Machine Translation **14:** 113–157, 1999.

Review Article: Example-based Machine Translation

HAROLD SOMERS
Centre for Computational Linguistics, UMIST, PO Box 88, Manchester M60 1QD, England
(E-mail: harold@fs1.ccl.umist.ac.uk)

Abstract. In the last ten years there has been a significant amount of research in Machine Translation within a "new" paradigm of empirical approaches, often labelled collectively as "Example-based" approaches. The first manifestation of this approach caused some surprise and hostility among observers more used to different ways of working, but the techniques were quickly adopted and adapted by many researchers, often creating hybrid systems. This paper reviews the various research efforts within this paradigm reported to date, and attempts a categorisation of different manifestations of the general approach.

Key words: example-based MT, hybrid methods, corpora, translation memory

1. Background

In 1988, at the Second TMI conference at Carnegie Mellon University, IBM's Peter Brown shocked the audience by presenting an approach to Machine Translation (MT) which was quite unlike anything that most of the audience had ever seen or even dreamed of before (Brown et al. 1988). IBM's "purely statistical" approach, inspired by successes in speech processing, and characterized by the infamous statement "Every time I fire a linguist, my system's performance improves" flew in the face of all the received wisdom about how to do MT at that time, eschewing the rationalist linguistic approach in favour of an empirical corpus-based one.

There followed something of a flood of "new" approaches to MT, few as overtly statistical as the IBM approach, but all having in common the use of a corpus of translation examples rather than linguistic rules as a significant component. This apparent difference was often seen as a confrontation, especially for example at the 1992 TMI conference in Montreal, which had the explicit theme "Empiricist vs. Rationalist Methods in MT" (TMI 1992), though already by that date most researchers were developing hybrid solutions using both corpus-based and theory-based techniques.

The heat has largely evaporated from the debate, so that now the "new" approaches are considered mainstream, in contrast though not in conflict with the older rule-based approaches.

In this paper, we will review the achievements of a range of approaches to corpus-based MT which we will consider variants of “example-based MT” (EBMT), although individual authors have used alternative names, perhaps wanting to bring out some key difference that distinguishes their own approach: “analogy-based”, “memory-based”, “case-based” and “experience-guided” are all terms that have been used. These approaches all have in common the use of a corpus or database of already translated examples, and involve a process of matching a new input against this database to extract suitable examples which are then recombined in an analogical manner to determine the correct translation.

There is an obvious affinity between EBMT and Machine Learning techniques such as Exemplar-Based Learning (Medin & Schaffer 1978), Memory-Based Reasoning (Stanfill & Waltz 1986), Derivational Analogy (Carbonell 1986), Case-Based Reasoning (Riesbeck & Schank 1989), Analogical Modelling (Skousen 1989), and so on, though interestingly this connection is only rarely made in EBMT articles, and there has been no explicit attempt to relate the extensive literature on this approach to Machine Learning to the specific task of translation, a notable exception being Collins’ (1998) PhD thesis.

Two variants of the corpus-based approach stand somewhat apart from the scenario suggested here. One, which we will not discuss at all in this paper, is the Connectionist or Neural Network approach. So far, only a little work with not very promising results has been done in this area (see Waibel et al. 1991; McLean 1992; Wang & Waibel 1995; Castaño et al. 1997; Koncar & Guthrie 1997).

The other major “new paradigm” is the purely statistical approach already mentioned, and usually identified with the IBM group’s Candide system (Brown et al. 1990, 1993), though the approach has also been taken up by a number of other researchers (e.g. Vogel et al. 1986; Chen & Chen 1995; Wang & Waibel 1997; etc.). The statistical approach is clearly example-based in that it depends on a bilingual corpus, but the matching and recombination stages that characterise EBMT are implemented in quite a different way in these approaches; more significant is that the important issues for the statistical approach are somewhat different, focusing, as one might expect, on the mathematical aspects of estimation of statistical parameters for the language models. Nevertheless, we will try to include these approaches in our overview.

2. EBMT and Translation Memory

EBMT is often linked with the related technique of “Translation Memory” (TM). This link is strengthened by the fact that the two gained wide publicity at roughly the same time, and also by the (thankfully short-lived) use of the term “memory-based translation” as a synonym for EBMT. Some commentators regard EBMT and TM as basically the same thing, while others – the present author included – believe there is an essential difference between the two, rather like the difference between computer-aided (human) translation and MT proper. Although they have in common the idea of reuse of examples of already existing translations, they differ in that TM is an interactive tool for the *human* translator, while EBMT is an essentially *automatic* translation technique or methodology. They share the common problems of storing and accessing a large corpus of examples, and of matching an input phrase or sentence against this corpus; but having located a (set of) relevant example(s), the TM leaves it to the human to decide what, if anything, to do next, whereas this is only the start of the process for EBMT.

2.1. HISTORY OF TM

One other thing that EBMT and TM have in common is the long period of time which elapsed between the first mention of the underlying idea and the development of systems exploiting the ideas. It is interesting, briefly, to consider this historical perspective. The original idea for TM is usually attributed to Martin Kay's well-known "Proper Place" paper (1980), although the details are only hinted at obliquely:

> ... the translator might start by issuing a command causing the system to display anything in the store that might be relevant to [the text to be translated] .... Before going on, he can examine past and future fragments of text that contain similar material. (Kay 1980: 19)

Interestingly, Kay was pessimistic about any of his ideas for what he called a "Translator's Amanuensis" ever actually being implemented. But Kay's observations are predated by the suggestion by Peter Arthern (1978)[1] that translators can benefit from on-line access to similar, already translated documents, and in a follow-up article, Arthern's proposals quite clearly describe what we now call TMs:

> It must in fact be possible to produce a programme [sic] which would enable the word processor to 'remember' whether any part of a new text typed into it had already been translated, and to fetch this part, together with the translation which had already been translated, ... Any new text would be typed into a word processing station, and as it was being typed, the system would check this text against the earlier texts stored in its memory, together with its translation into all the other official languages [of the European Community]. ... One advantage over machine translation proper would be that all the passages so retrieved would be grammatically correct. In effect, we should be operating an electronic 'cut and stick' process which would, according to my calculations, save at least 15 per cent of the time which translators now employ in effectively producing translations. (Arthern 1981: 318).

Alan Melby (1995: 225f) suggests that the idea might have originated with his group at Brigham Young University (BYU) in the 1970s. What is certain is that the idea was incorporated, in a very limited way, from about 1981 in ALPS, one of the first commercially available MT systems, developed by personnel from

BYU. This tool was called “Repetitions Processing”, and was limited to finding exact matches *modulo* alphanumeric strings. The much more inventive name of “translation memory” does not seem to have come into use until much later.

The first TMs that were actually implemented, apart from the largely inflexible ALPS tool, appear to have been Sumita & Tsutsumi’s (1988) ETOC (“Easy TO Consult”), and Sadler & Vendelman’s (1990) Bilingual Knowledge Bank, predating work on corpus alignment which, according to Hutchins (1998) was the prerequisite for effective implementations of the TM idea.

2.2. HISTORY OF EBMT

The idea for EBMT dates from about the same time, though the paper presented by Makoto Nagao at a 1981 conference was not published until three years later (Nagao 1984). The essence of EBMT, called “machine translation by example-guided inference, or machine translation by the analogy principle” by Nagao, is succinctly captured by his much quoted statement:

> Man does not translate a simple sentence by doing deep linguistic analysis, rather, Man does translation, first, by properly decomposing an input sentence into certain fragmental phrases ..., then by translating these phrases into other language phrases, and finally by properly composing these fragmental translations into one long sentence. The translation of each fragmental phrase will be done by the analogy translation principle with proper examples as its reference. (Nagao 1984: 178f)

Nagao correctly identified the three main components of EBMT: matching fragments against a database of real examples, identifying the corresponding translation fragments, and then recombining these to give the target text. Clearly EBMT involves two important and difficult steps beyond the matching task which it shares with TM.

To illustrate, we can take Sato & Nagao’s (1990) example in which the translation of (1) can be arrived at by taking the appropriate fragments – underlined – from (2a, b) to give us (3).[2] How these fragments are identified as being the appropriate ones and how they are reassembled varies widely in the different approaches that we discuss below.

(1) He buys a book on international politics.

(2) a. <u>He buys</u> a notebook.
<u>Kare wa</u> nōto <u>o kau</u>.
HE topic NOTEBOOK obj BUY.

b. I read <u>a book on international politics</u>.
Watashi wa <u>kokusai seiji nitsuite kakareta hon</u> o yomu.
I topic INTERNATIONAL POLITICS ABOUT CONCERNED BOOK obj READ.

(3) *Kare wa kokusai seiji nitsuite kakareta hon o kau.*

It is perhaps instructive to take the familiar pyramid diagram, probably first used by Vauquois (1968), and superimpose the tasks of EBMT (Figure 1). The source-text analysis in conventional MT is replaced by the matching of the input against the example set (see Section 3.6). Once the relevant example or examples have been selected, the corresponding fragments in the target text must be selected. This has been termed "alignment" or "adaptation" and, like transfer in conventional MT, involves contrastive comparison of both languages (see Section 3.7). Once the appropriate fragments have been selected, they must be combined to form a legal target text, just as the generation stage of conventional MT puts the finishing touches to the output. The parallel with conventional MT is reinforced by the fact that both the matching and recombination stages can, in some implementations, use techniques very similar (or even identical in hybrid systems – see Section 4.4) to analysis and generation in conventional MT. One aspect in which the pyramid diagram does not really work for EBMT is in relating "direct translation" to "exact match". In one sense, the two are alike in that they entail the least analysis; but in another sense, since the exact match represents a perfect representation, requiring no adaptation at all, one could locate it at the top of the pyramid instead.

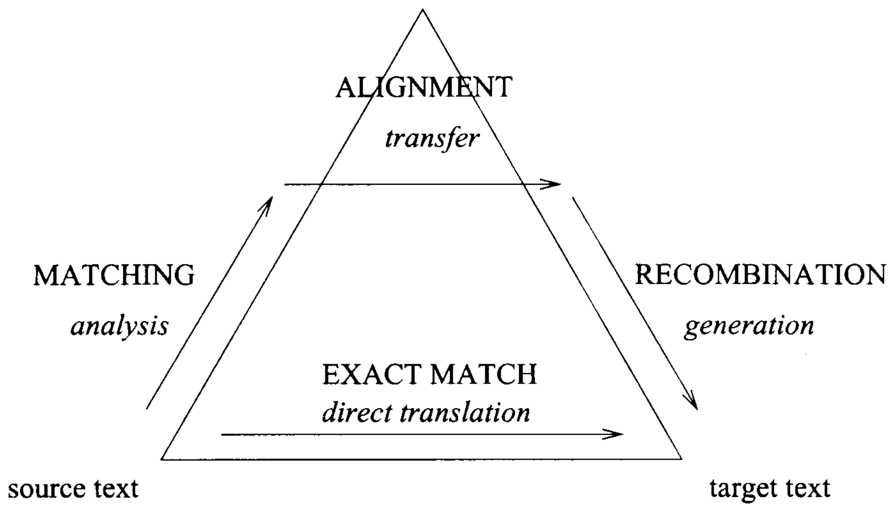

Figure 1. The "Vauquois pyramid" adapted for EBMT. The traditional labels are shown in *italics*; those for EBMT are in CAPITALS.

To complete our history of EBMT, mention should also be made of the work of the DLT group in Utrecht, often ignored in discussions of EBMT, but dating from about the same time as (and probably without knowledge of) Nagao's work. The matching technique suggested by Nagao involves measuring the semantic proximity of the words, using a thesaurus. A similar idea is found in DLT's "Linguistic Knowledge Bank" of example phrases described in Pappegaaij et al. (1986a, b) and

Schubert (1986: 137f) – see also Hutchins & Somers (1992: 305ff). Sadler's (1991) "Bilingual Knowledge Bank" clearly lies within the EBMT paradigm.

3. Underlying problems

In this section we will review some of the general problems underlying example-based approaches to MT. Starting with the need for a database of examples, i.e. parallel corpora, we then discuss how to choose appropriate examples for the database, how they should be stored, various methods for matching new inputs against this database, what to do with the examples once they have been selected, and finally, some general computational problems regarding speed and efficiency.

3.1. PARALLEL CORPORA

Since EBMT is corpus-based MT, the first thing that is needed is a parallel aligned corpus.[3] Machine-readable parallel corpora in this sense are quite easy to come by: EBMT systems are often felt to be best suited to a sublanguage approach, and an existing corpus of translations can often serve to define implicitly the sublanguage which the system can handle. Researchers may build up their own parallel corpus or may locate such corpora in the public domain. The Canadian and Hong Kong parliaments both provide huge bilingual corpora in the form of their parliamentary proceedings, the European Union is a good source of multilingual documents, while of course many World Wide Web pages are available in two or more languages (cf. Resnik 1998). Not all these resources necessarily meet the sublanguage criterion, of course.

Once a suitable corpus has been located, there remains the problem of aligning it, i.e. identifying at a finer granularity which segments (typically sentences) correspond to each other. There is a rapidly growing literature on this problem (Fung & McKeown 1997, includes a reasonable overview and bibliography; see also Somers 1998) which can range from relatively straightforward for "well behaved" parallel corpora, to quite difficult, especially for typologically different languages and/or those which do not share the same writing system.

The alignment problem can of course be circumvented by building the example database manually, as is sometimes done for TMs, when sentences and their translations are added to the memory as they are typed in by the translator.

3.2. GRANULARITY OF EXAMPLES

As Nirenburg et al. (1993) point out, the task of locating appropriate matches as the first step in EBMT involves a trade-off between length and similarity. As they put it:

> The longer the matched passages, the lower the probability of a complete match (...). The shorter the passages, the greater the probability of ambiguity (one

> and the same S' can correspond to more than one passage T') and the greater the danger that the resulting translation will be of low quality, due to passage boundary friction and incorrect chunking. (Nirenburg et al. 1993: 48)

The obvious and intuitive “grain-size” for examples, at least to judge from most implementations, seems to be the sentence, though evidence from translation studies suggests that human translators work with smaller units (Gerloff 1987). Furthermore, although the sentence as a unit appears to offer some obvious practical advantages – sentence boundaries are for the most part easy to determine, and in experimental systems and in certain domains, sentences are simple, often monoclausal – in the real world, the sentence provides a grain-size which is too big for practical purposes, and the matching and recombination process needs to be able to extract smaller “chunks” from the examples and yet work with them in an appropriate manner. We will return to this question in Section 3.7.

Cranias et al. (1994: 100) make the same point: “the potential of EBMT lies [i]n the exploitation of fragments of text smaller than sentences” and suggest that what is needed is a “procedure for determining the best ‘cover’ of an input text ...” (1997: 256). This in turn suggests a need for parallel text alignment at a subsentence level, or that examples are represented in a structured fashion (see Section 3.5).

3.3. HOW MANY EXAMPLES

There is also the question of the *size* of the example database: how many examples are needed? Not all reports give any details of this important aspect. Table I shows the size of the database of those EBMT systems for which the information is available.

When considering the vast range of example database sizes in Table I, it should be remembered that some of the systems are more experimental than others. One should also bear in mind that the way the examples are stored and used may significantly effect the number needed. Some of the systems listed in the table are not MT systems as such, but may use examples as part of a translation process, e.g. to create transfer rules.

One experiment, reported by Mima et al. (1998) showed how the quality of translation improved as more examples were added to the database: testing cases of the Japanese adnominal particle construction (*A no B*), they loaded the database with 774 examples in increments of 100. Translation accuracy rose steadily from about 30% with 100 examples to about 65% with the full set. A similar, though less striking result was found with another construction, rising from about 75% with 100 examples to nearly 100% with all 689 examples. Although in both cases the improvement was more or less linear, it is assumed that there is some limit after which further examples do not improve the quality. Indeed, as we discuss in the next section, there may be cases where performance starts to decrease as examples are added.

Table I. Size of example database in EBMT systems

System	Reference(s)	Language pair	Size
PanLite	Frederking & Brown (1996)	Eng → Spa	726 406
PanEBMT	Brown (1997)	Spa → Eng	685 000
TDMT	Sumita et al. (1994)	Jap → Eng	100 000
CTM	Sato (1992)	Eng → Jap	67 619
Candide	Brown et al. (1990)	Eng → Fre	40 000
no name	Murata et al. (1999)	Jap → Eng	36 617
PanLite	Frederking & Brown (1996)	Eng → SCr	34 000
TDMT	Oi et al. (1994)	Jap → Eng	12 500
TDMT	Mima et al. (1998)	Jap → Eng	10 000
no name	Matsumoto & Kitamura (1997)	Jap → Eng	9 804
TDMT	Mima et al. (1998)	Eng → Jap	8 000
MBT3	Sato (1993)	Jap → Eng	7 057
no name	Brown (1999)	Spa → Eng	5 397
no name	Brown (1999)	Fre → Eng	4 188
no name	McTait & Trujillo (1999)	Eng → Spa	3 000
ATR	Sumita et al. (1990), Sumita & Iida (1991)	Jap → Eng	2 550
no name	Andriamanankasina et al. (1999)	Fre → Jap	2 500
Gaijin	Veale & Way (1997)	Eng → Ger	1 836
no name	Sumita et al. (1993)	Jap → Eng	1 000
TDMT	Sobashima et al. (1994), Sumita & Iida (1995)	Jap → Eng	825
TTL	Güvenir & Cicekli (1998)	Eng ↔ Tur	747
TSMT	Sobashima et al. (1994)	Eng → Jap	607
TDMT	Furuse & Iida (1992a, b, 1994)	Jap → Eng	500
TTL	Öz & Cicekli (1998)	Eng ↔ Tur	488
TDMT	Furuse & Iida (1994)	Eng → Jap	350
EDGAR	Carl & Hansen (1999)	Ger → Eng	303
ReVerb	Collins et al. (1996), Collins & Cunningham (1997), Collins (1998)	Eng → Ger	214
ReVerb	Collins (1998)	Irish → Eng	120
METLA-1	Juola (1994, 1997)	Eng → Fre	29
METLA-1	Juola (1994, 1997)	Eng → Urdu	7

Key to languages – Eng: English, Fre: French, Ger: German, Jap: Japanese, SCr: Serbo-Croatian, Spa: Spanish, Tur: Turkish

Considering the size of the example data base, it is worth mentioning here Grefenstette's (1999) experiment, in which the entire World Wide Web was used as a virtual corpus in order to select the best (i.e. most frequently occurring) translation of some ambiguous noun compounds in German–English and Spanish–English.

3.4. SUITABILITY OF EXAMPLES

The assumption that an aligned parallel corpus can serve as an example database is not universally made. Several EBMT systems work from a manually constructed database of examples, or from a carefully filtered set of "real" examples.

There are several reasons for this. A large corpus of naturally occurring text will contain overlapping examples of two sorts: some examples will mutually reinforce each other, either by being identical, or by exemplifying the same translation phenomenon. But other examples will be in conflict: the same or similar phrase in one language may have two different translations for no other reason than inconsistency (cf. Carl & Hansen 1999: 619).

Where the examples reinforce each other, this may or may not be useful. Some systems (e.g. Somers et al. 1994; Öz & Cicekli 1998; Murata et al. 1999) involve a similarity metric which is sensitive to frequency, so that a large number of similar examples will increase the score given to certain matches. But if no such weighting is used, then multiple similar or identical examples are just extra baggage, and in the worst case may present the system with a choice – a kind of "ambiguity" – which is simply not relevant: in such systems, the examples can be seen as surrogate "rules", so that, just as in a traditional rule-based MT system, having multiple examples (rules) covering the same phenomenon leads to over-generation.

Nomiyama (1992) introduces the notion of "exceptional examples", while Watanabe (1994) goes further in proposing an algorithm for identifying examples such as the sentences in (4) and (5a).[4]

(4) a. *Watashi wa kompyūtā o kyōyōsuru*.
I topic COMPUTER obj SHARE-USE.
'I share the use of a computer.'

b. *Watashi wa kuruma o tsukau.*
I topic CAR obj USE.
'I use a car.'

(5) *Watashi wa dentaku o shiyōsuru.*
I topic CALCULATOR obj USE.
a. 'I share the use of a calculator.'
b. 'I use a calculator.'

Given the input in (5), the system might incorrectly choose (5a) as the translation because of the closer similarity of *dentaku* 'calculator' to *kompyūtā* 'computer' than to *kuruma* ‘car’ (the three words for ‘use’ being considered synonyms; see Section 3.6.2), whereas (5b) is the correct translation. So (4a) is an exceptional example because it introduces the unrepresentative element of ‘share’. The situation can be rectified by removing example (4a) and/or by supplementing it with an unexceptional example.

Distinguishing exceptional and general examples is one of a number of means by which the example-based approach is made to behave more like the traditional rule-based approach. Although it means that “example interference” can be minimised, EBMT purists might object that this undermines the empirical nature of the example-based method.

3.5. HOW ARE EXAMPLES STORED?

EBMT systems differ quite widely in how the translation examples themselves are actually stored. Obviously, the storage issue is closely related to the problem of searching for matches, discussed in the next section.

In the simplest case, the examples may be stored as pairs of strings, with no additional information associated with them. Sometimes, indexing techniques borrowed from Information Retrieval (IR) can be used: this is often necessary where the example database is very large, but there is an added advantage that it may be possible to make use of a wider context in judging the suitability of an example. Imagine, for instance, an example-based dialogue translation system, wishing to translate the simple utterance *OK*. The Japanese translation for this might be *wakarimashita* ‘I understand’, *iidesu yo* ‘I agree’, or *ijō desu* ‘let’s change the subject’, depending on the context.[5] It may be necessary to consider the immediately preceding utterance both in the input and in the example database. So the system could broaden the context of its search until it found enough evidence to make the decision about the correct translation.

Of course if this kind of information was expected to be relevant on a regular basis, the examples might actually be stored with some kind of contextual marker already attached. This was the approach taken in the MEG system (Somers & Jones 1992).

3.5.1. *Annotated Tree Structures*

Early attempts at EBMT – where the technique was often integrated into a more conventional rule-based system – stored the examples as fully annotated tree structures with explicit links. Figure 2 (from Watanabe 1992) shows how the Japanese example in (6) and its English translation is represented. Similar ideas are found in Sato & Nagao (1990), Sadler (1991), Matsumoto et al. (1993), Sato (1995), Matsumoto & Kitamura (1997) and Meyers et al. (1998).

(6) *Kanojo wa kami ga nagai.*
SHE topic HAIR subj IS-LONG
'She has long hair.'

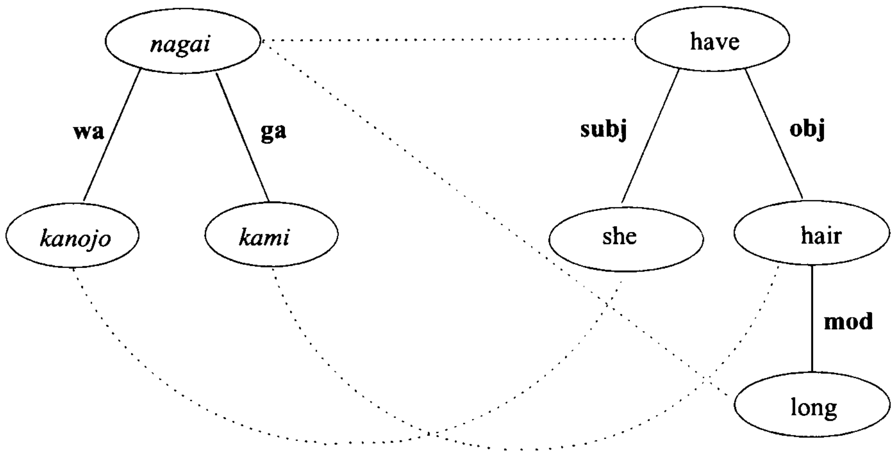

Figure 2. Representation scheme for (6). (Watanabe 1992: 771).

More recently a similar approach has been used by Poutsma (1998) and Way (1999): here, the source text is parsed using Bod's (1992) DOP (data-oriented parsing) technique, which is itself a kind of example-based approach, then matching subtrees are combined in a compositional manner.

In the system of Al-Adhaileh & Kong (1999), examples are represented as dependency structures with links at the structural and lexical level expressed by indexes. Figure 3 shows the representation for the English–Malay pair in (7).

(7) a. He picks the ball up.
b. *Dia kutip bola itu.*
HE PICK-UP BALL THE

The nodes in the trees are indexed to show the lexical head and the span of the tree of which that item is the head: so for example the node labelled "ball(1)[n](3-4/2-4)" indicates that the subtree headed by *ball*, which is the word spanning nodes 3 to 4 (i.e. the fourth word) is the head of the subtree spanning nodes 2 to 4, i.e. *the ball*. The box labelled "Translation units" gives the links between the two trees, divided into "Stree" links, identifying subtree correspondences (e.g. the English subtree 2-4 *the ball* corresponds to the Malay subtree 2-4 *bola itu*) and "Snode" links, identifying lexical correspondences (e.g. English word 3-4 *ball* corresponds to Malay word 2-3 *bola*).

Planas & Furuse (1999) represent examples as a multi-level lattice, combining typographic, orthographic, lexical, syntactic and other information. Although their proposal is aimed at TMs, the approach is also suitable for EBMT. Zhao & Tsujii

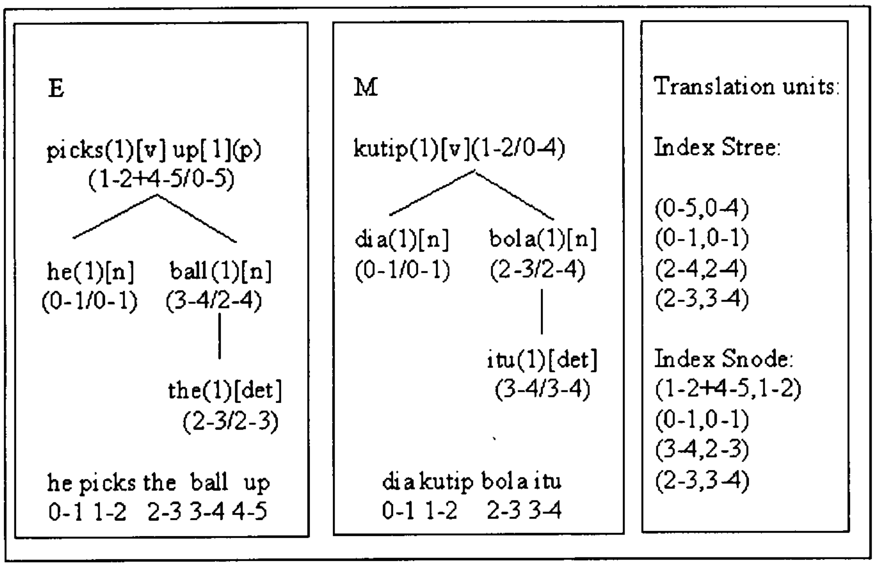

Figure 3. Representation scheme for (7). (Al-Adhaileh & Kong 1999: 247).

(1999) propose a multi-dimensional feature graph, with information about speech acts, semantic roles, syntactic categories and functions and so on.

Other systems annotate the examples more superficially. In Jones (1996) the examples are POS-tagged, carry a Functional Grammar predicate frame and an indication of the sample's rhetorical function. In the ReVerb system (Collins & Cunningham 1995; Collins 1998), the examples are tagged, carry information about syntactic function, and explicit links between "chunks" (see Figure 5 below). Andriamanankasina et al. (1999) have POS tags and explicit lexical links between the two languages. Kitano's (1993) "segment map" is a set of lexical links between the lemmatized words of the examples. In Somers et al. (1994) the words are POS-tagged but not explicitly linked.

3.5.2. *Generalized Examples*

In some systems, similar examples are combined and stored as a single "generalized" example. Brown (1999) for instance tokenizes the examples to show equivalence classes such as "person's name", "date", "city name", and also linguistic information such as gender and number. In this approach, phrases in the examples are *replaced* by these tokens, thereby making the examples more general. This idea is adopted in a number of other systems where general rules are derived from examples, as detailed in Section 4.4. Collins & Cunningham (1995: 97f) show how examples can be generalized for the purposes of retrieval, but with a corresponding precision–recall trade-off.

The idea is taken to its extreme in Furuse & Iida's (1992a, b) proposal, where examples are stored in one of three ways: (a) literal examples, (b) "pattern examples" with variables instead of words, and (c) "grammar examples" expressed as context-sensitive rewrite rules, using sets of words which are concrete instances of each category. Each type is exemplified in (8–10), respectively.

(8) *Sochira ni okeru* ⇒ We will send it to you.
Sochira wa jimukyoku desu ⇒ This is the office.

(9) *X o onegai shimasu* ⇒ may I speak to the X′
(X = *jimukyoku* 'office', . . .)
X o onegai shimasu ⇒ please give me the X′
(X = *bangō* 'number', . . .)

(10) *N1 N2 N3* ⇒ the N3′ of the N1′
(N1 = *kaigi* 'meeting', N2 = *kaisai* 'opening', N3 = *kikan* 'time')
N1 N2 N3 ⇒ N2′ N3′ for N1′
(N1 = *sanka* 'participation', N2 = *mōshikomi* 'application', N3 = *yōshi* 'form')

As in previous systems, the appropriate template is chosen on the basis of distance in a thesaurus, so the more appropriate translation is chosen as shown in (11).

(11) a. *jinjika o onegai shimasu* (*jinjika* = 'personnel section') ⇒ may I speak to the personnel section
b. *kenkyukai kaisai kikan* (*kenkyukai* = 'workshop') ⇒ the time of the workshop
c. *happyō mōshikomi yōshi* (*happyō* = 'presentation') ⇒ application form for presentation

What is clear is the hybrid nature of this approach, where the type (a) examples are pure strings, type (c) are effectively "transfer rules" of the traditional kind, with type (b) half-way between the two. A similar idea is found in Kitano & Higuchi (1991a, b), who distinguish "specific cases" and "generalized cases", with a "unification grammar" in place for anything not covered by these, though it should be added that their "memory-based" approach lacks many other features usually found in EBMT systems, such as similarity-based matching, adaptation, realignment and so on.

Several other approaches in which the examples are reduced to a more general form are reported together with details of how these generalizations are established in Section 4.5 below.

3.5.3. *Statistical Approaches*

At this point we might also mention the way examples are "stored" in the statistical approaches. In fact, in these systems, the examples are not stored at all, except inasmuch as they occur in the corpus on which the system is based. What is stored is the precomputed statistical parameters which give the probabilities for bilingual word pairings, the "translation model". The "language model" which gives the probabilites of target word strings being well-formed is also precomputed, and the translation process consists of a search for the target-language string which optimises the product of the two sets of probabilities, given the source-language string.

3.6. MATCHING

The first task in an EBMT system is to take the source-language string to be translated and to find the example (or set of examples) which most closely match it. This is also the essential task facing a TM system. This search problem depends of course on the way the examples are stored. In the case of the statistical approach, the problem is the essentially mathematical one of maximising a huge number of statistical probabilites. In more conventional EBMT systems the matching process may be more or less linguistically motivated.

3.6.1. *Character-based Matching*

All matching processes necessarily involve a distance or similarity measure. In the most simple case, where the examples are stored as strings, the measure may be a traditional character-based pattern-matching one. In the earliest TM systems as mentioned above (ALPS' "Repetitions Processing", cf. Weaver 1988), only exact matches, *modulo* alphanumeric strings, were possible: (12a) would be matched with (12b), but the match in (13) would be missed because the system has no way of knowing that *small* and *large* are similar.

(12) a. This is shown as A in the diagram.
     b. This is shown as B in the diagram.

(13) a. The large paper tray holds up to 400 sheets of A3 paper.
     b. The small paper tray holds up to 300 sheets of A4 paper.

There is an obvious connection to be made here with the well-known problem of sequence comparison in spell-checking (the "string-correction" or "string-edit" problem, cf. Wagner & Fischer 1974), file comparison, speech processing, and other applications (see Kruskal 1983). Interestingly, few commentators make the connection explicitly, despite the significant wealth of literature on the subject.[6]

In the case of Japanese–English translation, which many EBMT systems focus on, the notion of character-matching can be modified to take account of the fact that certain "characters" (in the orthographic sense: each Japanese character is represented by two bytes) are more discriminatory than others (e.g. Sato 1992). This introduces a simple linguistic dimension to the matching process, and is akin to the well-known device in IR, where only keywords are considered.

3.6.2. *Word-based Matching*

Perhaps the "classical" similarity measure, suggested by Nagao (1984) and used in many early EBMT systems, is the use of a thesaurus or similar means of identifying word similarity on the basis of meaning or usage. Here, matches are permitted when words in the input string are replaced by near synonyms (as measured by relative distance in a hierarchically structured vocabulary, or by collocation scores such as mutual information) in the example sentences. This measure is particularly effective in choosing between competing examples, as in Nagao's examples, where, given (14a, b) as models, we choose the correct translation of *eat* in (15a) as *taberu* 'eat (food)', in (15b) as *okasu* 'erode', on the basis of the relative distance from *he* to *man* and *acid*, and from *potatoes* to *vegetables* and *metal*.

(14) a. A man eats vegetables. *Hito wa yasai o taberu*.
b. Acid eats metal. *San wa kinzoku o okasu*.

(15) a. He eats potatoes. *Kare wa jagaimo o taberu*.
b. Sulphuric acid eats iron. *Ryūsan wa tetsu o okasu*.

Another nice illustration of this idea is provided by Sumita et al. (1990) and Sumita & Iida (1991) who proposed EBMT as a method of addressing the notorious problem of translating Japanese adnominal particle constructions (*A no B*), where the default or structure-preserving translation (*B of A*) is wrong 80% of the time, and where capturing the wide variety of alternative translation patterns – a small selection of which is shown in (16) – with semantic features, as had been proposed in more traditional approaches to MT, is cumbersome and error-prone. Note that the Japanese is also underspecified for determiners and number, as well as the basic structure.

(16) a. *yōka no gogo*
8TH-DAY adn AFTERNOON
the afternoon of the 8th

b. *kaigi no mokuteki*
CONFERENCE adn SUBJECT
the subject of the conference

c. *kaigi no sankaryō*
CONFERENCE adn APPLICATION-FEE
the application fee for the conference

d. *kyōto-de no kaigi*
KYOTO-IN adn CONFERENCE
a conference in Kyoto

e. *kyōto-e no densha*
KYOTO-TO adn TRAIN
the Kyoto train

f. *isshukan no kyuka*
ONE-WEEK adn HOLIDAY
one week's holiday

g. *mittsu no hoteru*
THREE adn HOTEL
three hotels

Once again, a thesaurus is used to compare the similarity of the substituted items in a partial match, so that in (17)[7] we get the appropriate translations due to the similarity of *Kyōto* and *Tōkyō* (both place names), *kaigi* 'conference' and *kenkyukai* 'workshop', and *densha* 'train' and *shinkansen* 'bullet train'.

(17) a. *tōkyō-de no kenkyukai*
a workshop in Tokyo

b. *tōkyō-e no shinkansen*
the Tokyo bullet-train

Examples (14)–(17) show how the idea can be used to resolve both lexical and structural transfer ambiguity.

3.6.3. *Carroll's "Angle of Similarity"*

In a little-known research report, Carroll (1990) suggests a trigonometric similarity measure based on both the relative length and relative contents of the strings to be matched. The basic measure, like others developed later, compares the given sentence with examples in the database looking for similar words and taking account of deletions, insertions and substitutions. The relevance of particular mismatches is reflected as a "cost", and the cost can be programmed to reflect linguistic generalizations. For example, a missing comma may incur a lesser cost than a missing adjective or noun. And a substitution of like for like – e.g. two dissimilar alphanumerics as in (12) above, or a singular for a plural – costs less than a more significant replacement. The grammatical assignment implied by this was effected by a simple stem analysis coupled with a stop-word list: no dictionary as such was needed (though a re-implementation of this nowadays might, for example, use a tagger of the kind that was not available to Carroll in 1990). This gives a kind of "linguistic distance" measure which we shall refer to below as δ.

In addition to this is a feature which takes into account, unlike many other such similarity measures, the important fact illustrated by the four sentences in (18): if we take (18a) as the given sentence, which of (18b–d) is the better match?

(18) a. Select ‘Symbol’ in the Insert menu.
b. Select ‘Symbol’ in the Insert menu to enter a character from the symbol set.
c. Select ‘Paste’ in the Edit menu.
d. Select ‘Paste’ in the Edit menu to enter some text from the clip board.

Most similarity metrics will choose (18c) as the better match for (18a) since they differ by only two words, while (18b) has eight additional words. But intuitively, (18b) is a better match since it entirely includes the text of (18a). Further, (18b) and (18d) are more similar than (18a) and (18c). Carroll captures this with his notion of the “angle of similarity”: the distance δ *between* two sentences is seen as one side of a triangle, with the “sizes” of the two sentences as the other two sides. These sizes are calculated using the same distance measure, δ, but comparing the sentence to the null sentence, which we represent as ø. To arrive at the “angle of similarity” between two sentences x and y, we construct a triangle with sides of length $\delta(x, ø)$ (the size of x), $\delta(y, ø)$ (the size of y) and $\delta(x, y)$ (the difference between x and y). We can now calculate the angle θ_{xy} between the two sentences using the “half-sine” formula in (19).[8]

(19) $$\sin\frac{\theta_{xy}}{2} = \frac{\delta(x, y) - |\delta(x, ø) - \delta(y, ø)|}{2 \times \min\{\delta(x, ø), \delta(y, ø)\}}$$

We can illustrate this by assuming some values for the δ measure applied to the example sentences in (18), as shown in Table II. The angle of 0° in the first row shows that the difference between (18a) and (18b) is entirely due to length differences, that is, a quantitative difference but no qualitative difference. Similarly, the second and third rows show that there is both a qualitative and quantitative difference between the sentences, but the difference between (18b) and (18d) is less than that between (18a) and (18c).

Table II. Half-sine differences between sentences in (18)

Sentence pair x	y	Distance $\delta(x, y)$	Size x $\delta(x, ø)$	Size y $\delta(y, ø)$	Angle θ_{xy}
(18a)	(18b)	125	113	238	0°
(18a)	(18c)	103	113	125	47°
(18b)	(18d)	103	238	250	22°

3.6.4. *Annotated Word-based Matching*

The availability to the similarity measure of information about syntactic classes implies some sort of analysis of both the input and the examples. Cranias et al.

(1994, 1997) describe a measure that takes function words into account, and makes use of POS tags. Furuse & Iida's (1994) "constituent boundary parsing" idea is not dissimilar. Here, parsing is simplified by recognizing certain function words as typically indicating a boundary between major constituents. Other major constituents are recognised as part-of-speech bigrams.

Veale & Way (1997) similarly use sets of closed-class words to segment the examples. Their approach is said to be based on the "Marker hypothesis" from psycholinguistics (Green 1979) – the basis also for Juola's (1994, 1997) EBMT experiments – which states that all natural languages have a closed set of specific words or morphemes which appear in a limited set of grammatical contexts and which signal that context.

In the multi-engine Pangloss system, the matching process successively "relaxes" its requirements, until a match is found (Nirenburg et al. 1993, 1994): the process begins by looking for exact matches, then allows some deletions or insertions, then word-order differences, then morphological variants, and finally POS-tag differences, each relaxation incurring an increasing penalty.

3.6.5. *Structure-based Matching*

Earlier proposals for EBMT, and proposals where EBMT is integrated within a more traditional approach, assumed that the examples would be stored as structured objects, so the process involves a rather more complex tree-matching (e.g. Maruyama & Watanabe 1992; Matsumoto et al. 1993; Watanabe 1995; Al-Adhaileh & Tang 1999) though there is generally not much discussion of how to do this (cf. Maruyama & Watanabe 1992; Al-Adhaileh & Tang 1998), and there is certainly a considerable computational cost involved. Indeed, there is a not insignificant literature on tree comparison, the "tree edit distance" (e.g. Noetzel & Selkow 1983; Zhang & Shasha 1997; see also Meyers et al. 1996, 1998) which would obviously be of relevance.

Utsuro et al. (1994) attempt to reduce the computational cost of matching by taking advantage of the surface structure of Japanese, in particular its case-frame-like structure (NPs with overt case-marking). They develop a similarity measure based on a thesaurus for the head nouns. Their method unfortunately relies on the verbs matching exactly, and also seems limited to Japanese or similarly structured languages.

3.6.6. *Partial Matching for Coverage*

In most of the techniques mentioned so far, it has been assumed that the aim of the matching process is to find a single example or a set of individual examples that provide the best match for the input. An alternative approach is found in Nirenburg et al. (1993) (see also Brown 1997), Somers et al. (1994) and Collins (1998). Here, the matching function decomposes the cases, and makes a collection of – using these authors' respective terminology – "substrings", "fragments" or "chunks" of matched material. Figure 4 illustrates the idea.

```
danger/NN0 of/PRP NN0 < > above/PRP
danger/NN0 of/PRP
of/PRP NN0 < > above/PRP
above/PRP CRD m/NP0
there/PNP is/VVV a/AT0
avalanche/NN0 < > above/PRP
there/PNP is/VVV
is/VVV a/AT0
danger/NN0 of/PRP avalanche/NN0
avalanche/NN0 above/PRP CRD m/NP0
avalanche/NN0 above/PRP
of/PRP avalanche/NN0
there/PNP is/VVV < > a/AT0
is/VVV < > a/AT0
there/PNP is/VVV a/AT0 < > danger/NN0 < > of/PRP
there/PNP is/VVV < > danger/NN0 < > of/PRP
there/PNP is/VVV a/AT0 < > danger/NN0
a/AT0 < > danger/NN0
there/PNP is/VVV < > danger/NN0
```

Figure 4. Fragments extracted for the input *there is a danger of avalanche above 2000m*. The individual words are tagged; the matcher can also match tags only, and can skip unmatched words, shown as < >. The fragments are scored for relevance and frequency, which determines the order of presentation. From Somers et al. (1994).

Jones (1990) likens this process to "cloning", suggesting that the recombination process needed for generating the target text (see Section 3.7 below) is also applicable to the matching task:

> If the dataset of examples is regarded as not a static set of discrete entities but a permutable and flexible interactive set of process modules, we can envisage a control architecture where each process (example) attempts to clone itself with respect to (parts of) the input. (Jones 1990: 165)

In the case of Collins, the source-language chunks are explicitly linked to their corresponding translations, but in the other two cases, this linking has to be done at run-time, as is the case for systems where the matcher collects whole examples. We will consider this problem in the next section.

3.7. ADAPTABILITY AND RECOMBINATION

Having matched and retrieved a set of examples, with associated translations, the next step is to extract from the translations the appropriate fragments ("alignment" or “adaptation”), and to combine these so as to produce a grammatical target output (“recombination”). This is arguably the most difficult step in the EBMT process: its difficulty can be gauged by imagining a source-language monolingual trying to use a TM system to compose a target text. The problem is twofold: (a) identifying which portion of the associated translation corresponds to the matched portions of the source text, and (b) recombining these portions in an appropriate manner. Compared to the other issues in EBMT, they have received considerably less attention.

We can illustrate the problem by considering again the first example we saw (1), reproduced here (slightly simplified) as (20).

(20) a. He buys a notebook ⇒ *Kare wa nōto o kau*
b. I read a book on politics ⇒ *Watashi wa seiji nitsuite kakareta hon o yomu*
c. He buys a book on politics ⇒ *Kare wa seiji nitsuite kakareta hon o kau*

To understand how the relevant elements of (20a, b) are combined to give (20c), we must assume that there are other examples such as (21a, b), and a mechanism to extract from them the common elements (underlined here) which are assumed to correspond. Then, we have to make the further assumption that they can be simply pasted together as in (20c), and that this recombination will be appropriate and grammatical. Notice for example how the English word *a* and the Japanese word *o* are both common to all the examples: we might assume (wrongly as it happens) that they are mutual translations. And what mechanism is there which ensures that we do not produce (21c)?

(21) a. <u>He buys a</u> pen ⇒ <u>*Kare wa*</u> *pen* <u>*o kau*</u>
b. She wrote <u>a book on politics</u> ⇒ *Kanojo* <u>*wa seiji nitsuite kakareta hon o*</u> *kaita*
c. * *Kare wa wa seiji nitsuite kakareta hon o o kau*

In some approaches, where the examples are stored as tree structures, with the correspondences between the fragments explicitly labelled, the problem effectively disappears. For example, in Sato (1995), the recombination stage is a kind of tree unification, familiar in computational linguistics. Watanabe (1992, 1995) adapts a process called “gluing” from Graph Grammars, which is a flexible kind of graph unification. Al-Adhaileh & Tang (1999) state that the process is “analagous to top-down parsing” (p. 249).

Even if the examples are not annotated with the relevant information, in many systems the underlying linguistic knowledge includes information about correspondence at word or chunk level. This may be because the system makes use of a bilingual dictionary (e.g. Kaji et al. 1992; Matsumoto et al. 1993) or existing MT lexicon, as in the cases where EBMT has been incorporated into an existing rule-based architecture (e.g. Sumita et al. 1990; Frederking et al. 1994). Alternatively some systems extract automatically from the example corpus information about probable word alignments (e.g. Somers et al. 1994; Brown 1997; Veale & Way 1997; Collins 1998).

3.7.1. *Boundary Friction*

The problem is also eased, in the case of languages like Japanese and English, by the fact that there is little or no grammatical inflection to indicate syntactic function. So for example the translation associated with *the handsome boy* extracted, say, from (22), is equally reusable in either of the sentences in (23). This however is not the case for a language like German (and of course many others), where the form of the determiner, adjective and noun can all carry inflections to indicate grammatical case, as in the translations of (23a, b), shown in (24).

(22) The handsome boy entered the room.

(23) a. The handsome boy ate his breakfast.
b. I saw the handsome boy.

(24) a. *Der schöne Junge aß seinen Frühstück.*
b. *Ich sah den schönen Jungen.*

This is the problem sometimes referred to as "boundary friction" (Nirenburg et al. 1993: 48, Collins 1998: 22). One solution, in a hybrid system, would be to have a grammar of the target language, which could take the results of the gluing process and somehow smooth them over. Where the examples are stored as more than simple text strings, one can see how this might be possible. There is however no report of this approach having been implemented, as far as we know.

Somers et al. (1994) make use of the fact that the fragments have been extracted from real text, and so there is some information about contexts in which the fragment is known to have occurred:

> 'Hooks' are attached to each fragment which enable them to be connected together and their credibility assessed. The most credible combination, i.e. the one with the highest score, *should* be the best translation. (Somers et al. 1994:[8]; emphasis original)

The hooks indicate the words and POS tags that can occur before and after the fragment, with a weighting reflecting the frequency of this context in the corpus. Competing proposals for target text can be further evaluated by a process the authors call "disalignment", a kind of back-translation which partly reverses the process: if the proposed target text can be easily matched with the target-language part of the example database, this might be seen as evidence of its grammaticality.

3.7.2. *Adaptability*

Collins & Cunningham (1996, 1997; Collins 1998) stress the question of whether all examples are equally reusable with their notion of "adaptability". Their example-retrieval process includes a measure of adaptability which indicates the similarity of the example not only in its internal structure, but also in its external context. The notion of "adaptation-guided retrieval" has been developed in Case-Based Reasoning (CBR) (Smyth & Keane 1993; Leake 1995): here, when cases are retrieved from the example-base, it is not only their similarity with the given input, but also the extent to which they represent a good model for the desired output, i.e. to which they can be adapted, that determines whether they are chosen. Collins (1998: 31) gives the example of a robot using a "restaurant" script to get food at Macdonald's, when buying a stamp at the post-office might actually be a more appropriate, i.e. adaptable, model. Their EBMT system, ReVerb, stores the examples together with a functional annotation, cross-linked to indicate both lexical and functional equivalence. This means that example-retrieval can be scored on two counts: (a) the closeness of the match between the input text and the example, and (b) the adaptability of the example, on the basis of the relationship between the representations of the example and its translation. Obviously, good scores on both (a) and (b) give the best combination of retrievability and adaptability, but we might also find examples which are easy to retrieve but difficult to adapt (and are therefore bad examples), or the converse, in which case the good adaptability should compensate for the high retrieval cost. As the following example (from Collins, 1998: 81) shows, (25) has a good similarity score with both (26a) and (27a), but the better adaptability of (27b), illustrated in Figure 5, makes it a more suitable case.

(25) Use the Offset Command to increase the spacing between the shapes.

(26) a. Use the Offset Command to specify the spacing between the shapes.
b. *Mit der Option Abstand legen Sie den Abstand*
WITH THE OPTION SPACING MAKE YOU THE SPACING
zwischen den Formen fest.
BETWEEN THE SHAPES FIRM

(27) a. Use the Save Option to save your changes to disk.
b. *Mit der Option Speichern können Sie ihre Anderungen auf Diskette*
WITH THE OPTION SAVE CAN YOU YOUR CHANGES TO DISK
speichern.
SAVE

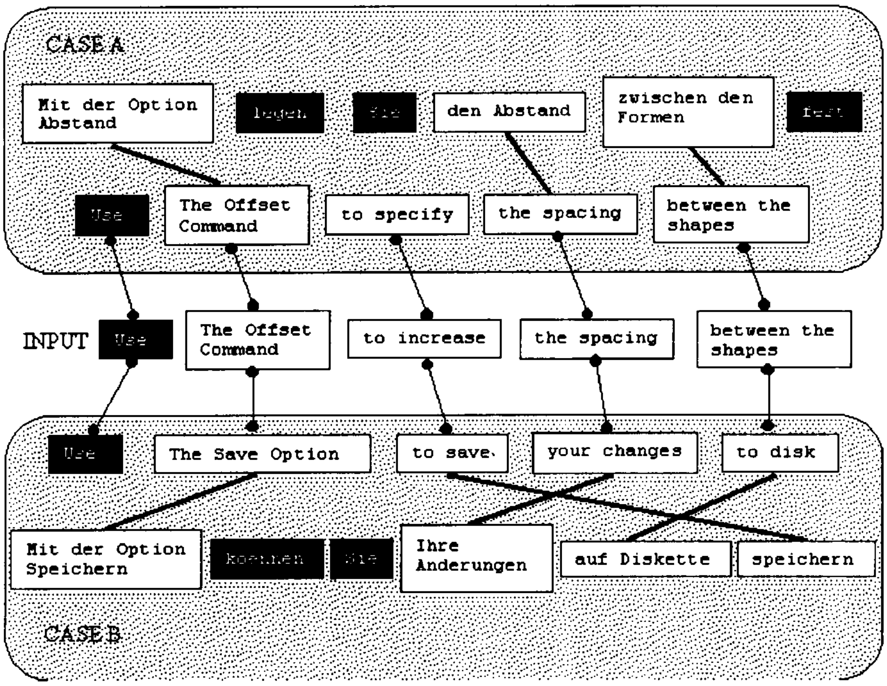

Figure 5. Adaptability versus similarity in retrieval (Collins 1998: 81).

3.7.3. *Statistical Modelling*

One other approach to recombination is that taken in the purely statistical system: like the matching problem, recombination is expressed as a statistical modelling problem, the parameters having been precomputed. This time, it is the "language model" that is invoked, with which the system tries to maximise the product of the word-sequence probabilities.

This approach suggests another way in which "recombined" target-language proposals could be verified: the frequency of co-occurrence of sequences of 2, 3 or more words (n-grams) can be extracted from corpora. If the target-language corpus (which need not necessarily be made up *only* of the aligned translations of the examples) is big enough, then appropriate statistics about the probable "correctness" of the proposed translation could be achieved. There are well-known techniques for calculating the probability of n-gram sequences, and a similar idea is found in Grefenstette's (1999) experiment, mentioned above, in which alternative translations of ambiguous noun compounds are verified by using them as search terms on the World Wide Web.

By way of example, consider again (23b) above, and its translation into German, (24b), repeated here as (28a). Suppose that an alternative translation (28b), using the substring from (24a), was proposed. In an informal experiment with AltaVista®, we used `"Ich sah den"` and `"Ich sah der"` as search terms, stipulating German web pages. The former gave 341 hits while the latter only 17. With `ich` rather than `Ich` in either case, the hits were 467 and 28 respectively. Other search engines produced similar or better results.

(28) a. *Ich sah den schönen Jungen*.

b. * *Ich sah der schöne Junge*.

3.8. COMPUTATIONAL PROBLEMS

All the approaches mentioned so far of course have to be implemented as computer programs, and significant computational factors influence many of them. One criticism to be made of the approaches which store the examples as complex annotated structures is the huge computational cost in terms of creation, storage and matching/retrieval algorithms. This is particularly problematic if such resources are difficult to obtain for one (or both) of the languages, as Güvenir & Cicekli (1998) report, relating to earlier work by Güvenir & Tunç (1996) on Turkish. Sumita & Iida (1995) is one of the few papers to address this issue explicitly, turning to parallel processing for help, a solution also adopted by Kitano (1994) and Sato (1995). Utsuro et al.'s (1994) approach has been described in Section 3.6.5 above.

A further criticism is that the complexities involved detract from some of the alleged advantages of EBMT, particularly the idea that the system's linguistic knowledge can be extended "simply" by increasing the size of the example-set (cf. Sato & Nagao, 1990: 252): adding more examples involves a significant overhead if these examples must be parsed, and the resulting representations possibly checked by a human. In the same vein, another advantage of the EBMT approach is said to be the ability to develop systems despite a lack of resources such as parsers, lexicons and so on, a key difference between the so-called rationalist and empiricist approaches to MT: a good example of this is Li et al.'s (1999) corpus-based Portuguese–Chinese MT system, a language pair whose development is enabled (and, in a circular manner, made necessary) by the particular situation in Macao.

One important computational issue is speed, especially for those of the EBMT systems that are used for real-time speech translation. Sumita et al. (1993) address this problem with the use of "massively parallel processors". With a small example base (1,000 cases) they achieved processing speeds almost 13 times faster than a more conventional architecture. For a more significant database, say 64,000 examples, the improvement would be 832 times. They warn however that speed advantages can be lost if the communication between the parallel processors and other processors is inefficient. It is understandable that some researchers are looking at ways of maximising the effect of the examples by identifying and making explicit significant generalizations. In this way the hybrid system has emerged, assuming the advantages of both the example-based and rule-based approaches.

4. Flavours of EBMT

So far we have looked at various solutions to the individual problems which make up EBMT. In this section, we prefer to take a wider view, to consider the various different contexts in which EBMT has been proposed. In many cases, EBMT is used as a component in an MT system which also has more traditional elements: EBMT may be used in parallel with these other "engines", or just for certain classes of problems, or when some other component cannot deliver a result. Also, EBMT methods may be better suited to some kinds of applications than others. And finally, it may not be obvious any more what exactly is the dividing line between EBMT and so-called "traditional" rule-based approaches. As the second paragraph of this paper suggests, EBMT was once seen as a bitter rival to the existing paradigm, but there now seems to be a much more comfortable coexistence.

4.1. SUITABLE TRANSLATION PROBLEMS

Let us consider first the range of translation problems for which EBMT is best suited. Certainly, EBMT is closely allied to *sublanguage* translation, not least because of EBMT's reliance on a real corpus of real examples: at least implicitly, a corpus can go a long way towards defining a sublanguage. On the other hand, nearly all research nowadays in MT is focused on a specific domain or task, so perhaps all MT is sublanguage MT.

More significant is that EBMT is often proposed as an antidote to the problem of "structure-preserving translation as first choice" (cf. Somers 1987: 84) inherent in MT systems which proceed on the basis of structural analysis. Because many EBMT systems do not compute structure, it follows that the source-language structure cannot be imposed on the target language. Indeed, some of the early systems in which EBMT is integrated into a more traditional approach explicitly use EBMT for such cases:

> When one of the following conditions holds true for a linguistic phenomenon, [rule-based] MT is less suitable than EBMT.
> (a) Translation rule formation is difficult.
> (b) The general rule cannot accurately describe [the] phenomen[on] because it represents a special case.
> (c) Translation cannot be made in a compositional way from target words. (Sumita & Iida 1991: 186)

One obvious question is whether any particular language pairs are more or less well suited to EBMT. Certainly, a large number of EBMT systems have been developed for Japanese–English (or vice versa) – cf. Table I – and it is sometimes claimed that the EBMT methodology favours typologically distinct languages, in that it distances itself from the structure-preserving approach that serves such language pairs so badly. But the fact that this language-pair is well represented could of course just be an accident of the fact that much of the research has been done in Japan. The availability of corpus material is also a factor, enabling for example an otherwise unlikely (for commercial reasons) language pair such as Portuguese–Chinese to be developed (Li et al. 1999). In fact, the range of languages for which EBMT systems – albeit experimental – have been developed is quite extensive.

4.2. PURE EBMT

Very few research efforts have taken an explicitly "purist" approach to EBMT. One exception is our own effort (Somers et al. 1994), where we wanted to push to the limits a "purely non-symbolic approach" in the face of, we felt, a premature acceptance that hybrids were the best solution. Not incorporating any linguistic information that could not be derived automatically from the corpus became a kind of dogma.

The other non-linguistic approach is of course the purely statistical one of Brown et al. (1988, 1990, 1993). In fact, their aspirations were much less dogmatic, and in the face of mediocre results, they were soon resorting to linguistic knowledge (Brown et al. 1992); not long afterwards the group broke up, though other groups have taken up the mantle of statistics-based MT (Vogel et al. 1986; Wang & Waibel 1997; etc.).

Other approaches, as we have seen above, while remaining more or less true to the case-based (rather than theory-based) approach of EBMT, accept the necessity to incorporate linguistic knowledge either in the representation of the examples, and/or in the matching and recombination processes. This represents one kind of hybridity of approach; but in this section we will look at hybrids in another dimension, where the EBMT approach is integrated into a more conventional system.

4.3. EBMT FOR SPECIAL CASES

One of the first uses envisaged for the EBMT approach was where the rule-based approach was too difficult. The classical case of this, as was shown above, example (16), was the translation of Japanese adnominal particle constructions. In the ATR system (Sumita et al. 1990; Sumita & Iida 1991), a traditional rule-based system, the EBMT module was invoked just for this kind of example (and a number of other similarly difficult cases). In a similar way, Katoh & Aizawa (1994) describe how only "parameterizable fixed phrases" in economics news stories are translated on the basis of examples, in a way very reminiscent of TM systems. Yamabana et al. (1997) integrate rule-based MT with a corpus-based statistical model for lexical selection, and an example-based method for structures such as compound nouns and noun phrases, which have a simple (and therefore sometimes syntactically and semantically idiosyncratic) structure, and also idiomatic expressions.

4.4. EXAMPLE-BASED TRANSFER

One can describe a number of systems where examples are stored as trees or other complex structures as "example-based transfer" systems: Sato & Nagao (1990), Sato (1991), Sadler (1991), Watanabe (1992, 1993, 1994, 1995), Matsumoto et al. (1993), Matsumoto & Kitamura (1997), Meyers et al. (1998), Al-Adhaileh & Tang (1998, 1999), Zhao & Tsujii (1999). In these systems, source-language inputs are analysed into structured representations in a conventional manner, only transfer is on the basis of examples rather than rules, and then generation of the target-language output is again done in a traditional way.

Watanabe (1993) provides a detailed description of how "translation patterns" are actually extracted from examples. Crucial to the process is the comparison of incorrect translations produced by the normal structure-preserving technique with the correct translation, as illustrated in (29).

(29) a. *Kare wa kuruma o kuji de ateru.*
HE topic CAR obj LOTTERY inst STRIKES
Lit. 'He strikes a car with the lottery.'
He wins a car as a prize in the lottery.

b. *Watashi no seibutsugaku no chishiki wa hinjaku da.*
I adn BIOLOGY adn KNOWLEDGE topic WEAK IS.
Lit. 'My knowledge of biology is weak.'
I have little knowledge of biology.

Taking the case of (29a), Figure 6 shows (a) the translation produced by existing rules or patterns, and (b) the correct translation. The parts of (b) which are different from (a) are highlighted, and provide the new pattern (c).

4.5. DERIVING TRANSFER RULES FROM EXAMPLES

Some researchers take this scenario a step further, using EBMT as a research technique to build the rule base rather than a translation technique *per se*. We can see this in the case of Furuse & Iida's (1992a, b) distinction of three types of "example" (8–10) in Section 3.5.2 above: they refer to "string-level", "pattern-level" and "grammar-level" transfer knowledge, and it seems that the more abstract representations are derived from examples by a process of generalization. The authors do not go into detail about how these generalized rules are constructed, though they *do* give some indication of how and where they are distributed (p. 146): from an analysis of the corpus on which their system is based, they have about 500 string-level rules covering the 50 most frequent sentences, frequent compound nouns, and single lexical items. About 300 pattern-level rules cover "frequent sentence patterns" and "A particle B patterns such as *A no B*", while there are about 20 grammar-level rules covering "continuation of nouns" (this term is not further explained). Remaining translation problems are handled in the traditional manner.

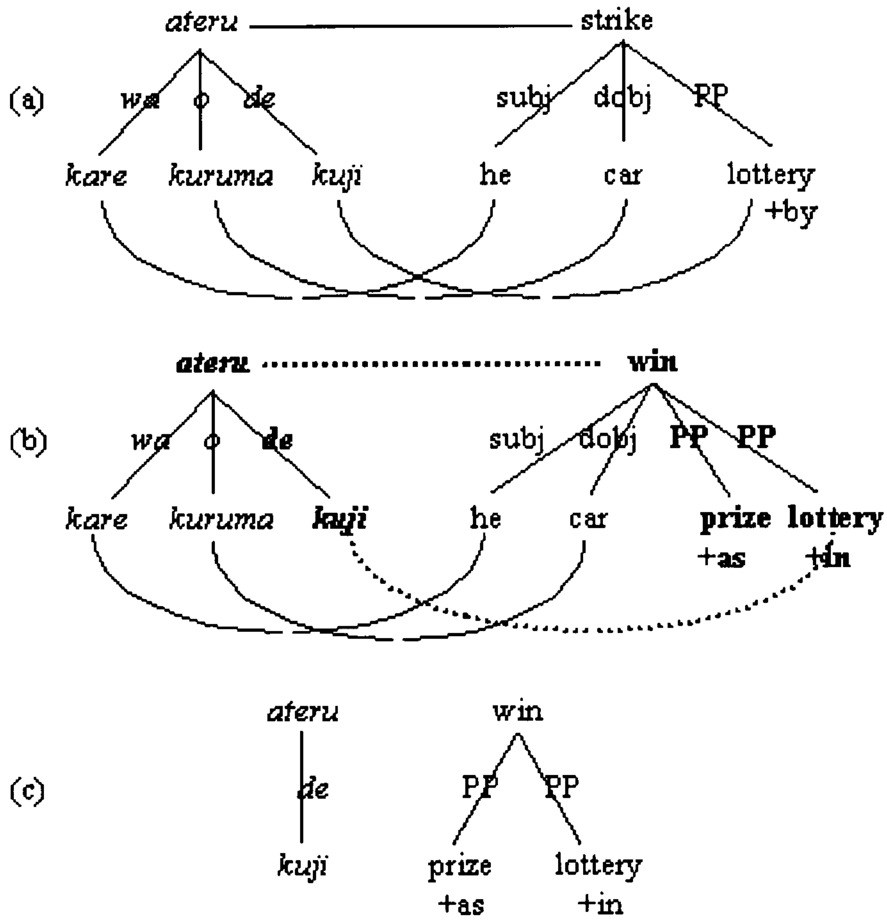

Figure 6. Extraction of translation pattern from example. (Watanabe 1993: 293).

4.5.1. *Generalization by Syntactic Category*

Kaji et al. (1992) describe their “two phase” EBMT methodology, the first phase involving “learning” of templates (i.e. transfer rules) from a corpus. Each template is a “bilingual pair of pseudo sentences”, i.e. example sentences containing variables. The translation templates are generated from the corpus first by parsing the translation pairs and then aligning the syntactic units with the help of a bilingual dictionary, resulting in a translation template as in Figure 7. This can then be generalized by replacing the coupled units with variables marked for syntactic category, as shown in (30).

(30) a. *X[NP] no nagasa wa saidai 512 baito de aru.*
X[NP] adn LENGTH topic MAXIMUM 512 BYTES cop BE
The maximum length of X[NP] is 512 bytes.

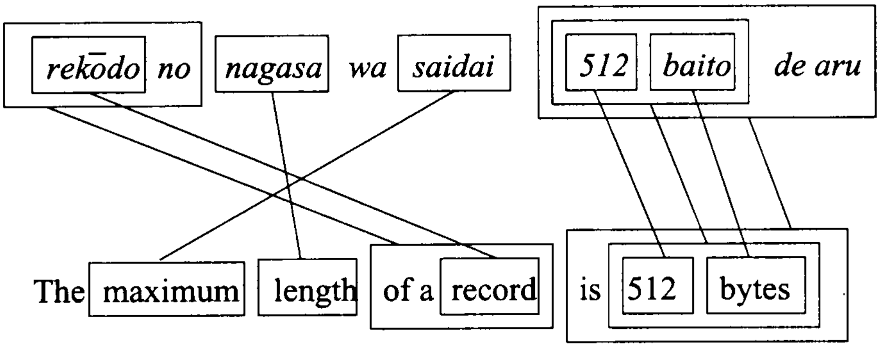

Figure 7. Generation of translation templates from aligned example. (Kaji et al. 1992: 673).

b. *X[NP] no nagasa wa saidai Y[N] baito de aru.*
The maximum length of X[NP] is Y[N] bytes.

Kaji et al. do not make explicit the criteria for choosing the units: any coupled unit pair can be replaced by variables. However, they do discuss the need to eliminate or refine templates which give rise to a conflict, as in (31–32).

(31) a. play baseball → *yakyu o suru*
b. play tennis → *tenisu o suru*
c. play X[NP] → *X[NP] o suru*

(32) a. play the piano → *piano o hiku*
b. play the violin → *baiorin o hiku*
c. play X[NP] → *X[NP] o hiku*

If possible, the template is "refined" by the addition of "semantic categories" which are "extracted from the original translation examples and attached to variables in the template", as shown in (33). The features are apparently determined manually.

(33) a. play X[NP/sport] → *X[NP] o suru*
b. play X[NP/instrument] → *X[NP] o hiku*

Carl (1999) similarly refines examples to give generalizations based on syntactic categories and morphological features. Likewise, Langé et al. (1997) describe their "skeleton-sentences" approach to TMs, where candidates for generalization are term pairs or "transwords" (roughly, alphanumerics and proper names which are not translated). Jain et al. (1995) report a similar approach.

4.5.2. *Generalization by Semantic Features*

Matsumoto & Kitamura (1995) describe how acquisition of general rules centres on individual words. A word must first be chosen explicitly as a possible source of translation rules. Then examples of the word in use are found in the corpus, these being then parsed with "LFG-like grammars" into dependency structures. From these, matching subgraphs are extracted. If the graphs consist just of a single word, then a word-level rule is generated. Otherwise, it is regarded as a "phrasal expression". The elements that differ are generalized according to similarities as determined by thesauri. Matsumoto & Kitamura work through an example with the Japanese verb *ataeru*. The most frequent translation found was *give*, when the subject is in one of the semantic classes [substance], [school], [store] or [difference], the direct object one of [difference], [unit], [chance], [feeling], [number] or [start end], and so on. A number of phrasal rules are also identified, as shown in (34), where the number of occurrences in the corpus is also indicated. It is not clear from the text how a rule can be generalized from a single occurrence, as in the last two cases.

(34) a. A[store, school] *ga* B[store, school, cause, . . .] *ni eikyō o ataeru* (17) ↔ A affect B
b. A[store, school] *ga* B[store, school] *ni hōshū o ataeru* (2) ↔ A compensate B
c. A[store, school] *ga* B[store, school] *ni doi o ataeru* (2) ↔ A assent to B
d. A[store] *ga* B[store] *ni shōnin o ataeru* (1) ↔ A authorize B
e. A[store] *ga* B[store] *ni* C[substance] *no hitsuyōryō o ataeru* (1) ↔ A furnish B with C

The authors note that "the quality of the translation rules depends on the quality of the thesaurus" (p. 415), and also note that their method works best with non-idiomatic text. Furthermore, their method is restricted to simple active declarative sentences.

Nomiyama (1992) similarly describes how examples ("cases") can be generalized into rules by combining them when similar segments occur in similar environments, this similarity being based on semantic proximity as given by a hierarchical thesaurus.

Almuallim et al. (1994) and Akiba et al. (1995) report much the same idea, though they are more formal in their description of how the process is implemented, citing the use of two algorithms from Machine Learning. Interestingly, these authors make no claim that their system is therefore "example-based". Also, many of the examples that they use to induce the transfer rules are artificially constructed. Watanabe & Takeda (1998) come to much the same position from the other direction: their "pattern-based MT" is essentially a variant of the traditional transfer-rule approach, but they propose extending the rule set by incorporating more specific rules (i.e. with fewer variables), and by treating existing rules as if they were examples, so that the rule used to translate a phrase like *take a bus* can be used, in the familiar manner, to translate *take a taxi*.

4.5.3. *Aligned Parse Trees*

Grishman (1994) and Meyers et al. (1996, 1998) are also quite formal in their description of how transfer rules are derived from aligned parse trees. Their representation is very similar to the dependency structures seen in many EBMT papers, though they restrict themselves to "alignments which preserve the dominance relationship" (unlike, for example, the well-known *long hair* example shown in Figure 2 above), stating that they see no need to consider violations of this constraint as "there are [none] in our corpus and many hypothetical cases can be avoided by adopting the appropriate grammar" (Meyers et al. 1998: 843).

4.5.4. *Automatic Generalization*

A simpler approach, requiring less initial analysis of the corpora is described by Cicekli & Güvenir (1996), Güvenir & Tunç (1996) and Güvenir & Cicekli (1998) for Turkish–English, and by McTait et al. (1999) and McTait & Trujillo (1999) for English–Spanish. Similar examples of translation sentence pairs are discovered and then combined into more general rules in the following way. Consider the pairs of sentences in (35) or (36):

(35) a. I took a ticket from Mary ⇔ *Mary'den bir bilet aldım*
b. I took a pen from Mary ⇔ *Mary'den bir kalem aldım*

(36) a. The Commission gave the plan up ⇔ *La Comisión abandonó el plan*
b. Our Government gave all laws up ⇔ *Nuestro Govierno abandonó todas las leyes*

From the sentence pairs can be identified the common elements, which are supposed to be mutual translations (37).

(37) a. I took a ... from Mary ⇔ *Mary'den bir ... aldım*
b. ... gave ... up ⇔ *abandonó*

This generalization can be stored as a translation "template". For Turkish, an agglutinative language, many such generalizations can be missed if examples are considered in their surface form. Therefore, the examples are subjected to a simple morphological analysis. This permits pairs like (38a) to be matched, with lexical representations as in (38b), where *H* is a morphophonemic representation capturing the Turkish vowel harmony.

(38) a. I am coming ⇔ *geliyorum*
I am going ⇔ *gidiyorum*
b. I am come+ing ⇔ *gel+Hyor+yHm*
I am go+ing ⇔ *gid+Hyor+yHm*

In both the approaches, the complementary elements in the matched sentences can be supposed to correspond as shown in (39).

(39) a. ticket ⇔ *bilet*; pen ⇔ *kalem*
b. The Commission . . . the plan ⇔ *La Comisión* . . . *el plan*; Our Government . . . all laws ⇔ *Nuestro Govierno* . . . *todas las leyes*

While the Turkish examples shown here involve a single correspondence, the Spanish examples leave more work to be done, since it is not obvious which of *La Comisión* and *el plan* correspond to *The Commission* and *the plan* (notwithstanding knowledge of Spanish, or recognition of cognates, which is not part of McTait & Trujillo's approach). Güvenir & Cicekli (1998) also face this problem, which they refer to as a "Corresponding Difference Pair" (CDP), as in (40), where, taking morphological alternation into account, the common, corresponding, elements are underlined, leaving non-unique CDPs.

(40) a. I gave the book ⇔ *Kitabi verdim*
b. You gave the pen ⇔ *Kalemi verdin*

Güvenir & Cicekli solve this problem by looking for further evidence in the corpus. For example, the pair already seen as (35) suggests that *kalem* corresponds to *pen*. McTait & Trujillo suggest an alternative method in which the elements of the "complement of collocation" are aligned according to their relative string lengths, as in Gale & Church's (1993) corpus alignment technique.

A further refinement is added by Öz & Cicekli (1998), who associate with each translation template derived in the above manner a "confidence factor" or weight based on the amount of evidence for any rule found in the corpus.

4.6. EBMT AS ONE OF A MULTI-ENGINE SYSTEM

One other scenario for EBMT is exemplified by the Pangloss system, where EBMT operates in parallel with two other techniques: knowledge-based MT and a simpler lexical transfer engine (Frederking & Nirenburg 1994; Frederking et al. 1994). Nirenberg et al. (1994) and Brown (1996) describe the EBMT aspect of this work in most detail. Frederking & Brown (1996) describe the PanLite implementation which covers four language pairs: English–Spanish, English–Serbo-Croatian and the inverse. What is most interesting is the extent to which the different approaches often mutually confirm each other's proposed translations, and the comparative evidence that the multi-engine approach offers.

Yamabana et al. (1997) also propose a multi-engine system, combining EBMT with rule-based and corpus-based approaches. An important feature of this system is its interactive nature: working bottom up, the system uses a rule-based approach to attempt to derive the syntactic structure, and proposes translations for the structures so determined. These translations are determined in parallel by the different modules of the system, i.e. rule-based transfer, statistics-based lexical selection, and an example-based module. These are then presented to the user who can modify the result of the analysis, intervene in the choice of translation, or directly edit the output.

Chen & Chen (1995) offer a combination of rule-based and statistical translation. Their approach differs from the previous two in that the translation method chosen is determined by the translation problem, whereas in the other two typically all the different engines will be activated in all cases, and their results compared.

5. Evaluation

An important feature of MT research in recent years has been evaluation, and this is no less the case for EBMT systems. A number of papers report usually small-scale evaluations of their proposals. As with all evaluations, there are the usual questions of what to evaluate and how. Nowhere in the literature so far, as far as we can ascertain, is there a paper exclusively reporting an evaluation of EBMT: so the evaluations that have been reported are usually added on as parts of papers describing the authors' approach. Some papers describe an entire EBMT translation system and so the evaluation section addresses overall translation quality. Other papers describe just one part of the EBMT method, often the matching part, occasionally other aspects.

5.1. EVALUATING EBMT AS A WHOLE

Where papers describe an entire EBMT translation system and include an evaluation section, this will be an evaluation of the translation quality achieved. As is well known, there are many different ways to evaluate translation quality, almost all of them beset with operational difficulties. The small-scale evaluations described as part of papers reporting broader issues are inevitably informal or impressionistic in nature.

A common theme is to use part of an available bilingual corpus for "training" the sytem, and then another part of the same corpus for testing. The translations proposed by the system are then compared to the translations found in the corpus. This is the method famously used by Brown et al. (1990) with their statistical MT system: having estimated parameters based on 117,000 sentences which used only the 1,000 most frequent words in the corpus, they then got the system to translate 73 sentences from elsewhere in the corpus. The results were classified as "identical", "alternate" (same meaning, different words), "different" (legitimate translation but not the same meaning), "wrong" and "ungrammatical". 30% of the translations came in the first two categories, with a further 18% possible but incorrect translations. This figure of 48% provided the baseline from which the authors strove to improve statistical MT until it came close to matching the performance of more traditional MT systems.

A simpler, binary, judgment was used by Andriamanankasina et al. (1999), who initially set up an example-base of 2,500 French–Japanese examples from conversation books, and then tested their system on 400 new sentences taken from the same source. The result was 62% correct translations. In a further experiment, these translations were edited and then added to the database. The success rate rose to 68.5%, which they took to be a very promising result.

Not so rigorous is the evaluation of Furuse & Iida (1992a, b), who claim an 89% success rate (their notion of "correct translation" is not defined) for their TDMT system, though it seems possible that their evaluation is using the same material from the ATR corpus that was used to construct the model in the first place. This also appears to be the case with Murata et al. (1999), who use a corpus of 36,617 sentences taken from a Japanese-English dictionary for the translation of tense, aspect and modality; they then take 300 randomly selected examples from the same source and compare their system with the output of commercially available software. The problem of needing test data independent of the training data is solved by Sumita & Iida (1991) with their "jackknife" evaluation method: the example database of 2,550 examples is partitioned into groups of 100. One group is taken as input and the remaining examples are used as data. This is then replicated 25 times. They report success rates (on the translation of *A no B* noun phrases – see above) between 70% and 89%, with an average of 78%.

Frederking & Nirenburg (1994) compare the translation performance of the EBMT module with that of the other translation systems in Pangloss, their multi-engine MT system. Their evaluation consisted of counting the number of editing key-strokes needed to convert the output into a "canonical" human translation. The results of a test using a 2,060-character text showed the multi-engine configuration to require 1,716 keystrokes, compared to 1,829 for simple dictionary look-up, 1,876 for EBMT and 1,883 for KBMT, with phrasal glossary look-up worst at 1,973 key-strokes. The authors admit that there are many flaws in this method of evaluation, both in the use of a single model translation (a human translator's version differed from the model by 1,542 key-strokes), and in the way that key-strokes are counted. Brown's (1996) evaluation focusses on the usefulness of the proposals made by the EBMT engine, rather than their accuracy. He talks of 70% "coverage", meaning that useful translation chunks are identified by the matcher, and 84% for which some translation is produced.[9]

Carl & Hansen (1999) compare translation performance of three types of EBMT system: a string-based TM, a lexeme-based TM and their structure-based EBMT system, EDGAR. Each of the systems is trained on a 303-sentence corpus and then tested on 265 examples taken from similar material. The evaluation metric involves comparison of the proposed translation with a manually produced "ideal", and measures the number of content words in common, apparently taking no account of word-order or grammatical correctness. The evaluation leads to the following conclusions:

> [T]he least generalizing system ... achieved higher translation precision when near matches can be found in the data base. However, if the reference corpus does not contain any similar translation example, EDGAR performed better.... We therefore conclude that the more an MT system is able to decompose and generalize the translation sentences, translate parts or single words of it and to recompose it into a target language sentence, the broader is its coverage and the more it loses translation precision. (Carl & Hansen 1999: 623)

5.2. EVALUATING THE MATCHER

Some papers on EBMT concentrate on the matching function of their system, a feature which is obviously of relevance also for TM systems. In each case, an attempt is made to quantify not only the number of examples retrieved, but also their usefulness for the translator in the case of a TM, or the effort needed by the next part of the translation process in the case of EBMT. Most evaluations exclude from the test suite any exact matches with the database, since identifying these is recognised as trivial.

Some evaluations involve rating the matches proposed. Both Sato (1990) and Cranias et al. (1994) use 4-point scales. Sato's "grades" are glossed as follows: (A) exact match, (B) "the example provides enough information about the translation of the whole input", (C) "the example provides information about the translation of the whole input", (F) "the example provides almost no information about the translation of the whole input". Sato apparently made the judgments himself, and so was presumably able to distinguish between the grades. More rigorously, Cranias et al. (1994) asked a panel of five translators to rate matches proposed by their system on a scale ranging from "a correct (or almost) translation", "very helpful", "[it] can help" and "of no use". Of course both these evaluations could be subject to criticism regarding subjectivity and small numbers of judges.

Matsumoto et al. (1993) could evaluate their structure-based matching algorithm by comparing the proposed structure with a target model. Their reported success rate of 89.8% on 82 pairs of sample sentences randomly selected from a Japanese–English dictionary conceals the fact that 23 of the examples could not be parsed, and of the remaining 59, 53 were correctly parsed. Of these 53, 47 were correctly matched by their algorithm, uniquely in the case of 34 of the examples. So this could be construed as 34 out of 82 unique correct matches: a success rate of 41.4%.

Collins (1998) uses a classification of the errors made by the matcher to evaluate her "adaptation-guided retrieval" scheme on 90 examples taken from an unused part of the corpus which she used to train her system.

Nirenburg et al.'s (1993) matching metrics include a self-scoring metric which can be used to evaluate matches. But an independent evaluation is also needed: they count the number of keystrokes required to convert the closest match back intop the input sentence. Counting keystrokes is a useful measure because it relates to the kind of task (post-editing) that is relevant for an example-matching algorithm. As Whyman & Somers (1999) discuss, however, arriving at this apparently simple measure is not without its difficulties, since mouse moves and clicks must also be counted, and also there are often alternatives ways of achieving the same post-editing result, including simply retyping. Their proposal is a general methodology, based on variants of the standard precision and recall measures, for determining the "fuzzy matching" rate at which a TM performs most efficiently, and is illustrated with a case study. A simpler variant on keystroke counting is found in Planas & Furuse (1999), who evaluate their proposed retrieval mechanism for TM by comparing its performance against a leading commercial TM system. Again taking sentences from an unused part of the training corpus, they quantify the difference between the input and the matched sentence by simply counting number of words needing to be changed.

5.3. OTHER EVALUATIONS

Almuallim et al. (1994) and Akiba et al. (1995) describe how examples are used to "learn" new transfer rules. Their approach, which is in the framework of Machine Learning, includes a "cross-validation" evaluation of the rules proposed by their technique.

Juola's (1994, 1997) small-scale experiments with "self-organizing" MT are accompanied by detailed evaluations, both "black-box" and "glass-box".

McTait & Trujillo (1999) applied their algorithm for extracting translation patterns to a corpus of 3,000 sentence pairs, and evaluated the "correctness" of 250 of the proposed templates by asking five bilinguals to judge them. The patterns align 0, 1 and 2 words in the source and target languages in various combinations. The 1:1 patterns, which were the most frequent (220) were 84% correct. The 146 2:2 patterns were 52% correct. 2:1 and 1:2 patterns were the next most accurate (35% of 26 and 21% of 72), while patterns involving alignments with no words (0:1, 0:2 and the converse) were frequently incorrect.

6. Conclusions

In this review article, we have seen a range of applications all of which might claim to "be" EBMT systems. So one outstanding question might be, What counts as EBMT? Certainly, the use of a bilingual corpus is part of the definition, but this is not sufficient. Almost all research on MT nowadays makes use at least of a "reference" corpus to help to define the range of vocabulary and structures that the system will cover. It must be something more, then.

EBMT means that the main knowledge-base stems from examples. However, as we have seen, examples may be used as a device to shortcut the knowledge-acquisition bottleneck in rule-based MT, the aim being to generalize the examples as much as possible. So part of the criterion might be whether the examples are used at run-time or not: but by this measure, the statistical approach would be ruled out; although the examples are not used to derive rules in the traditional sense, still at run-time there is no consultation of the database of examples.

The original idea for EBMT seems to have been couched firmly in the rule-based paradigm: examples were to be stored as tree structures, so rules must be used to analyse them: only transfer was to be done on the basis of examples, and then only for special, difficult cases. This was apparent in Sumita et al.'s reserved comments:

> [I]t is not yet clear whether EBMT can/should deal with the whole process of translation. We assume that there are many kinds of phenomena: some are suitable for EBMT and others are not. . . . Thus, it is more acceptable . . . if [rule-based] MT is first introduced as a base system which can translate totally, then its translation performance can be improved incrementally by attaching EBMT components as soon as suitable phenomena for EBMT are recognized. (Sumita et al. 1990: 211)

Jones (1992) discusses the trend towards "pure" EBMT research, which was motivated both by the comparative success of Sumita et al.'s approach, and also as a reaction to the apparent stagnation in research in the conventional paradigm. So the idea grew that EBMT might be a "new" paradigm altogether, in competition with the old, even. As we have seen, this confrontational aspect has quickly died away, and in particular EBMT has been integrated into more traditional approaches (and *vice versa*, one could say) in many different ways.

We will end this article by mentioning, for the first time, some of the advantages that have been claimed for EBMT. Not all the advantages that were claimed in the early days of polemic are obviously true. But it seems that at least the following do hold, inasmuch as the system design is primarily example-based (e.g. the examples may be "generalized", but corpus data is still the main source of linguistic knowledge):

– Examples are real language data, so their use leads to systems which cover the constructions which really occur, and ignore the ones that do not, so over-generation is reduced.
– The linguistic knowledge of the system can be more easily enriched, simply by adding more examples.
– EBMT systems are data-driven, rather than theory-driven: since there are therefore no complex grammars devised by a team of individual linguists, the problem of rule conflict and the need to have an overview of the "theory", and how the rules interact, is lessened. (On the other hand, as we have seen, there is the opposite problem of conflicting examples.)
– The example-based approach seems to offer some relief from the constraints of "structure-preserving" translation.
– Depending on the way the examples are used, it is possible that an EBMT system for a new language pair can be quickly developed on the basis of (only) an aligned parallel corpus. This is obviously attractive if we want an

MT system involving a language for which resources such as parsers and dictionaries are not available.

EBMT is certainly here to stay, not as a rival to rule-based methods but as an alternative, available to enhance and, sometimes, replace it. Nor is research in the purely rule-based paradigm finished. As I mentioned in Somers (1997:116), the problem of scaling up remains, as do a large number of interesting translation problems, especially as new uses for MT (e.g. web-page and e-mail translation) emerge. The "new" paradigm is now approaching its teenage years: the dust has settled, and the road ahead is all clear.

Acknowledgments

A version of this paper was originally prepared for the Workshop on Machine Translation at the 10th European Summer School in Logic, Language and Information, Saarbrücken, August 1998. I am grateful to Frank van Eynde for inviting me to participate. The present version is considerably enhanced. The organization of the refereeing and editing of this article has been done by Graeme Hirst on behalf of the editor of this journal who is also the author of this article. I am grateful to him and to the other members of the Editorial Board for their support in this, and also to the referees whose suggestions have helped me to improve the paper considerably. I am also grateful to Nick In 't Veen for useful discussions about EBMT.

Notes

1 I am indebted to John Hutchins for bringing this article to my attention. Early proposals for a TM, and other aspects of the idea of a Translator's Workstation are described in Hutchins (1998).

2 In this and subsequent Japanese examples, the widely used Hepburn transcription (Hepburn 1872) is adopted. This is a phonetic rather than phonemic or literal transcription, and differs from the transcription often used by Japanese authors – sometimes referred to as *wāpurō* 'word-processor' transcription – in several respects, including the transcription of the topic and object particles as *wa* and *o*, rather than *ha* and *wo*, respectively.

3 By "parallel" we mean a text together with its translation. By "aligned", we mean that the two texts have been analysed into corresponding segments; the size of these segments may vary, but typically corresponds to sentences. It is of interest to note that for some corpus linguists, the term "translation corpus" is used to indicate that the texts are mutual translations, while "parallel corpus" refers to any collection of multilingual texts of a similar genre. Other researchers prefer the term "comparable corpus" (cf. McEnery & Wilson 1996: 60n).

4 I have adapted Watanabe's transcription here, and corrected an obvious misprint in (5a).

5 Examples are from Somers et al. (1990: 274).

6 Exceptions include Kitano (1993) and Somers et al. (1994), both of which make reference to Dynamic Programming.

7 The examples have been adapted slightly.

8 This formula is preferred over the simpler cosine formula to overcome a small problem to do with the fact that the Euclidean hypothesis implicit in the use of the cosine formula does not in fact hold for our problem. See Carroll (1990) for details.

9 The text is garbled at this point, so we cannot be sure what was meant.

References

Akiba, Y., M. Ishii, H. Almuallim and S. Kaneda: 1995, 'Learning English Verb Selection Rules from Hand-made Rules and Translation Examples', in TMI (1995), pp. 206–220.

Al-Adhaileh, M. H. and E. K. Tang: 1998, 'A Flexible Example-Based Parser Based on the SSTC', in Coling-ACL (1998), pp. 687–693.

Al-Adhaileh, M. H. and E. K. Tang: 1999, 'Example-Based Machine Translation Based on the Synchronous SSTC Annotation Schema', *Machine Translation Summit VII*, Singapore, pp. 244–249.

Almuallim, H., Y. Akiba, T. Yamazaki, A. Yokoo and S. Kaneda: 1994, 'Two Methods for Learning ALT-J/E Translation Rules from Examples and a Semantic Hierarchy', in Coling (1994), pp. 57–63.

Andriamanankasina, T., K. Araki and K. Tochinai: 1999, 'Example-Based Machine Translation of Part-Of-Speech Tagged Sentences by Recursive Division', *Machine Translation Summit VII*, Singapore, pp. 509–517.

Arthern, P. J.: 1978, 'Machine Translation and Computerized Terminology Systems: A Translator's Viewpoint', in B. M. Snell (ed.), *Translating and the Computer: Proceedings of a Seminar, London, 14th November 1978*, pp. 77–108, Amsterdam (1979), North-Holland.

Arthern, P. J.: 1981, 'Aids Unlimited: The Scope for Machine Aids in a Large Organization', *Aslib Proceedings* **33**, 309–319.

Bod, R.: 1992, 'A Computational Model of Language Performance: Data Oriented Parsing', in Coling (1992), pp. 855–859.

Brown, P., J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer and P. Roossin: 1988, 'A Statistical Approach to French/English Translation', in TMI (1998).

Brown, P. F., J. Cocke, S. A. Della Pietra, V. J. Della Pietra, F. Jelinek, J. D. Lafferty, R. L. Mercer and P. S. Roossin: 1990, 'A Statistical Approach to Machine Translation', *Computational Linguistics* **16**, 79–85.

Brown, P. F., S. A. Della Pietra, V. J. Della Pietra, J. D. Lafferty and R. L. Mercer: 1992, 'Analysis, Statistical Transfer, and Synthesis in Machine Translation', in TMI (1992), pp. 83–100.

Brown, P. F., S. A. Della Pietra, V. J. Della Pietra and R. L. Mercer: 1993, 'The Mathematics of Statistical Machine Translation: Parameter Estimation', *Computational Linguistics* **19**, 263–311.

Brown, R. D.: 1996, 'Example-Based Machine Translation in the Pangloss System', in Coling (1996), pp. 169–174.

Brown, R. D.: 1997, 'Automated Dictionary Extraction for "Knowledge-Free" Example-Based Translation', in TMI (1997), pp. 111–118.

Brown, R. D.: 1999, 'Adding Linguistic Knowledge to a Lexical Example-based Translation System', in TMI (1999), pp. 22–32.

Carbonell, J. G.: 1986, 'Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition', *Machine Learning* **2**.

Carl, M.: 1999, 'Inducing Translation Templates for Example-Based Machine Translation', *Machine Translation Summit VII*, Singapore, pp. 250–258.

Carl, M. and S. Hansen: 1999, 'Linking Translation Memories with Example-Based Machine Translation', *Machine Translation Summit VII*, Singapore, pp. 617–624.

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

Castaño, M. A., F. Casacuberta and E. Vidal: 1997, 'Machine Translation Using Neural Networks and Finite-State Models', in TMI (1997), pp. 160–167.

Chen, K. and H-H. Chen: 1995, 'Machine Translation: An Integarted Approach', in TMI (1995), pp. 287–294.

Cicekli, I. and H. A. Güvenir: 1996, 'Learning Translation Rules From A Bilingual Corpus', *NeMLaP-2: Proceedings of the Second International Conference on New Methods in Language Processing*, Ankara, Turkey, pp. 90–97.

Coling: 1986, *11th International Conference on Computational Linguistics: Proceedings of Coling '86*, Bonn, West Germany.

Coling: 1990, *COLING-90, Papers Presented to the 13th International Conference on Computational Linguistics*, Helsinki, Finland.

Coling: 1992, *Proceedings of the fifteenth [sic] International Conference on Computational Linguistics, Actes du quinzième colloque international en linguistique informatique, COLING-92*, Nantes, France.

Coling: 1994, *COLING 94, The 15th International Conference on Computational Linguistics*, Kyoto, Japan.

Coling: 1996, *COLING-96: The 16th International Conference on Computational Linguistics*, Copenhagen, Denmark.

Coling-ACL: 1998, *COLING-ACL '98: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Montreal, Quebec.

Collins, B.: 1998, *Example-Based Machine Translation: An Adaptation-Guided Retrieval Approach*. PhD thesis, Trinity College, Dublin.

Collins, B. and P. Cunningham: 1995, 'A Methodology for Example Based Machine Translation', *CSNLP 1995: 4th Conference on the Cognitive Science of Natural Language Processing*, Dublin, Ireland.

Collins, B. and P. Cunningham: 1996, 'Adaptation-Guided Retrieval in EBMT: A Case-Based Approach to Machine Translation', in I. Smith and B. Faltings (eds), *Advances in Case-Based Reasoning: Third European Workshop, EWCBR-96*, pp. 91–104, Berlin: Springer.

Collins, B. and P. Cunningham: 1997, 'Adaptation Guided Retrieval: Approaching EBMT with Caution', in TMI (1997), pp. 119–126.

Collins, B., P. Cunningham and T. Veale: 1996, 'An Example-Based Approach to Machine Translation', *Expanding MT Horizons: Proceedings of the Second Conference of the Association for Machine Translation in the Americas*, Montreal, Quebec, pp. 1–13.

Cranias, L., H. Papageorgiou and S. Piperidis: 1994, 'A Matching Technique in Example-Based Machine Translation', in Coling (1994), pp. 100–104.

Cranias, L., H. Papageorgiou and S. Piperidis: 1997, 'Example Retrieval from a Translation Memory', *Natural Language Engineering* **3**, 255–277.

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

Frederking, R. and R. Brown: 1996, 'The Pangloss-Lite Machine Translation System', *Expanding MT Horizons: Proceedings of the Second Conference of the Association for Machine Translation in the Americas*, Montreal, Quebec, pp. 268–272.

Frederking, R. and S. Nirenburg: 1994, 'Three Heads are Better than One', *4th Conference on Applied Natural Language Processing*, Stuttgart, Germany, pp. 95–100.

Frederking, R., S. Nirenburg, D. Farwell, S. Helmreich, E. Hovy, K. Knight, S. Beale, C. Domashnev, D. Attardo, D. Grannes and R. Brown: 1994, 'Integrating Translations from Multiple Sources within the Pangloss Mark III Machine Translation System', *Technology Partnerships for Crossing the Language Barrier: Proceedings of the First Conference of the Association for Machine Translation in the Americas*, Columbia, Maryland, pp. 73–80.

Fung, P. and K. McKeown: 1997, 'A Technical Word- and Term-Translation Aid Using Noisy Parallel Corpora across Language Groups', *Machine Translation* **12**, 53–87.

Furuse, O. and H. Iida: 1992a, 'An Example-Based Method for Transfer-Driven Machine Translation', in TMI (1992), pp. 139–150.

Furuse, O. and H. Iida: 1992b, ‘Cooperation between Transfer and Analysis in Example-Based Framework’, in Coling (1992), pp. 645–651.

Furuse, O. and H. Iida: 1994, ‘Constituent Boundary Parsing for Example-Based Machine Translation’, in Coling (1994), pp. 105–111.

Gale, W. A. and K. W. Church: 1993, ‘A Program for Aligning Sentences in Bilingual Corpora’, *Computational Linguistics* **19**, 75–102.

Gerloff, P.: 1987. ‘Identifying the Unit of Analysis in Translation’, in C. Færch and G. Kasper (eds), *Introspection in Second Language Research*, Clevedon: Multilingual Matters, pp. 135–158.

Green, T. R. G.: 1979, ‘The Necessity of Syntax Markers: Two Experiments with Artificial Languages’, *Journal of Verbal Learning and Verbal Behavior* **18**, 481–496.

Grefenstette, G.: 1999, ‘The World Wide Web as a Resource for Example-Based Machine Translation Tasks’, *Translating and the Computer 21*, London: Aslib/IMI.

Grishman, R.: 1994, ‘Iterative Alignment of Syntactic Structures for a Bilingual Corpus’, *Second Annual Workshop on Very Large Corpora (WVLC2)*, Kyoto, Japan, pp. 57–68.

Güvenir, H. A. and I. Cicekli: 1998, ‘Learning Translation Templates from Examples’, *Information Systems* **23**, 353–363.

Güvenir, H. A. and A. Tunç: 1996, ‘Corpus-Based Learning of Generalized Parse Tree Rules for Translation’, in G. McCalla (ed.), *Advances in Artificial Intelligence*, Berlin: Springer Verlag, pp. 121–132.

Hepburn, J. C.: 1872. *A Japanese-English and English-Japanese Dictionary*. Shanghai: American Prestbyterian Mission Press.

Hutchins, J. 1998: ‘The Origins of the Translator’s Workstation’, *Machine Translation* **13**, 287–307.

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

Jain, R., R. M. K. Sinha and A. Jain: 1995, ‘Role of Examples in Translation’, *1995 IEEE International Conference on Systems, Man and Cybernetics*, Vancouver, BC, pp. 1615–1620.

Jones, D.: 1992, ‘Non-hybrid Example-based Machine Translation Architectures’, in TMI (1992), pp. 163–171.

Jones, D.: 1996, *Analogical Natural Language Processing*, London: UCL Press.

Jones, D. and H. Somers (eds): 1997, *New Methods in Language Processing*, London: UCL Press.

Juola, P.: 1994, ‘A Psycholinguistic Approach to Corpus-Based, Machine Translation’, *CSNLP 1994: 3rd Conference on the Cognitive Science of Natural Language Processing*, Dublin, Ireland.

Juola, P.: 1997, ‘Corpus-Based Acquisition of Transfer Functions Using Psycholinguistic Principles’, in Jones & Somers (1997), pp. 207–218.

Kaji, H., Y. Kida and Y. Morimoto: 1992, ‘Learning Translation Templates from Bilingual Text’, in Coling (1992), pp. 672–678.

Katoh, N. and T. Aizawa: 1994, ‘Machine Translation of Sentences with Fixed Expressions’, *4th Conference on Applied Natural Language Processing*, Stuttgart, Germany, pp. 28–33.

Kay, M.: 1980, ‘The Proper Place of Men and Machines in Language Translation’, Research Report CSL-80-11, Xerox PARC, Palo Alto, Calif. Reprinted in *Machine Translation* **12**, 3–23 (1997).

Kitano, H.: 1993, ‘A Comprehensive and Practical Model of Memory-Based Machine Translation’, *Proceedings of the International Joint Conference on Artificial Intelligence*, Chambéry, France, pp. 1276–1282.

Kitano, H.: 1994, *Speech-to-Speech Translation: A Massively Parallel Memory-Based Approach*, Boston: Kluwer.

Kitano, H. and T. Higuchi: 1991a, ‘High Performance Memory-Based Translation on IXM2 Massively Parallel Associative Memory Processor’, *AAAI-91 Proceedings Ninth National Conference on Artificial Intelligence*, Menlo Park: AAAI Press, pp. 149–154.

Kitano, H. and T. Higuchi: 1991b, ‘Massive Parallel Memory-Based Parsing’, *IJCAI-91 Proceedings of the Twelfth International Conference on Artificial Intelligence*, Sydney, Australia, pp. 918–924.

Koncar, N. and G. Guthrie: 1997, 'A Natural-Language-Translation Neural Network', in Jones & Somers (1997), pp. 219–228.

Kruskal, J. B.: 1983, 'An Overview of Sequence Comparison', in Sankoff & Kruskal (1983), pp. 1–44.

Langé, J. M., É. Gaussier and B. Daille: 1997, 'Bricks and Skeletons: Some Ideas for the Near Future of MAHT', *Machine Translation* **12**, 39–51.

Leake, D.: 1995, 'Adaptive Similarity Assessment for Case-based Explanation', *International Journal of Expert Systems* **8**, 165–194.

Li, Y., C. Pu and F. Wu: 1999, 'Portuguese-Chinese Machine Translation in Macao', *Machine Translation Summit VII*, Singapore, pp. 236–243.

Maruyama, H. and H. Watanabe: 1992, 'Tree Cover Search Algorithm for Example-Based Translation', in TMI (1992), pp. 173–184.

Matsumoto, Y., H. Ishimoto and T. Utsuro: 1993, 'Structural Matching of Parallel Texts', *31st Annual Meeting of the Association for Computational Linguistics*, Columbus, Ohio, pp. 23–30.

Matsumoto, Y. and M. Kitamura: 1995, 'Acquisition of Translation Rules from Parallel Corpora', in Mitkov & Nicolov (1995), pp. 405–416.

McEnery, T. and A. Wilson: 1996, *Corpus Linguistics*, Edinburgh: Edinburgh University Press.

McLean, I. J.: 1992, 'Example-Based Machine Translation Using Connectionist Matching', in TMI (1992), pp. 35–43.

McTait, K., M. Olohan and A. Trujillo: 1999, 'A Building Blocks Approach to Translation Memory', *Translating and the Computer 21*, London: Aslib/IMI.

McTait, K. and A. Trujillo: 1999, 'A Language-Neutral Sparse-Data Algorithm for Extracting Translation Patterns', in TMI (1999), pp. 98–108.

Medin, D. L. and M. M. Schaffer: 1978, 'Context Theory of Classification Learning', *Psychological Review* **85**, 207–238.

Melby, A. K.: 1995, *The Possibility of Language: A Discussion of the Nature of Language*, Amsterdam: John Benjamins.

Meyers, A., R. Yangarber and R. Grishman: 1996, 'Alignment of Shared Forests for Bilingual Corpora', in Coling (1996), pp. 459–465.

Meyers, A., R. Yangarber, R. Grishman, C. Macleod and A. Moreno-Sandeval: 1998, 'Deriving Transfer Rules from Dominance-Preserving Alignments', in Coling-ACL (1998), pp. 843–847.

Mima, H., H. Iida and O. Furuse: 1998, 'Simultaneous Interpretation Utilizing Example-based Incremental Transfer', in Coling-ACL (1998), pp. 855–861.

Mitkov, R. and N. Nicolov (eds): 1997, *Recent Advances in Natural Language Processing: Selected Papers from RANLP'95*, Amsterdam: John Benjamins.

Murata, M., Q. Ma, K. Uchimoto and H. Isahara: 1999, 'An Example-Based Approach to Japanese-to-English Translation of Tense, Aspect, and Modality', in TMI (1999), pp. 66–76.

Nagao, M.: 1984. 'A Framework of a Mechanical Translation between Japanese and English by Analogy Principle', in A. Elithorn and R. Banerji (eds), *Artificial and Human Intelligence*, Amsterdam: North-Holland, pp. 173–180.

Nirenburg, S., S. Beale and C. Domashnev: 1994, 'A Full-Text Experiment in Example-Based Machine Translation', *International Conference on New Methods in Language Processing (NeMLaP)*, Manchester, England, pp. 78–87.

Nirenburg, S., C. Domashnev and D. J. Grannes: 1993, 'Two Approaches to Matching in Example-Based Machine Translation', in TMI (1993), pp. 47–57.

Noetzel, A. S. and S. M. Selkow: 1983, 'An Analysis of the General Tree-Editing Problem', in Sankoff & Kruskal (1983), pp. 237–252.

Nomiyama, H.: 1992, 'Machine Translation by Case Generalization', in Coling (1992), pp. 714–720.

Oi, K., E. Sumita, O. Furuse, H. Iida and T. Higuchi: 1994, 'Real-Time Spoken Language Translation Using Associative Processors', *4th Conference on Applied Natural Language Processing*, Stuttgart, Germany, pp. 101–106.

Öz, Z. and I. Cicekli: 1998, 'Ordering Translation Templates by Assigning Confidence Factors', in Farwell et al. (1998), pp. 51–61.

Pappegaaij, B. C., V. Sadler and A. P. M. Witkam (eds): 1986a, *Word Expert Semantics: An Interlingual Knowledge-Based Approach*, Dordrecht: Reidel.

Pappegaaij, B. C., V. Sadler and A. P. M. Witkam: 1986b, 'Experiments with an MT-Directed Lexical Knowledge Bank', in Coling (1986), pp. 432–434.

Planas, E. and O. Furuse: 1999, 'Formalizing Translation Memories', *Machine Translation Summit VII*, Singapore, pp. 331–339.

Poutsma, A.: 1998, 'Data-Oriented Translation', *Computational Linguistics in the Netherlands: Ninth CLIN Meeting*, Leuven, Belgium; abstract at `http://www.ccl.kuleuven.ac.be/clin98/abstr/poutsma.html`.

Resnik, P.: 1998, 'Parallel Strands: A Preliminary Investigation into mining the Web for Bilingual Text', in Farwell et al. (1998), pp. 72–82.

Riesbeck, C. and R. Schank: 1989, *Inside Case-Based Reasoning*, Hillsdale, NJ: Lawrence Erlbaum.

Sadler, V.: 1991, 'The Textual Knowledge Bank: Design, Construction, Applications', *International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP)*, Kyoto, Japan, pp. 17–32.

Sadler, V. and R. Vendelmans: 1990, 'Pilot Implementation of a Bilingual Knowledge Bank, in Coling (1990), Vol. 3, pp. 449–451.

Sankoff, D. and J. Kruskal (eds): 1983, *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, Reading, Mass.: Addison-Wesley.

Sato, S.: 1992, 'CTM: an Example-Based Translation Aid System', in Coling (1992), pp. 1259–1363.

Sato, S.: 1993, 'Example-Based Translation of Technical terms', in TMI (1993), pp. 58–68.

Sato, S.: 1995, 'MBT2: A Method for Combining Fragments of Examples in Example-Based Machine Translation', *Artificial Intelligence* **75**, 31–49.

Sato, S. and M. Nagao: 1990, 'Toward Memory-Based Translation', in Coling (1990), Vol. 3, pp. 247–252.

Schubert, K.: 1986, 'Linguistic and Extra-Linguistic Knowledge: A Catalogue of Language-related Rules and their Computational Application in Machine Translation', *Computers and Translation* **1**, 125–152.

Skousen, R.: 1989, *Analogical Modeling*, Dordrecht, Kluwer.

Smyth, B. and M. T. Keane: 1993, 'Retrieving Adaptable Cases', in I. Smith and B. Faltings (eds), *Topics in Case-Based Reasoning: Proceedings of the First European Workshop on Case-Based Reasoning, EWCBR-93*, Vol. 1, pp. 76–82, Berlin: Springer.

Sobashima, Y., O. Furuse, S. Akamine, J. Kawai and H. Iida: 1994, 'A Bidirectional, Transfer-driven Machine Translation System for Spoken Dialogues', in Coling (1994), pp. 64–68.

Somers, H.: 1987, 'Some Thoughts on Interface Structure(s)', in W. Wilss and K.-D. Schmitz (eds), *Maschinelle Übersetzung – Methoden und Werkzeuge*, Tübingen: Niemeyer, pp. 81–99.

Somers, H. L.: 1997, 'The Current State of Machine Translation', *MT Summit VI: Machine Translation Past Present Future*, San Diego, California, pp. 115–124.

Somers, H.: 1998, 'Further Experiments in Bilingual Text Alignment', *International Journal of Corpus Linguistics* **3**, 1–36.

Somers, H. L. and D. Jones: 1992, 'Machine Translation Seen as Interactive Multilingual Text Generation', *Translating and the Computer 13: The Theory and Practice of Machine Translation – A Marriage of Convenience?*, London: Aslib, pp. 153–165.

Somers, H., I. McLean and D. Jones: 1994, 'Experiments in Multilingual Example-Based Generation', *CSNLP 1994: 3rd Conference on the Cognitive Science of Natural Language Processing*, Dublin, Ireland.

Somers, H. L., J. Tsujii and D. Jones: 1990, 'Machine Translation without a Source Text', in Coling (1990), Vol. 3, pp. 271–276.

Stanfill, C. and D. Waltz: 1986, 'Toward Memory-Based Reasoning', *Communications of the ACM* **29**, 185–192.

Sumita, E. and H. Iida: 1991, 'Experiments and Prospects of Example-Based Machine Translation', *29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, California, pp. 185–192.

Sumita, E. and H. Iida: 1995, 'Heterogeneous Computing for Example-Based Translation of Spoken Language', in TMI (1995), pp. 273–286.

Sumita, E., H. Iida and H. Kohyama: 1990, 'Translating with Examples: A New Approach to Machine Translation', in TMI (1990), pp. 203–212.

Sumita, E., N. Nisiyama and H. Iida: 1994, 'The Relationship Between Architectures and Example-Retrieval Times', *AAAI-94 Proceedings of the Twelfth National Conference on Artificial Intelligence*, Menlo Park: AAAI Press, pp. 478–483.

Sumita, E., K. Oi, O. Furuse, H. Iida, T. Higuchi, N. Takahashi and H. Kitano: 1993, 'Example-Based Machine Translation on Massively Parallel Processors', *IJCAI-93 Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, Chambéry, France, pp. 1283–1288.

Sumita, E. and Y. Tsutsumi: 1988, 'A Translation Aid System Using Flexible Text Retrieval Based on Syntax-Matching', in TMI (1988), Proceedings Supplement.

TMI: 1988, *Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Pittsburgh, Pennsylvania.

TMI: 1990, *The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, Austin, Texas.

TMI: 1992, *Quatrième colloque international sur les aspects théoriques et méthodologiques de la traduction automatique; Fourth International Conference on Theoretical and Methodological Issues in Machine Translation. Méthodes empiristes versus méthodes rationalistes en TA; Empiricist vs. Rationalist Methods in MT. TMI-92*, Montréal, Québec.

TMI: 1993, *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation TMI '93: MT in the Next Generation*, Kyoto, Japan.

TMI: 1995, *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*, Leuven, Belgium.

TMI: 1997, *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation*, Santa Fe, New Mexico.

TMI: 1999, *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI 99)*, Chester, England.

Utsuro, T., K. Uchimoto, M. Matsumoto and M. Nagao: 1994, 'Thesaurus-Based Efficient Example Retrieval by Generating Retrieval Queries from Similarities', in Coling (1994), pp. 1044–1048.

Vauquois, B.: 1968, 'A Survey of Formal Grammars and Algorithms for Recognition and Transformation in Machine Translation', *IFIP Congress-68*, Edinburgh, 254–260; reprinted in Ch. Boitet (ed.) *Bernard Vauquois et la TAO: Vingt-cinq Ans de Traduction Automatique – Analectes*, pp. 201–213, Grenoble (1988): Association Champollion.

Veale, T. and A. Way: 1997, '*Gaijin*: A Bootstrapping Approach to Example-Based Machine Translation', *International Conference, Recent Advances in Natural Language Processing*, Tzigov Chark, Bulgaria, pp. 239–244.

Vogel, S., H. Ney and C. Tillmann: 1996, 'HMM-Based Word Alignment in Statistical Translation', in Coling (1996), pp. 836–841.

Wagner, R. A. and M. J. Fischer: 1974, 'The String-to-String Correction Problem', *Journal of the Association for Computing Machinery* **21**, 168–173.

Waibel, A., A. N. Jain, A. E. McNair, H. Saito, A. G. Hauptmann and J. Tebelskis: 1991, 'JANUS: A Speech-to-Speech Translation System Using Connectionist and Symbolic Processing Strategies', *Proceedings of the 1991 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Ontario, pp. 793–796.

Wang, Y-Y. and A. Waibel: 1995, ‘Connectionist F-structure Transfer’, in Mitkov & Nicolov (1995), pp. 393–404.

Wang, Y-Y. and A. Waibel: 1997, ‘Decoding Algorithm in Statistical Machine Translation’, *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, Madrid, Spain, pp. 366–372.

Watanabe, H.: 1992. ‘A Similarity-Driven Transfer System’, in Coling (1992), pp. 770–776.

Watanabe, H.: 1993. ‘A Method for Extracting Translation Patterns from Translation Examples’, in TMI (1993), pp. 292–301.

Watanabe, H.: 1994. ‘A Method for Distinguishing Exceptional and General Examples in Example-Based Transfer Systems’, in Coling (1994), pp. 39–44.

Watanabe, H.: 1995, ‘A Model of a Bi-Directional Transfer Mechanism Using Rule Combinations’, *Machine Translation* **10**, 269–291.

Watanabe, H. and K. Takeda: 1998, ‘A Pattern-Based Machine Translation System Extended by Example-Based Processing’, in Coling-ACL (1998), pp. 1369–1373.

Way, A.: 1999, ‘A Hybrid Architecture for Robust MT using LFG-DOP’, *Journal of Experimental and Theoretical Artificial Intellignece* **11**, 441–471.

Weaver, A.: 1988, ‘Two Aspects of Interactive Machine Translation’, in M. Vasconcellos (ed.), *Technology as Translation Strategy*, pp. 116–123, Binghamton, NY: State University of New York at Binghamton (SUNY).

Whyman, E. K. and H. L. Somers: 1999, ‘Evaluation Metrics for a Translation Memory System’, *Software—Practice and Experience* **29**, 1265–1284.

Yamabana K., Kamei S., Muraki K., Doi S., Tamura S. and Satoh K.: 1997, ‘A Hybrid Approach to Interactive Machine Translation – Integrating Rule-based, Corpus-based, and Example-based Method’, *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, Nagoya, Japan, pp. 977–982.

Zhang, K. and D. Shasha: 1997, ‘Tree Pattern Matching’, in A. Apostolico and Z. Galil (eds), *Pattern Matching Algorithms*, pp. 341–371, New York: Oxford University Press.

Zhao, G. and J. Tsujii: 1999, ‘Transfer in Experience-Guided Machine Translation’, *Machine Translation Summit VII*, Singapore, pp. 501–508.