[*Translating and the computer 13: The Theory and Practice of Machine Translation – A Marriage of Convenience?* Papers presented at a conference … 28-29 November 1991 (London: Aslib, 1992)]

MACHINE TRANSLATION SEEN AS INTERACTIVE MULTILINGUAL TEXT GENERATION

Harold L. Somers and *Danny Jones*

UMIST

INTRODUCTION

In this paper we describe the design of what we hope is a radically new Machine Translation (MT) system, for use by a monolingual user wishing to construct texts in a specific and fairly restricted domain, which will then be generated multilingually[1]. The envisaged user is the original author of a text to be composed on-line (and interactively). The system is rooted in a specific domain, and this aspect permits the system to interpret the user's input, and even make predictions about message content, both of which will contribute to high quality of translation.

The domain that we are working in at present is the generation of multilingual job advertisements in semi-technical fields such as computer design, management or graphic arts, where the new career flexibility offered by the forthcoming open market suggests an obvious application. The domain has the advantage of moderately fixed text formats with some degree of flexibility in content, a combination which suits the particular approach we are taking. Other domains with similar properties might include:

- in a multinational conglomerate, exchange of messages between e.g. suppliers concerning orders for and availability of goods, spare parts, etc., or offers and tenders for services, equipment and so on.
- customers' and/or field engineers' initial maintenance/break-down reports on equipment (one can think of major manufacturers of electronic appliances based in an English-speaking country, but exported world-wide).
- communication between military or police forces (e.g. in NATO, or Interpol, or the need for Anglo-French communication in the Channel Tunnel).

The mode of operation is as an interactive system with a flexible human-machine interface (system- and/or user-driven) where the system cooperates with the user in order to produce a guaranteed quality translation for transmission. The system's linguistic and domain-specific knowledge bases are closely related, in that translation equivalents are tied to situational contexts. Furthermore, its domain knowledge serves to guide the user towards acceptable and standardized message structures. The system is nevertheless flexible enough to permit more than a rigid 'phrase-book' type of message generation. The system will accept input in English and generate texts in English, Spanish and Greek.

[1] The research described here is partly supported by SERC grant no. GR/G43546. The authors wish to acknowledge the contribution to the ideas expressed herein of Joseba Abaitua, Melina Alexa, Elena Barcena, Nuria Bel, John Darzentas, Anne de Roeck, Rod Johnson, Ian McLean and Mike Rosner. We remain of course solely responsible for any errors that remain.

BACKGROUND

In this section, we attempt to identify weaknesses in the classical approach to MT, as still found in so-called state-of-the-art developments. Then, we discuss a number of recent research directions which might be described as 'alternative', which we find promising.

What is wrong with state-of-art MT?

For a long time, it has been known that fully automatic high quality MT (FAHQT) of unrestricted texts is an unattainable goal. Amongst the strategies proposed to mitigate this, the most popular have been interaction with a user, and restrictions on the input. Coupled with certain typical computational and linguistic design features, these approaches have come to be known as '2nd generation MT' (Vauquois 1976). But there is now a feeling that 2nd generation MT has in some sense failed, or at least reached an impasse (cf. Somers 1990).

The interactive solution comes in the form of the widely promoted 'Translator's Workbench' idea (e.g. Kay 1980, Melby 1982), the main aims of which are to help translators to translate texts. Typically, the system proposes a translation, and then has a more or less user-friendly interactive post-editing phase; in some more ambitious cases, the system attempts to translate the input text with assistance from the human user in the form of interactions to help disambiguate source text or make lexical and stylistic selections in the target text. In this scenario, both the system and the user have knowledge about both source and target language, and it is sometimes difficult to see where the most appropriate division of labour should occur: indeed, there is sometimes a conflict between what the system offers the translator-user, and what the user already knows, or between the extent to which the system or the user should take the initiative, which might differ from occasion to occasion. Furthermore, the system has no idea of its own limitations, and so repeatedly makes the same mistakes, which become an irritation to the users who feel in the end that they could do better themselves. Even with the best of the interactive human-assisted MT systems, users who are the most positively pre-disposed towards MT soon become frustrated by the repetitive nature of the interactions, a problem at last recognised by at least one of the manufacturers of CAT systems (Seal 1992).

The use of MT with restricted input was probably first reported by Elliston (1979), and gained respectability through the great success of the Météo project (Chandioux & Guérard 1981), where MT took over a translation task too boring for any human doing it to last more than a few months, yet sufficiently constrained to allow an MT system to be devised which only makes mistakes when the input is ill-formed. Nevertheless, it is only recently that the obvious connection with 'sublanguage' has been made (Kittredge 1987, Isabelle et al. 1988, Kosaka et al. 1988, Luckhardt 1991). On the negative side, it is still the case that typically the MT system dictates the restrictions rather than vice versa: as recently as two years ago at this conference, Pym (1990) described the use of prescriptive writing rules to enable Weidner's MicroCat system to be used, and Sager talked positively of "'Systran French' or 'Logos German' ... system-specific target language forms ... based on the designers' simplified conception of the source language" (Sager 1990:7). Frankly, we do not see this in such a positive light.

The typical computational and linguistic design features of the '2nd generation' have been the use of linguistic rule-writing formalisms, independent software, stratificational analysis and generation, and an intermediate linguistically motivated representation; these innovations were admirable insofar as they seemed to address the perceived problems of '1st generation' MT system-design (i.e. before the 1966 ALPAC Report), but they nevertheless incorporate several general or specific deficiencies: they assume that the author's intention is recoverable from text alone; they reflect the preferred computational and linguistic techniques of the late 1960s and early 1970s, which have to a great extent been superseded; and worst of all, they encourage an approach to translation which can be called 'structure preserving translation as first choice', a strategy which embodies the motto "Let's produce translations that are as literal as we can get away with" (cf. Somers 1987:84).

What recent innovations have we found useful/inspiring?

Several recent innovations have nevertheless provided encouraging signs that there are new avenues to explore in the field of MT.

Although first suggested by Kay (1973), the idea of an MT system for monolingual user seems not to have been followed up until quite recently (unless one considers TITUS (Ducrot 1985) to be an MT system). Several proposals for interactive MT for monolingual users appear to have been initiated at about the same time, including DLT (Schubert 1986), the CMU system (Carbonell & Tomita 1987), Ntran (Whitelock et al. 1986, Johnson & Whitelock 1987), Zajac's (1988) system, and XTRA (Huang 1990). In systems for monolingual users it is the system which knows about translation into the target language, and the user who knows about the global context of the source text, and about the finer subtleties of the source language. The system has linguistic knowledge of the source language, but relies on the user to supply contextual and real-world knowledge in the form of an interaction. On the other hand, once the source text has been sufficiently analyzed, the system takes over completely. There is no conflict between what the system assumes to be the extent of the user's knowledge, nor in the user's expectations. The idea to develop this sort of interaction in the direction of a more sophisticated clarification *dialogue is* now gaining currency (Boitet 1989, 1990, Somers et al. 1990).

Another influential direction is the 'phrasebook' approach of Steer & Stentiford (1989). In this speech translation prototype system, set phrases are stored, as in a holidaymaker's phrasebook; they are retrieved by the fairly crude, though effective, technique of recognising keywords in a particular order in the input speech signal. The main disadvantage of this system is its inflexibility: if the phrase you want is not in the phrasebook, you cannot say anything.

Research here at UMIST for British Telecom (Jones & Tsujii 1990) on interactive generation of stereotypical texts has also been most promising: in this work, we propose a system which has stereotypical target texts in certain restricted domains (e.g. business correspondence in specific areas), retrieves appropriate texts through dialogues with users and reformulates them to fulfil the specific requirements expressed by users. In this scenario, the MT system becomes a kind of multilingual text generation system and adds a lot of information not contained in the 'source text' at all (a similar idea for 'automated text composition' in Japanese has been suggested by Saito & Tomita 1986, while Ahmad Zaki & Noor 1991 describe a system for composing official letters in Malay). Such a system has a good division of both knowledge and labour, since the system knows about translation, while the user knows only about the desired communicative content of the message. In this connection, we should also mention the idea of multilingual text generation 'without a source text' (Somers et al. 1990), an approach shared by the FoG system (Bourbeau et al. 1990, Kittredge & Polguère 1991a,b).

Finally, we should mention 'Example-Based MT' (EBMT) (Sumita et al. 1990, Jones 1991, Sato 1991, Sumita & Iida 1991) where a corpus of texts with their translations serves as a knowledge base for MT: previously translated text fragments are seen as examples on which to base the translation of a given text. This idea stems from a general proposal made by Kay in 1987 (Kay 1989:195), and also seems to have been in Nagao's mind in his (1984) article. The idea is to use a multilingual corpus of texts already translated (by humans) as a knowledge base in an MT system, so that the existing (and guaranteed) translations serve as a model for new translations. EBMT proceeds by finding suitable examples in the database and then recombining them appropriately.

Key factors are therefore the efficient retrieval of texts from the database which are sufficiently similar to the given text, a problem that we have been working on independently at UMIST (Carroll 1990) cf. also Kay & Röscheisen (1988) and Kitano & Higuchi (1991). The advantages of EBMT are that translation quality is assured, because the example translations are real. The system knows its limitations: if a suitable example cannot be found, the system will not translate on a word-for-word basis as in rule-based MT. This approach does not depend on structure preservation as a first choice, mentioned above. And perhaps most interesting of all, it is easy to extend an EBMT system: we simply add more examples to the database. Unlike in rule-based MT, there is not the overhead of 'entropy' of performance, where the addition of a new rule has unforeseen repercussions on the rest of the system, which sometimes do not surface until many months after the change was made, and therefore are extremely difficult to trace.

The use of existing translations also underlies the approach called 'Memory-Based MT' (Sato & Nagao 1990, Kitano et al. 1991), and the use of the 'Bilingual Knowledge Bank' in DLT (Sadler 1989, 1991, Sadler & Vendelmans 1990), while EBMT can be regarded as a special case of Corpus-Based MT (cf. Brown et al. 1990, 1991; Chen et al. 1991; Gale & Church 1991; Sebba 1991).

THE PROPOSED SYSTEM

How do these new research directions fit together?

Each area of research outlined in above can be classified in terms of five features:

monolingual, i.e. is the system designed for monolingual users?
dialogue-based, i.e. does the system assume a system-user dialogue?
example-based, i.e. does the system use example translations?
context-based, i.e. does the system use contextual knowledge?
sublanguage-based, i.e. does the system use domain knowledge?

However, not all features are common to any one piece of research in a generalised and coherent manner. In contrast, the proposed project will utilise information related to, or employ system design criteria associated with *each* feature listed above.

First we assume a system which is to be used by *monolinguals,* not translators. This leads us immediately to the realisation that we cannot at any time query the user as to the correctness of a translation (as nearly all current MT systems do). How then can we failsafe the translation process? This question leads naturally to the utilization of the examples, system-user dialogue, and contextual and domain knowledge sources.

The use of *example translations* has the advantage of maximally exploiting surface regularities in the given sublanguage. Such examples not only encode the correct grammatical target language constructs but also capture stylistic factors appropriate for the text-type. However, although there are major advantages in using example translations there are two problems: first, the flexible matching of input with examples, and second, the coverage of examples.

The first problem can be paraphrased as "Which examples in the database best match this input?", and can be addressed initially by the method of representation of the examples. Examples are represented at various syntactic levels differing in their degree of abstraction, e.g. as a literal string, a syntactic pattern with slots for restricted phrasal alternatives, or as a functional formula where actual and preferred surface realisations of the example semantic frame can be represented (with the associated target language equivalences). With these types of surface level representation, matching will be possible in most cases although the number of candidates will increase as the level of abstraction increases.

In order to reduce the search space, *contextual knowledge* is to be employed. As well as a syntactic description, each example translation is also associated with a contextual representation derived from the relations that hold between the example and the rest of the text from which it originated. In order to utilise this information, an ' intentional model' is required to allow the system to understand the nature of the text it is trying to compose (and consequently translate). For instance, certain contextual relations will occur together in certain circumstances (within the sublanguage). If an example with a specific contextual marker is matched, the intentional model will lead the system to expect some contextual relations but not others, thereby reducing the search space in which new input is processed. Also in cases where an input matches several candidate examples, this contextual knowledge determines the correct choice.

The second problem of coverage can be paraphrased as "Are there any examples in the database which match this input at all?", and is the basic problem with EBMT in that, in the extreme, EBMT ignores the relevance of linguistic generalities. It can be overcome by making the source text and target text composition process *recombinant.* In order to accomplish this the representations and the contextual relation definitions of the examples, and the global intentional model all play their part. It is important to ascertain the grammatical, pragmatic and stylistic legality of recombining examples and to maintain textual cohesion. In some cases, the system may require the user to rephrase the input so as to match more closely the expected input.

If recombination or rephrasing is required during the composition process, this implies *consultation with the user.* Therefore, the system must be able to interact in an intelligent manner with the user and, consequently, an additional module in the form of a human-machine interaction dialogue model is required. In order for the system to interact 'intelligently', it must understand the communicative intent of the user and therefore must have knowledge of the domain.

A consistent thread throughout this discussion has been the requirements of the *sublanguage* approach. This is by now the least controversial of features of MT research, though it is important that a commitment to sublanguage is not merely an excuse to limit the vocabulary and syntax that the system can translate. Rather we see sublanguage as an essential factor binding and defining the contextual knowledge expressed in the intentional model, the linguistic knowledge in the representations, as well as pertaining to the domain knowledge for the system overall. The system becomes portable only inasmuch as the sublanguage and its ramifications are parameterizable.

The overall design of the system is shown in Figure 1.

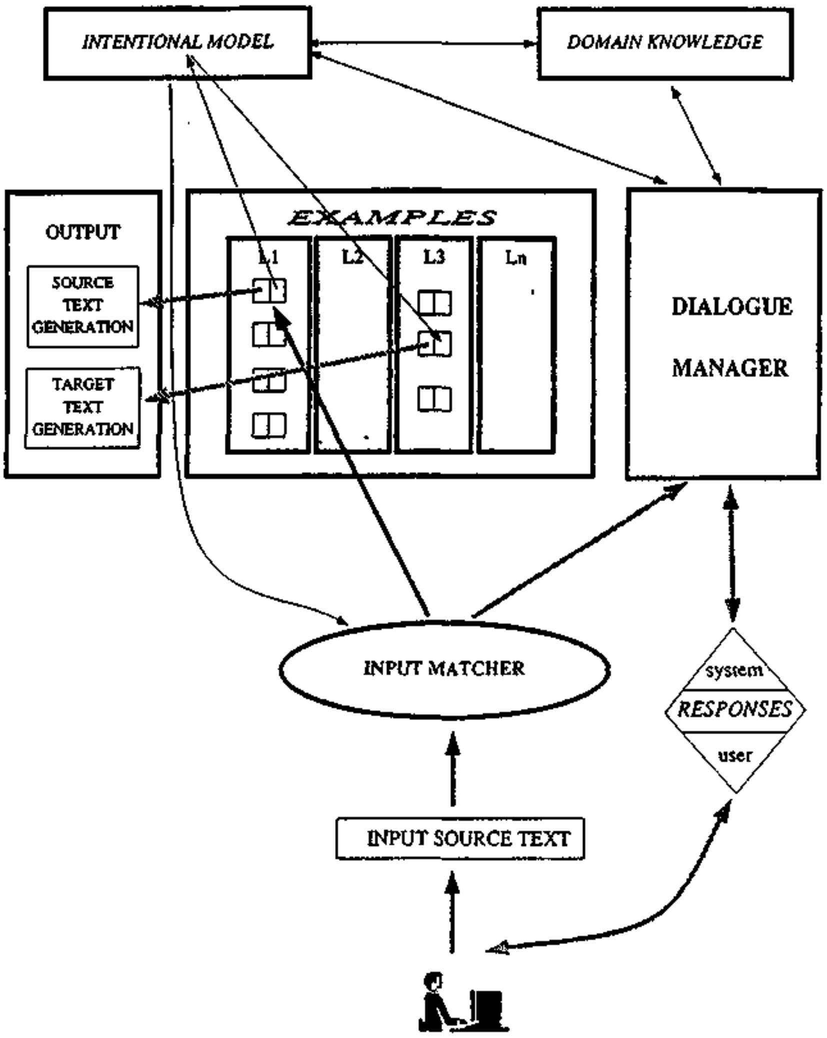
INTENTIONAL MODEL
DOMAIN KNOWLEDGE
OUTPUT
SOURCE TEXT GENERATION
TARGET TEXT GENERATION
EXAMPLES
L1
L2
L3
Ln
DIALOGUE MANAGER
INPUT MATCHER
system
RESPONSES
user
INPUT SOURCE TEXT

Figure 1

Details of the overall design

The system has two main components: the multilingual text generator, and the dialogue manager. However, these two components interact continuously, and share several knowledge sources. The operation of the system depends somewhat on the expertise of the user, but we can distinguish two extreme scenarios, with variations in between the two.

In the first scenario, appropriate with an experienced user, the user inputs a draft text. This text is input to a matching algorithm, which will attempt to find in the corpus of example texts in the user's own language a set of examples closely matching the user's proposed input. Examples are stored as pairs of syntactic and pragmatic descriptions of text: the 'syntactic descriptions' range from whole text fragments, through partial text templates to more abstract linguistic representations. The 'pragmatic descriptions' capture the pragmatic or contextual information associated with the linguistic descriptions, which serve to locate the text fragments in a more abstract 'intentional model' which defines the possible structures of the texts in general. The intentional model thus serves to restrict the search space of examples which the input matcher addresses. Since the corpora are not parallel texts as such, but merely collections of pragmatically similar texts, the intentional model also serves to give the bi- and multi-lingual links between the languages, although of course there will also be some paired examples e.g. of fixed phrases and single lexical items which provide, where necessary, 'translations' in the conventional sense. The results of the matching phase are passed to the dialogue manager, which takes into consideration how closely the user's input matches the allowable inputs given the current textual situation: the dialogue manager's knowledge about this conies from two sources, namely, the intentional model, and a domain knowledge source which contains information about the general domain. The dialogue manager is able to interact with the user in order to confirm the acceptability of the input, or to negotiate ways in which it can be changed to conform more closely to the expected range of inputs. Once the 'source' text has been successfully negotiated, there remains the task of generating corresponding target texts. This task involves locating corresponding example fragments in the target language corpora, and again because the corpora are not parallel recombining them to form appropriate target texts.

In the second scenario, the user is less experienced, and it is the system which takes the initiative. In this case, the intentional model and domain knowledge together form a database which enables the system to have a predictive model of the text, which it can then use to prompt the user more or less explicitly for appropriate input. In the extreme case, the system could almost derive for itself a series of menus or hypertexts which it would present to the user. In a less extreme case, the system can prompt the user with more general suggestions outlining the functionality of each portion of text, while permitting the user to propose a first draft for each segment. The role of the dialogue manager, and the task of generating target texts, are as in the first scenario.

Discussion

The system will use a wide variety of techniques, many of which are at the forefront of research. For the task of matching of user's input with the stored examples, parsing of a traditional nature may be employed, but the primary technique will involve stochastic or other pattern matching techniques. Because the aim is to match inputs with similar, but usually not identical examples in the database, the matching techniques must be flexible enough to locate a *range* of candidate matches against a given input. For this reason, pattern matching incorporating a similarity measure is indicated, such as the techniques proposed by Carroll (1990), Kay & Röscheisen (1988) or Kitano & Higuchi (1991).

The data forming the multilingual corpus of examples are derived empirically from real-life examples of job adverts. For practical reasons, we cannot expect to have available a truly parallel corpus of texts in this domain. The contrastive linguistic knowledge of the system cannot therefore be captured by paired examples of translational equivalents as in the IBM statistical approach for example (Brown et al. 1988, 1990, 1991), so the more abstract intentional model is relied on as a kind of mediator, where it is the functional rather than formal properly of the text fragment that gives its target language counterpart. The analysis of the multilingual corpora provides us with data influencing the design of the linguistic representations to be used in the example database, as well as determining the content and form of both the intentional model(s) where the functional and pragmatic aspects of the job adverts are defined and providing information to enable the domain knowledge of the system to be defined. This last is based on the propositional content of the corpus, which determines not only what are the commonalities of the language of the domain, but also enabling illegal phrases to be identified, as well as revealing problems of non-equivalences, especially of job titles and qualifications, etc. An additional point of interest arising especially from the text-type and domain chosen is the likelihood of cultural differences being reflected in differences in the examples and hence in the intentional models. These may be superficial, such as the typical order of presenting the information, or may pose more serious problems (see below).

Concerning the use of examples for generation by 'recombination', since the system is example-based, the issue here is not so much generation from representations which are largely given *a priori* by the corpus of examples but the capability of such a mechanism to generate texts which are not directly represented in the database of examples. The general advantages of example-based natural language processing have already been discussed. However, it is very much appreciated that in order for example-based systems to have any real degree of flexibility they must be afforded some degree of generative capacity above and beyond that supported by 'static' individual examples. This increased flexibility is gained by matching against subcomponents of more than one example across the example database. This may occur when an input text does not match against one complete example but several examples match against parts of the input. Obviously it is important not to reject the input text outright as 'ill formed' in some way, but attempt to generate a corresponding 'clone' of the input based on the highest scoring matches returned by the matching process. There is a need for information to guide this process, and this information comes from the intentional model and domain knowledge.

The domain knowledge itself includes pragmatic/linguistic knowledge captured by the intentional models, but also more general knowledge about the task of writing a job advert, which the dialogue manager will need to access in interpreting the user's input and responses to its questions, and in forming a model of the user's beliefs about the task in hand, since this is a typical example of a task-oriented dialogue. Additionally and this is of particular interest the domain knowledge must include the contrastive cultural knowledge about differences in job advert writing conventions, some of which may be essentially linguistic, but some of which might even have legal repercussions. For example, it may be acceptable to include stipulations of age and sex in an advert in one language which would actually be illegal in the corresponding advert in the target language (e.g. Spanish adverts distinguish between *senora, senorita, chica, muchacha* with a clear implication for the potential applicant which would quite possibly, if translated into the equivalent English terms, render the text illegal). A similar problem is caused by the non-equivalence of qualifications in different member states, while a third example of a cultural difference is that for a Greek national, permanence (tenure) and prestige of position appeal more than fringe benefits such as a tax-free car which would perhaps be a more significant attraction for the British candidate. These very practical problems have to be captured in the domain knowledge, and used by the dialogue manager where appropriate, to ask the user to rephrase. In the design of the system, the knowledge needs a structure, which identifies what are the components of a typical advert.

There are several other interesting aspects to this research project of which we cannot go into detail here: notably, these revolve around the dialogue part of the system. We find the project interesting both because it gives us a chance to explore some new avenues of basic research, and also because the ultimate aim of the project is to implement a small demonstrator system which, in its functionality, will represent a novel approach to translating and the computer.

REFERENCES

Ahmad Zaki A.B. and Noor H.M. 1991. Malay official letters translation. Proceedings of the International Conference on Current Issues in Computational Linguistics (Penang, Malaysia), 413-420.

ALPAC (Automatic Language Processing Advisory Committee). 1966. *Languages and Machines: computers in translation and linguistics.* NCR Publication no. 1416. Division of Behavioral Sciences, National Academy of Sciences, Washington, DC.

Boitet, C. 1989. Speech synthesis and Dialogue Based Machine Translation. ATR Symposium on Basic Research for Telephone Interpretation (Kyoto), Preprints, 6/5/1-9.

Boitet, C. 1990. Towards Personal MT: general design, dialogue structure, potential role of speech. COLING-90 (Helsinki), Vol.2, 30-35.

Bourbeau, L., D. Carcagno, E. Goldberg, R. Kittredge and A. Polguère. 1990. Bilingual generation of weather forecasts in an operations environment. COLING-90 (Helsinki), Vol.3, 318-320.

Brown, P., J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer and P. Roossin. 1988. A statistical approach to language translation. Coling Budapest, 71-76.

Brown, P.F., J. Cocke, S.A. Della Pietra, V.J. Della Pietra, F. Jelinek, J.D. Lafferty, R.L. Mercer and P.S. Roossin. 1990. A statistical approach to Machine Translation. *Computational Linguistics* 16, 79-85.

Brown, P.F., J.C. Lai and R.L. Mercer. 1991. Aligning sentences in parallel corpora. Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (Berkeley), 169-176.

Carbonell, J.G. and M. Tomita. 1987. Knowledge-based Machine Translation, the CMU approach. In Nirenburg (1987), 68-89.

Carroll, J.J. 1990. Repetitions processing using a Metric Space and the Angle of Similarity. Report No. 90/3, Centre for Computational Linguistics, UMIST, Manchester.

Chandioux, J. and M.F. Guérard. 1981. Météo: un système à 1'épreuve du temps. *Meta* 26, 18-22.

Chen, S-C., J-S. Chang, J-N. Wang and K-Y. Su. 1991. ArchTran: a corpus-based statistics-oriented English-Chinese Machine Translation system. Proceedings of Machine Translation Summit III (Washington DC), 33-40.

Ducrot, J.M. 1985. TITUS IV: System zur automatischen und gleichzeitigen Übersetzung in vier Sprachen. *Sprache und Datenverarbeitung* 9.1, 28-36.

Elliston, J.S.G. 1979. Computer-aided translation: a business viewpoint. In B.M. Snell (ed.) *Translating and the Computer,* Amsterdam: North Holland. 149-158.

Gale, W.A. and K.W. Church. 1991. A program for aligning sentences in bilingual corpora. Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (Berkeley), 177-184.

Huang, X. 1990. Machine Translation in a monolingual environment. Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages (Austin, Texas), 33-42.

Isabelle, P., M. Dymetman and E. Macklovitch. 1988. CRITTER: a translation system for agricultural market reports. Coling Budapest, 261-266.

Johnson, R.L. and P. Whitelock. 1987. Machine Translation as an expert task. In Nirenburg (1987), 136-144.

Jones, D. 1991. The processing of natural language by analogy with specific reference to Machine Translation. PhD thesis, UMIST, Manchester.

Jones, D. and J. Tsujii. 1990. Interactive high quality translation for monolinguals. Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages (Austin, Texas), 43-46.

Kay, M. 1973. The MIND system. In R. Rustin (ed.) *Natural Language Processing*, New York: Algorithmic Press. 155-188.

Kay, M. 1980. The proper place of men and machines in language translation. Research Report CSL-80-11. Xerox Palo Alto Research Center, Palo Alto, California.

Kay, M. 1989. Closing Remarks 2. In M. Nagao *(ed.) Machine Translation Summit*, Tokyo: Ohmsha. 192-196.

Kay, M. and M. Röscheisen. 1988. Text-translation alignment. Research Memo, Xerox Palo Alto Research Center, Palo Alto, California.

Kitano, H. and T. Higuchi. 1991. Massively parallel memory-based parsing. Proceedings of IJCAI-91 (Sydney).

Kittredge, R. and A. Polguère. 1991a. Dependency grammars for bilingual text generation: inside FoG's stratificational models. Proceedings of the International Conference on Current Issues in Computational Linguistics (Penang, Malaysia), 318-330.

Kittredge, R. and A. Polguère. 1991b. Generating extended bilingual texts from application knowledge bases. Proceedings of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (Kyoto), 147-160.

Kosaka, M., V. Teller and R. Grishman. 1988. A sublanguage approach to Japanese-English Machine Translation. In D. Maxwell, K. Schubert and T. Witkam (eds) *New Directions in Machine Translation,* Dordrecht: Foris. 109-120.

Luckhardt, H.-D. 1991. Sublanguages in Machine Translation. Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics (Berlin), 306-308.

Mayorcas, P. (ed.) 1990. *Translating and the Computer 10: the translation environment 10 years on,* London: Aslib.

Melby, A.K. 1982. Multi-level translation aids in a distributed system. Proceedings of Coling '82 (Prague), 215-220.

Nagao, M. 1984. A framework of a mechanical translation between Japanese and English by analogy principle. In A. Elithorn (ed.) *Artificial and Human Intelligence,* Amsterdam: Elsevier. 173-180.

Nirenburg, S. (ed.) 1987. *Machine Translation: theoretical and methodological issues,* Cambridge: Cambridge University Press.

Pym, P.J. 1990. Pre-editing and the use of simplified writing for MT: an engineer's experience of operating an MT system. In Mayorcas (1990), 80-96.

Sadler, V. 1989. *Working with Analogical Semantics: disambiguation techniques in DLT,* Dordrecht: Foris.

Sadler, V. 1991. The Textual Knowledge Bank: design, construction, applications. Proceedings of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (Kyoto), 17-32.

Sadler, V. and R. Vendelmans. 1990. Pilot implementation of a Bilingual Knowledge Bank. COLING-90 (Helsinki), Vol.3, 449-451.

Sager, J.C. 1990. Ten years of Machine Translation design and application: from FAHQT to realism. In Mayorcas (1990), 3-10.

Saito, H. and M. Tomita. 1986. On automatic composition of stereotypic documents in foreign languages. 1st Intemational Conference on Applications of Artificial Intelligence to Engineering Problems, (Southampton). Research Report CMU-CS-86-107, Department of Computer Science, Carnegie-Mellon University.

Sato, S. 1991. Example-based translation approach. Proceedings of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (Kyoto), 1-16.

Sato, S. and M. Nagao. 1990. Toward memory-based translation. COLING-90 (Helsinki), Vol.3, 247-252.

Schubert, K. 1986. Linguistic and extra-linguistic knowledge. *Computers and Translation* 1, 125-152.

Seal, T. 1992. Alpnet and TSS: the commercial realities of using a computer-aided translation service. This volume.

Sebba, M. 1991. The adequacy of corpora in Machine Translation. *Applied Computer Translation* 1, 15-27.

Somers, H. 1987. Some thoughts on interface structure(s). In W. Wilss and K.-D. Schmitz (Hgg.) *Maschinelle Übersetzung – Methoden und Werkzeuge,* Tübingen: Niemeyer. 81-99.

Somers, H.L. 1990. Current research in Machine Translation. Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages (Austin, Texas). 1-12

Somers, H.L., J. Tsujii and D. Jones. 1990. Machine Translation without a source text, COLING-90 (Helsinki), 271-276.

Steer, M.G. and F.W.M. Stentiford. 1989. Speech language translation. In J. Peckham (ed.) *Recent Developments and Applications of Natural Language Processing,* London: Kogan Page. 129-140.

Sumita, E. and H. Iida. 1991. Experiments and prospects of Example-Based Machine Translation. Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (Berkeley), 185-192.

Sumita, E., H. Iida and H. Kohyama. 1990. Translating with examples: a new approach to Machine Translation. Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages (Austin, Texas). 203-212

Tsujii, J. and M. Nagao. 1988. Dialogue translation vs. text translation interpretation based approach. Coling Budapest, 688-693.

Vauquois, B. 1976. Automatic translation – a survey of different approaches. *Statistical Methods in Linguistics* 1, 127-135.

Whitelock, P.J., M.M. Wood, B.J. Chandler, N. Holden and H.J. Horsfall. 1986. Strategies for interactive Machine Translation: the experience and implications of the UMIST Japanese project. Proceedings of Coling '86 (Bonn), 329-334.

Wood, M.M. and B.J. Chandler. 1988. Machine Translation for monolinguals. Coling Budapest, 760-763.

Zajac, R. 1988. Interactive translation: a new approach. Coling Budapest, 785-790.

AUTHORS

Harold L Somers and Danny Jones, Centre for Computational Linguistics, UMIST, PO Box 88, Manchester, M60 1QD, UK