

Computer-Assisted Language Learning and the Revolution in Computational Linguistics

Pius ten Hacken (Swansea)

Abstract

For a long period, Computational Linguistics (CL) and Computer-Assisted Language Learning (CALL) have developed almost entirely independently of each other. A brief historical survey shows that the main reason for this state of affairs was the long preoccupation in CL with the general problem of Natural Language Understanding (NLU). As a consequence, much effort was directed to fields such as Machine Translation (MT), which were perceived as incorporating and testing NLU. CALL does not fit this model very well so that it was hardly considered worth pursuing in CL. In the 1990s the realization that products could not live up to expectations, even in the domain of MT, led to a crisis. After this crisis the dominant approach to CL has become much more problem-oriented. From this perspective, many of the earlier differences disadvantaging CALL with respect to MT have now disappeared. Therefore the revolution in CL offers promising perspectives for CALL.

1 Introduction

Computer-Assisted Language Learning (CALL) is the field concerned with the use of computer tools in second language acquisition. Somewhat surprisingly, perhaps, this field has never been closely related to Computational Linguistics (CL). Until recently, the two fields were almost completely detached. Despite occasional attempts to apply techniques of Natural Language Processing (NLP) to the recognition of errors, NLP in CALL has long remained in a very small minority position while CALL was hardly if at all recognized as a part of CL. In this contribution, I intend to show how CL could remain largely irrelevant to CALL for such a long time and why there is a good prospect that this will change in the near future. Section 1 describes the situation of CL before the revolution. In section 2, the crisis leading to the revolution in CL is outlined. The revolution itself is the topic of section 3. The implications for the field are then sketched in section 4. Finally, section 5 summarizes the conclusions.

2 Computational Linguistics as Natural Language Understanding

CL is almost as old as the first working computer. In fact, at a time when computer science was still in its infancy, Weaver (1955 [1949]) had already proposed the use of computers for translation, thus initiating research in Machine Translation (MT). Weaver considered two approaches to MT, one based on linguistic analysis and the other on information theory.

Neither of these could be implemented at the time of Weaver's proposal. Information theory had been more or less fully developed by Shannon (1948), but its application to MT required computational power of a magnitude that would not be available for several decades. Linguistic analysis appeared more promising, because it can be performed with considerably less computational power, but the theoretical elements necessary for its successful application were still missing. Thus much work in early CL was devoted to developing the basic mechanisms required for linguistic analysis.

One of the first types of knowledge to be developed concerns the computational properties of formalisms to be used in the description of languages. In response to this requirement, the theory of formal grammars was developed, mainly in the course of the 1950s. Noam Chomsky played an active role in systematizing and extending this knowledge and Chomsky (1963) provides an early, fairly comprehensive overview of the properties of grammars consisting of rewrite rules of the general type as in (1).

$$(1) \quad \alpha \rightarrow \beta$$

In this approach, a formal description of a language consists of a set of rules in which α and β in (1) are replaced by strings of symbols. When designed properly, such a system of rules is able to generate sentences. If we consider a language as a set of sentences, we can see the grammar as a definition of the language. Different types of grammar impose different conditions on α and β . Thus, if in all rules of a grammar β is not shorter than α , it can always be determined by a finite procedure whether a given sentence belongs to the grammar or not. For Context-Free Grammars (CFGs), in which α in (1) is a single symbol in each rule, the structure can be represented as a tree diagram.

The next step on the road to linguistic analysis in CL was the development of parsers. A parser is an algorithm to determine for a given sentence x and a grammar G whether G can generate x and which structure(s) G assigns to x . Ground-breaking work in this area was done in the 1960s with the development of the chart parser (cf. Varile 1983 for an overview), Earley's (1970) efficient parser for CFGs, and the more powerful Augmented Transition Networks of Woods (1970).

With a grammar formalism and a number of parsing algorithms in place, the only missing link to successful linguistic analysis was the description of the relevant languages. As it turned out, however, this problem was more recalcitrant than the other two. Chomsky developed a theory of grammar using formal rules of the type in (1), but his theory is less congenial to CL than may appear at first sight. Chomskyan linguistics has often been considered as based on a concept of language as a set of sentences and some remarks by Chomsky (1957) can be taken to support this view. At least from the early 1960s onwards, however, Chomsky has consistently and explicitly rejected such a view in favour of language as a knowledge component in the speaker's mind. Chomsky (1988) gives an accessible explanation and justification of the assumptions underlying this general approach and the type of linguistic theory it leads to.

Given this approach to language, there is no convergence in goals between Chomskyan linguistics and CL. Whereas the former is interested in describing and explaining a human being's knowledge of language, the latter is interested in processing the products of language use on a computer. An example of this divergence is the reaction to the realization that transformational rules of the type used in Chomsky (1965) are excessively powerful. This excessive power appears both in language acquisition on the basis of input sentences and in language processing leading to the understanding of sentences and utterances. In Chomskyan linguistics it was not the processing complexity but only the learnability requirement of the grammar which drove the restriction of transformations. Chomsky's linguistic theory continued to involve movement operations defined over nodes in a tree structure. In analysis, this requires the 'undoing' of movement, which is a computationally complex operation. Processing complexity of grammars produced in the Chomskyan framework has remained a major problem for their computational implementation, but this does not and need not inconvenience Chomskyan linguists. From the perspective of Chomskyan linguistics, as language is a typically human property, it is quite plausible that the human mind is structured so as to facilitate processing of the type necessary for human language. A computer does not have this structure.

From the 1970s onwards, a number of alternative linguistic theories have been developed with the computational implementation in mind. At present, the most influential ones are Lexical-Functional Grammar (LFG, cf. Bresnan 2001) and Head-Driven Phrase Structure Grammar (HPSG, cf. Pollard/Sag 1994). They still use rewrite rules of type (1) to some extent, but their actual formal basis is the unification of feature structures. Feature structures can be seen as sets of attribute-value pairs describing individual nodes in a tree structure. The formal device of feature structures and the operations on them were developed in full only in the 1980s. An early overview is Shieber (1986). By applying operations such as unification to feature structures, movement of nodes in a tree can be dispensed with. This is important for CL, because operations of this type are much more computer-friendly than undoing movement.

Given this historical development, it is understandable why for a long time research in CL, a significant part of which was at least in name devoted to MT, largely coincided with research in natural language analysis, i.e. parsing techniques and formal linguistic description. Work on different applications (e.g. MT, dialogue systems, text summarization) did not lead to major divisions in the CL research community, because in all such applications analysis was considered as the logical first step. This attitude is reflected in Kay's (1973) proposal of a modular system of natural language understanding, the parts of which could be connected in different ways depending on the requirements of the application.

If major divisions in the CL research community could not be identified on the basis of different applications, one might wonder whether there was any other source of major divisions. Most of the discussions in CL turned on issues such as the choice of linguistic theory, formalism, and parsing strategy. Although in the perception of people working in the field, different positions on these issues led to a division into competing currents of research,

they should not be confused with major divisions in the field. All of these currents were basically geared towards the same task and their success could be compared directly. This contrasts with the situation in theoretical linguistics as described in ten Hacken (1997), where Chomskyan linguistics and LFG propose different, competing research programmes, whose results are often incompatible in a way that defies an evaluative comparison.

In this context it is interesting to see that in the perception of many computational linguists, work in CL was not essentially different from work in theoretical linguistics. Thus Thompson (1983) states that theoretical linguistics aims to characterize a language and CL proper aims to do so computationally. These were especially anti-Chomskyan linguists interested in grammar and language processing. Rather than concentrating on MT for its own merits, they were working on natural-language understanding (NLU). Concrete applications, among which MT was prominent, served on the one hand as a test of whether the goal of NLU, i.e. making a computer understand human language, had been achieved and on the other hand to convince funding sources of the practical use of their enterprise.

At this stage there was little interest in CALL among CL-practitioners, which can be explained by the orientation to NLU. Whereas the translation into another language reflects the degree of understanding of a sentence achieved by the computer fairly directly, the relationship between NLU and CALL is much more complex. Conversely CALL could not readily incorporate results obtained in CL. Work in NLU starts from the assumption that the sentences to be analysed are grammatical. Much of the analysis in CALL is actually concerned with establishing whether sentences are grammatical and appropriate and, if not, how they can be corrected. Advances in NLU were thus largely irrelevant to CALL.

The use of the computer in CALL in this period, as described by Levi (1997) in his historical overview, was determined to a considerable extent by general-purpose computing and text editing. Two types of application illustrating typical techniques are the vocabulary trainer and the generator of cloze tests. A vocabulary trainer is a system for the management of a bilingual vocabulary list. It presents a word in one language and prompts the user to enter the corresponding word in the other language. It checks whether the word entered is correct, gives the appropriate feedback, and stores the result. The order of presentation can be randomized and made to take into account the user's progress in vocabulary acquisition. Nesselhauf/Tschichold (2002) give an evaluative overview of a number of commercially available products of this type. The techniques involved are restricted to general pattern matching and database management, without any specifically linguistic components.

A cloze test is a sequence of sentences with spaces for the language learner to fill in. Examples are exercises for the endings of articles and adjectives in German or the translation of ambiguous words in context. Their generation on the basis of a full text can be done by an authoring tool which prompts the teacher to import a text, indicate the words or parts of words to be deleted, and if necessary add a question or hint for the learner as to the word to be entered. The test can then be taken and corrected electronically. Interface design and pattern matching are again the basic techniques used.

3 The Crisis in CL and the Emergence of Information Theory

In its NLU orientation, CL was closer to theoretical linguistics than to actual applications. This was not only evident in areas such as CALL, considered peripheral to mainstream NLU, but also in core applications such as MT. With the gradual advance of CL research as sketched above, the gap between mainstream research and practical applications increased.

In the 1950s and early 1960s, before the availability of advanced parsing technology for natural language, state-of-the-art MT systems were in a similar position to CALL until much more recently. They were based on word-for-word substitution supplemented by a number of general computational techniques. Among these systems, Systran is no doubt the best documented and most commercially successful example. Although Toma (1977) and Trabulsi (1989) emphasize the difference between Systran and early word-for-word MT systems, an important similarity is the absence of a full parse in the translation process. Instead, local pattern matching is used to identify relationships between words which influence the operations to be carried out on them. Dictionary lookup introduces the target language (TL) equivalents to the source language (SL) words as one of the first steps. Before any syntactic analysis is attempted, a homography resolution component tries to reduce ambiguity resulting from bilingual dictionary lookup on the basis of the words immediately preceding or following the ambiguous word. Subsequent steps aim to reduce ambiguity, but only at the very end do such operations as readjusting adjective placement in English to French translation take place. There is no representational level or operational component of NLU in Systran or similar systems.

With the increasing sophistication of parsers and linguistic theory, expectations as to the performance of MT were raised. It became increasingly embarrassing for people working in MT to claim that success was just around the corner, pending some further advances in parsing technology and linguistic analysis. Yet, in terms of practical output, the results of NLU-based MT were disappointing.

A good illustration of the gap between research effort and practical use in linguistically based MT is the experience of the Commission of the European Communities (CEC). As described by Wheeler (1987), the CEC purchased its first Systran translation system, English to French, in 1975 and gradually increased the number of language pairs and the size of the dictionaries for them. As described by Maegaard/Perschke (1991), the planning phase for the CEC's own MT project, Eurotra, started soon after the purchase of Systran, in 1977. Eurotra was intended from the start to eventually replace Systran. Using state-of-the-art linguistic analysis, the idea was to represent the SL sentence at such an abstract level that the transfer to the corresponding abstract TL representation would be minimal. This means that strictly monolingual analysis is maximized in order to reduce bilingual transfer. This is an advantage in a multilingual environment because with nine languages there are nine analysis modules, but 72 transfer components.

Despite these theoretical advantages and significant investment in Eurotra over the period 1982-1992, in their overview of current translation work in the European institutions Wagner et al. (2002) do not mention Eurotra or any of its offshoots. The index has an entry "machine

translation. *See Systran*" and on the relevant pages it is described how a proprietary version of Systran is now used for information scanning on the basis of raw translation, as a source of rapid post-editing, and as an optional aid to full quality translation.

In the late 1980s the divergence between expectations and the possibility of delivery reached a point where the field entered into a state of crisis. As is typical in such a situation, one finds a number of conflicting views of the basic assumptions of the field. Lehrberger/Bourbeau (1988) represent the traditional view in (2).

- (2) The obstacles to translating by means of the computer are primarily linguistic. [...] the computer must be taught to *understand* the text - a problem in artificial intelligence. (Lehrberger/Bourbeau 1988: 1, original emphasis)

Landsbergen (1989) expresses himself much more cautiously. He suggests that machines cannot translate in the sense that translation is understood by a professional translator, but they can be of help and are able to provide all linguistically possible translations. At the same time, a group of researchers at IBM had started exploring the use of information theory in MT. The first published presentation of this project was by Brown et al. (1988). The defiant attitude of this group is reflected in the probably apocryphal quote from the research director that "Each time I fire a linguist, the system's performance improves". This attitude is a reaction to the protracted failure of linguistic theory to live up to the expectations and constitutes a radical rejection of (2).

Information theory, developed by Shannon (1948), is a branch of mathematics. Applied to MT, it requires a large parallel corpus, i.e. a number of SL texts with their TL translations. The first task is to align the SL and TL corpora, i.e. state which word(s) in the SL corpus correspond to which word(s) in the TL corpus. For large corpora, it is practically impossible to do this manually and a major part of the effort in this approach to MT is devoted to automatic alignment procedures. Once the aligned parallel corpus is in place, information theory provides the formulae to calculate the most likely translation of a word in context on the basis of the most probable TL correspondence to the SL word and the most probable TL word in the TL context.

In a sense, the IBM project described by Brown et al. (1988, 1990, 1993) eventually implemented Weaver's (1955 [1949]) suggestion to use information theory in MT, following the development of computer technology beyond a certain threshold. This happened at a time when many scholars in MT were looking for a new way of solving their problems. The impact of this approach can be measured by the fact that in 1992, the fourth TMI conference (Theoretical and Methodological Issues in Machine Translation) was entirely devoted to "Empiricist vs. Rationalist Methods in MT". Here *empiricist* refers to the use of information theory and *rationalist* to the use of linguistic theories.

To many people in the field, the emergence of these empiricist methods seemed like a revolution. As argued in more detail in ten Hacken (2001a, b), this perception is not correct. Without repeating the entire argument, let me mention two indications that the emergence of the use of information theory in MT was less revolutionary than often thought. First, the evaluation criteria for MT results remained basically the same. In both cases, the percentage

of correctly translated sentences was considered an adequate measure of the performance of the MT system, so that the results of the two approaches could be compared directly. Second, it did not take long before a merger between the two approaches could be observed, e.g. Klavans/Resnik (1996). While Gazdar (1996) calls this a paradigm merger, under a standard Kuhnian interpretation of paradigm, a merger of paradigms is a contradiction in terms, because different paradigms are incommensurable, cf. Kuhn (1970).

The influence on the field of CALL of the emergence of approaches based on information theory is also less than revolutionary. Independently of the question of how successful the approach was in MT, the problem of CALL is of a significantly different nature. Before the first public presentation of information theory-based MT, Nirenburg (1987) formulated the problem of MT as in (3).

- (3) 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 crucial point of (3) is that it allows a view of the task of MT as the mapping between two corpora. This is the central condition for the successful application of information theory. CALL cannot be modelled as a mapping of this type in any plausible way. The goal of CALL in general can be described as contributing to second language acquisition. The starting point and the end point of second language acquisition are not corpora, but knowledge states in the learner's mind. Instead of a mapping between an SL text and a TL text, CALL is meant to improve the learner's knowledge.

From the perspective of CALL, a potentially interesting property distinguishing information theory from NLU is its robustness, because dealing with non-well-formed input is one of the problems of CALL. Lack of robustness is a major problem of NLU components. When they encounter a sentence which cannot be parsed with their grammar, they reject it. It is very difficult to turn the reason for rejection as perceived by the parser into sensible feedback to the learner. In practice, in many cases sentences are rejected although native speakers would accept them. Therefore many CALL applications resorted to pattern matching, searching for particular types of error. The question is then to what extent the robustness in information theory can be put to use in CALL.

Information theory always gives the best possible match. In MT, rather than crashing or stating that the sentence cannot be translated, the system will give the most probable translation, regardless of how low this probability is. In dealing with ungrammatical input in CALL, the advantage would be that the system does not crash, but the disadvantage is that errors are not recognized. While robustness is definitely a great advantage in MT, which deals with the meaning of its input, it is of doubtful value in CALL, where discovering and describing errors is one of the aims.

Even if individual tasks in CALL could be modelled as a mapping of the type represented by (3), the performance of components based on information theory requires a knowledgeable

user. While 90% correctness seems a reasonable score, one would not like to expose a learner to a module which gives the wrong feedback in ten percent of cases.

4 The Revolution in Computational Linguistics

A revolution is a much more far-reaching change of orientation than the replacement of one approach by another. It involves not only the means by which problems are approached, but the problems themselves. The rejection of (2), the view that when NLU does not yet work, more linguistics has to be added, is not necessarily a revolution. An actual revolution occurs when statements such as (3) are no longer considered valid.

In the 1990s, the insight gained ground that MT as formulated in (3) is not a well-formed problem, i.e. a problem for which a unified solution is possible. Instrumental in this development was the Verbmobil project, carried out in Germany from 1994 to 2000. In the preparation of this project, Kay et al. (1994) argued in detail that translation of the type implied by (3) has little in common with translation as conceived of by human translators. Even though they do not refer to modern theories of translation, they insist on the fact that what is a correct translation of a sentence crucially depends on the linguistic context and the pragmatic situation.

Although (3) mentions the text as the level of translation, MT prototypes developed in the 1970s and 1980s invariably took the sentence as the basic unit. Much attention was paid to the treatment of syntactic differences across languages of the type illustrated in (4).

- (4) a. All bishops like her.
b. Zij bevalt alle bisschoppen.

The example in (4) is taken from Leermakers/Rous (1986). It is a pair of an English sentence and its Dutch translation. The Dutch verb *bevallen* (as well as its equivalents *gefallen* in German and *plaire* in French) is the most natural translation of *like*, but the two arguments are reversed. In discussions of the translation part of MT, which gradually gained importance in the late 1980s, structural divergences such as (4) and lexical divergences, illustrated by the translation of *put* into Dutch (*zetten* when the result is vertical, *leggen* when it is horizontal), occupied a central position. Rosetta (1994) and Dorr (1993) present two opposing views, with some discussion in the reviews by Dorr (1995) and Arnold (1996).

As had become common ground in translation theory in the meantime, e.g. in the skopos theory of Reiß/Vermeer (1984), the situation of a text supplies essential information for its translation. Nord (1989) even proposes a detailed analysis of the translation contract as a first step in the translation process. Given this background, the study of divergences such as (4), with its deliberately baroque lexical choice, or the Dutch translation of *put* are at most relevant in linguistics, not in translation.

If translation is situationally determined, a general solution of (3) is inherently impossible. Instead, Verbmobil went on to define a particular setting for which it would provide a solution. This setting, translation of spoken dialogues about fixing details of a meeting between German, English, and Japanese, seems rather more difficult than text translation. In

fact, however, it was chosen quite cleverly as a problem for which a reasonably successful solution is possible. The essential point is that, in the type of dialogue chosen, the two sides are cooperating to achieve a common goal. They are trying to understand each other, which implies that they accept imperfect input, ask for clarification if necessary, and confirm intermediate or final results as appropriate. The Verbmobil system, as described by Wahlster (2000), exploits these features of normal cooperative dialogue to make up for any failure to produce the best translation of individual dialogue turns.

As argued in ten Hacken (2001a), the transition from an attitude as reflected in Rosetta (1994) and Dorr (1993) to the approach found in Verbmobil constitutes a revolution. The former attitude is marked by the adoption of a holistic approach to the field of MT reflected in (3) and the central concern with linguistic theories and problems. The Verbmobil approach is marked by a focus on a special problem of practical communication. The choice of knowledge to be applied is made subordinate to the choice of a genuine, practical problem. As shown in ten Hacken (2001b), this revolution in the domain of MT is only a special case of the more general revolution in CL.

One of the implications of the nature of the revolution is that, even if we restrict our attention to MT, the Verbmobil approach as such cannot and need not be generalized to the entire field. It is precisely the dissolution of the general problem of MT which constitutes the revolution. The question is then which aspects of the approach can be generalized. A straightforward candidate seems to be the use of sublanguage, a subset of language used in a specific situation. Kay et al. (1994) are adamant, however, that Verbmobil is not based on a sublanguage. Although the vocabulary required for arranging business meetings was entered, they emphasize that no architectural decision based on this particular sublanguage should be admitted in the system. The problem for which Verbmobil was designed is the translation of cooperative spoken dialogues. Success should not be measured in terms of the percentage of correctly translated turns, but as a percentage of successfully concluded dialogues.

It is instructive to compare the Verbmobil approach with domain-specific systems developed in Artificial Intelligence (AI) and with sublanguage systems. Barr/Feigenbaum (1981) describe a number of AI systems for NLU and MT. Their starting point is not a particular situational or communicative setting, but a specific conceptual domain. The main problem they encountered is the delimitation of the domain. This is a serious problem, because only the domain is modelled in the system so that the system cannot handle input referring to entities outside the domain. In Verbmobil, it is not the domain but the communicative setting which is covered. Whereas the conceptual domain of AI systems does not correspond to a real-life problem, as their poor performance indicates, the communicative setting chosen by Verbmobil does.

Sublanguage systems are similar to the domain-dependent AI systems, but their domain is defined by a naturally occurring subset of the language rather than by a conceptual domain. The use of sublanguage was considered by many to be the only road to success in MT. In their overview of (pre-revolutionary) MT, Lehrberger/Bourbeau (1988) formulate this as in (5).

- (5) the success of FAMT in the immediate future can be expected to be limited to domain dependent systems. (Lehrberger/Bourbeau 1988: 51)

In (5), "FAMT" stands for Fully Automatic MT and "domain dependent" means taking a sublanguage as its basis. The main case in point illustrating the success of such systems is *Météo*. As described by Chandoux (1976, 1989), *Météo* is a system for the translation of weather forecasts from English into French, which has been operational since 1976. Although the success of *Météo* was generally attributed to its using a sublanguage and triggered extensive research into the use of sublanguages in MT, cf. Kittredge/Lehrberger (1982), Kittredge (1987), it has never been possible to replicate the success. Even sceptics such as Landsbergen (1989) accept that the source of *Météo*'s success is its use of a sublanguage. He attributes the success of *Météo* as opposed to the lack of success with other sublanguages to a special property of the sublanguage of weather forecasts.

The nature of the revolution in MT suggests a different interpretation. Originally, as described by Isabelle (1987), TAUM-*Météo* was developed as a simplification of a more general approach to MT. The subsequent extension of the same method to the domain of aircraft maintenance manuals was less successful, although Isabelle suggests that funding was stopped before a definitive judgement could be passed on the success of the method. Chandoux (1989) describes how the system as used successfully at the Canadian weather forecast centre is no longer the stripped down general MT system, but a completely redesigned new version, specifically geared to the situation at the Canadian translation bureau. Rather than a successful sublanguage MT system, *Météo* can therefore be seen as a forerunner of the approach taking a concrete practical situation as a basis and aiming at communicative success. This explains why the record of successful sublanguages was basically restricted to *Météo*: even the success of *Météo* should not be attributed to its choice of a successful sublanguage, but to its accidentally finding a good problem to solve. Not the sublanguage, but the communicative situation (including a fairly rigid style sheet) was responsible for its success.

The revolution in MT, of which *Verbmobil* is perhaps the first explicit reflection, can therefore be seen to have its forerunners. This is not surprising in view of the general theory of the nature of such revolutions. Before the Copernican revolution as described by Kuhn (1957), Aristarchos of Samos claimed that the Earth moved around the Sun rather than the other way around, cf. Heath (1981 [1913]). A common property of Aristarchos and *Météo* is that neither they themselves nor the field at large realized the revolutionary nature of their achievement. The revolution had to wait until the field had entered a crisis. In the case of MT, and CL in general, it can be argued that only with the revolution did it turn into a genuinely applied science, cf. ten Hacken (2001b).

5 Implications of the Revolution

The story of CL as presented so far seems very much the story of MT. In fact, MT was a very important part of CL during the first few decades of its history. Starting as the earliest application in CL, MT was then reanalysed as the most obvious test for NLU. MT was also central in the revolution in CL. It was here that the clash between expectations and actual performance led to the deepest sense of crisis. MT is a task which can be explained easily to a wide audience, so that it has generated extensive general interest, wide publicity, and substantial funding. When the expectations raised in this way are not fulfilled, such a recognition will equally get wide coverage.

It is to be expected that the most immediate implications of the revolution will also concern MT. By choosing dialogue translation, Verbmobil takes a promising subfield of MT. The project can hardly be blamed for choosing this problem and exploiting its structure, but the transfer of insights gained from dialogue translation to the translation of texts, to which most effort had been devoted before the revolution, is not straightforward. It is a useful exercise, however, to investigate how the revolution affects an established field, modifying its habits of thinking as studied in the sciences by Margolis (1993).

A post-revolutionary approach to the translation of text does not start from the assumption that text translation is a unified field. A central question to ask is why we want a particular text to be translated. Depending on the answer, different fields can be distinguished. These fields are at the same level of generalization as dialogue translation.

Ten Hacken (2003) distinguishes three types of text translation, each with a different reason why the text should be translated. The first type has the translated text as an authority, as in contracts or legal documents. Here support for human translators by means of translation memories, terminology management, and similar tools as foreshadowed by Kay (1997 [1980]) is the most promising prospect.

The second type is translation in order to know what is in a text. In such cases, translating the text is a possible, but not the most efficient approach. A summary of the text is in general more useful than the full version. Text summarization, described by Mani (2001), is a relatively new field, but it can already claim a number of successes. Depending on the setting, summarization can be followed by human translation, foreign language reading, or machine translation, but in any case the problem is much reduced.

The third type of text translation is aimed at finding the answer to a specific question. It can be a *how-to*-question ('How can I insert a photo into my web page?') or a question about a particular fact ('Who is in charge of official translation services in the canton of Fribourg?'). If we find a foreign-language manual, book, or web-site where this information is supposed to be, translation is a particularly inefficient method to find the answer, because most of what will be translated is not part of the answer. Here, different types of information retrieval can be considered. From the point of view of the end user, Answer Extraction as described by Molla et al. (1996) is probably the most efficient approach, but much depends on the concrete setting. In the same way as for summarization, further steps are necessary for the

interpretation of the result, which is still in a foreign language, but the problem is much reduced in a fairly literal sense.

The new approach to choosing and analysing problems overcomes the habit of seeing any problem involving foreign languages as a translation problem to be treated as in (3). A consequence of the revolution is that for the first time in CL the question as to *why* we want a translation is raised. This approach is not specific to MT. Ten Hacken (2001b) argues that it reflects a more general revolution covering CL as a whole. In fact there is nothing specific to MT in the move from language-oriented to problem-oriented work.

Let us now return to CALL and consider the implications of the revolution in CL. A comparison of CALL and MT first of all shows how different the two fields are in the relevant aspects. Whereas for MT it has long been claimed that, given a set of languages, there should be a single system solving the problem once and for all, cf. (3) above, such a claim has always seemed ridiculous for CALL. The 'dissolution' of MT in the revolution in CL has removed this difference.

In the new approach, the choice and analysis of a particular problem takes a central position. In the same way as translation can be thought of as a 'framework problem' defining a broad field, for which MT investigates how computers can best be applied, CALL has second language acquisition (SLA) as its framework problem. Similarly to translation studies, SLA also has its practical and theoretical branches. Unlike translation theory, much theoretical work on SLA is oriented to the use of SLA data in determining the nature of language, e.g. Hawkins (2001), Flynn/Lust (2002). The major difference on the practical side appears to be that the result of translation is a text, whereas SLA results in cognitive abilities. In view of the revolution in CL, the former view has to be adjusted. The text is not in any absolute sense the optimal result of translation, but only the most common side effect. What we are trying to achieve with translation is also ultimately cognitive and need not involve a TL text with the same meaning as the original.

Typical applications in CALL such as vocabulary trainers and authoring tools for cloze tests, discussed in section 1, are not at the same level of abstraction as the problem of dialogue translation chosen by Verbmobil. Given a setting in which dialogues of the relevant type occur, Verbmobil can provide a complete solution. Vocabulary trainers and cloze test generators do not provide a full solution to a type of SLA. They are dedicated systems for specific tasks in the domain of SLA, which depending on the theory of SLA and the teaching strategy chosen can be part of a solution to SLA in a particular setting.

As an example of a CALL system illustrating some of the typical features of the more comprehensive, problem-oriented approach I will take here the ELDIT system (<http://www.eurac.edu/eldit>). This choice is not entirely arbitrary, although it is of course not the only modern CALL system. For a recent critical overview of intelligent CALL systems, cf. Gamper/Knapp (2002). ELDIT has been developed at the European Academy in Bolzano for the specific bilingual situation found in South Tyrol. In this bilingual region (German and Italian), few people grow up as true bilinguals, but a certain level of bilingualism is required for people employed in public administration. The ELDIT system originated as an advanced

type of electronic learner's dictionary, cf. Abel/Weber (2000). Rather than simply mapping the structure of a paper learner's dictionary on a screen, the system exploited the additional possibilities offered by a computer. An example is the presentation of paradigmatic semantic relationships between words as a diagram with the base word in the middle and labelled relationships to other words, linked to explanations on semantic and stylistic characteristics and contrasts between them and the base word. Moreover, the entire learner's dictionary has been developed in parallel for German and Italian, so that at any point the user can switch between languages, for instance in order to check the equivalent in the native language. Importantly, the interface is an essential part of the system. At each point, the needs of a user with limited experience of computers are considered. Thus straightforward, self-explanatory visual representations are preferred to the use of a help module.

Two steps in the extension of ELDIT to a full-fledged learning environment are the inclusion of verb valency and morphology. The system for the encoding of verb valency is described by Abel (2002). The extensive use of insightful diagrams is intended to make the information accessible to the linguistically naive user. For the integration of morphology, knowledge from the Word Manager (WM) system for morphological dictionaries was used. WM databases include both inflection and word formation. The structure of WM is described in ten Hacken/Domenig (1996). The specific use of the organization of information in WM for CALL was recognized by ten Hacken/Tschichold (2001). The use of this information in the context of ELDIT is described by Knapp et al. (2003). With the integration of WM into ELDIT, words in a text can be analysed in terms of inflection and word formation and mapped to their dictionary entry. It is also possible to start from a dictionary entry and obtain the inflectional paradigm or the word formation relationships of the base word. The latter are represented in diagrams comparable to the paradigmatic semantic relationships and allow successive exploration of relationships in the dictionary. Finally, it is possible to inspect the rules for inflection and word formation, presented in an appropriate way for the chosen group of users.

Although ELDIT at the moment does not include NLU modules, it definitely uses CL. The computational modelling of syntactic, semantic, and morphological knowledge related to the lexicon goes beyond simple listing and pattern matching. The use of a model of (aspects of) language qualifies it as using CL in CALL.

6 Conclusion

The discussion in this paper started from two questions, one of them directed to the past (why are CL and CALL traditionally separated?) and one to the future (why are the prospects of collaboration now better?). The relevant distinction between the past and the future is that a revolution in the field of CL has thoroughly changed the general approach. Before the revolution CL concentrated on NLU. CALL is not useful as a test for NLU and NLU components are at most marginally useful in CALL. After the revolution, CL has turned to the detailed analysis of practical problems. CALL provides an interesting set of such practical problems. A revolution does not mean that all earlier knowledge is lost. In fact, researchers

try to save as much of it as possible by reinterpreting it in the new framework. Parsing techniques and theories of grammar are still used, but in a more interesting way than before. CALL is likely to be among the typical fields of application of CL in the future. In ten Hacken (2001a) it is argued that the revolution in MT took place between 1988 and 1998. The post-revolutionary future with its bright prospect of the collaboration of CALL and CL has already begun.

References

- Abbou, André (ed.) (1989): *La Traduction Assistée par Ordinateur*. Paris.
- Abel, Andrea/Weber, Vanessa (2000): "ELDIT - A Prototype of an Innovative Dictionary". In: Heid, Ulrich/Evert, Stefan/Lehmann, Egbert/Rohrer, Christian (eds.): *Proceedings of the Ninth Euralex International Congress, Euralex 2000* (2 vol.). Stuttgart: 807-818.
- Abel, Andrea (2002): "Ein neuer Ansatz der Valenzbeschreibung in einem elektronischen Lern(er)wörterbuch Deutsch-Italienisch (ELDIT)". *Lexicographica* 18: 147-167.
- Arnold, Douglas (1996): "Parameterizing Lexical Conceptual Structure for Interlingual Machine Translation: A Review of 'Machine Translation: A View from the Lexicon' by Bonnie Jean Dorr". *Machine Translation* 11: 217-241.
- Barr, Avron/Feigenbaum, Edward A. (eds.) (1981): *The Handbook of Artificial Intelligence, Volume 1*. Los Altos, Calif.
- Bresnan, Joan (2001): *Lexical-Functional Syntax*. Oxford.
- Brown, Peter/Cocke, John/Della Pietra, Stephen/Della Pietra, Vincent J./Jelinek, Fredrick/Mercer, Robert L./Roossin, Paul S. (1988): "A Statistical Approach to Language Translation". In: Vargha, Dénes (ed.): *Coling Budapest: Proceedings of the 12th International Conference on Computational Linguistics*. (2 vol.). Budapest: 71-76.
- Brown, Peter/Cocke, John/Della Pietra, Stephen/Della Pietra, Vincent J./Jelinek, Fredrick/Lafferty, John D./Mercer, Robert L./Roossin, Paul S. (1990): "A Statistical Approach to Machine Translation". *Computational Linguistics* 16: 79-85.
- Brown, Peter F./Della Pietra, Stephen A./Della Pietra, Vincent J./Mercer, Robert L. (1993): "The Mathematics of Statistical Machine Translation: Parameter Estimation". *Computational Linguistics* 19: 263-311.
- Chandioux, John (1976): "MÉTÉO: un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public". *Meta* 21: 127-133.
- Chandioux, John (1989): "10 Ans de MÉTÉO". In: Abbou, André (ed.): 169-175.
- Chomsky, Noam (1957): *Syntactic Structures*. Den Haag.
- Chomsky, Noam (1963): "Formal Properties of Grammars". In: Luce, R. Duncan et al. (eds.): Vol. 2: 323-418.
- Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. Cambridge, Mass.
- Chomsky, Noam (1988): *Language and Problems of Knowledge*. Cambridge, Mass.
- Dorr, Bonnie J. (1993): *Machine Translation: A View from the Lexicon*. Cambridge, Mass.
- Dorr, Bonnie J. (1995): "Review of Rosetta, M.T. (1994), *Machine Translation*. Dordrecht: Kluwer". *Computational Linguistics* 21: 582-589.

- Earley, Jay (1970): "An Efficient Context-Free Parsing Algorithm". *Communications of the ACM* 13: 94-102.
- Flynn, Suzanne/Lust, Barbara (2002): "A Minimalist Approach to L2 Solves a Dilemma of UG". In: Cook, Vivian (ed.): *Portraits of the L2 User*. Clevedon: 95-120.
- Gamper, Johann/Knapp, Judith (2002): "A review of intelligent CALL systems". *Computer Assisted Language Learning* 15: 329-342.
- Gazdar, Gerald (1996): "Paradigm Merger in Natural Language Processing". In: Wand, Ian/Milner, Robin (eds.): *Computing Tomorrow: Future research directions in computer science*. Cambridge: 88-109.
- ten Hacken, Pius/Domenig, Marc (1996): "Reusable Dictionaries for NLP: The Word Manager Approach". *Lexicology* 2: 232-255.
- ten Hacken, Pius (1997): "Progress and Incommensurability in Linguistics". *Beiträge zur Geschichte der Sprachwissenschaft* 7: 287-310.
- ten Hacken, Pius (2001a): "Has There Been a Revolution in Machine Translation ?". *Machine Translation* 16: 1-19.
- ten Hacken, Pius (2001b): "Revolution in Computational Linguistics: Towards a Genuinely Applied Science". In: Daelemans, Walter/Sima'an, Khalil/Veenstra, Jorn/Zavrel, Jakob (eds.): *Computational Linguistics in the Netherlands 2000: Selected Papers from the Eleventh CLIN Meeting*. Amsterdam: 60-72.
- ten Hacken, Pius/Tschichold, Cornelia (2001): "Word Manager and CALL: Structured access to the lexicon as a tool for enriching learners' vocabulary". *ReCALL* 13: 121-131.
- ten Hacken, Pius (2003): "From Machine Translation to Computer-Assisted Communication". In: Giacalone Ramat, Anna/Rigotti, Eddo/Rocci, Andrea (eds.): *Linguistica e nuovi professioni*. Milano: 161-173.
- Hawkins, Roger (2001): *Second Language Syntax: A Generative Introduction*. Oxford.
- Heath, Thomas (1981 [1913]): *Aristarchus of Samos: The Ancient Copernicus*. Oxford, reprint New York 1981.
- Isabelle, Pierre (1987): "Machine Translation at the TAUM Group". In: King, Margaret (ed.): 247-277.
- Kay, Martin (1973): "The MIND System". In: Rustin, Randall (ed.): *Natural Language Processing: Courant Computer Science Symposium 8, December 20-21, 1971*. New York: 155-188.
- Kay, Martin/Gawron, Jean Mark/Norvig, Peter (1994): *Verbmobil: A Translation System for Face-to-Face Dialog*. Stanford, Calif.
- Kay, Martin (1997 [1980]): "The Proper Place of Men and Machines in Language Translation". *Machine Translation* 12:3-23. [unchanged reprint of the 1980 ms.]
- Knapp, Judith/Pedrazzini, Sandro/ten Hacken, Pius (2003): "ELDIT and Word Manager: A Powerful Partnership". *Proceedings of Ed-Media 2003, Hawaii*: 1309-1310.
- King, Margaret (ed.) (1987): *Machine Translation Today: The State of the Art*. Edinburgh.
- Kittredge, Richard/Lehrberger, John (eds.) (1982): *Sublanguage: Studies of Language in Restricted Semantic Domains*. Berlin.
- Kittredge, Richard I. (1987): "The significance of sublanguage for automatic translation". In: Nirenburg, Sergei (ed.): 59-67.

- Klavans, Judith L./Resnik, Philip (eds.) (1996): *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. Cambridge, Mass.
- Kuhn, Thomas S. (1957): *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*. Cambridge, Mass.
- Kuhn, Thomas S. (1962/1970): *The Structure of Scientific Revolutions*. Second Edition, Enlarged. Chicago.
- Landsbergen, Jan (1989): *Kunnen machines vertalen ? Oratie*. Eindhoven.
- Leermakers, René/Rous, Joep (1986): "The Translation Method of Rosetta". *Computers and Translation* 1: 169-183.
- Lehrberger, John/Bourbeau, Laurent (1988): *Machine Translation: Linguistic characteristics of MT systems and general methodology of evaluation*. Amsterdam.
- Levi, Michael (1997): *Computer-Assisted Language Learning: Context and Conceptualization*. Oxford.
- Luce, R. Duncan/Bush, Robert R./Galanter, Eugene (eds.) (1963-1965): *Handbook of Mathematical Psychology* (3 Vol.). New York.
- Maegaard, Bente/Perschke, Sergei (1991): "Eurotra: General System Design". *Machine Translation* 6: 73-82.
- Mani, Inderjeet (2001): *Automatic Summarization*. Amsterdam.
- Margolis, Howard (1993): *Paradigms and Barriers: How Habits of Mind Govern Scientific Beliefs*. Chicago.
- Mollá Aliod, Diego/Schwitter, Rolf/Hess, Michael/Fournier, Rachel (2000): "ExtrAns, An Answer Extraction System". *T.A.L.* 41: 495-522.
- Nesselhauf, Nadja/Tschichold, Cornelia (2002): "Collocations in CALL: An Investigation of Vocabulary-Building Software for EFL". *Computer-Assisted Language Learning* 15: 251-279.
- Nirenburg, Sergei (1987): "Knowledge and Choices in Machine Translation". In: Nirenburg, Sergei (ed.): 1-21.
- Nirenburg, Sergei (ed.) (1987): *Machine Translation: Theoretical and Methodological Issues*. Cambridge.
- Nord, Christiane (1989): "Textanalyse und Übersetzungsauftrag". In: Königs, Frank G. (ed.): *Übersetzungswissenschaft und Fremdsprachenunterricht: Neue Beiträge zu einem alten Thema*. München: 95-119.
- Pollard, Carl/Sag, Ivan A. (1994): *Head-Driven Phrase Structure Grammar*. Chicago/Stanford, CA.
- Reiß, Katharina/Vermeer, Hans J. (1984): *Grundlegung einer allgemeinen Translations-theorie*. Tübingen.
- Rosetta, M.T. (1994): *Compositional Translation*. Dordrecht.
- Shannon, Claude S. (1948): "The mathematical theory of communication". *Bell Systems Technical Journal* 27: 379-423/27: 623-656.
- Shieber, Stuart M. (1986): *An Introduction to Unification-Based Approaches to Grammar*. Stanford.

- Thompson, Henry S. (1983): "Natural Language Processing: a critical analysis of the structure of the field, with some implications for parsing". In: Sparck Jones, Karen/Wilks, Yorick (eds.): *Automatic Natural Language Parsing*. Ellis Horwood: 22-31.
- Toma, Peter (1977): "SYSTRAN - Ein maschinelles Übersetzungssystem der 3. Generation". *Sprache und Datenverarbeitung* 1: 38-46.
- Trabulsi, Sami (1989): "Le Système Systran". In: Abbou, André (ed.): 15-34.
- Varile, Giovanni B. (1983): "Charts: a Data Structure for Parsing". In: King, Margaret (ed.): *Parsing Natural Language*. London: 73-87.
- Wagner, Emma/Bech, Svend/Martínez, Jesús M. (2002): *Translating for the European Union Institutions*. Manchester.
- Wahlster, Wolfgang (2000): "Mobile Speech-to-Speech Translation of Spontaneous Dialogs: An Overview of the Final Verbmobil System". In: Wahlster, Wolfgang (ed.): *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin: 3-21.
- Weaver, Warren (1955 [1949]): "Translation", ms. reprinted in: Locke, William N./Booth, A. Donald (eds.) (1955): *Machine Translation of Languages: Fourteen Essays*. Cambridge, Mass./New York: 15-23.
- Wheeler, Peter J. (1987): "Systran". In: King, Margaret (ed.): 192-208.
- Woods, William A. (1970): "Transition Network Grammars for Natural Language Analysis". *Communications of the ACM* 13: 591-606.