Inferring Maximally Invertible Bi-grammars for Example-Based Machine Translation

Michael Carl and Dekai Wu
Human Language Technology Center
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
lrac@cs.ust.hk

Abstract

This paper discusses inference strategies of context-free bi-grammars for example based machine translation (EBMT). The EBMT system EDGAR is discussed in detail. The notion of invertible context-free feature bi-grammar is introduced in order to provide a means to decide upon the degree of ambiguity of the inferred bi-grammar. It is claimed that a maximally invertible bi-grammar can enhance the precision of the bilingual alignment process, reduce the complexity of the inferred grammar, and uncover inconsistencies in bi-corpora. This paper describes preliminary reflections and thus no empirical evaluation of the method is provided.

1 Introduction

Machine Translation can be considered as the permutation of word translations from a source language string into the target language via some internal representations. As has been shown in Carl (1999b), the number of possible prepresentations for a sentence of length $n, 0 < n < 20$ is greater than $n!$. Due to untractable representational complexity, it is unthinkable that any MT-system performs exhaustive search even on a short string.

Fortunately, parsing *strategies* select among the possible representations a small subset which can be derived from the input strings. Thus, to reduce complexity, the ReVerb EBMT system (Collins, 1999), for instance, only computes trees of depth 1 while it allows an arbitrary number of chunks. The optimal length for each chunk is calculated by statistic means. The inversion transduction grammar (Wu, 1995), in contrast, only allows binary trees of arbitrary depth, where all internal nodes are non-terminal symbols.

In addition representations may be based on linguistically motivated assumptions. Some systems try to generate representations which reflect the constituent structure of the input string. Other systems e.g. ReVerb focus on the ease of transfer and generation.

In inductive MT systems, one seeks to infer context-free bi-grammars from a corpus of example translations. The inferred bi-production rules are then applied in the translation phase to generate a target language string from a source language input.

Some systems allow all possible representations of the input strings. These systems make use of some heuristics to find the most specific bi-production rules in the translation process. In the so-called generalized exemplar-based MT-system (Güvenir and Cicekli, 1998), the heuristic consists in mapping the new sentence to be translated onto all bi-production rules and only consider those rules which contain the greatest number of shared terminal symbols. This process is top-down iteratively repeated until all items in the input sentence are translated. In a variant of this system (Öz and Cicekli, 1998), translation templates are weighted and confidence values are computed, thus, to keep only the most reliable structures. However, due to the Translation Template Learner (TTL) algorithm, Carl (1999b) argues that this method may be computational expensive as the number of translation templates grows with the power of 2 to the number of translation examples.

This paper discusses yet another EBMT approach (cf. (Carl, 1999a)) which makes use of three methods to restrict the number of representations. From a set of translation examples, a set of non-terminal bi-production rules (i.e. translation templates) is generated, which is in size linear to the set of translation examples. In addition to this, derivation trees are only generated if the feature structures unify at runtime. While these mechanisms have already been described elsewhere (Carl, 1999a), the potentials of invertible context-free bi-grammars shall be examined in this paper. Apart from the computational gain of restricted grammars, it shall be argued that invertible context-free bi-grammars may uncover inconsistencies in the bi-corpora and enhance bi-lingual alignment precision. As a consequence of this, the generated translations are likely to be more consistent and reliable.

The concept of invertible grammars is not new. It has been shown that invertible grammars can be updated in polynomial time in the size of the input (Mäkinen, 1992) and that for each context-free grammar, there exists an invertible context-free grammar such that both grammars generate the same language (Harrison, 1978). As yet, to our knowledge, no research has been undertaken which applies the invertibleness condition to bi-grammars inference.

This paper is structured as follows. In the next section a brief introduction to EBMT EDGAR shall be given. The third section outlines in more detail how translation templates are inferred. The fourth section introduces the concept of invertible context-free bi-grammars and discusses their expected benefits.

2 EBMT EDGAR

The EBMT system EDGAR is described in more detail in (Carl, 1999a). Here, only a brief overview shall be given.

EDGAR is an Example-Based Machine Translation (EBMT) system which integrates linguistic (i.e. morphological) knowledge, reference translations, simple syntactic rules for analysis and generation, and a component which infers generalized translation templates from translation examples. EDGAR morphologically analyzes (cf. (Maas, 1996)) the input sentence, decomposes and generalizes it by matching it against the set of reference translations and reducing the matching chunks into one node. The generalized input sentence is then specified i.e. the correct linguistic information is gathered, and 'refined' in the target language.

The new sentence to be translated is decomposed according to the source language parts of the examples contained in an example base (EB). On a number of levels, the input sentence is thereby reduced by applying a set of reduction rules until no further generalization can be computed. The generalized input sentence is then specified and refined in the target language. Specification of the generalization retrieves the target language parts of the translation examples from the EB and refines them by applying a set of refinement rules.

In order to percolate features in the derivation trees, a set of linguistically motivated rules may apply (cf. (Carl et al., 1997) for the rule formalism). The so-called set of reduction rules applies when analysing the input string in order to percolate features from the daughter nodes into the parents. A set of refinement rules applies when generating the target language string and percolates features from the parents into the specified daughter nodes. These rules allow the following operations on nodes in the derivation trees:

- Unification and deletion of features.
- Concatenation and replacement of values.
- Insertion and deletion of nodes.

While the decomposition of the input string and thus the structure of the derivation tree is guided by the contents of the EB, the percolation of features in the derivation trees is achieved through an independent rule system. The shape of the translation examples contained in the EB is thus most important for a correct decomposition, to derive a reasonable derivation tree and thus to generate a reliable translation.

3 Inferring Translation Templates in EDGAR

In this section we show how EDGAR infers translation templates from a set of reference translations. We first define a context-free feature bi-grammar.

A context-free feature bi-grammar G is an equivalence of two context-free grammars enriched with a source language and a target language feature set:

$G = (N, \Sigma^s, \Sigma^t, T^s, T^t, EB)$ where
N : set of non-terminal symbols
Σ^s : source language terminal symbols
Σ^t : target language terminal symbols
T^s : source language features
T^t : target language features
EB : set of bi-production rules

Bi-production rules are of the form $(b)_\alpha \leftrightarrow (b)_\beta$ where $b \in (\{N^s \cup \Sigma^s\} \times T^s)^+$, $y \in (\{N^t \cup \Sigma^t\} \times T^t)^+$, $\alpha \in T^s$ and $\beta \in T^t$. For each bi-production rule in EB, there is an equal number of unambiguous, mutually linked left-hand side non-terminal symbols connected to one non-terminal right-hand side symbols and vice versa.

Translation templates are inferred from translations examples which are contained in the initial EB. A translation template is a bilingual generalization of a translation example where compositionally translatable sequences are replaced by non-terminal symbols. While a translation template contains at least one non-terminal symbol, a translation example contains only terminal symbols.

A translation template needs to hold the following template correctness criteria TCC (cf. (Carl, 1999a) for more information).

1. A translation template contains at least two symbols in both the source and the target language sides.
2. There is an equal number > 0 of non-terminal reductions in both the source and the target language sides of the translation template.
3. Reductions in the source and the target language sides are based on the same examples.

The initial EB is sorted by the length of their translation examples and then incrementally extended

Figure 1: From an initial example base containing the translation examples 1 to 5, the four translation templates 3', 4',5' and 5" are generated. The translation template 5" is a second order translation template generated from translation template 3', 4' and 5'.

1	$(\text{automatic})_{adj}$	$\leftrightarrow$	$(\text{Automatik})_{adj}$
2	$(\text{external})_{adj}$	$\leftrightarrow$	$(\text{äusseres})_{adj}$
3	$(\text{automatic transmission})_{noun}$	$\leftrightarrow$	$(\text{Automatik/getriebe})_{noun}$
4	$(\text{external control})_{noun}$	$\leftrightarrow$	$(\text{äusseres Bedien/element})_{noun}$
5	$(\text{Automatic transmission external control})_{noun}$	$\leftrightarrow$	$(\text{Äusseres Bedien/element des Automatik/getriebes})_{noun}$
3'	$(\mathcal{X}_{adj}\ \text{transmission})_{noun}$	$\leftrightarrow$	$(\mathcal{X}_{adj}\ \text{getriebes})_{noun}$
4'	$(\mathcal{X}_{adj}\ \text{control})_{noun}$	$\leftrightarrow$	$(\mathcal{X}_{adj}\ \text{Bedien/element})_{noun}$
5'	$(\mathcal{X}_{adj}\ \text{transmission}\ \mathcal{Y}_{adj}\ \text{control})_{noun}$	$\leftrightarrow$	$(\mathcal{Y}_{adj}\ \text{Bedien/element des}\ \mathcal{X}_{adj}\ \text{getriebes})_{noun}$
5"	$(\mathcal{X}_{noun}\ \mathcal{Y}_{noun})_{noun}$	$\leftrightarrow$	$(\mathcal{Y}_{noun}\ \text{des}\ \mathcal{X}_{noun})_{noun}$

Figure 2: Inference of translation templates is achieved by replacing the matching parts in a translation example through non-terminals.

$$\left.\begin{array}{rcl}(b)_{\alpha^1} & \leftrightarrow & (y)_{\beta^1}\\ (abc)_{\alpha^2} & \leftrightarrow & (xyz)_{\beta^2}\end{array}\right\} \rightarrow (a\mathcal{X}_{\alpha^1}c)_{\alpha^2} \leftrightarrow (x\mathcal{X}_{\beta^1}z)_{\beta^2}$$

$$\begin{array}{rcl}a, c & \in & (\{N \cup \Sigma^s\} \times 2^{T^s})^*,\\ b & \in & (\{N \cup \Sigma^s\} \times 2^{T^s})^+,\\ x, z & \in & (\{N \cup \Sigma^t\} \times 2^{T^t})^*,\\ y & \in & (\{N \cup \Sigma^t\} \times 2^{T^t})^+,\\ \mathcal{X} & \in & N,\ \alpha \in T^s,\ \beta \in T^t\end{array}$$

by non-terminal bi-production rules i.e. translation templates containing one or more linked non-terminals on each language side. This is done by means of the following simple inference rule. Given two translation examples E^1 and $E^2 \in EB$ and E^1 is contained in E^2. A translation template $E^{2'}$ is generated by replacing the matching parts in E^2 through annotated non-terminal symbols carrying the indices of the matching rule E^1. This inference rule is shown in Figure 3.

In case there are possible overlapping substitutions in E^2 due to two bi-production rules E^a and E^b, the first longest segmentation is chosen. To avoid recursive non-reducing production rules, those translation templates are filtered out which contain only one symbol, i.e. one non-terminal only, on any language side according to TCC 1.

Different non-terminal symbols are taken in case more than one substring is substituted. Assuming that translation examples 6 and 7 are in the EB, translation template 8' may be generated from translation example 8, where the word order of the generalized substrings changes in the source and the target language:

6	$(b)_{\alpha^1}$	$\leftrightarrow$	$(y)_{\beta^1}$
7	$(d)_{\alpha^2}$	$\leftrightarrow$	$(w)_{\beta^2}$
8	$(abcde)_{\alpha^3}$	$\leftrightarrow$	$(vwxyz)_{\beta^3}$
8'	$(a\ \mathcal{X}_{\alpha^1}\ c\ \mathcal{Y}_{\alpha^2}\ e)_{\alpha^3}$	$\leftrightarrow$	$(v\ \mathcal{Y}_{\beta^1}\ x\ \mathcal{X}_{\beta^2}\ z)_{\beta^3}$

The indices α^1, α^2, β^1 and β^2 are copied from the matching rules 6 and 7 into the non-terminal symbols in translation template 8'.

The EB undergoes several inference cycles until no more substitutions are performed. At each inference cycle a maximum number of substrings are replaced in each bi-production rule in order to obtain the shortest possible generalization. Non-terminal translation templates of order one or higher may be generated as the inference process runs iteratively over previously inferred translation templates. Translation template 5" in Figure 2 is a second order translation template, because it is, in fact, derived from three translation templates 3', 4' and 5'[1]. The maximum number of translation templates that can be inferred from one translation example is $n - 1$ where n is the number of words on either of the two language sides. $n - 1$ translation templates are generated if at each iteration cycle, only one further word is generalized. The maximun number of inferred translation templates is thus $m \times n$ where m is the size of the initial EB and n is the maximum translation example length.

There are a number of drawbacks to this simple inference method. These are due to the fact that each of the translation examples is generalized on its own. It does not handle well the fact that there can be ambiguous translation examples in the initial example base and, consequently, ambiguous translation templates might be generated. Although the inference mechanism makes sure that a non-redundant, minimal number of translation templates is generated from each translation example, the union of the

---

[1] In the examples I have used surface forms although the internal representation is based on lemmata. The genitive 's' in 'des' and 'getriebes' is thus internally invisible.

Figure 3: Two invertibleness conditions and a chunking condition apply on a intitial EB as a whole.

1. Invertibleness condition for translation examples.

$$\left.\begin{array}{rcl} d & \leftrightarrow & w \\ abc & \leftrightarrow & xyz \end{array}\right\} \Longrightarrow \left\{ ((d \neq b) \wedge (w \neq y)) \quad OR \quad ((d = b) \wedge (w = y)) \right.$$

2. Invertibleness condition for translation templates.

$$\left.\begin{array}{rcl} (b)_{\alpha^1} & \leftrightarrow & (y)_{\beta^1} \\ (d)_{\alpha^2} & \leftrightarrow & (w)_{\beta^2} \\ (abc)_{\alpha^3} & \leftrightarrow & (xyz)_{\beta^3} \\ (adc)_{\alpha^3} & \leftrightarrow & (xwz)_{\beta^3} \end{array}\right\} \Longrightarrow \left\{ \begin{array}{l} ((\alpha^1 \neq \alpha^2) \wedge (\beta^1 \neq \beta^2)) \quad OR \\ ((\alpha^1 = \alpha^2) \wedge (\beta^1 = \beta^2)) \end{array}\right.$$

3. Chunking condition ensures unambiguous chunking.

$$\left.\begin{array}{rcl} bc & \leftrightarrow & wx \\ cd & \leftrightarrow & xy \\ abcde & \leftrightarrow & vwxyz \end{array}\right\} \Longrightarrow \left\{ \begin{array}{l} ((c = \emptyset) \wedge (x = \emptyset)) \quad OR \\ ((a = \emptyset) \wedge (e = \emptyset) \wedge (v = \emptyset) \wedge (z = \emptyset)) \end{array}\right.$$

$a, c, e \in \{\Sigma^s \times T^s\}^*$, $v, x, z \in \{\Sigma^t \times T^t\}^*$, $b, d \in \{\Sigma^s \times T^s\}^+$, $w, y \in \{\Sigma^t \times T^t\}^+$,
$\alpha \in 2^{T^s}$, $\beta \in 2^{T^t}$ i.e. α and β are subsets out of T^s and T^t respectively.

produced translation templates contained in the resulting EB might be a universal grammar (see Carl (1999b) for a discussion). We, therefore, need to put more constraints on the initial EB to avoid ambiguity and redundancy of inferred translation templates.

4 Invertible feature bi-grammars

One way to obtain a homogeneous initial EB is to look at the way how example translations are obtained. Similar to a number of other EBMT approaches (Collins, 1999; Güvenir and Cicekli, 1998), we use pre-aligned bi-texts from which finer grained translation examples are extracted.

The extracted translation examples are weighted according to the probability of the known lexical translation they contain and the reliability of the morpho-syntactic pattern they describe. To achieve this, we use a morphological processor (Maas, 1996), a shallow syntactical processor (Carl et al., 1997) and a bilingual lexicon. From the set of weighted translation examples, a suitable maximally invertible bi-grammar is extracted according to the criteria given below.

For instance, there are 34 entries in the lexicon which have the stem *work* on the source side *or* which have the stem *arbeit* on the target side. From these 34 lexicon entries, there are four entries which match the translation *work* ⟷ *arbeit* such that the lexical translation weight $W_{lex}(work,\ arbeit) = 0.117$.

In a similar way the weight of the morpho-syntactic pattern is computed. Roughly 800 morpho-syntactic patterns of bilingual phrase translations are used where each of the pattern is manually annotated with the part of speech it represents[2]. These 800 patterns include translation phrases for compound nouns (*noun*), determiner phrases (*dp*), prepositional phrases (*pp*) and simple sentences (*s*).

Due to inconsistencies in the usually available bi-texts, it is unlikely to extract a set of consistent translation examples without any further constraints. The invertibleness conditions, however, seem to be appropriate for this matter. Invertibleness conditions apply to the EB as a whole and, thus, make sure that the union of the extracted translation examples and generated translation templates are consistent with respect to the set of reference translations.

We define now more formally the notion of invertibleness in bi-grammars. An invertible context-free feature bi-grammar is a context-free bi-grammar

$$G = (N, \Sigma^s, \Sigma^t, T^s, T^t, EB)$$

where for each bi-production rule in EB there is an equal number of unambiguous, mutually linked left-hand side non-terminal symbols connected to one non-terminal right-hand side symbols and vice versa. In addition, there are two invertibleness conditions. The first invertibleness condition in Figure 3 requires that if any one language side of a bi-production rule is a substring of another bi-production rule, then the other language side must be a substring of the second bi-productin rule as well. As an example, consider Figure 4 where the entries are not conformed to

---

[2] This process shall be automated in the future using the shallow parser KURD (Carl et al., 1997).

Figure 4: Bi-lingual aligned translation examples. These translation examples are not conform to the first invertibleness condition. The translation examples (1), (2) and (3) do not fulfill the chunking condition.

$$\begin{array}{rcl}
(\text{Gang/wahl und})_{noun} & \longleftrightarrow & (\text{Shift Selection \&})_{noun} \qquad (1)\\
(\text{Schalt/position})_{noun} & \longleftrightarrow & (\text{Gear Position})_{noun} \qquad (2)\\
(\text{wahl und Schalt})_{noun} & \longleftrightarrow & (\text{Selection \& Gear})_{noun} \qquad (3)\\
(\text{Gang/wahl und Schalt/position})_{noun} & \longleftrightarrow & (\text{Selection \& Gear Position})_{noun} \qquad (4)\\
\hline
(\text{Gestänge/hebel})_{noun} & \longleftrightarrow & \left\{\begin{array}{l}(\text{Gear shift lever})_{noun}\\(\text{Transmission Unit Gear Selector})_{noun}\end{array}\right. \qquad (5)\\
\hline
(\text{Aussen/seil befestigen.})_{s} & \longleftrightarrow & \left\{\begin{array}{l}(\text{Locate the outer cable})_{s}\\(\text{Secure the outer cable})_{s}\end{array}\right. \qquad (6)
\end{array}$$

Figure 5: The inferred translation templates $3'$ and $4'$ do not fulfill the second invertibleness conditions. Changing the label of translation example 1 from *pp* into *adv* makes possible the inference of consistent translation templates $3'_a$ and $4'$.

$$\begin{array}{lrcl}
1: & (\text{in the morning})_{pp} & \leftrightarrow & (\text{am morgen})_{adv}\\
2: & (\text{into the room})_{pp} & \leftrightarrow & (\text{in den Raum})_{pp}\\
3: & (\text{John came (in the morning})_{pp})_{s} & \leftrightarrow & (\text{Hans kam (am morgen})_{adv})_{s}\\
4: & (\text{John came (into the room})_{pp})_{s} & \leftrightarrow & (\text{Hans kam (in den Raum})_{pp})_{s}\\
3': & (\text{John came } \mathcal{X}_{pp})_{s} & \leftrightarrow & (\text{Hans kam } \mathcal{X}_{adv})_{s}\\
4': & (\text{John came } \mathcal{X}_{pp})_{s} & \leftrightarrow & (\text{Hans kam } \mathcal{X}_{pp})_{s}\\
\\
1_a: & (\text{in the morning})_{adv} & \leftrightarrow & (\text{am morgen})_{adv}\\
3_a: & (\text{John came (in the morning})_{adv})_{s} & \leftrightarrow & (\text{Hans kam (am morgen})_{adv})_{s}\\
3_a{}': & (\text{John came } \mathcal{X}_{adv})_{s} & \leftrightarrow & (\text{Hans kam } \mathcal{X}_{adv})_{s}
\end{array}$$

the invertibleness condition. The upper part of Figure 4 shows four translation examples from which (1) and (4) are of the form $a \leftrightarrow xy$ and $ab \leftrightarrow yz$. The left-hand side "Gang/wahl und" in (1) is contained in the translation example (4), but not so the right-hand side and hence invertibleness condition 1 is violated. The two example pairs in the lower part of Figure 4 (5) and (6) are of the form $a \leftrightarrow x$ and $a \leftrightarrow y$. For each of the conflicting pairs of translation examples we would have to eliminate one from the initial example base in order to fulfill the invertibleness condition

The second condition insures invertibleness of translation templates. Given the template inference algorithm as described in the previous section, the following translation templates are generated from an example base as shown in the Figure 3.

$$(a\mathcal{X}_{\alpha^1}c)_{\alpha^3} \leftrightarrow (x\mathcal{X}_{\beta^1}z)$$
$$(a\mathcal{X}_{\alpha^2}c)_{\alpha^3} \leftrightarrow (x\mathcal{X}_{\beta^2}z)$$

If we assume that $(\alpha^1 = \alpha^2) \wedge (\beta^1 \neq \beta^2)$ than we have two identical left-hand sides mapped onto different right-hand sides. Similarly, if $(\alpha^1 \neq \alpha^2) \wedge (\beta^1 = \beta^2)$ then two identical right-hand sides are mapped into different left-hand sides.

Consider for instance, the English-German translation examples 3 and 4 in Figure 5 and assume that "in the morning ↔ am morgen" and "into the room ↔ in den Raum" have correctly been extracted as shown in 1. However, the erroneous phrase tag *pp* has been assigned to the English string *in the morning*. The translation templates $3'$ and $4'$ are inferred given the mechanism described in the previous section. Although there is no contradiction on the level of translation examples, the inferred translation templates $3'$ and $4'$ are not in accordance with the second invertibleness condition. Two similar left-hand sides (John came $\mathcal{X}_{pp})_s$ in the translation templates have two different right-hand sides. (Hans kam $\mathcal{X}_{adv})_s$ and (Hans kam $\mathcal{X}_{pp})_s$

There are many ways to change segmentation and/or labeling such that the union of the inferred translation templates become invertible. One obvious modification, in this case, is to change the *pp* label of the English phrase *in the morning* into an adverbial phrase as shown in translation example 1_a. The union of the inferred translation templates $3'_a$ and $4'$ thus suffices the invertibleness condition. Another possibility is to enrich the feature system by adding a directional tag *dir*. It could thus be feasible to extract the English phrase (into the room)$_{pp,dir}$.

Which of the ways are to be taken in order to achieve an invertible set of bi-production rules depends on a number of factors such as the shape of the bi-text and the richness of the morphological tagger.

The chunking condition 3 in Figure 3 is aimed to insure unambiguous chunking of the input sentence. The condition says that if a prefix from one translation example is a suffix of another translation example, then there must be a further translation example that consists of the concatenation of the prefix and suffix only. In this way, if we would have to chose among two ambiguous segmentations *a(bc)de* and *ab(cd)e* then there should be a bi-production rule that allows for the segmentation *a(bcd)e*.

An example is given in Figure 4 in translation example (1), (2) and (3). In the left-hand sides of the translation examples "wahl und" is a suffix of (1) and a prefix of (3) while "Schalt" is a suffix in (3) and a prefix in (2). In the right-hand sides, the same applies for "Selection &" and "Gear" which are respectively prefixes of (3) and (2) and suffixes of (1) and (3). Eliminating the translation example (3) "wahl und Schalt ↔ Selection & Gear" from the example base would make it conform to the chunking condition.

5 Conclusion

This paper examines a strategy to infer bi-production rules in example based machine translation (EBMT) systems. The inference strategy of EBMT EDGAR is discussed in detail. Grammar inference is based on lexical and morphological knowledge of the languages involved. First translation examples are extracted from a bi-text. From these translation examples, translation templates are generated. The union of the translation examples and the inferred translation templates constitute the final example base. It is sought to generate the least ambiguous example base such that only one derivation trees for any example translation is generated. For this matter, the notion of invertible feature bi-grammars is introduced. However, invertible (bi-) grammars are not closed with respect to union. That is, the union of two or mor) invertible grammars is not necessarily again an invertible grammar. When learning a bi-grammar from unseen text, we can, therefore, at best generate maximally invertible bi-grammars where the number of non-invertible bi-production rules is minimal.

There are two distinct sources for inferring non-invertible bi-production rules from bi-texts. One source is due to erroneous segmentation and/or erroneous phrase tag assignment during alignment of bi-lingual texts. Another source of inferring non-invertible grammars is due to inconsistencies in the reference bi-text. In a German-English bi-text describing repair instructions as delivered from a car manufacturer containing 303 translation examples, we have discovered, among others, inconsistent translations as shown in the lower two examples in Figure 4. The upper example in Figure 4 is due to an erroneous alignment.

These latter non-invertible terminal bi-production rules can be seen as inconsistencies in the learn corpus. Such information may be valuable, e.g. in a multilingual editing system, where consistent use of terminology among several translators is a major challenge. It can help, as well, to upgrade translation products or to normalize the use of terminology and translation equivalences.

References

Michael Carl, Antje Schmidt-Wigger, and Munpyo Hong. 1997. KURD - a Formalism for Shallow Post-morphological Processing. In *Proceedings of the NL-PRS*, Phuket, Thailand.

Michael Carl. 1999a. Inducing Translation Templates for Example-Based Machine Translation. In *MT-Summit VII.*

Michael Carl. 1999b. Tackling the complexity of context-free representations in example-based machine translation. In *NLPRS 99.*

Bróna Collins. 1999. *Example-Based Machine Translation: An Adaptation-Guided Retrieval Approach.* Ph.D. thesis, Trinity College, Dublin.

Halil Altay Güvenir and Ilyas Cicekli. 1998. Learning Translation Templates from Examples. *Information Systems*, 23(6):353–363.

Michael A. Harrison. 1978. *Introduction to Formal Language Theory.* Addison-Wesley, Reading.

Heinz-Dieter Maas. 1996. MPRO - Ein System zur Analyse und Synthese deutscher Wörter. In Roland Hausser, editor, *Linguistische Verifikation, Sprache und Information.* Max Niemeyer Verlag, Tübingen.

Erkki Mäkinen. 1992. On the structural grammatical inference problem for some classes of context-free grammars. *Information Processing Letters*, 42:1–5.

Zeynep Öz and Ilyas Cicekli. 1998. Ordering Translation Templates by Assigning Confidence Factors. In *Third Conference of AMTA*, pages 51–62.

Dekai Wu. 1995. Grammarless extraction of phrasal translation examples from parallel texts. In *TMI-95.*