Translation*, 10th European Summer School in Logic, Language and Information, Saarbrücken, pp. 22–33.

Steiner, Erich H. and Jutta Winter-Thielen: 1988, 'On the Semantics of Focus Phenomena in Eurotra', in Coling (1988), pp. 630–635.

Stolze, Radegundis: 1994, *Übersetzungstheorien: Eine Einführung* [Translation Theories: An Introduction], Tübingen: Narr.

ten Hacken, Pius: 1997, 'Progress and Incommensurability in Linguistics', *Beiträge zur Geschichte der Sprachwissenschaft* **7**, 287–310.

ten Hacken, Pius: 1998, 'Chomskyan Linguistics and HPSG as Competing Research Programmes', in Bernard Caron (ed.), *Actes du 16^e Congrès International des Linguistes*, Oxford: Pergamon, Paper 0042.

ten Hacken, Pius: 1998, 'Review of Melby, Alan K. (1995), *The Possibility of Language: A Discussion of the Nature of Language, with Implications for Human and Machine Translation*, Amsterdam: Benjamins', *Language* **74**, 232–233.

ten Hacken, Pius: 2000, *Reusability and Research Programmes in Computational Linguistics*, Habilitationsschrift, Universität Basel.

Thorne, James Peter: 1965, 'Review of Postal, Paul (1964), *Constituent Structure: A Study of Contemporary Models of Syntactic Description*, The Hague: Mouton', *Journal of Linguistics* **1**, 73–76.

Van Eynde, Frank: 1988, 'The Analysis of Tense and Aspect in Eurotra', in Coling (1988), pp. 699–704.

Van Eynde, Frank: 1998, 'Review of Rosetta, M. T. (1994), *Machine Translation*, Dordrecht: Kluwer', *Journal of Logic, Language and Computation* **7**, 107–110.

Weaver, Warren: 1949, 'Translation', ms. reprinted in William N. Locke and A. Donald Booth (eds), *Machine Translation of Languages, Fourteen Essays*, Cambridge, MA: Technology Press of MIT (1955), pp. 15–23.

Woszczyna, Monika, Matthew Broadhead, Donna Gates, Marsal Gavaldà, Alon Lavie, Lori Levin and Alex Waibel: 1998, 'A Modular Approach to Spoken Language Translation for Large Domains', in Farwell et al. (1998), pp. 31–40.

Yang, Jin and Elke D. Lange: 1998, 'SYSTRAN on AltaVista: A User Study on Real-Time Machine Translation on the Internet', in Farwell et al. (1998), pp. 275–285.