Linking Example-Based and Rule-Based Machine Translation

Michael Carl, Catherine Pease and Oliver Streiter

Institut für Angewandte Informationsforschung,

Martin-Luther Str. 14

(carl,cath,oliver)@iai.uni-sb.de

66111 Saarbrücken, Germany

January 20, 1999

Abstract

In order to improve the quality of translation and to make the MT systems easier to tune to the needs of different users, two machine translation ideologies are combined, in order to investigate the consequences of this linkage and to determine what kind of linguistic entities (syntactic constructions, lexicographic types, collocations etc...) can be dynamically transferred between the different components without introducing new translation errors.

1 Introduction

All of the individual MT approaches which have been employed so far have their strengths and weaknesses and it is unlikely that an new, "ideal" approach may be proposed and implemented on a sizeable scale in the foreseeable future. Substantial progress in the field can therefore be achieved only by combining the strengths of different approaches. A discussion of the expected benefits of such a linkage in terms of a recall, coverage, adaptability, translation quality, reliability and user-orientation is beyond the scope of this paper and can be found in ESSLI (Carl et al., 1998).

Starting from this assumption, we investigated how different MT paradigms can be integrated in one framework. The paradigms we consider here are rule-based MT (RBMT) and example-based MT (EBMT). Experiments of linkage have been carried out using the RBMT system CAT2 and the EBMT system EDGAR. More information about theses system than can be provided here can be found in (Streiter, 1996), (Streiter, 1998), (Carl, 1998).

```
                             s
   __________________________|______
   g                               s
___|__            _________________|____________
f    g            s                           pred
|    |      ______|___           ______________|___
|    |      s       mod         mod              pred
|    |      |    ____|_          |     ___________|____
|    |      |   mod   g          |    pred            t
|    |      |    |  __|___       |     |    __________|___
|    |      |    |  f    g       |     |    f            t
|    |      |    |  |    |       |     |    |        ____|_______
|    |      |    |  |    |       |     |    |       mod         t
d sprache haben rel d arbeiten gross angst rel unterscheiden morphem
```

```
                             s
   __________________________|__________________
   g      s           mod                      t
   |      |   _________|_____              ____|_________
   |      |   mod  t        g             mod           t
   |      |    |   |     ___|____       __|_______      |
   |      |    |   |     g      a       mod      t      |
sprache angst rel pro arbeiten pro unterscheiden pro morphem
```

2 CAT2

CAT2 is a unification-based MT formalism used at IAI for the development of various MT related applications. In order to explain the operation of the system, we follow the translation of a illustrative (although fictive) German sentence into English step by step: *Der Sprachwissenschaftler hat bei der Arbeit grosse Angst vor ununterscheidbaren Morphemen.* Syntactic and semantic analysis is performed following HPSG-like schemes of composition. Besides the syntactic functions, semantic roles are identified (a=agent, t=theme, g=goal). In our example, *Angst* (fear) is recoginzed as predicative noun of the support verb *haben* (have), *der Sprachwissenschaftler* (the linguist) as the person who experiences the fear and *ununterscheidbare Morpheme* (undistinguishable morphemes) as the content of the fear:

Before transfer, function words (determiners, case marking prepositions, degree words, auxiliaries, magnifiers etc...) are structurally removed. Pronouns are introduced as internal arguments of modifier relations. Support verbs (e.g. *haben*) and copula verbs are replaced by the element bearing the argument structure (e.g. *Angst*).

The lexical transfer triggers successively different translational options until an overall integration of these options becomes possible:

After transfer, support verbs and copulative verbs (here *be*) are inserted. Word order is rearranged and function words are generated as necessary:

```
{slex=sprachwissenschaftler} =>{lex=language,slex=linguist}
{slex=angst}                 =>{lex=anxious,slex=anxious}
{slex=bei}                   =>{lex=rel,slex=during}
{slex=angst}                 =>{lex=pro,slex=?}
{slex=arbeit}                =>{lex=operate,slex=operation}
{slex=someone}               =>{lex=pro,slex=someone}
{slex=ununterscheidbar}      =>{lex=differ,slex=different/differ/differently}
                               {lex=distinguish,slex=distinguishable}
{slex=someone}               =>{lex=pro,slex=someone}
{slex=morphem}               =>{lex=morpheme,slex=morpheme}
```

```
                              s
   ___________________________|__________________
   g        s           mod                    t
   |        |     _______|_____            ____|________
   |        |    mod  t       g            mod          t
   |        |     |   |     __|____      __|______     |
   |        |     |   |    g     a      mod      t     |
language anxious rel pro operate pro distinguish pro morpheme
```

```
                       s
   ____________________|______
   g                         s
___|__               ________|_____________________________
f    g               s                                   mod
|    |    ___________|__                              ____|_
|    |    s           pred                           mod   g
|    |    |   ________|___                            | ___|_
|    |    |   f           pred                        |  f   g
|    |    |   |   __________|___                      |  |   |
|    |    |   | pred           t                      |  |   |
|    |    |   |   |     _______|___                   |  |   |
|    |    |   |   |     f         t                   |  |   |
|    |    |   |   |     |    _____|_____              |  |   |
|    |    |   |   |     |    t        mod             |  |   |
|    |    |   |   |     |    |    _____|___           |  |   |
|    |    |   |   |     |    |    |  _____|__         |  |   |
|    |    |   |   |     |    |    | mod    pred       |  |   |
|    |    |   |   |     |    |    |  |   ___|____     |  |   |
|    |    |   |   |     |    |    |  |   f     pred   |  |   |
d langua. be deg anxi. rel morph. pro be deg distin. rel d operation
"The linguist is very anxious about morphemes that are not distinguishable
during the operation."
```

3 EDGAR

EDGAR is an experimental EBMT system which decomposes and ”generalises” a morphologically analyzed input sentence by matching it against a case base (henceforth CB). In this CB the morphological analysis of translation examples is stored together with translation templates which are automatically generalized from the translation examples. The generation of translation templates is seen as a sort of example driven grammar induction. In the absence of a complete match of a sentence and a reference translation, a simple subject-verb-object sentence, for example, can be translated using a subject-verb-object template via linguistic generalisation. The generalized input sentence is then specified (i.e. the correct linguistic information is gathered) and ”refined” in the target language.

Each example is divided into two feature sets: fixed and variable features. The fixed feature include the lexical description (i.e. the lemmatised form of the word) LU, gender G and type (i.e. syntactic category) C of the input. The variable features describes the morpho-syntactic features, these include type C, tense TNS, verbal type VTYP, number NB and prepositional form PFORM as shown in the table below.

lexical features:	LU, G, C
morphological features:	C, TNS, VTYP, NB, PFORM

Generalisation consists in replacing sub-sequences of an example with reductions (i.e. variables constraint by a set of features). Reductions in translation templates disregard the fixed features while keeping track of the variable ones. For instance, from French/English translation examples (1) and (2) below a generalised case (2g) can be inferred.

	French expression		English expression
1	$(ski)_{NOUN}$	$\longleftrightarrow$	$(ski)_{NOUN}$
2	$(station\ de\ ski)_{NOUN}$	$\longleftrightarrow$	$(ski\ station)_{NOUN}$
2g	$(station\ de\ \mathcal{X}_{NOUN})_{NOUN}$	$\longleftrightarrow$	$(\mathcal{X}_{NOUN}\ station)_{NOUN}$

The translation template (2g) will match a number of chunks such as *station de sport, station de taxi* etc... where the fillers of the slot $\mathcal{X}_{NOUN}$ are constraint by a set of features to be shared with *ski*, e.g. the feature *NOUN.*

These sequences would be translated in the absence of full matching cases into *sport station, taxi station* etc....

More than one reduction within a translation template is possible if different sequences can be reduced. The generalisation 3a, for instance has reductions $\mathcal{X}_{DP,NOM}$ and $\mathcal{Y}_{DP,ACC}$.

3a. $(\mathcal{X}_{DP,NOM}\ \textit{eat}\ \mathcal{Y}_{DP,ACC})_S \longleftrightarrow (\mathcal{X}_{DP,NOM}\ \textit{essen}\ \mathcal{Y}_{DP,ACC})_S$

3b. $(\mathcal{X}_{DP,NOM}\ \textit{eat}\ \mathcal{Y}_{DP,ACC})_S \longleftrightarrow (\mathcal{Y}_{DP,ACC}\ \textit{essen}\ \mathcal{X}_{DP,NOM})_S$

During **translation**, the morphological analysis of a new sentence is matched against the examples in the CB. Those sequences of the sentence which match one or more translation examples are reduced to one node. The newly created reduction keeps track of the of the matching example(s) number i and their properties such as *NOUN* or *ACC*. Whereas the properties of the matching example are visible in the generalisation, the number(s) i only serve in the refinement step to determine the internal structure of the target language chunk to be generated. The input sentence – thus generalised – is then matched (again) against the CB until no more reductions can be performed or the entire sentence is reduced to one node only. Depending on the type of the example, different features are percolated into the external constraints of the reduced nodes as shown in the following table.

phrase type	tag type	external constraint
adverbial phrase	*ADV*	–
adjective phrase	*A*	NB, CASE
noun phrase	*NOUN*	NB, CASE
determiner phrase	*DP*	NB, CASE, SPEC
prepositional phrase	*PP*	NB, CASE, SPEC, PFORM
sentence	*S*	TNS, VFORM

4 Linkage of CAT2 and EDGAR

In the CAT2-EDGAR Experiment we linked the CAT2 dynamically to EDGAR in such a way that EDGAR comes first into play after the morphological analysis and before the syntactic analysis performed by CAT2 and secondly, during generation, after the syntactic generation and before the morphological generation. In such an architecture, EDGAR serves for CAT2 as an intelligent multi-word and phrase translation front end, whereas CAT2 performs translation of linguistic structures which are beyond the capabilities of EDGAR.

EDGAR uses only morphological and syntactic information of translation examples which can be acquired automaticallys; as a consequence, it is easy to tune and to extend to new domains. On the other hand, CAT2 focuses on syntactic and semantic principles which underlie the languages involved. When EDGAR fails to find an appropriate translation example, CAT2 comes into play and covers the text with a default translation. If a user prefers a translation different from the default translation, a suitable translation example can be added to the CB. Subsequent translations will then use this translation example instead of the default translation. According to the contens of the CB, the interaction of EDGAR and CAT2 may have one of the following shapes:

- EDGAR segments the entire input text into autonomous chunks. In this case, the (reduced) chunks need not pass through CAT2 at all.
- EDGAR cannot find any segments to be reduced in the input text. In this case, the source text is transmitted to CAT2 to be processed as usual.
- EDGAR can partially reduce the input text by matchin it against the CB. In this case, both the reduced chunks and the remaining unrecognised text elements are transmitted to CAT2. In generation, EDGAR re-generates only those target language parts that it has reduced previously.

In this architecture, an MT system adapts itself dynamically to the data which the user enters into the CB and the texts encountered: while a complete match of cases in a sentence converts the system into a TM, in the next sentence the system may return to a purely rule-based treatment, or combine the two approaches.

As for the chunks obtained from EDGAR, they remain "lexically sealed" for CAT2. This means that CAT2 considers the TUs that come from EDGAR as single nodes, disregarding their internal lexical structures. CAT2 may or may not assign some grammatical features to the target side of the chunks in order to guide adaptation. The lexical content of these units remains unchanged and thus does not affect their reliability. CAT2, on the other hand, is bound to operate faster and in a more robust way, if for no other reason than simply because it has fewer units to handle.

5 Example Translations

In this section we show how the hybrid EDGAR-CAT2 system translates some phrase types, given the following CB:

CB1	$(\textit{man})_{NOUN}$	↔	$(\textit{Mann})_{NOUN}$
CB2	$(\textit{newspaper})_{NOUN}$	↔	$(\textit{Zeitung})_{NOUN}$
CB3	$(\textit{a man})_{DP}$	↔	$(\textit{Ein Mann})_{DP}$
CB3G	$(\textit{a}\ \mathcal{X}_{NOUN})_{DP}$	↔	$(\textit{Ein}\ \mathcal{X}_{NOUN})_{DP}$
CB4	$(\textit{The newspaper})_{DP}$	↔	$(\textit{Die Zeitung})_{DP}$
CB4G	$(\textit{The}\ \mathcal{X}_{NOUN})_{DP}$	↔	$(\textit{Der}\ \mathcal{X}_{NOUN})_{DP}$
CB5	$(\textit{The old man})_{DP}$	↔	$(\textit{Der alte Mann})_{DP}$
CB5G	$(\textit{The old}\ \mathcal{X}_{NOUN})_{DP}$	↔	$(\textit{Der alte}\ \mathcal{X}_{NOUN})_{DP}$
CB6	$(\textit{for the man})_{PP}$	↔	$(\textit{für den Mann})_{PP}$
CB6G	$(\textit{for}\ \mathrm{DP})_{PP}$	↔	$(\textit{für}\ \mathcal{X}_{DP})_{PP}$
CB7	$(\textit{The old women})_{DP}$	↔	$(\textit{die alten Frauen})_{DP}$
CB8	$(\textit{secretary of state})_{NOUN}$	↔	$(\textit{Staatsminister})_{NOUN}$
CB9	$(\textit{on the table})_{PP}$	↔	$(\textit{auf dem Tisch})_{PP}$
CB10	$(\textit{day after day})_{ADV}$	↔	$(\textit{Tag für Tag})_{ADV}$
CB11	$(\textit{The man reads the newspaper every day.})_{S}$	↔	
	$(\textit{Der Mann liest jeden Tag die Zeitung.})_{S}$		

As for notational conventions, we underline units which EDGAR recongnises and reduces to one node during the chunking step (C). The resulting generalisation (G) is translated by CAT2 into German. The reduced nodes in CAT2's translation output (T) are then refined by EDGAR (R).

Example 1

C: $\underline{\textit{The old man}}$ is selling $\underline{\textit{the secretary of state's}}$ car .

G: $\mathcal{X}^{6}{}_{DP,NOM;ACC,DEF,SG}$ is selling $\mathcal{Y}^{5g/8}{}_{DP,GEN,DEF,SG}$ car .

T: $\mathcal{X}^{6}{}_{DP,NOM,DEF,SG}$ verkauft den Pkw $\mathcal{Y}^{5g/8}{}_{DP,GEN,DEF,SG}$.

R: $\underline{\textit{Der alte Mann}}$ verkauft den Pkw $\underline{\textit{des Staatsministers}}$.

The chunk $\underline{\textit{The old man}}$ matches CB5 $(\textit{the old man})_{DP}$ and is reduced into the node $\mathcal{X}^{6}{}_{DP,NOM;ACC,DEF,SG}$. The chunk $\underline{\textit{the secretary of state's}}$ is recognised in two successive steps of generalisation. On a first level $\underline{\textit{secretary of state's}}$ matches CB8 $(\textit{secretary of state})_{NOUN}$ and is reduced into the node $\mathcal{Y}^{8}{}_{NOUN,GEN,SG}$. Notice that *state* and *state's* differ only in their morphological analysis by the case feature. While *state's* can only be genitive, *state* can not. As outlined above, the CASE feature – here having the value *GEN* – is taken from the matching chunk and percolated into the reduction. On a second level of generalisation, the chunk $\underline{\textit{the } \mathcal{X}^{8}{}_{NOUN,GEN,SG}}$ matches CB4G *(the* X_{NOUN}*)*$_{DP}$. Given the CB, no more reductions can be computed. The resulting generalisation is thus passed to CAT2 for translation.

Because only the first constituent can be the subject of the sentence, CAT2 is able to disambiguate the CASE feature. CAT2 parses the $\mathcal{Y}$ node as a pre-nominal modifier which in German can be realised as a post-nominal genitive. CAT2 recognises the progressive *is selling* and translates it into the German presence *verkauft.* The resulting structure is then, again, passed to EDGAR for specification and refinement of the reduced nodes.

Example 2

C: $\underline{\textit{The old men}}$ sell cars .

G: $\mathcal{X}^{6-7}{}_{DP,NOM;ACC,DEF,PLU}$ sell cars .

T: $\mathcal{X}^{6-7}{}_{DP,NOM,DEF,PLU}$ verkaufen Autos.

R: $\underline{\textit{Die alten Männer}}$ verkaufen Autos.

In contrast to Example 1 the chunk $\underline{\textit{The old men}}$ matches CB5 *(The old man)*$_{DP}$ only lexically due to the different number of *men* and *man.* The chunk is, however, identical with CB7 *(The old women)*$_{DP}$ with respect to morphological features, and so both CB6 and CB7 are used as reference translations. CAT2 translates the remaining items $\mathcal{X}^{6-7}{}_{DP,NOM;ACC,DEF,PLU}$ *sell cars.* and dictates the case of the object chunk *NOM.* EDGAR then merged together the lexical and the morphological features of the target language reference translations and refines the merged chunk according to the dictated case.

Example 3

C: $\underline{\textit{The old woman}}$ is waiting $\underline{\textit{for the old man}}$.

G: $\mathcal{X}^{7-6}{}_{DP,NOM;ACC,DEF,SG}$ is waiting $\mathcal{Y}^{6g/6}{}_{PP,NOM;ACC,DEF,SG}$

T: $\mathcal{X}^{7-6}{}_{DP,NOM,DEF,SG}$ wartet $\mathcal{Y}^{6g/6}{}_{PP,ACC,DEF,SG,auf}$

R: $\underline{\textit{Die alte Frau}}$ wartet $\underline{\textit{auf den alten Mann}}$.

The chunk *The old woman* is found in a similar way to *The old men* in Example 2. Morphological features are matched onto CB5 *(The old man)*$_{DP}$ and lexical features are matched onto CB7 *(The old women)*$_{DP}$. *For the old man* is chunked in two generalisation steps. First *the old man* is matched onto CB5 which yields the reduction $\mathcal{X}^{5}{}_{DP,NOM;ACC,DEF,SG}$. In a second step of generalisation the chunk matches CB6G *(for* $\mathcal{X}_{DP}$*)*$_{PP}$. A sentence reduced to four nodes is passed to CAT2.

CAT2 translates the progressive *is waiting* into simple German presence as in Example 2. Further, the node $\mathcal{Y}^{6g/6}{}_{PP,NOM;ACC,DEF,SG}$ which represents the prepositional phrase *for the old man* is assigned the semantic role *THEME* as argument of *wait*. The German translation requires for the theme of this verb the preposition *auf* (i.e. *warten auf*) and the accusative case. When refining the target language side of the CB6G *(für* $\mathcal{X}_{DP}$*)*$_{PP}$, EDGAR replaces the preposition *für* with the preposition *auf* based on the information provided from CAT2. In this way the correct preposition can be assigned to an argument PP.

Example 4

C: *The man* put the book *on the table* .
G: $\mathcal{X}^{4g/1}{}_{DP,NOM;ACC,DEF,SG}$ put the book $\mathcal{Y}^{9}{}_{PP,NOM;ACC,DEF,SG}$.
T: $\mathcal{X}^{4g/1}{}_{DP,NOM;ACC,DEF,SG}$ stellt das Buch $\mathcal{Y}^{9}{}_{PP,ACC,DEF,SG}$.
R: *Der Mann* stellt das Buch *auf den Tisch*.

The chunk *The man* is reduced to one node in two generalisation steps: First *man* is reduced based on CB1 and then CB4G matches the entire chunk. *On the table* has a complete match in CB9. The reduced sentence is then translated in CAT2 where the node $\mathcal{Y}$ receives the semantic role *LOCATION* and the case *ACC*. In contrast to the role *THEME*, no specific preposition is dictated from CAT2. The default preposition is thus taken from the CB while specifying the target example.

Example 5

C: *Day after day* *the man* buys *a newspaper*.

G: $\mathcal{X}^{10}{}_{ADV}\ \mathcal{Y}^{4g/1}{}_{DP,NOM;ACC,DEF,SG}$ buys $\mathcal{Z}^{3g/2}{}_{DP,NOM;ACC,INDEF,SG}$.

T: $\mathcal{X}^{10}{}_{ADV}\ \mathcal{Y}^{4g/1}{}_{DP,NOM,DEF,SG}$ buys $\mathcal{Z}^{3g/2}{}_{DP,ACC,INDEF,SG}$.

R: *Tag für Tag* kauft *der Mann* *eine Zeitung*.

Day after day is recognised as an adverbial phrase *ADV*s and underline*the man* and *a newspaper* are recognised as determiner phrases *DP* as they respectively match CB10, CB4G/1 and CB3G/2. CAT2 translates the unreduced item in the sentence and generates an appropriate word order of the elements in the target language.

Example 6

C: *The man reads the newspaper every day .*

G: $\mathcal{X}^{11}{}_{S}$

T: $\mathcal{X}^{11}{}_{S}$

R: *Der Mann liest jeden Tag die Zeitung.*

There is a complete match for the whole sentence in CB11, and there is no unreduced part left for CAT2 to translate. The entire translation is thus taken from the CB with no adaptation required.

6 Conclusion

The translation examples in the previous section show a possible linkage of a RBMT and an EBMT system: prerequisite for the fruitful integration of EDGAR and CAT2 is an appropriate adaptation ability of both systems. As shown in the example translations, in the refinement step EDGAR performs adaptation on agreement features in determiner and prepositional phrases and replaces prepositions in a prepositional phrase according to the values dictated by CAT2. Such a minimal adaptation capacity is required if sub-sentential translation examples (such as determiner or noun phrases) are used at one time in the subject position and another time in the object position or, as in Example 3 and 4 are once an argument of the verb and another time are a modifier.

On the other hand CAT2 needs to adapt to differently chunked input sentence and accordingly dictate EDGAR the appropriate adaptation features. In one extreme, EDGAR recognises none of the words in the input sentence and the entire translation process is carried out by CAT2, or the whole sentence can be recognised as one chunk as in Example 6.

In this case the reduction is just passed through CAT2 who need nothing add to it. This architecture ensures an optimal interaction of the two components where the full reliability of the EBMT component is enhanced by the adaptative power of the RBMT system, which ensures a high coverage even if the text is beyond the scope of the CBMT component.

References

Michael Carl, Leonid L. Iomdin, and Oliver Streiter. 1998. Towards dynamic linkage of example-based and rule-based machine translation. In *Proceedings of the ESSLLI '98 Machine Translation Workshop*.

Michael Carl. 1998. A constructivist approach to Machine Translation. In *Proceedings of NeMLaP3/CoNLL98*, pages 247–256, Sydney.

Oliver Streiter. 1996. *Linguistic Modeling for Multilingual Machine Translation*. Informatik. Shaker Verlag, Aachen.

Oliver Streiter. 1998. A semantic description language for multilingual NLP. In *Paper presented at the Tuscan Word Centre–Institut für Deutsche Sprache Workshop on Multilingual Lexical Semantics*, 19-21 June 1998. URL: http://www.iai.uni-sb.de/en/cat-docs.html.