Learning Translation Templates from Bilingual Texts[1]

H. Altay Güvenir and Ilyas Cicekli

Department of Computer Engineering and Information Sciences
Bilkent University, Ankara 06533, Turkey
e-mail: {guvenir,ilyas}@cs.bilkent.edu.tr

Abstract

This paper proposes a mechanism for learning structural correspondences between two languages from a corpus of translated sentence pairs. The proposed mechanism uses analogical reasoning between two translations. Given a pair of translations, the similar parts of the sentences in the source language must correspond the similar parts of the sentences in the target language. Similarly, the different parts should correspond to the respective parts in the translated sentences. The correspondences between the similarities, and also differences are learned in the form of translation templates. The system is tested on a small training dataset and produced promising results for further investigation.

1 Introduction

Traditional approaches to machine translation (MT) suffer from tractability, scalability and performance problems due to the necessary extensive knowledge of both the source and the target languages. Corpus-based machine translation is one of the alternative directions that have been proposed to overcome the difficulties of traditional systems. Two fundamental approaches in corpus-based MT have been followed. These are *statistical* and *example-based* machine translation (EBMT), also called *memory-based* machine translation (MBMT). Both approaches assume the existence of a bilingual parallel text (an already translated corpus) to derive a translation for an input. While statistical MT techniques use statistical metrics to choose the most probable structures in the target language, EBMT techniques employ pattern matching techniques to translate subparts of the given input [1].

Exemplar-based representation has been widely used in Machine Learning (ML). According to Medin and Schaffer [7], who originally proposed exemplar-based learning as a model of human learning, examples are stored in memory without any change in the representation. The characteristic examples stored in the memory are called exemplars. The basic idea in exemplar-based learning is to use past experiences or cases to understand, plan, or learn from novel situations [4, 6, 10].

EBMT has been proposed by Nagao [8] as *Translation by Analogy* which is in parallel with memory based reasoning [14], case-based reasoning [11] and derivational analogy [2]. Example-based translation relies on the use of past translation examples

[1]This research has been supported in part by NATO Science for Stability Program Grant TU-LANGUAGE.

to derive a translation for a given input [3, 9, 12, 13, 15]. The input sentence to be translated is compared with the example translations analogically to retrieve the *closest* examples to the input. Then, the fragments of the retrieved examples are translated and recombined in the target language. Prior to the translation of an input sentence, the correspondences between the source and target languages should be available to the system; however this issue has not been given enough consideration by the current EBMT systems. Kitano has adopted the manual encoding of the translation rules, however this is a difficult and an error-prone task for a large corpus [5]. In this paper, we formulate this acquisition problem as a machine learning task in order to automate the process.

In this paper, we propose a technique which stores exemplars in the form of *templates* that are *generalized exemplars*. A template is an example translation pair where some components (e.g., words stems and morphemes) are generalized by replacing them with variables in both sentences, and establishing bindings between the variables. We will refer this technique as GEBMT for *Generalized Example Based Machine Translation*.

The algorithm we propose here, for learning such templates, is based on a heuristic to learn the correspondences between the patterns in the source and target languages, from two translation pairs. The heuristic can be summarized as follows: Given two translation pairs, if the sentences in the source language exhibit some similarities, then the corresponding sentences in the target language must have similar parts, and they must be translations of the similar parts of the sentences in the source language. Further, the remaining differing constituents of the source sentences should also match the corresponding differences of the target sentences. However, if the sentences do not exhibit any similarity, then no correspondences are inferred. Consider the following translation pair given in English and Turkish to illustrate the heuristic:

I gave the ticket to Mary $\leftrightarrow$ Mary'e bileti verdim
I gave the pencil to Mary $\leftrightarrow$ Mary'e kurşun kalemi verdim .

Similarities between the translation examples are shown as underlined. The remaining parts are the differences between the sentences. We represent the similarities in the source language as [I gave the X^S to Mary], and the corresponding similarities in the target language as [Mary'e X^T+i verdim]. According to our heuristic, these similarities should correspond each other. Here, X^S denotes a component that can be replaced by *any* appropriate structure in the source language and X^T refers to its translation in the target language. This notation represents an *abstraction* of the differences "ticket" vs. "pencil" and "bilet" vs. "kurşun kalem" in the source and target languages, respectively. Continuing even further, we infer that "ticket" should correspond to "bilet" and "pencil" should correspond to "kurşun kalem"; hence learning further correspondences between the examples.

Our learning algorithm based on this heuristic is called TTL (for *Translation Template Learner*). Given a corpus of translation cases, TTL infers the correspondences between the source and target languages in the form of templates. These templates can be used for translation in both directions. Therefore, in the rest of the paper we will refer these languages as L^1 and L^2. Although the examples and experiments herein are on English and Turkish, we believe the model is equally applicable to other language pairs.

The rest of the paper is organized as follows. Section 2 explains the representation in the form of translation templates. The TTL algorithm is described in Section 3.

Section 4 illustrates the TTL algorithm on some example translation pairs. Section 5 describes how these translation templates can be used in translation. Section 6 concludes the paper.

2 Translation Templates

A template is a generalized translation exemplar pair, where some components (e.g., words stems and morphemes) are generalized by replacing them with variables in both sentences, and establishing bindings between these variables. For example, the translation template that would be learned from the example translations given above is:

$$\begin{aligned} \texttt{[I gave the } X^1 \texttt{ to Mary]} &\leftrightarrow \texttt{[Mary'e } X^2\texttt{+i verdim] if} \\ [X^1] &\leftrightarrow [X^2] \ . \end{aligned}$$

This translation template is read as the sentence "I gave the X^1 to Mary." in L^1 and the sentence "Mary'e X^2+i verdim." in L^2 are translations of each other, where X^1 in L^1 and X^2 in L^2 are translations of each other. Therefore, for example, if it has already been acquired that "basket" in L^1 and "sepet" in L^2 are translations of each other, i.e., `[basket]` $\leftrightarrow$ `[sepet]` then the sentence "I gave the basket to Mary." can be translated into L^2 as "Mary'e sepeti verdim."

Since the TTL algorithm is based on finding the similarities and differences between translation examples, the representation of sentences plays an important role. As it is, the TTL algorithm may use the sentences exactly as they can be found in a regular text. That is, no grammatical information or no preprocessing, e.g., bracketing, on the bilingual parallel corpus is needed. Therefore, it is a grammarless extraction algorithm for phrasal translation templates from bilingual parallel texts.

For agglutinative languages such as Turkish, this surface level representation limits the generality of the templates to be learned. For example, the translation of the sentence "I am coming" in Turkish is a single word "geliyorum." When surface level representation is used, it is not possible to find a template from that translation and "I am going" $\leftrightarrow$ "gidiyorum." Therefore, we will represent a word in its lexical level representation, that is its stem and its morphemes. For example, the translation pair "I am coming" $\leftrightarrow$ "geliyorum" will be represented as

$$\texttt{I am come+ing} \leftrightarrow \texttt{gel+Hyor+yHm} \ .$$

Here, the letter "H" in the morphemes represents a vowel whose surface level realization is realized according to vowel harmony rules of the Turkish language. According to this representation, the first two translation pairs would be given as

$$\begin{aligned} &\texttt{I give+p the ticket to Mary} \leftrightarrow \texttt{Mary'e bilet+yH ver+DH+m} \\ &\texttt{I give+p the pencil to Mary} \leftrightarrow \texttt{Mary'e kurşun kalem+yH ver+DH+m} \ . \end{aligned}$$

The translation template learned is

$$\begin{aligned} \texttt{[I give+p the } X^1 \texttt{ to Mary]} &\leftrightarrow \texttt{[Mary'e } X^2\texttt{+yH ver+DH+m] if} \\ [X^1] &\leftrightarrow [X^2] \ . \end{aligned}$$

This representation allows an abstraction over technicalities such as vowel and/or consonant harmony rules, as in Turkish and also, different realizations of the same verb according to tense, as in English. We assume that the generation of surface level representation of words from their lexical level representations is trivial.

3 Learning Translation Templates

The TTL algorithm infers translation templates using similarities and differences between a pair of translation examples (E_i, E_j) from a bilingual parallel corpus. Formally, a translation example $E_i \ : \ E_i^1 \leftrightarrow E_i^2$ is composed of two sentences, E_i^1 and E_i^2, that are translations of each other in L_1 and L_2, respectively.

Given a pair of translation examples (E_i, E_j), we try to find similar constituents between E_i and E_j. A sentence is considered as a sequence of lexical items (i.e., words or morphemes). If no similarities can be found, then no templates from this examples is learned. If there are similar constituents then a *match sequence* in the following form is generated.

$$S_0^1, D_0^1, S_1^1, \cdots, D_{n-1}^1, S_n^1, \leftrightarrow S_0^2, D_0^2, S_1^2, \cdots, D_{m-1}^2, S_m^2 \text{ for } 1 \ \leq \ n, \ m.$$

Here, S_k^1 represents a similarity (a sequence of common items) between E_i^1 and E_j^1. Similarly, $D_k^1 \ : \ (D_{i,k}^1, \ D_{j,k}^1)$ represents a difference between E_i^1 and E_j^1, where $D_{i,k}^1$ and $D_{j,k}^1$ are non-empty differing items between two similar constituents S_k^1 and S_{k+1}^1. Corresponding differences do not contain common items. That is, for a difference D_k, $D_{i,k}$ and $D_{j,k}$ do not contain any common item. Also, no lexical item in a similarity S_i appear in any previously formed difference D_k for $k \leq i$. Any of S_0^1, S_n^1, S_0^2 or S_m^2 can be empty, however, S_i^1 for $0 < i < n$ and S_j^2 for $0 < j < m$ must be non-empty. There exists either a unique match or no match between a pair of translation examples.

For instance, the match sequence obtained for the translation examples given above is

```
I give+p the (ticket, pencil) to Mary ↔
Mary'e (bilet, kurşun kalem)+yH ver+DH+m
```

That is, S_0^1= "`I give+p the`", D_0^1=("`ticket`", "`pencil`"), S_1^1= "`to Mary`", S_0^2= "`Mary'e`", D_0^2=("`bilet`", "`kurşun kalem`"), S_1^2= "`+yH ver+DH+m`".

If there exist only single differences in both sides of a match sequence, e.i., $n \ = m \ = \ 1$, then these differing constituents must be translations of each other. Therefore, from the match sequence given above the following translation template can be inferred:

$$\begin{aligned} \texttt{[I give+p the } X^1 \texttt{ to Mary]} \ &\leftrightarrow \texttt{[Mary'e } X^2\texttt{+yH ver+DH+m] if} \\ \texttt{[}X^1\texttt{]} \ &\leftrightarrow \texttt{[}X^2\texttt{]} \ . \end{aligned}$$

If, on the other hand, the number of differences are equal on both sides, but more than one, e.i., $1 \ \leq \ n, \ m$, without prior knowledge, it is impossible to determine which differences in one side correspond to which differences on the other side. Therefore, learning depends on previously acquired translation templates. For example, the following translation examples have two differences on both sides.

```
    I give+p the book  ↔ Kitab+yH ver+DH+m
You give+p the pencil  ↔ Kurşun kalem+yH ver+DH+n .
```

Without prior information, we cannot determine if `I` corresponds to `Kitab` or `+m`. However, if it has already been learned that `i` corresponds to `+m` and `You` corresponds to `+n`, then the following three translation templates can be inferred:

procedure TTL(*Training_Set*)
begin
for each pair of translation examples E_i and E_j in *Training_Set* **do**
Let the match sequence be
$M_{i,j} = S_0^1, D_0^1, \cdots, D_{n-1}^1, S_n^1, \leftrightarrow S_0^2, D_0^2, \cdots, D_{m-1}^2, S_m^2$.
if $n = m = 1$ **then** generate the follwoing rules:
$[S_0^1\ X^1,\ S_1^1] \leftrightarrow [S_0^2\ X^2,\ S_1^2]$ `if`
$[X^1] \leftrightarrow [X^2]$,
$[D_{0,i}^1] \leftrightarrow [D_{0,i}^2]$ and
$[D_{0,j}^1] \leftrightarrow [D_{0,j}^2]$.
else if $1 \leq n = m$ **and** for all differences in $M_{i,j}$ except possibly one, D_k^1 and D_l^2 the differences can be reduced **then**
generate the follwoing rules:
$[S_0^1 \cdots X^1,\ S_n^1] \leftrightarrow [S_0^2 \cdots X^2,\ S_m^2]$ `if`
$[X^1] \leftrightarrow [X^2]$,
$[D_{k,i}^1] \leftrightarrow [D_{l,i}^2]$ and
$[D_{k,j}^1] \leftrightarrow [D_{l,j}^2]$.
end.

Figure 1: The TTL algorithm.

$[X_1^1$ `give+p the` $X_2^1] \leftrightarrow [X_2^2$`+yH ver+DH`$X_1^2]$ `if`
$[X_1^1] \leftrightarrow [X_1^2]$ `and`
$[X_2^1] \leftrightarrow [X_2^2]$
and
`[book]` $\leftrightarrow$ `[kitab],`
`[pencil]` $\leftrightarrow$ `[kurşun kalem]`

In general, when the number of differences in both sides of a match sequences is greater than 1, e.i., $1 \leq n = m$, the TTL algorithm learns new translation templates only if at least $n - 1$ of the differences have already been learned. Otherwise, the current version of the algorithm cannot learn new rules. A formal description of the TTL algorithm is summarized in Fig. 1.

4 Examples

In order to evaluate the TTL alogorithm we have implemented it in PROLOG and tested on a sample bilingual parallel text. In this section, we will illustrate the behavior of TTL on that sample text.

Example 1: Given the example translations "I saw you at the garden" $\leftrightarrow$ "Seni bahçede gördüm" and "I saw you at the party" $\leftrightarrow$ "Seni partide gördüm", their lexical level representations are

`I see+p you at the garden` $\leftrightarrow$ `Sen+yH bahçe+DA gör+DH,+m`
`I see+p you at the party` $\leftrightarrow$ `Sen+yH parti+DA gör+DH,+m`

Form these examples with one pair of differences in both sides, the following transla-

tion templetes are learned:

$$
\begin{array}{rcl}
\texttt{[i see+p you at the } X^1\texttt{]} & \leftrightarrow & \texttt{[sen+yH } X^2\texttt{+DA gör+DH+m] if} \\
\texttt{[}X^1\texttt{]} & \leftrightarrow & \texttt{[}X^2\texttt{],} \\
\multicolumn{3}{c}{\text{and}} \\
\texttt{[garden]} & \leftrightarrow & \texttt{[bahçe],} \\
\texttt{[party]} & \leftrightarrow & \texttt{[parti]}
\end{array}
$$

Example 2: Given the example translations "It falls" ↔ "Düşer", "I will take the car" ↔ "Arabayı alacağım", "If a pencil is dropped then it falls" ↔ "Bir kurşun kalem bırakılırsa, düşer" and "If he brought then I will take car" ↔ "getirdiyse arabayı alacağım", their lexical level representations are

$$
\begin{array}{rcl}
\texttt{It fall+s} & \leftrightarrow & \texttt{düş+Ar} \\
\texttt{I will take the car} & \leftrightarrow & \texttt{araba+yH al+yAcAk+yHm} \\
\texttt{if a pen is drop+pp} & \leftrightarrow & \texttt{Bir kalem bırak+Hl+Hr+ysA,} \\
\texttt{then it fall+s} & & \texttt{düş+Ar} \\
\texttt{if he bring+p} & \leftrightarrow & \texttt{getir+DH+ysA,} \\
\texttt{then I will take the car} & & \texttt{araba+yH al+yAcAk+yHm}
\end{array}
$$

The match sequence between the last two example translations contains two similatities for `if` and `then`, and two differences. Since there are more than one differences, no translations templetes can be learned directly. However, with the help of the first two translation examples, the following translation templetes are learned:

$$
\begin{array}{rcl}
\texttt{[if } X_1^1 \texttt{ then } X_2^1\texttt{]} & \leftrightarrow & \texttt{[}X_1^2\texttt{+ysA, }X_2^2\texttt{] if} \\
\texttt{[}X_1^1\texttt{]} & \leftrightarrow & \texttt{[}X_1^2\texttt{] and} \\
\texttt{[}X_2^1\texttt{]} & \leftrightarrow & \texttt{[}X_2^2\texttt{],} \\
\multicolumn{3}{c}{\text{and}} \\
\texttt{[a pencil is drop+pp]} & \leftrightarrow & \texttt{[Bir kurşun kalem bırak+Hl+Hr],} \\
\texttt{[he bring+p]} & \leftrightarrow & \texttt{[getir+DH]}
\end{array}
$$

Example 3. Given the example translations "I would like to look at it" ↔ "Ona bakmak isterim" and "Do not look at it" ↔ "Ona bakma" their lexical level representations are

$$
\begin{array}{rcl}
\texttt{I would like to look at it} & \leftrightarrow & \texttt{O+nA bak+mAk iste+Hr+yHm} \\
\texttt{Do not look at it} & \leftrightarrow & \texttt{O+nA bak+mA}
\end{array}
$$

Even from these structurally different translations examples, the following translation templetes are learned:

$$
\begin{array}{rcl}
\texttt{[}X^1\texttt{ look at it]} & \leftrightarrow & \texttt{[o,+nA bak } X^2\texttt{] if} \\
\texttt{[}X^1\texttt{]} & \leftrightarrow & \texttt{[}X^2\texttt{]} \\
\multicolumn{3}{c}{\text{and}} \\
\texttt{[i would like to]} & \leftrightarrow & \texttt{[+mAk iste+Hr+yHm],} \\
\texttt{[do not]} & \leftrightarrow & \texttt{[+mA]}
\end{array}
$$

Example 4. Given the example translations "he can read a book" ↔ "kitap okuyabilir", "do not talk" ↔ "konuşma", "he can read a book while he is walking" ↔ "yürürken kitap okuyabilir" and "do not talk while you are eating" ↔ "yemek yerken

konuşma", their lexical level representations are

he can read a book	$\leftrightarrow$	kitab oku+yAbil+Hr
do not talk	$\leftrightarrow$	konuş+mA
he can read a book while he is walk+ing	$\leftrightarrow$	yürü+Hr+yken kitab oku+yAbil+Hr
do not talk while you are eat+ing	$\leftrightarrow$	yemek ye+Hr+yken konuş+mA

From these translations examples, the following translation useful templetes are learned:

$[X_1^1$ `while` X_2^1`+ing]`	$\leftrightarrow$	$[X_2^2$`+Hr+yken` $X_1^2]$ `if`
$[X_1^1]$	$\leftrightarrow$	$[X_1^2]$ `and`
$[X_2^1]$	$\leftrightarrow$	$[X_2^2]$`,`
and		
`he is walk`	$\leftrightarrow$	`yürü`
`you are eat`	$\leftrightarrow$	`yemek ye`

The last two translation templetes may be used to fill in more complex translation templetes.

5 Translation

The translation templetes learned by the TTL algorithm can be used in the translation directly. These templetes can be used for translation in both directions. The outline of the translation process is given below:

1. First, the lexical level representation of the input sentence is derived.
2. The translation templetes with highest match score (total number of matching terminals) are collected. These templetes are those that are most similar to the sentence to be translated.
3. For each selected (most specific) templete, its variables are instantiated with the corresponding values in the source sentence. Then, templetes matching these bound values are sought. If they are found successfully, their values are replaced in the variables corresponding to the sentence in the target language.
4. The surface level representation of the sentence obtained in the previous step is generated.

Note that, if the sentence in the source language is ambiguous, then templates corresponding to each sense will be retrieved, and the sentences for each sense will be generated. Among the possible translations, a human user can choose the right one according to the context.

6 Conclusion

In this paper, we have presented a model for learning translation templetes between two languages. The model is based on a simple pattern matcher. We integrated this model with an example-based translation model into Generalized Exemplar-Based Machine Translation. We have implemented this model as the TTL (Translation Template Learner) algorithm. The TTL algorithm is illustrated in learning translation templetes between Turkish and English. It is clear that the approach is applicable to a pair of languages.

The major contribution of this paper is that the propsed TTL algorithm eliminates the need for manually encoding the translations, which is a difficult task for a large corpus. The TTL algorithm can work directly on surface level representation of sentences. However, in order to generate useful translation patterns, it is helpful to use the lexical level representations. It is usually trivial, at least for English and Turkish, to obtain the lexial level representations of words.

Our main motivation was that the underlying inference mechanism is compatible with one of the ways humans learn languages, i.e. learning from examples. We believe that in everyday usage, humans learn general sentence patterns, using the similarities and differences between many different example sentences that they are exposed to. This observation lead us to the idea that a computer can be trained similarly, using analogy within a corpus of example translations.

The accuracy of the translations learned by this approach is quite high with ensured grammaticality. Given that a translation is carried out using the rules learned, the accuracy of the output translation critically depends on the accuracy of the rules learned.

We do not require an extra operation to maintain the grammaticality and the style of the output, as in Kitano's EBMT model [5]. The information necessary to maintain these issues is directly provided by the translation templates.

The model that we have proposed in this paper may be integrated with an intelligent tutoring system (ITS) for second language learning. The template representation in our model provides a level of information that may help in error diagnosis and student modeling tasks of an ITS. The model may also be used in tuning the teaching strategy according to the needs of the student by analyzing the student answers analogically with the closest cases in the corpus. Specific corpora may be designed to concentrate on certain topics that will help in student's acquisition of the target language. The work presented by this paper provides an opportunity to evaluate this possibility as a future work.

References

[1] Arnold D., Balkan L., Humphreys R.Lee, Meijer S. Sadler L.: *Machine Translation*, NCC Blackwell (1994).

[2] Carbonell, J.G.: Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition. In Jude W. Shavlik and Thomas G. Dietterich (eds), *Readings in Machine Learning*, Morgan Kaufmann (1990) 636–646.

[3] Furuse, O. & Iida, H.: Cooperation between Transfer and Analysis in Example-Based Framework, *Proceedings of COLING-92* (1992).

[4] Hammond, K.J.: (Ed.) *Proceedings: Second Case-Based Reasoning Workshop.* Pensacola Beach, FL:Morgan Kaufmann (1989).

[5] Kitano, H.: A Comprehensive and Practical Model of Memory-Based Machine Translation. In Ruzena Bajcsy (Ed.) *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, Morgan Kaufmann V.2 (1993) 1276-1282.

[6] Kolodner, J.L.: (Ed.) *Proceedings of a Workshop on Case-Based Reasoning.* Clearwater Beach, FL: Morgan Kaufmann (1988).

[7] Medin, D.L. & Schaffer, M.M.: Context theory of classification learning. *Psychological Review, 85* (1978) 207-238.

[8] Nagao, M. A.: Framework of a Mechanical Translation between Japanese and English by Analogy Principle (1985).

[9] Nirenburg, S., Beale, S. & Domashnev, C.: A Full-Text Experiment in Example-Based Machine Translation. *Proceedings of the International Conference on New Methods in Language Processing, NeMLap* Manchester, UK (1994) 78-87.

[10] Ram, A.: Indexing, Elaboration and Refinement: Incremental Learning of Explanatory Cases. In Janet L. Kolodner (ed.), *Case-Based Learning*, Kluwer Academic Publishers (1993).

[11] Reisbech, C. & Schank, R.: *Inside the Case-Based Reasoning*, Lawrence Elbaum Associates (1990).

[12] Sato, S.: *Example-Based Machine Translation*, Ph.D. Thesis, Kyoto University (1991).

[13] Sato, S. & Nagao, M.: The Memory-Based Translation, *Proceedings of COLING-90* (1990).

[14] Stanfill, C. & Waltz, D.: Toward Memory-Based Reasoning. *CACM*, Vol.29, No.12 (1991) 185-192.

[15] Sumita, E. & Iida, H.: Experiments and Prospects of Example-Based Machine Translation, *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics* (1991).