MACHINE TRANSLATION REVIEW

The Periodical
of the
Natural Language Translation Specialist Group
of the
British Computer Society
Issue No. 6
October 1997

The *Machine Translation Review* incorporates the Newsletter of the Natural Language Translation Specialist Group of the British Computer Society and appears twice yearly.

The Review welcomes contributions, articles, book reviews, advertisements, and all items of information relating to the processing and translation of natural language. Contributions and correspondence should be addressed to:

Derek Lewis
The Editor
Machine Translation Review
School of Modern Languages
Queen's Building
University of Exeter
Exeter
EX4 4QH
United Kingdom

Tel.: +44 (0)1392 264330
Fax: +44 (0)1392 264377
E-mail: D.R.Lewis@exeter.ac.uk

ISSN 1358–8346

Contents

Group News and Information

Letter from the Chairman

I would like again to stress that we would welcome articles, papers and reports on the subject of machine translation and related subjects such as computer assisted language teaching, computer based dictionaries and aspects of multilinguality in computing, etc. We welcome papers from staff and students in linguistics and related disciplines, and from translators and any other users of MT software. We would also like more book reviewers. Below is a list of books that have been sent to the Editor for review. If you would like to review one or more of the items, please contact Derek Lewis.

We are also interested in reviews of some of the translation software being published on the Internet and in particular I am still looking for someone to review the Link parser offered by Carnegie Mellon at http://bobo.link.cs.cmu.edu/grammar/html/intro.html, which looks interesting.

If you are sufficiently interested in machine (assisted) translation to read this review you could well have some interesting knowledge or experiences to pass on to other members, so please do not be backward in coming forward with further contributions.

Perhaps I could remind members that they do not need to live near London to assist the Committee. We do not have sufficient funds to pay travel expenses for all Committee members to attend meetings, but we still welcome Correspondent members. Anyone interested in helping should contact me or any other Committee member. Correspondent committee members are treated as full members of the committee and kept advised of all committee business.

I would also like to remind you that there is a lot of MT related information on our web pages at the BCS at http://www.bcs.org.uk/siggroup/sg37.htm.

The Proceedings of the International Machine Translation Conference held at Cranfield in 1994 are now complete and negotiations are in hand to print them. We expect the cost to be about £25-£30; copies should be available in the New Year.

All opinions expressed in this Review are those of the respective writers and are not necessarily shared by the BCS or the Group.

Books for review

Daniel Jones and Harold Somers (eds), *New Methods in Language Processing*, UCL Press, 1997

Michael R. Brent (ed), *Computational Approaches to Language Acquisition*, MIT Press, 1997

Steve Young and Gerrit Bloothooft, *Corpus-based Methods in Language and Speech Processing*, Kluwer, 1997

Emmanuel Roche and Yves Schabes (eds), *Finite State Language Processing*, MIT Press, 1997

J.D.Wigg

The Committee

The telephone numbers and e-mail addresses of the Officers of the Group are as follows:

David Wigg (Chair)	Tel.: +44 (0)1732 455446 (H)
	Tel.: +44 (0)171 815 7472 (W)
	E-mail: wiggjd@vax.sbu.ac.uk
Monique L'Huillier (Secretary)	Tel.: +44 (0)1276 20488 (H)
	Tel.: +44 (0)1784 443243 (W)
	E-mail: m.lhuillier@vms.rhbnc.ac.uk
Ian Thomas (Treasurer)	Tel.: +44 (0)181 464 3955 (H)
	Tel.: +44 (0)171 382 6683 (W)
Derek Lewis (Editor)	Tel.: +44 (0)1404 814186 (H)
	Tel.: +44 (0)1392 264330 (W)
	Fax: +44 (0) 1392 264377
	E-mail: d.r.lewis@exeter.ac.uk
Catharine Scott (Assistant Editor)	Tel.: +44 (0)181 889 5155 (H)
	Tel.: +44 (0)171 607 2789 X 4008 (W)
	E-mail: c.scott@unl.ac.uk
Roger Harris (Rapporteur)	Tel.: +44 (0)181 800 2903 (H)
	E-mail: rwsh@dircon.co.uk
Correspondent Members:	
Gareth Evans (Minority Languages)	Tel.: +44 (0)1792 481144
	E-mail: g.evans@sihe.ac.uk
Ruslan Mitkov	Tel: +44 (0)1902 322471 (W)
	Fax: +44 (0)1902 322739
	E-mail: R.Mitkov@wlv.ac.uk

BCS Library

Books kindly donated by members are passed to the BCS library at the IEE, Savoy Place, London, WC2R 0BL, UK (tel: +44 (0)171 240 1871; fax: +44 (0)171 497 3557). Members of the BCS may borrow books from this library either in person or by post. All they have to provide is their membership number. The library is open Monday to Friday, 9.00 am to 5.00 pm.

Website

The website address of the BCS-NLTSG is: http://www.bcs.org.uk/siggroup/sg37.htm

Educational Implications of a Machine Translation System

by

Tatyana M. Gurina

Department of Computer Science, Kharkov Teacher Training University.

This paper is concerned with the technology of using machine translation systems in teaching foreign languages. This technology has no connection with the old form of computer-assisted language learning which uses *drill-and-practice* computer exercises and provides a kind of surrogate *electronic teacher*. This technology is intended to put the learner into the position of becoming a discoverer who discovers the meaning of the word from its context.

As is well known, a word comes to a life when it occurs in a sentence. The main objective of the two series of computer lessons which were developed as an educational spin-off of the PARS English-Russian machine translation system was to help the learner become familiar with the words in their normal contexts. Introduction of a machine translation system into teaching foreign languages was intended to obtain the most fruitful pedagogical results from the use of personal computers and to enable the learners to make use of up-to-date information technologies.

The first series of lessons is called *Homonymy* and deals with words that are pronounced like other words or word combinations but are different in meaning or spelling, such as *sum/some*.

The first lesson in this series of lessons is intended to help the learners master the technique of translating unfamiliar texts with the aid of the PARS machine translation system.

The student is given a short English text including pairs of homonyms. The first task is to delete the incorrect word in each pair, leaving the homonym that fits the context. If the learner encounters an unknown word, he/she is expected to consult the PARS electronic dictionary to obtain the translation.

The next step is to translate the text from English into Russian with the help of the translating program.

Then the learner analyses the quality of the translation performed by the computer, correcting it if necessary.

The last step is to fill in the table of homonyms:

Pair of English homonyms	Pair of Russian equivalents
... / ...	... / ...

The structure of the other lessons in the *Homonymy* series is quite different from that of the first lesson.

The text and the table of homonyms which the learner has worked with in the first lesson are available during the other lessons; the student can access them for reference.

Each lesson comprises two lists of sentences. One, called *Monolingual dictionary*, provides simple definitions of some homonyms which occur in the first text. However, the homonyms themselves are missing in these sentences. Instead of the homonym, there is a blank in each item of the *Monolingual dictionary*.

The other list is called *The list of contexts*. It provides illustrative sentences showing the meanings of the homonyms given in the *Monolingual dictionary*. The homonyms are missing in these sentences as well. The learner is given the following instructions:

- restore the *Monolingual dictionary* using the table of homonyms as a reference;
- discover which context matches each of the homonyms in the *Monolingual dictionary*;
- summarize the results of the work by filling in the following table (this should be done to make the results of the student's work easily checkable for the teacher):

No. of the item in the monolingual dictionary	Homonym explained in this item	Numbers of contexts in which this homonym occurs

Additional practical exercises are available on the basis of the lexis used in the lesson; these are designed to enhance learner's command of the language and develop flexibility of expression. The learner is expected to create some sentences of his/her own and adopt them to the MT system.

Another series of computer lessons has been developed for the topic *Phrasal verbs*.

For this series the lesson is divided into five stages:

1) translating the illustrative sentences from English into Russian by PARS;
2) choosing synonyms;
3) filling in the table;
4) extending the electronic dictionary;
5) creating new sentences.

At the beginning of the lesson, the learner is introduced to a list of 6–8 specially selected illustrative English sentences, each including a phrasal verb. There are several types of phrasal verbs. The verbs of the COME IN type consist of a verb + an adverb particle as two separate lexical units. In such phrases, IN, OUT, DOWN and the like have their basic meaning. The same pattern is frequently used with the adverb particle having a secondary meaning, as in GIVE UP. Combinations such as GIVE UP, GIVE IN (both meaning *surrender*) are idioms whose meaning is different from the sum of the expression's parts.

For the first stage, the learner performs the following tasks:

- translate the illustrative sentences from English into Russian using the machine translation system;
- analyse the machine translation output and find out in which way the program has translated each of the phrasal verbs (as an idiom or as consisting of separate lexical units);

- if the program has translated the two parts of the phrasal verb separately, check that the combined literal meaning of separate lexical items fits the illustrative sentences in which the phrasal verb occurs;
- if the phrasal verb translated by PARS is an idiom, enter both the phrasal verb and its idiomatic meaning into a special table.

The student should take into account that some phrasal verbs can have both literal and idiomatic meanings. Attention should be paid to the fact that the phrasal verbs having idiomatic meaning can be polysemantic and need illustrative sentences to show their meanings.

At the next stage of the lesson the student is given a list of verbs which are synonymous to the phrasal verbs that occur in the illustrative sentences. The synonyms in the list are jumbled up. The learner is supposed to choose a synonymous counterpart to each phrasal verb. To do this the learner should use not the original list of the illustrative sentences but its copy in which all the phrasal verbs are substituted with blanks.

If the learner does not know the meaning of a synonym, he/she can ask the MT program for help as all the verbs of the list of synonyms are stored in the PARS dictionary.

If the learner can guess the meaning of the phrasal verb after he/she has analysed the translation of the illustrative sentences performed by the program, he/she selects the synonym. Then the learner finds a suitable context, fills in the blank and asks PARS for a translation to make sure that his/ her choice is correct.

If the learner has no idea about the meaning of the phrasal verb, he/she has to use the trial-and-error method to make his/her choice.

The third stage is meant to help the student to keep in mind the new information he/she has acquired at the lesson.

After the learner has chosen the synonyms to all the phrasal verbs, he/she is supposed to fill in the table in which his/her work is summarized:

Phrasal Verb	Synonym	Russian equivalent

At the fourth stage, the learner is supposed to extend the electronic dictionary by adding to it the phrasal verbs having idiomatic meanings which were absent in the initial dictionary.

The last stage of the lesson is meant to stimulate the learner's creativity. The learner must create as many sentences as possible, using phrasal verbs he/she has learned duirng the lesson. Another task is to make them up in such a way that PARS is able to translate them in the best way possible.

These lessons have been used for several years at Kharkov Teacher Training University. They have proved to be effective and to make learning English an interesting combination of language-based and IT-related activity. PARS lends itself to this kind of work, as its dictionary updating routine is very easy to master and to use.

Tatyana M. Gurina may be contacted at 102, Prospekt Traktorostroiteley, apt.47, Kharkov 310118, Ukraine. Tel: (0572) 63 28 59, e-mail: blekhman@lotus.kpi.kharkov.ua

Terminology for Machine Translation: a Study

by

David Quinn
Centre for Applications of Advanced IT (CAIT)

Introduction

Terminology

Terminology is specialised vocabulary used to refer to the concepts of a particular domain. A useful method of improving machine translation (MT) of technical documents is to extend the coverage of the MT system's dictionary to include more terminology from that domain. This relates both to terms not in the dictionary, and to those without a suitable translation for that technical domain.

This article describes the terminology investigation and development undertaken by BICC as an associate partner in the OTELO research project.

The OTELO Project

OTELO is a research and technological development (RTD) project, partly funded by the European Commission under the Language Engineering sector of the Telematics Programme. The goal of OTELO is the integration of MT systems and other translation tools, such as translation memory, into a translation environment.

This includes the development of a common lexical resource format, to allow for the porting of lexicons between MT systems. The OTELO lexical interface and importation facility will allow the development and maintenance of a terminology database, and the export of these terms to supplement the existing dictionaries of the OTELO MT systems.

The rôle of BICC is as a user of the OTELO technologies, with translation between English and Spanish, particularly *into* Spanish, of sample technical documents in the *cables* domain. The work within the project includes a focus on the terminological aspects of this translation.

OTELO is part of a large portfolio of projects carried out by BICC CAIT through which new technologies are investigated and implemented, with the objective of improving business and manufacturing performance.

Development of Terminological Entries

Three main types of terminological entries were developed. In addition to bilingual (transfer) entries, detailed monolingual entries are needed for both the English and Spanish terms. These enable the analysis by an MT system of the source language text, the translation (transfer) of the terms, and the generation of the corresponding target language text.

Each type of entry consists of a particular set of fields (e.g. Part of Speech), to which values must be assigned (e.g. 'noun'). The OTELO common lexical resource format includes a division of lexical entries into partitions (which include the above three types), and a

specification of the fields in each partition and of the allowed values for each field (Ritzke 1997). The specification built on an analysis of the features used in four lexicon formats to be integrated within OTELO, and of their commonalities (McCormick 1997).

The fields used in our terminological development work are drawn from this specification. Terminological entries do not require all of the fields and values needed for lexical entries in general. The fields were selected on the basis of usefulness to the task, and feasibility of collection. Some are needed to enable machine translation, while others are for the usability and reusability of the terminology itself. This article describes the methods used to populate the selected fields.

Spanish

The work described involves translation from English into Spanish. Spanish has two main dialects, as used in Spain (Castilian) and in Latin America. It is a Romance language, part of the family of languages descended from Latin. This family includes, among a few others, French, Portuguese and Catalan.

Spanish is a major world language, with around 280 million native speakers. The language is 'demonstrably well equipped to express and define existing scientific fact' (Routledge 1997: xiii). Where new vocabulary is created, it uses productive Spanish patterns, especially a range of standard suffixes (Green 1990).

Monolingual English Terminology

The following three subsections discuss English-language cables terminology. The first presents the linguistic structure of English technical terminology, with examples from the cables domain. The next describes the glossaries and other resources used as sources of English cables terms, and includes a discussion of corpus-based terminology extraction. The third subsection discusses the development of monolingual entries for the English cables terms obtained from these sources.

The Structure of English Technical Terminology

Justeson and Katz (1995) analysed around 200 English technical terms from each of four domains. They found that about 96% of the terms were noun phrases (NPs), about 4% adjectives, and less than 0.5% (3) were verbs. The NPs had a mean length of 1.9 words — 70% were compound.

97% of the compound NPs contained only nouns and adjectives, and 3% included a preposition (almost always *of*). A terminology identification algorithm was presented, which includes the recognition of the standard patterns for terms. A candidate term is either 'a string of nouns and/or adjectives, ending in a noun', or 'two such strings, separated by a single preposition' (Justeson and Katz 1995: 17). This gives seven basic patterns for two or three word terms, containing only nouns (N), adjectives (Adj) and prepositions (Prep). 95% of the compound NPs analysed were of length two or three.

These results were found to be closely applicable to cables terms. The following examples, of terms drawn from the cables domain, illustrate the seven basic patterns:

Adj	N				metallic insulator
N	N				distribution cabinet

Adj	Adj	N			transducing passive quadripole
Adj	N	N			overhead conductor tap
N	Adj	N			tape armoured cable
N	N	N			anchor wire insulator
N	Prep	N			angle of wires

The longest term found, containing seven words, is *solid rubber insulated twisted pair copper wire*, with the pattern Adj N Adj Adj N N N. The grouping is as follows, with a top-level Adj N N N pattern:

$((\text{solid}_{\text{Adj}}\ \text{rubber}_{\text{N}})_{\text{N}}\ \text{insulated}_{\text{Adj}})_{\text{Adj}}\ (\text{twisted}_{\text{Adj}}\ \text{pair}_{\text{N}})_{\text{N}}\ \text{copper}_{\text{N}}\ \text{wire}_{\text{N}}$

Adjectives, though relatively rare, are the next most common element. Further to the analysis of NPs given in Justeson and Katz (1995), there are at least three main patterns for adjectives. One includes the part of speech adverb (Adv). The patterns are illustrated using examples from the cables domain:

Adv	Adj			externally operable
N	Adj			wire armoured
N	N	Adj		wire braid reinforced

Sources of Monolingual English Terminology

The main sources of monolingual English cables terms that were used are described below. They provided a collection of English terms which denote many of the cables concepts. Monolingual and then bilingual terminological entries were developed for most of these terms.

Internal Cables Glossary

An internal glossary of English cables terms (Blackburn 1965) was used as the main source of the central vocabulary for the cables domain. The glossary contains explanations of cables terms, and has approximately 600 main entries. The list of terms extracted from this glossary provided the concepts for which detailed terminological entries were developed.

The structure of the terms in the glossary is very much as described in the previous section, though the proportions of nouns and of compounds are a little lower. About 90% of the terms are nouns (compared to 96%), while a further 9% are adjectives. The mean length of the terms was 1.7 words, with about 60% of the terms being compound (compared to 70%). Again, an overwhelming proportion of the compounds (95%) contain either two or three words.

VERBA

The VERBA resource (Ediciones VERBA 1997) contains 644 Spanish – English entries in the area of cables and conductors (see section 0, on bilingual resources, for further details). The resource contains about 550 distinct English terms, and was the second main source of cables terms. Detailed terminological entries were developed for these terms.

The overlap with the internal glossary has turned out be very small, with only 19 terms in common. The mean length of the English terms was 2.1 words (compared to 1.7 in the glossary). A very high percentage (about 87%) of the English terms are compound (compared to 60% in the glossary).

Thus, the entries are more complex and specialised than those in the internal cables glossary, which accounts partly for the small overlap. In addition, the limited overlap suggests that the number of general and specialised cables terms is probably in the thousands.

BSI Glossary

The British Standards Institution has a number of glossaries which include cables terminology. BS 4727, parts 1 to 3, is a *Glossary of Electrotechnical, power, telecommunication, electronics, lighting and colour terms*. It is divided into about 50 groups, with a total of around 3000 pages.

One group was selected as the most relevant source of general cables terms:

BS 4727: Part 2: Group 08: 1994 (Part 2. Terms particular to power engineering. Group 08. Electric cables). It is 24 pages long, and contains definitions for 190 terms.

This glossary was used for reference purposes. Those terms occurring in it were marked in the monolingual term entries, pointing to this source of authoritative definitions.

Terminology Extraction from Corpora

Even where sources of terms, such as lexicons, glossaries and thesauri, are available, they are unlikely exactly to describe the terminology of a domain. The resources may have a different focus, not be sufficiently up-to-date, not be specialised enough, or just contain too few terms. The automatic extraction of candidate terms from a domain-specific electronic corpus, followed by a process of human validation, is an important means of supplementing such resources. A good-sized corpus of texts from the organisation itself would be particularly valuable. 'Serious terminology compilation is now firmly corpus-based, i.e. it relies on the analysis of textual evidence which covers the full range of usage of the special language in question' (Sager 1990: 154).

A small sample of electronic documents for translation from English to Spanish was collected, but did not constitute a corpus. However, some development and testing of terminology extraction algorithms was done. It is described below, along with further discussion.

Extraction of Single Word Terms

One method of extracting single word domain terminology is to find those words with the highest *relative frequency ratio* — the ratio of the word's relative frequency in the domain documents to that in general English text (for example, Damerau 1993). The idea is that domain terms occur much more frequently in documents from that domain.

The BNC (British National Corpus 1997) is a 100 million word electronic corpus of written and spoken British English, representing a wide range of current usage. Adam Kilgarriff of the Information Technology Research Institute, University of Brighton, has produced and

made available various frequency lists of words occurring in the BNC (Kilgarriff 1996). These include a list of each word occurring over 5 times in the written English BNC sub-corpus (90 million words), its frequency, part of speech, and how many files it appears in.

Corpus analysis programs (Quinn 1996), developed previously in the Construction Industry Specification Analysis and Understanding (CISAU) project (Douglas, Hurst and Imlah 1995), were used to extract the words and their frequencies (along with other information) from a 4000-word cables document. Further programs were written to calculate the relative frequency ratios, as compared to the above BNC frequencies. Reports were produced of the words with highest relative frequency ratios, and with lowest frequencies in the BNC sub-corpus. Potentially valuable candidate terms (e.g. *resistance*, *elastomeric*, and *IEC — International Electrotechnical Commission*) were retrieved in this test run.

Extraction of compound terms

Other CISAU corpus analysis programs produce all-word sequences of a given length (n-grams), and the frequencies of the sequence and of the component words. Candidate two-word terms (bigrams) are produced using a score that is a weighted combination of three factors. These are the relative frequency of the compound, the mutual information statistic (Church and Hanks 1990), and the conditional probabilities of each word of the compound, given the other. When run against the same cables document as above, the output included some useful terms, such as *tensile strength* and *radial thickness*.

A weakness of an approach such as the above is that it is purely statistical. The standard patterns for technical terms in English (described in section 0) utilise linguistic part of speech information, and are used in the TERMS program for terminology identification (Justeson and Katz 1995). They observe that noun phrases that are repeated exactly within a technical document are likely to be terminological. Thus, candidate terms produced by TERMS have a minimum frequency of 2. Raising this minimum improves precision: higher frequency candidates are more likely to be terms.

A hybrid approach, combining lexical statistics and linguistic filtering, is used in Daille (1995). She employs three patterns for French terminology consisting of two main constituents. They include the parts of speech noun (N), adjective (Adj), preposition (Prep), and (optionally) determiner (Det). As the patterns can be directly applied to Spanish, examples are given from the cables domain (interestingly, the N N pattern is rare in Spanish):

N	Prep	(Det)	N	abrazadera de cable — cable clamp
N	Adj			cable aéreo — aerial cable
N	N			cable portador — carrier cable

Daille empirically compares the effectiveness of a variety of statistical measures for extraction of French terminology. The study results in the selection of Dunning's log-likelihood coefficient (Dunning 1993) as the measure which best discriminates between terminology and non-terminology.

A further useful technique is the grouping for human inspection of candidate terms with the same head noun, as is done in the terminology identification tool *termight* (Dagan and Church 1994). This feature of the tool is very powerful. Related terms may be more readily examined

in a group than in isolation. 'Frequent head words are likely to generate a number of terms' (Dagan and Church 1994: 37). Infrequent items are less likely to be missed.

Development of Monolingual Terminological Entries (English)

The main fields for monolingual English terminological entries and the methods which were used for obtaining their values are discussed below.

In the context of monolingual English entries for cables terms, some fields always have a particular value, and are updated programmatically:

Language is 'English'; and

Subject Area is 'cables'. Term entries from given subject areas may be used by the MT system in preference to other entries, by allocation of priorities.

The **Term** (e.g. 'coaxial cable') appears in the terminological resource.

Source (e.g. 'Internal glossary') identifies the terminological resource. It is useful for terminology management.

Dialect (empty, 'UK' or 'US') is noted, for example, in the internal glossary. It allows an MT system to ignore the entries with dialect different from that selected for translating a particular document.

Type of Entry is available from the terminological resource (in the case of abbreviations, e.g. *PVC* for *polyvinylchloride*). The Type of Entry, if not 'abbreviation', is updated programmatically as either a single word or a multiword term. It allows special processing of abbreviations by the MT system.

Head Word has a value only for entries with Type of Entry 'multiword term'. It is defaulted programmatically to the last word (for English), for example in *coaxial **cable***. This is almost always correct, but may be reviewed manually. The most common case in which it may be incorrect is where the term includes a preposition, for example in ***angle** of wires* (see section 0, on the structure of English terms).

This field identifies the inflecting word for automatic pluralisation by the MT system. For example, *coefficient* is the head word, yielding *coefficient**s** of expansion*, as opposed to the usual case (e.g. *lever switch**es***).

Part of Speech, for the great majority of terms, has the value 'noun'. A few terms are adjectives, adverbs, or verbs. This field is updated manually. A standard ambiguity is between noun and verb. In order to provide broad coverage for machine translation, a terminological entry should be developed for the verb, as well as the noun.

Natural Gender has a value only for entries with Part of Speech 'noun'. It is updated programmatically with the value 'neuter', as the terms almost invariably refer to inanimate objects (e.g. *cable*), and it is then quickly reviewed manually. The Natural Gender allows an MT system to resolve (or select) English pronouns (e.g. *it* for the above case, as opposed to *he* or *she*).

Number has a value only when the Part of Speech is 'noun'. It has the empty value except where a term occurs in only the singular or the plural. In the usual case where a term may occur in both singular and plural forms, a lexical entry is created only for the singular form of

the term. The MT system generates the plural form, using standard English morphological rules for nouns.

This field is updated by manual inspection of the noun entries (i.e. almost all the entries), but programmatic assistance is possible. The first subtask is to identify plural terms and decide whether they may also occur in the singular. If yes, the value of the Term field should be altered to the singular form. If no (e.g. for the term *wiring sundries*), the value of Number should be set to 'plural'. Programmatic recognition of plural nouns is of assistance in this subtask (for example, using the CISAU corpus analysis programs).

The second subtask is to identify terms which occur only in the singular (e.g. *cable laying*). The most common cases of this for English are probably processes and mass nouns. Many processes have the present participle verb form with suffix *-ing* (e.g. *curing*, *annealing*). Terms ending in *-ing* may be programmatically extracted for manual review. Materials (e.g. *copper*) are commonly occurring mass nouns in technical domains.

Transitivity has a value only in the case where the Part of Speech is 'verb'. The verbs are usually transitive (e.g. to *insulate* a conductor) or intransitive. This field is updated manually for the small number of verbs.

Some resources include a **Definition**, but it is not intended to enter these. However, where there is a need to distinguish senses of homographs (e.g. *lead*), a brief explanation is entered ('the metal', or 'a cord or flex').

Comment on Definition is used to refer to the definition in the BSI Glossary.

Example/Context may contain multiple contexts (examples of a term's usage). It could be updated programmatically from electronic resources with some degree of manual review. However, not enough data has been collected.

Abbreviations, synonyms, and variant forms have their own lexical entries, and are cross referenced to the primary lexical entry.

Bilingual English - Spanish Terminology

Sources of Bilingual English - Spanish Terminology

European Language Resources Association

ELRA (European Language Resources Association 1997) promotes the collection and distribution of language resources, and is supported by the European Commission.

644 Spanish–English entries in the specific area of cables and conductors (Ediciones VERBA 1997) were purchased through ELRA. The terms are supplied by Ediciones VERBA (Madrid). The entries consist of pairs of English and Spanish terms. The English verbs are marked (with a parenthesised 'to'), but the rest of the terminological entries remained to be developed.

Further more general terminology is also available through ELRA, for example electrical engineering terms.

Technical Dictionaries

The process of choosing technical dictionar(ies) for use in the project included inspection of a good-sized catalogue of specialist dictionaries produced by a London bookseller (Grant & Cutler 1997).

The Routledge Spanish Technical Dictionary, volume 2: English – Spanish (Routledge 1997) contains over 110,000 entries, and is up-to-date. It covers over 70 subjects, including relevant areas such as electrical engineering and telecommunications. An electronic version of the dictionary is available on CD-ROM, but is bidirectional only, at a similar cost to the two volumes of the printed dictionary.

The two volumes of this dictionary were created from a database developed by translation into Spanish of an English term list. Thus, the dictionary is particularly appropriate for use in the current context, with Spanish as target language. The English term list was itself based on the Routledge French Technical Dictionary (1994), but the process of creation of the current dictionary included review of the adequacy of coverage of terms.

The English terms from the internal glossary were looked up in the dictionary in order to find appropriate Spanish translation(s) and the Spanish gender of each (for nouns only). The entry having the same Part Of Speech as that of the English term was selected. The subject area labels and sometimes the context (especially including the word *cable*) were used to select the translations most appropriate to the cables domain. The primary subject area of relevance is Electrical Engineering, with other important areas including Telecommunications, Plastics And Rubber, Electricity, and Mechanical Engineering.

Eurodicautom

Eurodicautom (Eurodicautom 1997) — *European Automatic Dictionary* — is a broad-coverage terminological database produced by the Translation Service of the European Commission. It contains over 400,000 scientific and technical terms, and over 100,000 abbreviations, in all 11 official languages of the European Union.

The use of Eurodicautom is free, and on-line access is available via the World Wide Web. It is used by selecting the database (terms or abbreviations), the source language, and one or more target languages, and then entering items for translation.

The main fields of interest in the current context were the English and Spanish Keywords (VE: French *vedette — term*), and the Reliability Code (CF: confidence factor, from 0 up to 5). Other fields, particularly the Subject Field (CM), the Definition (DF), and the Explanatory Note (NT) were used to help select the translation most appropriate to the cables domain.

The Eurodicautom database of terms is a very useful resource, containing many general terms, though irregular in its coverage. It includes for example the terms *vulcanisation*, *to vulcanise*, *degree of vulcanisation*, *over-vulcanisation*, *vulcanising agent*, and *vulcanised*, as well as further derivatives. On the other hand, for example, it includes *cable* (and *to cable*), but no compounds of *cable*.

The abbreviations database did not prove useful for the cables domain. Though a typical search retrieved several different uses of an abbreviation, the cables usage was not often among them. Where it was, there was no corresponding Spanish entry.

Materials Vocabulary

A common reference vocabulary (multilingual) for the European Materials Databanks Demonstrator Programme was compiled at UMIST on behalf of the European Commission (EEC 1988).

Spanish translations for 41 English terms from the internal cables glossary were found, as well as a few abbreviations and variants. Only about a third of these terms are materials (e.g. *aluminium*, *polystyrene*), with further examples including processes (e.g. *annealing*) and properties (e.g. *tensile strength*).

English – Spanish Samples

Possibly the best source of terms as used within a particular company is parallel documents produced within the company itself. Some internal newsletters, in parallel English and Spanish versions, were collected in printed form. Articles from the newsletters were used for test translations, and as part of an informal evaluation (see section 0).

As in the monolingual case, corpus-based methods are very important. Programs for alignment of parallel electronic texts, and extraction of likely translation pairs, provide valuable input to the development of bilingual terminology. See for example (Garside 1995) for discussion on matching words in a bilingual corpus.

Transfer Entries

The following are the main fields used for the bilingual English to Spanish transfer entries.

The **Source Word** (e.g. 'vulcanisation') is the English term.

The **Source Language** is 'English'.

The **Target Word** (e.g. 'vulcanización') is the Spanish translation of the English term.

The **Target Language** is 'Spanish'.

The **Source** is the source of the *translation* (e.g. 'Eurodicautom').

The **Type of Equivalence** (e.g. 'full').

The **Comment On Transfer Relation** is a comment field in which is entered, for example, the Eurodicautom reliability code (0–5).

Monolingual Spanish Terminology

Development of Monolingual Terminological Entries (Spanish)

The main fields for monolingual Spanish terminological entries and methods for obtaining their values are discussed below.

In the context of a monolingual Spanish term translated from English,

Language is 'Spanish'; and

Subject Area is copied from the English term of which it is a translation.

The **Term** (e.g. 'polietileno' — *polyethylene*) appears in the terminological resource.

Source (e.g. 'Eurodicautom') identifies the bilingual terminological resource which contains the Spanish translation.

Dialect (empty, 'Castilian' or 'Latin American') is entered, where available, from the terminological resource.

Type of Entry is available from the terminological resource (in the case of abbreviations, e.g. *CC* for *corriente continua — direct current*) or otherwise is updated programmatically as either a single word or a multiword term.

Head Word has a value only for entries with Type of Entry 'multiword term'. It is defaulted programmatically to the first word (for Spanish), for example in ***par*** *coaxial — coaxial pair*. This is almost always correct, but may be reviewed manually. It will be incorrect in the occasional occurrences in Spanish of premodification of a noun, for example in *mal* ***contacto*** *— poor contact*.

Spanish term inflection is more complicated than in English. The head noun is pluralised, and so is any postmodifying adjective (e.g. *conmutador**es** controlado**s** por silicio — silicon controlled switches*).

Part of Speech, for the great majority of terms, has the value 'noun'. The value may on the whole be copied from the English term of which this is a translation. The Spanish inflectional endings also allow tagging of the part of speech, by using morphological analysis.

Natural Gender is not needed for Spanish. The usage of pronouns is determined by the grammatical gender of the noun (see next paragraph).

Grammatical Gender has a value only for entries with Part of Speech 'noun'. It is 'masculine' or 'feminine', depending on the gender of the head noun. Spanish gender normally matches the sex of an animate entity, but otherwise is purely grammatical (Green 1990). For example *aluminio — aluminium*, is masculine, and *aleación — alloy* is feminine. The gender allows an MT system to produce correct agreement (between the noun and its modifiers), which is very important in Spanish. It further allows an MT system to select (or resolve) Spanish pronouns (e.g. masculine *él — he* or *it*, or feminine *ella —* s*he* or *it*).

This field may be updated programmatically to a great degree of accuracy, but each group of entries with the same head word should be reviewed manually. The gender is often indicated by inspection of the suffix, with masculine nouns usually ending in *-o*, *-r*, *-l*, *-y*, *-ma* or *-ión* (but not *-ción* or *-sión*, which are usually feminine), and around ten typical feminine endings, including *-a* (HarperCollins 1997). The suffixes *-o* and *-a*, in particular, are frequent and reliable markers (Green 1990).

The value of **Number** may sometimes be copied from the English term of which this is a translation.

The value of **Transitivity** may usually be copied from the English term of which this is a translation.

The **Definition** and **Comment on Definition** fields have not been used for Spanish terms. If there is a definition for the corresponding English term, it may be referred to, as the two terms are intended to denote the same *concept*.

Example/Context has not been used for Spanish terms.

Lexical Entry Manipulation Programs

The terminological entries have been developed using text files, with one entry to a line. Each entry consists of a sequence of pairs of a field and its value (e.g. `Source=`

`Eurodicautom;`). Each pair shows the field (e.g. Source) and its value (e.g. 'Eurodicautom').

Programs have been written to manipulate these lexical entries. The programs are written in Prolog (for example, Clocksin and Mellish 1994).

The main purpose of this manipulation is to facilitate batch updating, combining, and checking of these files, to supplement the manual processes. In particular, the values of fields can be automatically updated to a specified value (e.g. 'cables' for the field Subject Area) or a calculated value (e.g. 'single word' or 'multiword' for the field Type of Entry). The sections on development of terminological entries detailed the programmatic updating performed for individual fields.

The programs have the following basic functionality:

Read in all the entries from a specified file.

Perform specified manipulation(s) on each lexical entry in turn, using procedures to retrieve, modify, add or delete fields and their values.

Store the modified lexical entries from a particular file, indexing them on the value of a specified key field (e.g. Term).

Retrieve all the entries read from a particular input file, or individual lexical entries with a specified value for the index field (e.g. the entry for 'cable').

After modification, output to a file all or selected entries from a particular input file, in the original format, specifying a list of required fields.

These programs are to be extended to produce marked-up lexical entries in the format required for importation into the OTELO terminological database (Thurmair 1997).

Evaluation

An informal evaluation of the likely usefulness of the work was performed. Three articles from the parallel English and Spanish newsletters (section 0) were selected, on the basis of containing the highest numbers of technical terms. The three articles were together judged to contain 19 distinct cables terms.

8 of the terms were covered in the terminological development work, and Spanish translations had been found for 7. Of these, 5 exactly matched the human translation, but all 7 were judged probably correct. This is a success rate of approximately 35%.

The articles were translated from English to Spanish, using a trial version of the Transcend MT system (Transparent Language Inc. 1997). 7 of the terms were translated, with 3 exact matches, and 4 were judged probably correct (approximately 20%).

If the terminological entries were added to the Transcend user dictionary, which is given priority, the result would be 9 probably correct translations (approximately 50%).

The conclusions drawn from this informal evaluation were as follows:

Only a small proportion of the cables terms were correctly translated by Transcend, and this would be expected to hold true for other MT systems.

The use of the MT system along with the terminological entries that have been developed would yield a much better, and useful, success rate (in this case, approximately 50%).

The number of general and specialised cables terms is probably several thousand.

Conclusions

Developing terminological entries for a particular domain is an important method of improving machine translation of technical documents. This article described the terminology investigation and development work done as part of the OTELO project.

Terminological entries for cables terms were developed, yielding about 1000 translations for about 800 distinct English terms. These entries provide a core of terms for use in machine translation of documents in the cables domain, while the number of general and specialised cables terms is probably several thousand.

Acknowledgements

I thank Jason Duckers, of BICC, and acknowledge his contribution to this work, particularly on the terminology lookup and the coding of the lexical entry manipulation programs. I thank Grace McCarthy, of BICC, for supporting this work, and acknowledge her contribution to setting its direction. I acknowledge the contributions of the OTELO project members in specifying the OTELO common lexical resource format, which provides the structure at which this terminological development work is targetted. I acknowledge the support for the OTELO project (LE-2703) of the European Commission and the BICC Group. I thank Adam Kilgarriff, of ITRI, University of Brighton, for producing the frequency lists of words in the British National Corpus, and making them available for use. I further acknowledge the support for the CISAU project (IED4/1/5818) of the DTI, EPSRC, and the BICC Group.

References

Blackburn, R. (1965) *Glossary of Words and Expressions Associated with Electric Cables and Cablemaking*, BICC Cables

British National Corpus (1997) British National Corpus Home Page, URL: http://info.ox.ac.uk/bnc/

British Standards Institution (1994) BS 4727: Part 2: Group 08: 1994 (Part 2. Terms particular to power engineering. Group 08. Electric cables), British Standards Institution

Church, K. and Hanks, P. (1990) 'Word Association Norms, Mutual Information, and Lexicography', in *Computational Linguistics*, Vol. 16, No 1: 22–29

Clocksin, W. and Mellish, C. (1994) *Programming in Prolog*, 4th edition, Berlin: Springer-Verlag

Dagan, I. and Church, K. (1994) 'Termight: Identifying and Translating Technical Terminology', in *Proceedings of the Fourth Conference on Applied Natural Language Processing*, Stuttgart, Germany: 34–40

Daille, B. (1995) *Combined Approach for Terminology Extraction: Lexical Statistics and Linguistic Filtering*, Technical Report, Lancaster University, UCREL Technical Papers

Damerau, F. (1993) 'Generating and Evaluating Domain-Oriented Multi-Word Terms from Text', in *Information Processing & Management*, Vol 29, No 4: 433–47

Douglas, S., Hurst, M. and Imlah, G. (1995) 'Construction Industry Specification Analysis and Understanding', in *Proceedings of Language Engineering '95*, Montpellier, France: 291–300

Dunning, T. (1993) 'Accurate Methods for the Statistics of Surprise and Coincidence' in *Computational Linguistics*, Vol. 19, No 1: 61–74

Ediciones VERBA (1997) *VERBA Polytechnic and Plurilingual Terminological Database: Cables and Conductors Terms*, Madrid: Ediciones VERBA

EEC (1988) *European Materials Databanks Demonstrator Programme: Common Reference Vocabulary*, Luxembourg: EEC

Eurodicautom (1997) Eurodicautom – Home Page, URL: http://www2.echo.lu/echo/databases/eurodicautom/en/eu92home.html

European Language Resources Association (1997) The ELRA Home Page, URL: http://www.icp.grenet.fr/ELRA/home.html

Garside, R. (1995) 'Matching Words in a Bilingual Corpus', in *Machine Translation Review*, No 2, October 1995: 7–18

Grant & Cutler (1997) Foreign *Language DICTIONARIES Specialist and General*, London: Grant & Cutler Ltd, Language Booksellers

Green, J. (1990) 'Spanish', in *The World's Major Languages*, Comrie, B. (ed), Oxford: Oxford University Press: 236–59

HarperCollins (1997) *Collins Spanish Dictionary Plus Grammar*, HarperCollins Publishers

Justeson, J. and Katz, S. (1995) 'Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text', in *Natural Language Engineering*, Vol. 1, No 1: 9–27

Kilgarriff, A. (1996) British National Corpus Frequency Lists, URL: ftp://ftp.itri.bton.ac.uk/pub/bnc/

McCormick, S. (1997) Integration Requirements: Common Lexical Resource Format, OTELO Project Document, Walldorf, Germany: SAP

Quinn, D. (1996) 'Using Icon for Text Processing', in *Machine Translation Review*, No 3, April 1996: 21–31

Ritzke, J. (1997) Common Lexical Resource Format / CDB specification — Format Specifications: CDB Features and Values, OTELO Project Document, Munich: GMS

Routledge (1997) *Spanish Technical Dictionary*, Volume 2: English – Spanish, London: Routledge

Sager, J. (1990) A Practical Course in Terminology Processing, Amsterdam: John Benjamins

Thurmair, G. (1997) Common Lexical Resource Format — Format specifications: The Open Lexicon Interchange Format, OTELO Project Document, Munich: GMS

Transparent Language Inc. (1997) Transcend Home Page, URL: http://www.translc.com/

Fifty Years of the Computer and Translation

by

John Hutchins

In March this year we mark the fiftieth anniversary of the 'birth' of machine translation, or — more precisely, the beginnings of discussion about how electronic computers might be applied to the task of translating natural languages.

There had, of course, been earlier proposals for the mechanization of some aspect of the translation process, primarily devices for a mechanical dictionary which would substitute words of one language by words of another. Patents for such devices were issued in 1933 in France to Georges Artsrouni and in Russia to Petr Troyanskii (Hutchins 1993). In conception, it is possible to trace back the idea of mechanical dictionaries to the seventeenth century, but the mechanisms for implementation did not become reality until the twentieth.

However, these earlier suggestions were unknown to Weaver, and indeed to many other MT pioneers until the late 1950s. In any case, Weaver's inspiration was the potential of the newly developed US computers, whose awesome power gave them the popular name at the time of 'electronic brains'.

It was on 4 March 1947 that Warren Weaver, Director of the Natural Sciences Division of the Rockefeller Foundation, wrote to his mathematician friend Norbert Wiener, shortly to become famous for his writings on cybernetics. Weaver had met Wiener during the War when both were involved in military research — Weaver on ballistics, Wiener on radar and prediction theory. Weaver (1947) wrote as follows:

'One thing I wanted to ask you about is this. A most serious problem, for UNESCO and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples. Huxley has recently told me that they are appalled by the magnitude and the importance of the translation job.

'Recognising fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while.

'Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography – methods which I believe succeed even when one does not know what language has been coded – one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.

'Have you ever thought about this? As a linguist and expert on computers, do you think it is worth thinking about?'

The letter was reproduced by Weaver in his memorandum of July 1949, which effectively launched MT research in the United States — and the memorandum itself was included in the Locke and Booth collection (Weaver 1955).

Weaver's reference to the possible analogy of cryptography was fully explicable at the time. He had himself heard of an impressive feat of decipherment from Professor Prager at Brown University. It involved a message encoded in Turkish, which was taken to a mathematician ignorant of the language of the original. The text was decoded based on letter frequencies in English, which fortunately matched those of Turkish (after simplification of letters with diacritic marks not found in English). To Weaver, the feat seemed to demonstrate that there were statistical uniformities in all languages that could be used in some way in translation. Weaver was himself a prominent mathematician and was familiar with the work of Claude Shannon on cryptography; later, Weaver collaborated with Shannon on the statistical theory of communication (Shannon and Weaver 1949).

The idea of using computers for translation had evidently occurred to Weaver some time before writing to Wiener. In his autobiography (Weaver 1972) he states: 'Early in 1947, having pondered the matter for nearly two years, I started to formulate some ideas about using computers to translate ...' Quite possibly he spoke about it to others before writing to Wiener. Bar-Hillel (1952) says that 'as early as 1945 ... Dr Warren Weaver ... started thinking and talking about the possibility ...' One of those he may have mentioned it to was Desmond Bernal of Birkbeck College (University of London), a physicist much interested in numerical computation, who arranged for Andrew D. Booth to visit the United States to investigate developments in computing.

Booth's first visit was in June 1946. He met Weaver at the Rockefeller Foundation to discuss the possibility of the Foundation's funding a computer at London University. he then went to the Moore School in Pennsylvania and Princeton University to see and learn about progress on the ENIAC and EDVAC computers. In the following year, he returned for a three-month study period at Princeton, funded by a grant from the Rockefeller Foundation. He met Weaver again on 6 March 1947, just two days after Weaver's letter to Wiener. It was on this occasion that Weaver mentioned to Booth the possibility of using the London computer for 'non-numerical' applications — which were more likely to attract American funding support — and suggested in particular mechanical translation. Booth claimed later in his well-known 'historical introduction' to the collection he edited with William Locke (Booth and Locke 1955) that MT had been discussed by Weaver and himself during the 1946 meeting. However, there is no documentary evidence to support the assertion. For example, in the report of his 1946 visit which Booth submitted to the Rockefeller Foundation, he makes no reference to such an application (Booth 1946), whereas, by contrast, after his 1947 visit, his report to the Foundation in February 1948 includes the following passage:

'A concluding example, of possible application of the electronic computer, is that of translating from one language into another. We have considered this problem in some detail and it transpires that a machine of the type envisaged could perform this function without any modification in its design.' (Booth 1948)

In fact, in a number of articles on MT, Booth states explicitly that the discussion with Weaver took place in March 1947. It is unfortunate that Booth's historical account in the Locke/Booth collection of early MT articles has led so many later writers into error on this point.

There is, in short, no doubt. Although Weaver may have spoken to Bernal and others earlier about the possibility of using computers to translate, the first corroborated and definite mention was made by Warren Weaver in his letter to Norbert Wiener and in his discussion with Andrew Booth, in the early days of March 1947.

[This article first appeared in MT News International 14 and is reproduced here by kind permission of the Editor.]

References

Bar-Hillel, Y. (1952) 'Mechanical translation: needs and possibilities', paper presented at MIT conference, June 1952

Booth, A.D. (1946) Report on visit to American laboratories, 3 July 1946 (Rockefeller Foundation Archives)

Booth, A.D. (1948) Proposals for the establishment of a centre for the construction and use of an electronic computer, February 12, 1948. (Rockefeller Foundation Archives)

Booth, A.D. and Locke, W.N. (1955) 'Historical introduction', in Locke and Booth: 1–14

Hutchins, J. (1993) 'The first MT patents', *MT News International* 5, 14–16

Locke, W.N. and Booth, A.D. (eds) (1955) *Machine Translation of Languages: Fourteen Essays*, Cambridge, Mass.: MIT Press

Shannon, C. and Weaver, W. (1949) *The Mathematical Theory of Communication*, Urbana: University of Illinois Press

Weaver, W. (1947) Letter to Norbert Wiener, 4 March 1947

Weaver, W. (1955) 'Translation', in Locke and Booth: 15–23

Weaver, W. (1972) *Scene of Change: A Lifetime in American Science*, New York: Scribner

MT Evaluation: Science or Art?

by

Derek Lewis
School of Modern Languages, University of Exeter, Devon, UK

The Current State of MT Evaluation

The famous 1966 ALPAC Report, for which evaluators rated output MT on scales on scales of speed, cost and quality, was one of the first major comparative evaluations of MT systems. John Lehrberger's guidelines for evaluation, compiled in 1981, took account of output quality, cost, time, and system improvability (Lehrberger and Bourbeau 1987). In the early 1990s the Essex University MT evaluation group discussed various methodologies, tools and approaches, which appeared in a number of reports. An indication of the stage evaluation has now reached is the 1994 draft report by the EAGLES Subgroup on the evaluation of natural language processing systems. The report does not focus on MT, which is considered as a translation aid alongside multilingual dictionaries and thesauri, terminology management systems and translation memories. Indeed the only specific MT product mentioned is ALPnet, chosen because it was at the time (March 1994) the only package that was both available commercially and familiar to the report's compilers.

The report does not aim to supply a league table detailing the advantages and disadvantages of specific products. Instead it lists criteria for evaluating certain types of translation aids. It defines what is involved in the evaluation process and sets up feature checklists of translators' tools according to what it calls the 'consumer report paradigm', a set of specific questions asked of the software which may be answered by yes/no or by values on quantifiable scales. The result is a set of worked out checklists which corresponds to an approach suggested by Steven Krauwer (1993). Although the 1994 report devotes little attention to the specifics of MT evaluation (which is the object of attention of a different subgroup), it indicates the position of MT within an increased range of electronic tools for translations and shows how evaluation has become increasingly orientated to benchmark measures of users' requirements.

The report pays particular attention to user profiles as a factor in evaluation. These include the quantity and quality of translation work required, the nature of texts to be translated, and the characteristics of the organisation in which translation is carried out (e.g. its policy on language, or the number of languages it uses in-house). Various developments in the translation industry are noted. Firstly, the number of languages in which translations are required is increasing; at the same time certain languages are emerging as universal focal languages (e.g. a text is produced directly in English, and then translated into, say, Finnish and Japanese). Secondly, source texts are becoming more repetitive (mass produced manuals are a typical example); at the same time translation involves more revision, updating, and layouting than before and it is proving difficult to keep up with burgeoning terminology. Thirdly, organisations are tending either to outsource translation or to turn their translation departments into independent business units. The conclusion is that there is indeed no such thing as a typical user profile; the main determinants are an organisation's resources and its policy towards translation and languages.

What effect do these factors have on MT evaluation? Since MT is more often found in large, well resourced organisations, smaller companies and freelance translators are still less likely to use MT or to have the resources to evaluate, integrate, and customise MT to their needs. A sample of users and organisations undertaken by the EAGLES subgroup indicated that usage of high-end electronic tools such as memories, archives and dictionaries is currently low, although increasing. MT is seen to be at the very top of the scale of technical sophistication and is either used or being considered by large translation companies or concerns with internal translation units. The most likely environment for any translation help tool is an integrated document production and translation facility, a computer network, a high level of computer literacy, and a positive framework for exploring new systems; MT remains a sophisticated translation aid demanding significant resources for development and fine-tuning. The report still sees MT as the province of large scale users with extensive IT and human resources. At the same time it recognises the emergence of a market for small users as systems become more compact and affordable. While larger companies may increasingly outsource translations to freelance translators, these are less likely to have technical aids such as MT; very occasionally companies will make resources (such as standardised term banks and in-house dictionaries) available to the contracted translator in order to ensure consistency in the product.

Overall the EAGLES subgroup regards MT as a 'promising aid' with 'significant restrictions'. Since the systems are costly, they are viable only for text in a particular field involving thousands of pages of text every year. The style and vocabulary must not vary much, and the best results are achieved with restricted language. Significantly, MT is not intelligent: a system does not remember past mistakes but repeats them each time it runs; it is therefore deemed unsuitable for repetitive texts and updating. It seems that MT is very much a first stage, one-off step in the production of a translated documented. These general conclusions confirm what has been known about MT for some time.

If anything the EAGLES report underestimates the recent emergence of relatively low-cost MT packages and their use in translating electronic mail and texts on the internet. MT is now being interfaced with e-mail packages, web browsers and also voice recognition systems. These applications are for very basic communication and information extraction. Their products are ephemeral and are not designed for integration into an organisation's document production process. This will undoubtedly affect the evaluative criteria applied by users. For example, the nature of the interaction between the communication partners using e-mail may mean that sender and receiver will re-send or re-edit the same messages in order to clarify information as it is transmitted. Alternatively, MT may be increasingly used as a basic first-stage filter to eliminate junk mail on the internet. In both cases, key criteria for evaluation are likely to be operational speed and basic intelligibility rather than linguistic sophistication; such systems will be marketed more on the basis of their seamless interfaces and user-friendly windows than on the quality of language output. It is even more likely that the user will be, not a translator, but any (possibly monolingual) employee; in effect, the human translator is by-passed and the user is the target language reader. It is even possible that the low cost of systems may render large-scale evaluation unnecessary: the investment would not justify it.

There is a significant body of theoretical literature on MT evaluation and on the factors involved (see, for example, Hutchins and Somers 1992, Arnold et al 1994, Sparck Jones and Gailliers 1996). Typically, three types of evaluation are distinguished: *operational evaluation* assesses the economic benefits of using a particular system as part of a process in an organisation; *declarative evaluation* (the 'standard' method) looks at how systems perform according to criteria such as accuracy and intelligibility; *typological evaluation* provides information for developers on what linguistic constructions a system can handle and typically

employs test suites. Distinctions have also been drawn between black box and glass box evaluation, depending on whether or not the evaluator has access to operational modules of the system (Balkan et al 1994). The type of evaluation will often determine the methods employed and the information evaluated (e.g. whether translation quality, speed or cost is being assessed).

To assess the linguistic performance of MT systems (as opposed to external factors such as speed of translation, vendor support, extendibility and portability), a number of methods have been proposed. One is to count up the errors in a set of output sentences or a text, weighting them according to a scale of seriousness; the disadvantage with this is that what constitutes an error may depend on subjective judgment or the purpose of the translation. Another method is to rate output according to measures or scales of intelligibility and accuracy (or fidelity); ratings are usually based on the judgments of evaluators reading the output. This approach was used in the ALPAC evaluation, and underlies Makoto Nagao's 7-point scale of fidelity and van Slype's measures of information transfer (see Jordan et al 1993).

The problem with intelligibility or comprehensibility is not just the reliance on subjective judgment. Assigning a sentence a value on a scale of intelligibility tells us nothing about what intelligibility really is: to measure something is not necessarily to be clear about *what* we are measuring. This is true whether we measure intelligibility through Flesch scales or Cloze techniques, or whether we try to quantify intelligibility in terms of the physical time taken to post-edit a text in order to make it intelligible. Developers of the MELTRAN Japanese-English system proposed a framework which compared post-editing times with evaluators' classifications and weightings of linguistic errors in MT output (a difficult task) to see whether these times could be taken as a benchmark measure of fidelity (Minnes 1993). Often criteria of assessment are combined: thus the collection and statistical evaluation of post-editing times, linguistic errors and intelligibility ratings formed part of a scheme for evaluating the French-English CAT/ARIANE (Roudaud at al 1993). A different approach was adopted for an evaluation of SYSTRAN, performed for the US Air Force in 1979–80 (Wilks 1992). Applying stringent control conditions and statistical techniques this relied entirely on monolingual and bilingual evaluators' assessment of accuracy, comprehensibility, and naturalness (for monolingual speakers, coherence was taken to correlate strongly with correctness). But the evaluators were not asked to assign scores on an arbitrary scale (which would implicitly compare MT output with human translation): they simply decided whether, of two MT output sentences, they preferred one or the other (both sentences were raw SYSTRAN output, but one had been produced by the system after its dictionaries had been updated to handle a new text domain). What the designers were measuring was only whether SYSTRAN could be improved or not. Clearly, complex and time-consuming schemes for evaluation are more readily undertaken by large organisations, agencies, or in-house developers than by end-users or customers.

Recent attention has focused on tools for MT evaluation, especially test suites and corpora. In particular, it is felt, suites containing specially constructed sentences and structures in combination may allow developers to assess in a controlled way how systems behave and how they can be improved; running corpora through MT, on the other hand, may produce large amounts of data but cannot ensure that a particular linguistic phenomenon is tested. The TSNLP (Test Suites for Natural Language Processing) project reported in 1994 that it was in the process of developing methods of automatically constructing test suites for NLP applications such as MT (Balkan 1994, Balkan et al 1994). There are a number of issues in constructing and using test suites for evaluating MT.

Firstly, test suites are generally regarded as diagnostic tools for system developers but they can also be used to assess adequacy of output, which is what users are interested in. To do this,

however, test suites need to take into account the frequency of occurrence of a particular phenomenon and should include a variety of lexical, rather than syntactic, information. The data in a test suite may need to be tailored to an application, such as MT, or even to a domain or text type within MT itself. Tools may be developed to generate suites that test particular phrase types, combinations of structures, and even domain-specific vocabulary; the data is held in a database and the suite generated according to criteria specified by the user (see Nerbonne et al 1993). It has also been suggested that structures in test suites should be matched with structures found in corpora so that the contents of the suite reflect real-life texts. Finally, it is clear that general-purpose monolingual test suites are only of limited use in evaluating a particular MT system in a particular operational environment: they need to be adapted for user's needs and, more generally, for language pairs. As a very simple example, a test suite for English-to-German MT would have to test for nouns whose gender is known to vary in German; the gender of a noun will also determine pronominal constructions elsewhere in the sentence, including, for instance, relative pronouns introducing sub-clauses, as in the following:

English:	the table *which* is in the corner ...
	the lamp *which* is in the corner ...
	the television *which* is in the corner
German:	der Tisch, *der* in der Ecke steht ...
	die Lampe, *die* in der Ecke steht ...
	das Fernsehen, *das* in der Ecke steht ...

Constructing test suites for MT users is an expensive and complex task. Evaluating the results of test suites is likewise beyond the resources of most users (for references to developed test suites see Nerbonne et al 1993 and Way 1991).

Using MT in a University Curriculum: General Issues

This section addresses the question of how MT can be introduced in a university teaching curriculum. It is based on experiences of teaching students in a natural language processing applications module for undergraduate students of foreign languages at the University of Exeter. Although the need for language undergraduates to be familiar with the technological tools of translation has been highlighted (Clark 1994), a recent survey on tools and techniques for MT conducted by the University of Essex concluded that the time was not right to pursue any coordinated initiative to promote or develop MT teaching; the reasons included lack of interest on the part of suppliers, high cost, and the large amount of work involved. Despite such obstacles the integration of MT remains a worthwhile goal for a university modern languages programme. For a university student the central task should be to demonstrate that he understands the parameters of MT evaluation and can apply these in practice. For a future employer the student must show that he can apply those parameters to MT system in general, not just the package used on the course. On the other hand a university student is not in the position of being able to evaluate MT in a genuinely professional or practical context, i.e. for a company or organisation. He is not handling translation as part of a document production or handling process and does not have access to, say, large volumes of technical texts. Neither is he a professional translator with experience of translating domain-specific texts into his mother tongue. Most UK undergraduates translate into the target language as part of a language learning activity, with all the differences which this approach entails. A further constraint is time: a student has to complete a module in a limited period and alongside demands made by other areas of the course; in a culture-based combined honours modern languages curriculum

MT is unlikely to enjoy a high priority. It must also be noted that MT is new to students: unlike other software tools, they have almost certainly never encountered MT before. As a result they are likely to approach it with great curiosity but also with the expectations inherent in the title: a computer program which does the kind of translation that a human being performs, only fully automatically.

This may sound like a long list of reasons for not expecting students to evaluate MT. On the other hand students are not unlike many potential users of MT (who, as noted above, are unlikely to exhibit a standard profile). As a future employee with foreign language expertise in a company (whether or not a full-time translator), a student may well be expected to 'look at' or 'advise on' the wisdom of investing in MT and to produce a recommendation on a fairly informal basis; this, of course, may change when tools for professional evaluation of MT eventually become available. At the moment, however, the question is what are the most appropriate methods for student-based MT evaluation within the constraints outlined?

The first method is for students to evaluate a particular MT package using samples of different text types (e.g. journalistic, literary, technical). The evaluation can be informal (i.e. based on general impressions on the quality of output), or it can be based on surveys of fellow students' assessments, using scaled measures of intelligibility, accuracy, etc. The samples are usually small and manageable; most students in fact rely on their personal, i.e. intuitive, assessment of output. Another approach is for students to evaluate a system using part of a test suite of input sentences, selected according to various criteria (e.g. broad coverage of a range of basic structures, or detailed coverage of specific phenomena). The test suite used at Exeter is the Hewlett Packard suite for English (Flickinger 1987), a general-purpose NLP test suite with minimal annotation. An alternative suite has been produced by the Essex group as part of the EUROTRA project (Way 1991). Students may also construct their own test suite fragments; this is obligatory if they are working from German into English. Some of the results of this work are given in more detail below.

For the moment, however, it is useful to note what aspects of MT students/users themselves rate as important. For instance, documentation (or on-line help facilities) can strongly influence how a system is perceived. The user manual is consulted in cases of difficulty; topics must therefore be accessible, informative and helpful when consulted. Although this is clearly nothing to do with the performance of the MT engine as such, like any software tool, the manual is an essential part of the system in operation. The manual can be rated in terms of whether it is up-to-date with the software it is supposed to describe (curiously, not always the case); whether topics can be found easily or are buried in longer accounts; whether explanations are clear, concise and informative. Obviously these factors are not MT-specific. Although they are not intrinsic to the linguistic quality of MT output, they can determine the perception of a system's user-friendliness. Apart from basic documentation, users often comment on the quality of the user interface. To do justice to this topic, users should ideally receive guidelines on how to assess software ergonomy. This factor is important in view of the fact that many PC-based MT packages are being promoted as much on the basis of their user-friendly interfaces as on the quality of their output.

Many users have problems updating dictionaries (where this facility is available). The process is often seen as difficult and not always worth the effort when measured against improvement in the output. For example, systems typically distinguish single term entries from multiple word entries (called 'idioms' or 'semantic units'). The procedures for adding these items are often seen as opaque. This is despite the fact that most commercially available MT systems claim to offer user-friendly and context-free menu-driven interfaces which require only minimal syntactic and semantic information. However, before reaching the stage of exploring a

system's dictionaries, most users rely heavily on the quality of raw output to assess a system; a relatively small number of users will take the trouble to explore fully the facilities of customising a system through dictionary update. In other words: first impressions count a lot in evaluating MT. This may be countered by careful, tutor-directed introduction to the use and extension of dictionaries, but it raises the question of how far commercial, small-scale users are prepared to customise systems out of their own resources. There are also technical problems in providing a networked MT system in which individual students working on public machines update and maintain their own dictionaries (as they must do for assessment and evaluation purposes): most MT systems are not designed to be installed and used in this way.

Finally, one can raise the issue of how much linguistic knowledge is required by users in order properly to evaluate errors in output. Leaving aside matters of intelligibility and accuracy, it is clear that a clear knowledge of syntactic categories and phrase structures is a prerequisite; it is also important to be able to bring this knowledge to bear on MT output. As an example, consider the following example, taken from an article on fashion:

SL English phrase: ... time to hold off with the damsons and deep chocolates of last season.

Raw MT German translation: ... Zeit, um mit dem damsons und tiefen Schokoladen letzter Jahreszeit.

User's comment: 'This text has been translated fairly well. Even the genitive has been correctly employed. This is due to the shortness of the sentence. An adjective has been translated as a noun: chocolates. However, this is also unclear in English. It would be necessary to define it with the addition of a noun.'

What the user/student really should be saying is something like: '(a) the prepositional phrase (..) attached to the noun phrase (...) has been correctly translated, with the English preposition rendered by a German genitive; (b) the English noun 'chocolates' is, unusually, used as a noun referring to colour (here: shades of chocolate); this is because the text domain is fashion. The question is how the dictionary could be modified to handle this feature (which also occurs in the use of 'damsons').'

Assessing Text Suite Output

The following section contains samples of raw MT output of an extract from the HP test suite of (minimally) annotated sentences. The output of three systems (referred to S1, S2 and S3) is shown. The systems are relatively low-cost, commercially available packages for PC: S1 is the EASY TRANSLATOR system for translating on-line the contents of Web pages or the contents of a Windows clip-board; S2 is Langenscheidt's T1 system (version 3.0), and S3 is Globalink's POWER TRANSLATOR (version 2.0). The samples illustrate some of the problems of evaluating MT output, especially the relative merits of different systems, using a general purpose test suite. The language direction is English to German. In conclusion, the appropriateness of using text suite output in teaching MT is briefly discussed.

Firstly, consider the translation of examples of English restrictive relative clauses (wh-type clauses):

ENGLISH: (Annotation: Restrictive relative clauses (wh- type))

1. Abrams hired a woman who was competent.
2. Abrams hired women who were competent.
3. Abrams hired women whose manager was competent.
4. Abrams hired a woman who Browne interviewed.
5. Abrams hired a woman who Browne approved of.
6. Abrams hired a woman who Browne knew Chiang interviewed.
7. Abrams has a bookcase which is heavy.
8. Abrams has a bookcase that is heavy.
9. Abrams has an office that Browne showed Chiang.
10. Abrams has an office which Browne showed Chiang.
11. Abrams has an office Browne showed Chiang.
12. Abrams has an office that Browne showed to Chiang.
13. Abrams has an office which Browne showed to Chiang.

S1

1. Abrams hat eine Frau angestellt, die fähig war.
2. Abrams hat Frauen angestellt, die fähig waren.
3. Abrams hat Frauen angestellt, deren Verwalter fähig war.
4. Abrams hat eine Frau angestellt, die Browne interviewt hat.
5. Abrams hat eine Frau angestellt, die Browne gebilligt hat.
6. Abrams hat eine Frau angestellt, die Browne gewußt hat, daß Chiang interviewt hat.
7. Abrams hat ein Bücherregal, das schwer ist.
8. Abrams hat ein Bücherregal, das schwer ist.
9. Abrams hat ein Büro, das Browne Chiang gezeigt hat.
10. Abrams hat ein Büro, das Browne Chiang gezeigt hat.
11. Abrams hat ein BüroBrowne Chiang hat gezeigt.
12. Abrams hat ein Büro, das Browne zu Chiang gezeigt hat.
13. Abrams hat ein Büro, das Browne zu Chiang gezeigt hat.

S2

1. Abrams stellte eine Frau ein, die fähig war.
2. Abrams stellte Frauen ein, die fähig waren.
3. Abrams stellte Frauen ein, deren Manager fähig waren.
4. Abrams stellte eine Frau ein, die Browne interviewte.
5. Abrams stellte eine Frau ein, die Browne genehmigte.
6. Abrams mieten eine Frau die Browne kannte Chiang geinterviewte.
7. Abrams hat ein Bücherregal, das schwer ist.
8. Abrams hat ein Bücherregal, das schwer ist.
9. Abrams hat ein Büro, das Browne Chiang zeigte.
10. Abrams hat ein Büro, das Browne Chiang zeigte.
11. Abrams hat ein Büro Browne zeigte Chiang.
12. Abrams hat ein Büro, das Browne an Chiang zeigte.
13. Abrams hat ein Büro, das Browne an Chiang zeigte.

S3
1. Abrams einstellte eine Frau, die qualifiziert war.
2. Abrams einstellte Frauen, die qualifiziert waren.
3. Abrams einstellte Frauen, deren Leiter qualifiziert war.
4. Abrams einstellte eine Frau, die Browne interviewte.
5. Abrams einstellte eine Frau, von der Browne zustimmte.
6. Abrams einstellte eine Frau, die Browne wusste, dass Chiang interviewte.
7. Abrams hat ein Büchergestell, das schwer ist.
8. Abrams hat ein Büchergestell, das schwer ist.
9. Abrams hat ein Büro, das Browne Chiang zeigte.
10. Abrams hat ein Büro, das Browne Chiang zeigte.
11. Abrams hat ein Büro, das Browne Chiang zeigte.
12. Abrams hat ein Büro, das Browne zu Chiang zeigte.
13. Abrams hat ein Büro, das Browne zu Chiang zeigte.

The distinction between restrictive and non-restrictive relative clauses' is sometimes reflected in the use of 'which' or 'that'. As shown in sentences 7 and 8, this is 'flattened' in German. For another target language, however, it may be necessary to render the distinction overtly. The sample also shows variation across systems in the translation of the English simple past tense: either the German imperfect or the present perfect is used. While not necessarily significant, as has pointed out (Arnold et al 1994), the consistent use of the present perfect in German requires us also to test for the correct rendering of different forms of the auxiliary ('haben' or 'sein'); this suggests that test suites need to be customised to reveal information about specific language pairs. To illustrate this we see that S3 consistently ignores the rule for separable verb prefixes when translating into German: the prefix is never detached from the verb (it produces 'einstellte' for 'stellte .. ein'). Obviously a test suite should test for features of German verb behaviour so that the evaluator is able to assess its impact on translation quality. But our sample shows that it is the lexical choice of an English verb that triggers a particular feature; clearly our input suite must include verbs which will produce German equivalents exhibiting separable prefixes. Another language pair-specific phenomenon for which we may wish specifically to test the output of different systems is the translation of so-called active and static passives in German. Consider, for example, the following:

ENGLISH:	Two competent programmers were hired.
S1:	Zwei fähige Programmierer wurden angestellt.
S2:	Zwei fähige Programmierer wurden eingestellt.
S3:	Zwei qualifizierte Programmierer waren eingestellt.

S3 fails to translate the active passive. This might lead us to check more thoroughly the system's handling of passive structures. S2's translation of sentence 3 illustrates the difference between intelligibility and accuracy: 'deren Manager fähig waren' ('whose managers were') is syntactically impeccable but a mistranslation (the plural possibly conditioned by 'deren'). Even this limited sample shows how different systems handle identical lexical items. Consider the translations of English 'manager' and the verb 'approve of':

Manager = Verwalter/Manager/Leiter
approve of = billigen/genehmigen/zustimmen

Given the general absence of neat one-to-one lexical correspondences between languages, there is bound to be a high degree of lexical variation in output from different MT systems. In some

cases the variation will be relatively trivial (as here in the German words for 'manager'). In others it is more important, as in the German versions for 'approve of', where the choice of 'zustimmen' (among other things) makes the sense harder to understand. But it would be wrong, at least for language that is not domain-specific, to draw from isolated examples conclusions about the adequacy of the lexical choices made by a system as a whole. Such conclusions could only be drawn from test suites devised for particular domains of vocabulary. At the same time, syntax and vocabulary cannot be artificially separated. As seen in German separable verb prefixes, the lexical choice of a verb in the source language may trigger a particular syntactic feature in the target language; a single general purpose test suite may not be all that useful for evaluating MT output in different languages.

It is tempting but unwise to draw broad inferences about intelligibility from test suite output. Consider, for instance, sentence 11 above ('Abrams has an office (that) Browne showed Chiang', where the subject relative pronoun 'that' is deleted). Only the S3 reintroduces the relative in German, where is it obligatory: both S2 and S1 omit it, resulting in a significant loss of intelligibility. We could infer from this that S3's output for this feature will be consistently more intelligible. We would, however, have to be certain that the feature was translated similarly over similar constructions. Intelligibility is better assessed as a feature of texts rather than individual sentences.

In the following example, S1 is possibly less intelligible, but only because it has departed from the SL word order (often required when translating from English into German). In other structures, the tendency to stick to SL word order might hinder intelligibility. It is difficult to tell.

ENGLISH:	Abrams has an office Browne showed Chiang.
S1:	Abrams hat ein Büro Browne Chiang hat gezeigt.
S2:	Abrams hat ein Büro Browne zeigte Chiang.

Test suites are especially good at revealing precise information about structures in syntactic combination. A simple example is the use of 'mass/mass-creating partitives' in subject NPs:

ENGLISH:	Most of the staff is competent.
	Most of the program works.
	Almost all of the program works.
S1:	Am meisten ist vom Personal fähig.
	Am meisten von den Programmarbeiten.
	Fast alle Programmarbeiten.
S2:	Meiste vom Personal ist fähig.
	Meiste vom Programm funktioniert.
	Fast alle das Programm funktioniert.
S3:	Der meiste des Stabes ist qualifiziert.
	Die meisten der Programm-Werke.
	Fast all die Programm-Werke.

The partitives are translated tolerably well. But the outputs suggest that S1 and S3 are most likely to have problems disambiguating the English plural noun and 3rd person singular

present tense verb forms (here: 'works') when the partitive occurs in a subject NP. Since the suite is annotated, we know precisely what linguistic structure is being input and tested. As an MT user or developer, however, we might also like to know what output structure the MT system thinks it has produced. But only the developer might be able to tell the MT system to annotate the output; the user typically has only the output sentences to go on, from which it is often difficult to judge what the MT system thinks it has produced. Operational systems, moreover, do not necessarily operate with clear linguistic models through to the final output generation stage; there may be no abstract structure or 'linguistic annotation' at all that we can attach to the output.

Consider the following output from subject NPs containing the apostrophe marker for possessives.

ENGLISH:	The project's engineers work for Abrams.
	Abrams's engineers were interviewed by Browne.
	Abrams' engineers were interviewed by Browne.
S1:	Die Ingenieure des Projekts arbeiten für Abrams.
	Ingenieure Abrams wurden von Browne interviewt.
	Abrams wurden' Ingenieure von Browne interviewt.
S2:	Die Ingenieure des Projekts arbeiten bei Abrams.
	Daß Ingenieure von Browne geinterviewt wurden, Abrams ist.
	Die Ingenieure von Abrams wurden von Browne geinterviewt.
S3:	Des Projektes Ingenieure arbeiten für Abrams.
	Abrams's Ingenieure wurden von Browne interviewt.
	Abrams' Ingenieure wurden von Browne interviewt.

S2 has succeeded in transposing the Noun1's + Noun2 construction into Noun2 + von + Noun1, which indicates a superior transfer capability. At the same time, where Noun1 ends in 's' (as in 'Abrams's), S2's output is catastrophically garbled. This 3-sentence sample exemplifies an important dilemma in MT evaluation: how do we rate overall different systems in which 2/3 of the output of one system is very good and the remaining 1/3 very poor, while 3/3 of another system are neither very good nor very poor?

Finally, even relatively short test sentences can be syntactically complex and produce compounded errors which are difficult to measure except in general terms of intelligibility and comprehensibility. In the following sample it is hard to explain why S2 has gone wrong, but first impressions about the sentence are clear and immediate. S1 has added a relative 'dass' which aids clarity; S2 misparses the subject - verb construction in the main clause and misconstructs a past tense form (so does S1 incidentally, but not so badly); S2 compounds the syntactic failure with poor lexical choice (mieten).

ENGLISH: Abrams hired a woman who Browne knew Chiang interviewed.
S1: Abrams hat eine Frau angestellt, die Browne gewußt hat, daß Chiang interviewt hat.
S2: Abrams mieten eine Frau die Browne kannte Chiang geinterviewte.

In conclusion it would appear that MT evaluation has made serious efforts to become a science, with different modes of evaluation for different purposes and the prospect of clearer benchmark criteria for users. On the other hand, the limited resources available in higher education for teaching MT suggest that, while evaluation should be an essential component of MT-based courses, students are more likely to benefit from a discovery-based approach using less rigorous techniques based on relatively small volumes of text and sections of text suites, as illustrated above. The approach may be well correspond to how low cost MT systems are likely to be evaluated by users in the market.

References

Arnold, D. et al (1994) *Machine Translation. An Introductory Guide*, Oxford: NCC Blackwell

Balkan, L. (1994) 'Test suites: Some issues in their use and design', *Machine Translation Ten Years On, Conference at the University of Cranfield*, 12-14 November 1994, 26-1

Balkan, L. et al (1994) 'TSNLP: Test Suites for Natural Language Processing' in *Proceedings of Language Engineering Convention, CNIT,* La Défense, Paris, France: 17-22

Clark, R. (1994) 'Computer-assisted translation: the state of the art' in Dollerup: 301-308

Dollerup, C. and Lindegaard, A. (1994) *Teaching Translation and Interpreting 2*, *Papers from the Second Language International Conference*, Elsinore, 4-6 June 1993, Amsterdam: John Benjamins

EAGLES (Expert Advisory Group on Language Engineering Standards) Evaluation of Natural Language Processing Systems. Draft - Work in Progress (1994). EAGLES Document EAG-EWG-IR.2, Consiglio Nationale delle Richerche, Instituo di Linguistica Computazionale, Pisa

Flickinger D. et al (1987) *Toward Evaluation of NLP Systems, Hewlett Packard Laboratories,* Palo Alto

Hutchins, W. J. and Somers, H. L. (1992) *An Introduction to Machine Translation*, London: Academic Press

Jordan, P.A.W. et al (1993) 'A First-Pass Method for Evaluating Machine Translation Systems' in *Machine Translation 8*, Nos 1-2, Special Issue on Evaluation of MT Systems, 49-58

Krauwer, S. (1993) 'Evaluation of MT Systems: A Programmatic View' in *Machine Translation 8*, Nos 1-2, Special Issue on Evaluation of MT Systems, 59-66

Lehrberger J. and Bourbeau, L. (1987) *Machine Translation. Linguistic Characteristics of MT Systems and General Methodology of Evaluation*, Amsterdam: John Benjamins

Minnes, S. (1993) 'Constructive Machine Translation Evaluation' in *Machine Translation 8*, Nos 1-2, Special Issue on Evaluation of MT Systems, 67-75

Nerbonne, J. et al (1993) 'A Diagnostic Tool for German Syntax' in *Machine Translation 8*, Nos 1-2, Special Issue on Evaluation of MT Systems, 85-107

Newton, J. (1992) *Computers in Translation. A Practical Appraisal*, Routledge, London

Roudaud, B. et al (1993) 'A Procedure for the Evaluation and Improvement of an MT System by the End-User', in *Machine Translation 8*, Nos 1-2, Special Issue on Evaluation of MT Systems, 109-116

Sparck Jones, K. and Galliers, J. (1996) *Evaluating Natural Language Processing Systems. An Analysis and Review,* Heidelberg and Berlin: *S*pringer Verlag

Way, A. (1991) 'A Practical Developer-Oriented Evaluation of Two MT Systems', Working Papers in Language Processing 26.

Wilks, Y. (1992) 'SYSTRAN: it obviously works but how much can it be improved?' in Newton: 166-88

The above paper was presented at the ASLIB Conference *Translating and the Computer 19* on 13 November 1997 and is reproduced by kind permission of the organisers. Derek Lewis may be contacted at the School of Modern Lanuages, University of Exeter, Exeter, EX4 4QH; e-mail: D.R.Lewis@exeter.ac.uk

Book Reviews

Thierry Dutoit (1997) *An Introduction to Text-to-Speech Synthesis*, Dordrecht: Kluwer. Hardback £63. 312 pages. ISBN 0-7923-4498-7.

The term 'an introduction ..' in the title of this most excellent book is too modest. Rather, it is a comprehensive review of the state-of-the- art of text-to-speech (TTS) processing linked to an equally comprehensive survey of the foundations of the subject. It is claimed to be the first book to treat speech synthesis from the perspective of two different approaches. Another remarkable feature is that it is based on an internet project, the MBROLA Project, which won the IT European Award in 1996. See the *Appendix* (below) for more details on MBROLA.

There are many delightful things about this book. Firstly, it is thoroughly European and shows how well the concept of Europe can work. The author is Belgian; his English is a pleasure to read; French, German, English and Dutch language examples are used regularly; the publishers are Dutch; it is liberally sprinkled with rather apt quotations from Lewis Carroll's *Alice in Wonderland*, not inappropriate to computer-based TTS; expert systems and neural networks have a place; digital signal processing is also prominent which is not surprising as the author is an electrical engineer.

Much use is made of well-presented state-transition diagrams — which points up the author's engineering background again. There are a number of excellent block-diagrams which serve to illustrate the systems being discussed. The List of Figures takes up five pages which gives some idea of the wealth of illustrations. A particularly useful feature is some nineteen pages of up-to-date references. These follow chapters so that a form of classification applies. Humour is evident as in the example (on page 58) where Antoine Rivarol is quoted as saying 'ce qui n'est pas clair n'est pas francais', which hardly needs to be translated as ''what is not clear, is not French!' — to which the author adds 'if only it were true'.

The author suggests that ' .. very few people associate a good knowledge of signal processing with a comprehensive insight into language processing, synthesis mostly remains unclear'. He also says that the book is not self-contained in these respects, and that 'some understanding of language processing and, more importantly, of digital signal processing will help'. Although the reviewer has both of these talents, he believes that a reader without them will still find the book extremely interesting and a useful reference and text book.

Chapter One also serves as an Introduction, and provides a very clear and easily read outline of the acoustic, phonetic, phonological, syntactic, semantic, and pragmatic background to speech processing. One might say that the physiological and systems engineering aspects are also included. Then follow two parts with five chapters each; entitled 'From Text to its Narrow Phonetic Transcription', Part One deals with natural language processing and its inherent problems in just 170 pages.

Part Two takes a digital signal processing approach, with an emphasis on concatenation, and is entitled 'From Narrow Phonetic Transcription to Speech' (100 pages). It would be tedious to list the contents of all eleven chapters as these take up five Contents pages, so here is a brief outline based on chapter headings.

Part One: From Text to its Narrow Phonetic Transcription

2: Grammars, Inference, Parsing, and Transduction.
There are some brief examples of Prolog programs and philosophy.
3: NLP Architectures for TTS Synthesis.
This covers various formalisms, data structures, grammars, and numerous other structures.
4: Morpho-Syntactic Analysis.
Here is where preprocessing and neural networks are mentioned.
5: Automatic Phonetization.
This is where text and phonemes come together, expert Systems, neural nets again, and phonetic postprocessing appear.
6: Automatic Prosody Generation.
This section is one of the most detailed.

Part Two: From Narrow Phonetic Transcription to Speech

7: Synthesis Strategies.
Another core chapter where rule-based and concatenation-based synthesizers are covered, also speech synthesis and quality issues.
8: Linear Prediction Synthesis.
Packed with good things such as the autoregressive model, autocorrelation, how that compares with covariance, algorithms such as Grahm-Schmidt, Schur, Split, and Levinson; quality, filtering, compression, and prosody, too.
9: Hybrid Harmonic/Stochastic Synthesis.
Models and spectral analysis.
10: Time-Domain Algorithms.
Here the TD-PSOLA system appears
11: Conclusions and Perspectives. Natural language and TTS, also digital signal processing and TTS are discussed here.

All chapters are equally good but Chapter 11, perhaps, with its tables showing applications and comparisons of various NLP techniques in TTS are especially noteworthy.

I have two reservations about the book, however. Firstly, there is an abundance of acronyms with which a reader can loose contact. It is a salutary experience to be bouncing along on a good read and to unexpectedly crash into a forgotten one; but these can usually be resolved by reference to the Index.

The other reservation is more serious. There is no mention throughout the book of the world-class work on TTS which has come out of BT Martlesham Heath. A thorough search of the references failed to show anything by Andrew Breen or Julian Page, who are both prominent in TTS. On NLP, Keith Preston and Sandra Williams might also have been cited. BTs involvement in TTS is not new. My first contact with that research station (when I was at National Engineering Laboratory) was nearly twenty years ago. At that time, BT had an excellent system based on the FAD (Fetch and Detect) digital signal processing chip, which, even then, had only a slight buzz. Since then, the Laureate system, considered to be quite revolutionary, has been developed and generates the highest quality of speech. A reader may

confirm this for himself by calling up the following URL:

http://www.labs.bt.com/innovate/speech/laurate/index.htm

Click on ‘pregenerated examples’ and choose the example ‘My dog has no nose’. The reviewer regards the female voice as the best quality of synthetic speech he has yet heard.

There is no doubt that the comprehensive nature of this book makes it the best in its field. and its very high quality, ensures that it is particularly relevant to anyone in the broad field of speech processing and phonetics. To management, researchers, stduents in both academia and industry, or anyone who wants to know what this rapidly emerging technology is all about, it can be thoroughly recommended.

Appendix

The EMBROLA (pronounced ‘em-brola’ (think of umbrella, suggests the author) Project is Internet-based. It is a multinational venture pioneered by the Faculté Polytechnique de Mons in Belgium. The aim, largely now achieved, is to obtain a set of speech synthesizers for as many languages as possible, and providing them free for non-commercial and non-military applications. Already available are French, German, Dutch, Spanish, Romanian, British English!, and Brazilian Portuguese.

MBROLA, currently on v2.05, is a speech synthesiser based on the concatenation of diphones. It is pointed out that MBROLA does not accept plain text as input, but a diphone database tailored to the EMBROLA format. All that, and more, can be found on the Internet at the URL:

http://tcts.fpms.ac.be/synthesis/mbrola.html

Various synthesis demonstrations samples are available here in WAV format for direct re-production with an appropriate sound card.

John W. Bruce

Conferences and Workshops

The following is a list of recent (i.e. since the last edition of the MTR) and forthcoming conferences and workshops. Telephone numbers and e-mail addresses are given where known (please check area telephone codes).

11–13 September 1997
Architectures and Mechanisms for Language Processing
Apex Hotel, Edinburgh, Scotland
Fax: +44 131 6504587, e-mail: amlap@cogsci.ed.ac.uk
http://www.cogsci.ed.ac.uk/~amlap/

22–24 September 1997
LACL97: Logical Aspects of Computational Linguistics
Nancy, France
Tel: +33 4 76 82 78 52, fax: +33 4 76 82 56 65, e-mail: lecomte@shm.upmf-grenoble.fr
http://www.loria.fr/~bechet/LACL.html

29 October–1 November 1997
MT Summit VI: Machine Translation: Past, Present, Future
Catamaran Resort Hotel, San Diego. USA
Tel/fax: +1 703/716 912; e-mail: AMTA@clark.net
http://www.isi.edu/natural-language/mtsummit.html

2–4 December 1997
NLPRS97: Natural Language Processing Pacific Rim Symposium 1997
Cape Panwa Hotel, Phuket, Thailand
E-mail: links@nwg.nectec.or.th
http://www.links.nectec.or.th/NLPRS/nlprs.html

11–17 January 1998
NeMLaP: New Methods in Natural Language Processing
Sydney, Australia
E-mail: powers@uia.ua.ac.be

18–20 February 1998
PACLIC12: 12th Pacific Asia Conference on Language, Information and Computation
http://www.iscs.nus.sg/~colips/paclic92.html

28–30 May 1998
LREC: Ist International Conference on Language Resources and Evaluation
Granada, Spain
E-mail: lrec@ilc.pi.cnr.it

13–24 July 1998
ELSNET's 6th European Summer School on Language and Speech Communication.
Robustness: Real Life Applications in Language and Speech
Barcelona, Spain
Fax: +34 3401 6447, e-mail: summer98@gps.tsc.upc.es
http://gps-tsc.upc.es/veu/ess98/

10–14 August 1998
COLING-ACL98: 17th International Conference on Computational Linguistics
36th Annual Meeting of the Association for Computational Linguistics
University of Montreal, Canada
E-mail: coling-acl98@iro.umontreal.ca
http://coling-acl98.iro.umontreal.ca

23–28 August 1998
ECAI98: 13th Biennial European Conference on Artificial Intelligence
Brighton, BN1 9QH, UK
Tel: +44 1273 678448, fax: +44 1273 671320, e-mail: ecai98@cogs.susx.ac.uk
http://www.cogs.susx.ac.uk/ecai98

30 November–4 December 1998
ICSLP98: 5th International Conference on Spoken Language Processsing
Sydney Convention Centre, Sydney Australia
E-mail: icslp98@tourhosts.com.au
http://cslab.anu.edu.au/icslp98

MEMBERSHIP: CHANGE OF ADDRESS

If you change your address, please advise us on this form, or a copy, and send it to the following (this form can also be used to join the Group):

Mr. J.D.Wigg
BCS-NLTSG
72 Brattle Wood
Sevenoaks, Kent TN13 1QU
U.K.

Date: ...../...../.....

Name: .....
Address: .....
.....
Postal Code: ..... Country: .....
E-mail: ..... Tel.No: .....
Fax.No: .....

Note for non-members of the BCS: your name and address will be recorded on the central computer records of the British Computer Society.

Questionnaire

We would like to know more about you and your interests and would be pleased if you would complete as much of the following questionnaire as you wish (please delete any unwanted words).

1. a. I am mainly interested in the computing/linguistic/user/all aspects of MT.
   b. What is/was your professional subject? .....
   c. What is your native language? .....
   d. What other languages are you interested in? .....
   e. Which computer languages (if any) have you used? .....

2. What information in this Review (No.6, October '97) or any previous Review, have you found:
   a. interesting? Date .....
      .....
      .....
   b. useful (i.e. some action was taken on it)? Date .....
      .....
      .....

3. Is there anything else you would like to hear about or think we should publish in the *MT Review*?
   .....
   .....
   .....
   .....

4. Would you be interested in contributing to the Group by,

   a. Reviewing MT books and/or MT/multilingual software
   b. Researching/listing/reviewing public domain MT and MNLP software .....
   c. Designing/writing/reviewing MT/MNLP application software .....
   d. Designing/writing/reviewing general purpose (non-application specific) MNLP .....
      procedures/functions for use in MT and MNLP programming .....
   e. Any other suggestions? .....
      .....
      .....
      .....

Thank you for your time and assistance.