

# What is Example-Based Machine Translation?

**Davide Turcato, Fred Popowich**

*gavagai* Technology Incorporated  
420-6450 Roberts Street  
Burnaby, British Columbia, Canada V5G 4E1  
{turk, popowich}@gavagai.net

## Abstract

We maintain that the essential feature that characterizes a Machine Translation approach and sets it apart from other approaches is the kind of knowledge it uses. From this perspective, we argue that Example-Based Machine Translation is sometimes characterized in terms of inessential features. We show that Example-Based Machine Translation, as long as it is linguistically principled, significantly overlaps with other linguistically principled approaches to Machine Translation. We make a proposal for translation knowledge bases that make such an overlap explicit.

## Introduction

In an excellent review article about Example-Based Machine Translation (EBMT), Harold Somers (1999) provides a comprehensive classification of the broad variety of MT research falling within the example-based paradigm, and makes an attempt at capturing the essential features that make an MT system an example-based one.

The present paper takes Somers' discussion as its starting point and tries to take further steps in answering the questions posed therein. We acknowledge at this point that we also draw heavily from Somers' paper in terms of citations of previous works in EBMT.

In the broad and diversified panorama of MT, we believe that this definition task, far from being a pedantic exercise, is an important step towards separating essential differences among MT approaches from inessential ones. This effort may lead to uncovering overlaps between approaches that at first sight seem quite far apart, or conversely it may bring to light significant differences between approaches that are superficially similar. We believe that a better understanding of the relations among different approaches provides valuable insight that can guide MT researchers in their decisions about further directions to take.

## Classification Criteria

In his apparently provisional conclusions about a definition of EBMT, Somers (1999) discusses three increasingly specific criteria for defining EBMT:

1. EBMT uses a bilingual corpus.
2. EBMT uses a bilingual corpus as its main knowledge base.
3. EBMT uses a bilingual corpus as its main knowledge base, at run-time.

Somers (1999) states that the first two criteria are too broad, but he argues that the third criterion may be too strict, as it rules out, for instance, statistical MT, where all the corpus-driven probabilities are computed in advance. While agreeing with Somers on the inadequacy of the first two criteria, we would like to suggest that the third criterion might also be too broad (disregarding here whether it is at the same time too strict for the reasons put forward by Somers). In the following sub-sections we discuss the proposed criteria.

## Implicit vs. Explicit knowledge

One of Somers' criteria is that an example database be used at run-time. As far as we can see, there are two reasons why a corpus is used at run-time in an MT system:

1. The system uses knowledge that can only be dynamically acquired at run-time by accessing an entire corpus, or sections of it whose extent cannot be determined in advance.
2. The system uses knowledge that could be extracted in advance, but is instead left implicit in the corpus, and extracted as needed at run-time.

We argue that only the former case is relevant to the above-mentioned criterion for characterising EBMT. We adopt here the software engineering perspective of separating processes from data, and we focus on data. Obviously, there may be reasons for preferring to leave knowledge implicit rather than making it explicit (in terms of efficiency, memory requirements, time/space trade-off, etc.). However, from the point of view of characterising an approach to MT, such as taken here, what we regard as essential is ascertaining what kind of knowledge a given approach uses, as opposed to whether the same body of knowledge is explicitly or implicitly encoded.

Analogously, once a body of knowledge is explicitly encoded, we regard the source from which it was acquired as secondary, for classification purposes. Of course, we do not intend to overlook the issue of knowledge acquisition, both in terms of cost-effectiveness and system coverage. However, we maintain that this issue is not crucial in classifying an MT system. To give an example coming from direct experience, in our English-Spanish lexicalist transfer system (Popowich et al., 1997) we initially handcrafted our bilingual lexicon. Subsequently, we developed tools for the automatic acquisition of relevant terms from corpora and for the automatic or semi-automatic generation of bilingual lexicons (Turcato et al., 2000a, and references therein). Although we obviously considered this a major achievement, nevertheless we did not feel that this made our MT system more example-based than it was before.

## Single vs. Multiple Knowledge Sources

Another criterion stated by Somers is that the example database be the main knowledge base used by a system. This is a point often emphasised in the EBMT literature. One of its corollaries is that a system's accuracy can be increased by simply adding more examples.

One preliminary remark is that accounts of EBMT systems tend sometimes to overlook or understate the use they make of other resources, besides an example database. For example, most EBMT systems assume the existence of a bilingual lexicon to perform substitutions in examples.

To give another example, one of the most characteristic operations performed in EBMT, the *similarity comparison*, is usually driven by a thesaurus. Of course, the availability of thesauri does not make an MT approach more data-driven or knowledge-free than if a thesaurus had to be specifically developed by hand. Moreover, such resources do not readily lend themselves to the porting of an MT system to a specific domain. Work on semantic databases like WordNet has shown that much of their information can be misleading in specific domains. For example, an MT system dealing with weather reports would have serious problems using a thesaurus where very frequent words like *snow* and *C* (for *Celsius*) were considered semantically similar because they are both synonyms for *cocaine* (Turcato et al., 2000b).

Finally, several other kinds of linguistic processing are performed in EBMT, ranging from named entity recognition (Brown, 1999), morphological analysis, and tagging to full parsing. In other words, most of the linguistic processing techniques used in conventional MT systems have been proposed in EBMT systems. Each of these processes requires some sort of linguistic resource.

A further and perhaps more crucial remark is that EBMT approaches tend to use the same resource (i.e. an example database) for different purposes, while traditional MT systems tend to use different resources. E.g., target sentences are used as sentence template for the recombination task, via some kind of substitution. The recombination task parallels generation, for which many other systems used specific target grammars. Analogously to what is argued about the implicit vs. explicit encoding of knowledge, we maintain that what is relevant here is to ascertain what kind of knowledge is used for the recombination/generation task. Again, there may be practical arguments for preferring a single resource to separate resources, or vice versa. However, what affects the output of a system is the knowledge used for a task, not the integration vs. segregation of this knowledge with respect to other knowledge sources.

### A Declarative Classification Criterion

To sum up, in classifying an MT approach, we believe it is useful to separately ask the following three questions:

1. What linguistic information is used?
2. Where is linguistic information acquired from?
3. When is linguistic information acquired?

We claim that only the first question should be the primary focus of a classification, while the differences in terms of the other two questions should be regarded as secondary. In other words, we suggest a declarative criterion that looks at the knowledge being used, rather than at the processes used to obtain that knowledge.

We illustrate this point with an example. In presenting their *Gaijin* EBMT system, Veale & Way (1997:239) claim that “the only linguistics employed by *Gaijin* is a psycholinguistic constraint – the marker hypothesis”. However, in describing the system they explain that they use a bilingual lexicon, statistically constructed by corpus word alignment. They also explain that “*Gaijin* employs

corpus-based statistics not as a translation strategy in themselves, but as a basis for inferring symbolic transfer rules” (p. 240). So, in answering our first question, we would say that *Gaijin* uses a bilingual lexicon and a set of symbolic transfer rules, besides knowledge about phrase markers. In addition, one might question whether the mere fact that *Gaijin* acquires its linguistic information from a corpus makes it an EBMT system.

Arguably, two approaches can be regarded as one and the same approach if they use the same knowledge in the same way, regardless of whether such knowledge is extracted in advance or at run-time, from a corpus or from other resources, or whether it is distributed over one or several resources. All variants ultimately behave in the same way.

To look at the same issue from a slightly different perspective, we tend to consider two systems that perform full syntactic analysis more similar among themselves, regardless of whether this information is encoded in a grammar or a tree-bank, than each of them is to a system that only performs, say, morphological analysis.

### EBMT Re-assessment

Equipped with this criterion that prioritises the knowledge content of a system over the way knowledge is expressed or acquired, we turn to a tentative review of EBMT based on this criterion. Our goal is not so much to give an absolute definition of what ‘true’ EBMT is, but rather to assess the proximity of approaches that look superficially distant, or conversely to assess the distance between systems that look superficially similar. We will draw a comparison between linguistically principled EBMT system and other linguistically principled approaches to MT. Our comparison will sometimes linger on a specific transfer approach, the lexicalist variant, particularly when the discussion concerns translation selection. This preference is contingently motivated by the fact that lexicalist transfer is the symbolic approach we are most familiar with. However, we think that such a comparison bears a general significance, for the following reasons:

1. Lexicalist transfer, in its turn, considerably overlaps with other symbolic approaches, not limited to the class of transfer approaches.
2. If we show that EBMT (or a subclass of it) is equivalent to (or significantly overlapping with) a symbolic approach that is customarily characterised in terms that make no reference to example bases, then we are led to either:
  - a) draw the somehow paradoxical conclusion that EBMT can be characterised in terms that make no reference to example bases; or
  - b) conclude that the characterisation of example-based approaches rests more upon knowledge acquisition aspects than upon knowledge usage aspects. This, in turn, suggests that the example-based characterisation could be reformulated as a transversal distinction between data-driven vs. theory-driven approaches that would cut across multiple linguistically principled approaches, rather than being a separate approach on a par with the others. So, one could imagine a data-driven lexicalist transfer vs. a theory-driven lexicalist transfer, a data-driven structural transfer vs. a theory-driven structural transfer, etc., depending on how knowledge is acquired.

We finally note that in our review we will try to leave out systems that are explicitly claimed to be hybrid by their authors, narrowing down our scope to systems that are claimed to be variants of the EBMT approach.

### Non-symbolic EBMT

A fundamental distinction exists between systems that use linguistic knowledge and systems that do not. Statistical MT falls in the latter class. We leave open the issue whether statistical MT is a kind of EBMT. In any case, it is clear that the methods of such approaches set them apart as much from linguistically principled EBMT as from other kinds of linguistically principled MT.

We illustrate this point by discussing a sample statistical MT system, the French-English system Candide (Berger et al., 1994). Candide is described in terms of three components resembling the traditional partition of transfer MT: analysis, transfer, and synthesis. Analysis maps source French sentences onto what the authors call “intermediate French”, i.e. normalised representations of sentences. Normalisation consists of case and spelling correction, name and number detection, segmentation, morphological analysis and word reordering. Conversely, synthesis maps “intermediate English” representations onto target sentences. Although these components do use some linguistic knowledge, the knowledge they use is low-level. Their task is to normalise the input and the output, rather than to add any linguistic knowledge to be used in transfer. The input and the output of transfer are just strings, and transfer is purely statistical. Its task is described as decoding an English sentence that was transmitted over a noisy communication channel, which “corrupted” it to a French sentence.

Transfer uses two knowledge sources: a *language model* of English, and a *translation model*. A language model is used to assign probabilities to English sequences of words, and it simply consists of a set of trigrams (i.e. triples of English words), with associated probabilities (i.e. numeric values). No other knowledge about words is used. A translation model is used to compute the conditional probability of an English sentence, given a French sentence and an alignment (an exhaustive mapping of word positions between two sentences). Several translation models are proposed. The simplest consists of a set of word translation probabilities (i.e. pairs of English and French words, to which numeric values are assigned). The other translation models are more refined in terms of more sophisticated parameter estimations, but none of them makes use of any additional linguistic knowledge.

Summing up, besides the result of alignment, which can be considered a word-based probabilistic bilingual lexicon (analogously to what was previously discussed for the Gaijin system), it can be seen here that transfer uses none of the other linguistic knowledge sources (either monolingual or bilingual) used by linguistically principled EBMT or other symbolic approaches.

### Symbolic EBMT

The present section expands and generalizes some remarks we made in (Turcato et al., 1999), where we attempted a comparison of different kinds of EBMT systems with lexicalist transfer MT. In that paper we remarked, as also mentioned above, that in EBMT an example database is used for different purposes at the same time: as a source of

sentence frame pairs, and as a source of sub-sentential translation pairs. Accordingly, Somers (1999) points out that EBMT comprises three phases (*matching*, *alignment*, and *recombination*), and draws a parallel with analogous phases in traditional transfer MT systems (*analysis*, *transfer*, and *generation*). Sentence frames are used in matching and recombination as aligned basic sentence structures to be combined with aligned sub-sentential translation pairs. In the following two sections we discuss the two key operations of EBMT, *sentence decomposition* and *translation selection*.

### Sentence Decomposition

Given the obvious fact that exact match would be too strict a requirement, a match between dissimilar sentences is generally obtained by decomposing them into some kind of constituents, in order to perform a partial match, in which dissimilar parts are substituted.

Proposals about how to decompose examples vary considerably. In most cases, some sort of linguistic analysis is used. The range of proposed techniques includes, just to name a few: segmenting sentences using markers as segment boundaries (Veale & Way, 1997), performing named entity recognition to obtain more abstract examples (Brown, 1999), using morphologically analyzed segments (Kitano, 1993), using tagged sequences as fragments (Somers et al., 1994), parsing sentences into dependency trees (Sato & Nagao, 1990). In brief, the whole range of available linguistic techniques is used for this purpose. In some cases it is proposed to explicitly store generalized examples (Furuse & Iida, 1992), in other cases examples are decomposed on the fly. In any case, the idea behind all proposals is that examples can be decomposed into smaller constituents to be processed independently.

While there is obviously a remarkable difference in analyzing power between performing a simple detection of segment boundaries or morphological analysis and performing full parsing, the common idea of decomposing sentences into constituents resembles the idea that underlies grammars. In fact, each of the proposed techniques for decomposing sentences parallels some corresponding kind of grammar in a conventional MT system. In each case, the knowledge used for decomposition could be explicitly stored as a separate set of rules. For example, in (Toole et al., 1999) we describe a grammar in terms of a flat list of tag sequences, used to cover a segmented input sentence.

In brief, we would like to argue that some of the techniques used in EBMT for sentence decomposition are farther apart from each other than each of them is from the corresponding technique used (or usable) in a conventional transfer system. This latter similarity would stand out more clearly if sentence templates were separately stored, as proposed in some EBMT systems, instead of being extracted on-the-fly. It is not apparent that there is any compelling reason ruling out this option.

We make a final remark about the advantage of easily handling structural mismatches, ascribed to EBMT. Although this is certainly true of EBMT, since no recursive transfer is performed, this is a property that EBMT shares with several other proposals not only in transfer approaches, such as lexicalist MT or semantic

transfer (Dorna et al., 1998), but also in the interlingua approach, such as proposed by Traum & Habash (2000).

### Translation Selection

Given that exact match of complete examples is the exception rather than the rule, and translation is usually done by decomposing a sentence into constituents, two questions arise: (i) to what extent the availability of complete examples is needed in translating constituents? (ii) how does EBMT translation of constituents differ from other symbolic approaches to MT (e.g. lexicalist transfer)? We discuss the two issues by putting forward a proposal for a translation knowledge base, then discussing how such a knowledge base would suit EBMT.

Paraphrasing the definition of grammars as descriptions of infinite sets of sentences by finite means, we can analogously define the translation task as the description of an infinite set of equivalencies by means of a finite set of equivalencies.

The translation task can be defined in terms of two fundamental properties:

1. Translation is *compositional*. The translation of an expression is a function of the translation of its constituents. E.g. the Spanish translation of *business trip* (*viaje de negocios*) is a function of the translations of *business* (*negocios*) and *trip* (*viaje*), when these appear in isolation. In turn, the translation of *a long business trip* (*un viaje de negocios largo*) is a function of the translations of *a* (*un*), *long* (*largo*) and *business trip*.
2. Translation is *non-monotonic*. In specific context, compositionality holding for narrower contexts is reversed. E.g. the Spanish translation of *field trip* (*viaje de estudio*) is not a function of the translations of *field* (*campo*) and *trip*, when these appear in isolation. However, the translation of *a long field trip* (*un viaje de estudio largo*) is indeed a function of the translations of *a* (*un*), *long* (*largo*) and *field trip*.

Accordingly, the knowledge that a bilingual knowledge base should contain can be declaratively stated as follows:

1. A repository of basic translation equivalencies, i.e. a bilingual lexicon of word-to-word equivalencies. This bilingual lexicon would define the base step in a recursive, compositional translation process. E.g.  
*business* ↔ *negocios*  
*viaje* ↔ *trip*  
*field* ↔ *campo*  
*long* ↔ *largo*
2. A repository of phrases that do not translate compositionally. This repository would define all and only translation ‘turning points’, which violate the monotonicity of compositional translation. E.g.  
*field trip* ↔ *viaje de estudio*

Note that the phrases in this repository could in turn be part of a compositional translation.

This proposed distinction resembles the distinction between *general examples* and *exceptional examples* proposed by Nomyiama (1992), cited by Somers (1999:121). The sum of the two knowledge bases would account for any possible translation. In other words, it would be a finite set of equivalencies that account for an infinite set of translations.

The repository of non-monotonic contexts would serve one of the main purposes of example databases, i.e.

identifying the appropriate translation of a particular word or phrase, based on its context. However, it would be more specific and explicit in two ways:

1. It would specify minimal contexts, i.e. just enough context as necessary to trigger a non-compositional translation. Consider an example database like the following:  
*I read an article about your business trip* ↔ *Leí un artículo sobre tu viaje de negocios*  
*I gave up my field trip* ↔ *Renuncié a mi viaje de estudio*  
An input sentences like *I read an article about your field trip* would probably match the first example, because of irrelevant context, and wrongly replace the translation of *business* with the translation of *field*.
2. It would eliminate redundancies (or, in EBMT terms, *example interference*). Each context would be specified only once.

Of course, one might argue that a knowledge base of this kind is a significant departure from an example database, and requires extra work. This may well be true. However, the point we are trying to make is that there is no inherent reason why example decomposition could not be done off-line and explicitly expressed in the knowledge base. If an example database defines a finite set of equivalencies, then this set of equivalencies can be explicitly stated (with each equivalence stated only once).

A knowledge base as we propose would be compatible with different approaches to MT. It would equally be equivalent to an EBMT example database (under some specified conditions), and to a lexicalist transfer bilingual lexicon. Declaratively stated, the linguistic information used by EBMT is indistinguishable from the information used by lexicalist MT. The evidence explicitly stated in a knowledge base is the same for the two approaches. The same knowledge base can be used by both.

Where would EBMT and lexicalist MT differ then? If their knowledge base was complete, they would not differ. Where the two perspectives would differ is in dealing with incompleteness. A lexicalist, bottom up approach, emphasizes compositionality, while an example-based, top down approach, emphasizes non-monotonicity. The former is more likely to miss relevant contexts that override compositionality. The latter is more likely to miss lexical generalizations that hold across different contexts. However, the evidence provided by the way either approach deals with incompleteness is a much weaker argument than the evidence either approach could put forward if they used different kinds of knowledge.

### Translation by analogy

We have intentionally postponed the discussion of translation by analogy until after the proposal put forward in the previous section. A classical example of translation by analogy is the one discussed by Nagao (1984), who shows a method for translating new sentences, based on their similarity with available examples. Similarity is measured by the distance of words in a thesaurus (although other methods could be devised). For instance, Nagao shows how the two English-Japanese examples

*A man eats vegetables* ↔ *Hito wa yasai o taberu*  
*Acid eats metal* ↔ *San wa kinzoku o okasu*

can be used to translate the new sentence

*He eats potatoes*

provided that a bilingual lexicon at the word level is available. The problem here is to choose one of two competing translations for *eat* (*taberu* vs. *okasu*). In Nagao's approach, the translation *taberu* is correctly selected, based on the greater semantic similarity of *he* and *potatoes* with *man* and *vegetables*, respectively, than with *acid* and *metal*. Conversely, the occurrence of *eat* in

*Sulfuric acid eats iron*

is correctly translated by *okasu*, based on the greater semantic similarity of *sulfuric acid* and *iron* with *acid* and *metal*, respectively, than with *man* and *vegetables*.

If we rephrase the translation selection problem in the terms described in the previous section, assuming one of the two translations as the "default" translation for *eat* (most likely *taberu*, which translates the more literal sense), the task is to identify the non-monotonic contexts that require a different translation from the default one. In doing this, it would be desirable to find the minimal context that triggers a given translation, so as to use it in a broader range of cases. It turns out that things stand differently, depending on whether exact match of contexts or match by analogy is performed.

When exact match of contexts is performed, an example (e.g. *a man eats vegetables*) only accounts for itself, i.e. it is only used when the same example, or a part of it, is input (e.g. *a man eats...* or *...eats vegetables*). In other words, given a set of examples, it is known in advance what input sentences they can be useful for. In this situation it is conceivable to use a contrastive method to reduce a set of overlapping examples to a set of minimal local contexts accounting for different translations of the same term. E.g. the set of examples

*A man eats vegetables* ↔ *Hito wa yasai o taberu*  
*Acid eats metal* ↔ *San wa kinzoku o okasu*  
*He eats potatoes* ↔ *Kare wa jagaimo o taberu*  
*Sulfuric acid eats iron* ↔ *Ryūsan wa tetsu o okasu*

could be reduced to the following set of minimal contexts (omitting function words for simplicity):

*man & eat* ↔ *hito & taberu*  
*acid & eat* ↔ *san & okasu*  
*he & eat* ↔ *kare & taberu*  
*sulfuric acid & eat* ↔ *ryūsan & okasu*

This would be done on the basis of a generalization that ascribes the selection of one or the other translation of *eat* to its subject (alternatively, one could choose to generalize over the objects, or perhaps to do no generalization). The set of contexts can be further reduced by identifying a default translation for *eat* and only explicitly encoding non-monotonic contexts:

*eat* ↔ *taberu*  
*acid & eat* ↔ *san & okasu*  
*sulfuric acid & eat* ↔ *ryūsan & okasu*

When contexts are matched by analogy, a given example (e.g. *a man eats vegetables*) may account not only for itself, but also for previously unseen contexts (e.g. *he eats potatoes*). Conversely to that discussed for exact matches,

given a set of examples, it is not known in advance what input sentences the examples can be useful for. In principle, an example might be used for an input sentence with no words in common, as long as there is a term-to-term semantic similarity between the example and the input sentence. Given an example like *a man eats vegetables*, how does one determine in advance what is the relevant context for an unforeseen translation? *A man eats* or *eats vegetables*? Or the two together? Because of the incompleteness assumption, extracting local contexts from examples becomes problematic. Given a set of examples, one can know what local contexts would adequately cover the examples at hand, but one cannot know whether such contexts would be adequate for all possibly relevant sentences. For instance, a sentence like *moths eat holes in clothes* may require that a larger context be taken into account, for a more informed similarity assessment. Therefore, it is advisable to have the largest possible contexts available (i.e. entire examples), so as to be able to use different portions of them depending on the input sentence to be translated.

A traditional counterpart of this selection mechanism would be to abstractly encode contexts by means of some sort of semantic features that would act as selectional restrictions. E.g. the distinction between the two senses of *eat* (respectively translated by *taberu* and *okasu*) might be captured by a [ $\pm$  animate] feature, which would select the appropriate subject. This approach would make a database of explicit contexts superfluous, and would also lend itself to a similarity-based approach, if selectional restrictions were used as preferences rather than hard constraints. However, this approach would be labor-intensive and, as Somers (1999:127) points out, it would be "cumbersome and error-prone".

From this informal discussion it appears that translation by analogy, which is the most characteristic technique of EBMT, is also the one where the use of entire examples is most motivated. As a final remark, we only note that an example database for the purpose of translating by analogy can be an additional resource to whatever other resources are used, along the lines discussed above. In principle, translation by analogy could also be an extension to a traditional transfer MT system, to solve cases of lexical ambiguity for which no direct evidence is found in a translation database.

## Conclusions

Classifications of EBMT systems often tend to bring to the foreground the fact that such systems use a bilingual corpus as the source of their linguistic knowledge, relegating to the background the assessment of what linguistic knowledge a system actually uses.

We propose a classification criterion that reverses the terms of the question, focusing primarily on the linguistic knowledge used by a system, and considering the source of this knowledge or the format in which it is represented as secondary. Accordingly, we make a distinction between the use of a corpus as: (i) a mere source for knowledge acquisition; (ii) a mere way of storing knowledge in a compact form; (iii) a genuine repository of knowledge that the system needs to access, and that could not be easily stored in other forms. A preliminary application of this criterion to linguistically principled EBMT approaches shows that:

- An example database is usually only one of the resources used by an EBMT system.
- The allegedly holistic approach of EBMT can be equivalently decomposed in a series of distinct knowledge sources and related tasks.
- The amount of linguistic knowledge required for any given component varies considerably for different approaches, ranging from low-level information like part of speech to full syntactic information.
- For most components a counterpart can be found in more conventional approaches to MT that performs an equivalent task without the need to access a database of full examples.
- The original idea of translation by analogy stands out as truly example based.

We feel that clarifying the linguistic knowledge required by an MT approach and making the overlap with other approaches explicit is important. MT is a complex task that requires a vast amount of knowledge, linguistic and possibly extra-linguistic, to be performed adequately. We feel there is no shortcut to overcome this requirement. Hence it is important for MT practitioners to emphasize commonalities among different approaches, and put some effort towards integrating and sharing resources.

### References

- Berger, A. L., Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., Gillett, J. R., Lafferty, G. D., Mercer, R. L., Printz, H. & Ureš, L. (1994). The Candide System for Machine Translation. In Proceedings of the 1994 ARPA Workshop on Human Language Technology, Plainsboro, NJ, USA.
- Brown, R. D. (1999). Adding Linguistic Knowledge to a Lexical Example-based Translation System. In Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (pp. 22--32). Chester, UK.
- Dorna, M., Frank, A., van Genabith, J. & Emele, M. C. (1998). Syntactic and Semantic Transfer with F-Structures. In Proceedings of the 36<sup>th</sup> Annual Meeting of the Association for Computational Linguistics and 17<sup>th</sup> International Conference on Computational Linguistics (pp. 341—347). Montréal, Canada.
- Furuse, O. & Iida, H. (1992). Cooperation between transfer and analysis in example-based framework. In Proceedings of the Fifteenth [sic] International Conference on Computational Linguistics (pp. 645--651). Nantes, France.
- Kitano, H. (1993). A comprehensive and practical model of memory-based Machine Translation. In Proceedings of the 13th International Joint Conference on Artificial Intelligence (pp. 1276--1282). Chambéry, France.
- Nagao, M. (1984). A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. In A. Elithorn & R. Banerji (Eds.), *Artificial and Human Intelligence: edited review papers at the International NATO Symposium on Artificial and Human Intelligence sponsored by the Special Programme Panel held in Lyon, France October, 1981* (173--180). Amsterdam, North-Holland, Elsevier Science Publishers.
- Nomyama, H. (1992). Machine Translation by Case Generalization. In Proceedings of the Fifteenth [sic] International Conference on Computational Linguistics (pp. 714--720). Nantes, France.
- Popowich, F., Turcato, D., Laurens, O., McFetridge P., Nicholson, J. D., McGivern, P., Corzo-Pena, M., Pidruchney, L. & McDonald, S. (1997). A Lexicalist Approach to the Translation of Colloquial Text. In Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation (pp. 76--86), Santa Fe, New Mexico, USA.
- Sato, S. & Nagao, M. (1990). Toward memory-based translation. In Proceedings of the 13th International Conference on Computational Linguistics (vol. 3, pp. 247--252). Helsinki, Finland.
- Somers H. (1999). Review Article: Example-based Machine Translation. *Machine Translation*, 14(2), 113--157.
- Somers, H., McLean, I. & Jones, D. (1994). Experiments in Multilingual Example-Based Generation. In Proceedings of the 3<sup>rd</sup> Conference on the Cognitive Science of Natural Language Processing. Dublin, Ireland.
- Toole, J., Popowich, F., Nicholson, D., Turcato, D. & McFetridge, P. (1999). Explanation-Based Learning for Machine Translation. In Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (pp. 161--171). Chester, UK.
- Traum, D. & Habash, N. (2000). Generation from Lexical Conceptual Structures. In Proceedings of the Third Workshop on Interlinguas and Interlingual Approaches. Seattle, Washington, USA.
- Turcato, D., McFetridge, P., Popowich, F. & Toole, J. (1999). A Unified Example-Based and Lexicalist Approach to Machine Translation. In Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (pp. 33--43). Chester, UK.
- Turcato, D., Popowich, F., McFetridge, P. & Toole, J. (2000a). Creating high-quality, large-scale bilingual knowledge bases using minimal resources. In Proceedings of the LREC-2000 Workshop on 'Developing Language Resources for Minority Languages: Reusability and Strategic Priorities' (pp. 92--99). Athens, Greece.
- Turcato, D., Popowich, F., Toole, J., Fass, D., Nicholson, D. & Tisher, G. (2000b). Adapting a synonym database to specific domains. In Proceedings of the ACL'2000 Workshop on Recent Advances in Natural Language Processing and Information Retrieval, Hong Kong.
- Veale, T. & Way, A. (1997). Gaijin: A Bootstrapping Approach to Example-Based Machine Translation. In Proceedings of the International Conference 'Recent Advances in Natural Language Processing' (239—244). Tzigov Chark, Bulgaria.