Back to the Future? The Case for English-Swedish Direct Machine Translation

Lars Ahrenberg and Maria Holmqvist
Department of Computer and Information Science
Linköpings universitet
{lah, x02marho}@ida.liu.se

Abstract

Direct MT systems, unless of the statistical kind, are nowadays considered out of vogue. Yet, for a given language pair and text type what kind of system is required is largely an empirical and a practical question. In this paper we argue that in many cases of interest to MT, the structural shifts in going from English to Swedish are sufficiently restricted in numbers and kinds to allow for word-based transfer followed by rules of reordering. We further present an implemented direct system, and an evaluation of it on a restricted domain. Finally, we discuss some of its shortcomings and ways to alleviate them.

1 Introduction

Direct MT systems are nowadays considered to be something of the past. There is one exception to this, which enjoys much interest and a high status at the moment: Statistical Machine Translation (SMT). This is sometimes presented as a separate approach, but it has much in common with the classical direct approaches, in particular the reliance on word correspondences. The main difference is that SMT uses probabilistic translation models, whereas direct MT uses rules to determine the best translation. Hybrid systems are possible, however, and the system we will present below includes a probabilistic ranking module.

Of the three classical models for machine translation, the interlingua approach clearly stands out as an extreme and easily defined model. The difference between a direct approach and a syntactic transfer approach is more difficult to tell, at least in practice if not in principle. There are however some traits that tend to reappear when direct systems are characterized (cf. Vauquois, 1976). They are:

- they are designed for a specific language-pair and direction of translation;
- they exploit word correspondences and similarities between the two languages as far as possible, and use syntactic and semantic analyses only to the extent that it is necessary for the translation quality;
- the central data source of a direct system is the dictionary, where correspondences are stored; rules for solving ambiguities are closely linked to entries in the dictionary;
- they tend to work in a stepwise fashion, performing sequential substitutions of the input, often on a word-for-word basis;
- in matching an input sentence to the dictionary, the longest possible match is chosen;
- differences in word order are handled by special rules of reordering;

In contrast, syntactic transfer approaches are based on an analysis phase that aims at capturing a complete formal syntactic representation for the sentence as a whole. Once this information has been acquired it is also used as the object of transfer. This difference in the object of transfer: a sentence analysis vs. a set of word analyses may actually be regarded as the main difference between syntactic transfer systems and direct systems. But it must be stressed that direct systems do not preclude syntactic or semantic analyses. There is a pragmatic constraint on the analysis, though, that it is subordinated to the translation task.

Another difference concerns generation. A pure transfer system relies on a grammar for the target language to derive target sentences, while a direct system uses the word order of the source sentence as the point of departure for deriving a proper word order for the translation.

General requirements on MT systems such as modularity, separation of data from processes, re-usability of resources and modules, robustness, corpus-based derivation of data and so on, do not, in our view, provide conclusive arguments for either one of the models. In particular, there is a trade-off relating to re-usability and robustness. An accurate high-coverage syntactic parser for a source language is a valuable resource that can be re-used in other systems, but such resources are quite rare and most practical syntactic parsers perform well below 100%, which means that special mechanisms must be developed to make the system robust (cf. Weijnitz et al., 2004). A direct system on the other hand, relying on word-based analysis and transfer, will usually be able to derive some output for every input. The real issue, therefore, is empirical.

2 The Case for Direct English-Swedish MT

The choice between direct models and transfer models for a given language pair and text type must consider a number of empirical and practical issues. If the two languages are structurally similar, in particular as regards lexical correspondences, morphology and word order, the case for abstract syntactic analysis seems less convincing. As Dyvik (1995) puts it:

"If we can establish simple pointers between corresponding expressions in source or target language or get from one to the other by means of a few simple constituent order adjustments / … / we do not want to waste time finding a lot of redundant grammatical and semantic information about the expressions"

The question, then, is whether English and Swedish are sufficiently similar to warrant a simpler, direct model. Generally speaking, translation units correspond quite well for these two languages in many text types, which makes it possible to make a comprehensive analysis into a set of word alignments of a given sentence pair. There are of course a number of translation shifts to be found, but a large majority of them can be handled at the lexical level. Divergences and convergences at the word level, i.e., 1-n or n-1 relations, are normally connected and can be handled by proper tokenization. Deletions can be regarded as relations to a null lexical entry. The sentence pair below is thus regarded as a case of a one-to-one mapping involving the three token alignments <delete, ta+bort>, <the, NULL> and <file, filen>:

E: Delete the file
E’: delete - the - file
S’: ta+bort – NULL - filen
S: Ta bort filen

Additions are somewhat more problematic, but in many cases they may be handled as divergences. The following sentence pairs illustrate how we extend the lexicon to provide for additions:

E: [To] view data from …
S: [Om du vill] se data från …

E: [Using] a connection file …
S: [Om du använder] en anslutningsfil …

E: About [customizing] the layout of …
S: Om [anpassning av] layouten för …

This is not without problems, however, as the number of alternative translations for a unit increases and the problem of selection becomes harder.

By treating additions in this way, the number of cases that remain analyzed as additions become fairly low. In Table 1 we give data on additions for the Access XP online help files that we've used as domain and compares it with figures from another corpus, which has been aligned without MT in mind.

It should also be observed that grammatical morphemes correspond fairly well in numbers and use in going from English to Swedish. In particular, there are no additions in translations that are caused by English lacking some grammatical distinction that is obligatory in a Swedish sentence.

Changes in word order are fairly regular and appears to be at least as common in our text type as in translations of ordinary prose (see Table 1). The most common cases are shifts of the finite verb to 2nd position of the Swedish main clause, shifts of sentence adverb and finite verb in subordinate clauses, and shift of position of the label and the common noun in NP:s such as the following:

E: the Minimize action
S: instruktionen Minimera

E: the Employees table
S: tabellen Anställda

It should be noted that all of the mentioned changes can be accomplished by moving a single word to a different position. The number of sentences in the sample where two constituents with two or more tokens each are involved are about 40, i.e., about one in every ten sentences.

In Table 2 we show how additions and inversions at the word level are distributed in the sample. It can be seen that about half of all the sentences in the sample have been translated in a token-for-token fashion, including deletions.

These facts indicate that a direct model, supplemented with the necessary rules for selection and reordering, could be quite successful in translating from English to Swedish. There is one more problem, however, in addition to the cases where reordering applies to complex constituents. This concerns interactions in the translation of tokens that are in construction with one another, e.g., when an active clause is translated with a passive clause or vice versa. While in general this is not an obligatory shift, it is not uncommon in our corpus and needs to be handled somehow. We return to this problem in Section 3.4.

	Access XP	**Harry Potter[1]**
Sentence pairs	405	1,768
Word links	4192	23,610
Additions	122 (2.5%)	1,095 (4.6%)
Inversions	203 (4.8%)	658 (2.8%)

Table 1. Absolute and relative frequencies of token additions and inversions in two text samples. An inversion occurs for each link pair $<s_i, t_j> <s_{i+1}, t_k>$ which are adjacent on the source side but inverted on the target side, i.e., where $k < j$.

Category	**Frequency**
One-to-one mapping	201
Additions only	45
Inversions only	112
Additions and inversions	47
Total	405

Table 2. Distribution of translations on different categories according to the occurrence of additions and inversions. The same sample as in Table 1 is used.

3 The T4F system

The T4F system is a direct translation system based on resources extracted from parallel corpora. The acronym T4F stands for the main modules in the system: *Tokenization*, *Tagging*, *Transfer*, *Transposition* and *Filtering*.

The basic design principles are those listed in the introduction for direct systems. Another way to put it is that we try to restrict analysis and computations as far as possible to what is needed for the task at hand and use as simple structures as possible for the purpose. The basic structure is an array of (categorized) tokens, to which dependency relations may be added. The system is modular with a clear separation of data from processes. In addition, by using alignment tools in combination with the system's analysis modules, we want to derive the system's data bases directly from relevant corpus data. In the current

[1] The alignment of this sample was made by Sofia Helgegren for a study on the Swedish translations of the Harry Potter novels. Gaps in the alignment amounting to some 1-3% of all tokens were filled by one of the authors so as to make the alignment of this sample complete.

version, this is implemented for the data bases used in transfer (see section 3.2 below), while other data are created manually.

The basic idea is to categorize each token with a supertag, a tag containing inherent grammatical features as well as relevant contextual information. The supertag can represent properties of surrounding words, functional information, semantic categories or other information that influence the choice of translation for the token.

From word level alignments of supertagged parallel texts we create not only a dictionary of tokens, but also a dictionary of supertags, or matching contexts, which is used in the transfer module.

The inherent grammatical information in a supertag is selected from Functional Dependency Grammar tags containing part-of-speech and morphological features (Tapanainen and Järvinen, 1997). Contextual information is added to each tag by applying supertag rules, which operate on the morpho-syntactical tags and dependency relations provided by the FDG parser[2].

3.1 Phases of translation

T4F uses the following modules for translation of a source sentence:

1. Tokenization and tagging: Words in the source sentence are provided with supertags using information from the FDG parser and rules for adding contextual information (supertag rules). Below is an example of supertags for the words in the phrase “is used” in “When a connection file is used”. The word “used” is supertagged with the contextual feature *present tense* taken from the preceding “is”.

Inherent tags:
is/pos:v-fin:fin-aux:yes-tmp:pres
used/pos:en-fin:inf-act:pass
Supertags:
is/pos:v-aux:yes-fin:fin-tmp:pres-**type:be-zero:yes**
used/pos:en-act:pass-fin:inf-**tns:pres**

[2] The parser is used under license from the developers, Connexor. Recent versions of the parser are branded Machinese Syntax. For more information on it, we refer the reader to http://www.connexor.com/.

2. Transfer: Lexical transfer of source words. Every source word is looked-up in the word dictionary and those target words with contexts that match the source word’s supertag are chosen as possible translations. The supertags of alternative translations are found in the Swedish lexicon, and corresponding source and target supertags are found in the supertag lexicon (see section 3.2 below). For the phrase “is used”, the word dictionary contains the following alternatives.

is → {finns, ska:_:vara, det:_:är, innebär, NULL, är}
used → {används, den:_:används, användas}

Only three of the alternative translations of ”is” have supertags matching the context of this “is”. This is the target alternatives after transfer:

is → {ska:_:vara, NULL, är}
used → {används}

3. Filtering: Application of filter rules on alternative target words. Filter rules reduce the set of alternative translations by removing translations that do not fit the target context. In our example, the filter rules successfully removes the finite verb phrases from the remaining translations of “is”, resulting in:

is → { NULL }
used → {används}

4. Transposition: Application of order rules and expansion rules. Order rules are applied to change the order of tokens where this is necessary. Expansion rules are used to split multi-word units to allow reordering.

Two examples of order rules are given below. The first one will replace a null form of a verb with a non-null form, as in the translation of “does not see” with “ser inte”. The second will move a particle from the position of the verb to the right of any occurring sentence adverb.

IF tmp:pres-null:yes X tmp:pres THEN 3 → 1
IF pos:prt X pos:sa THEN 1>>3

5. Ranking: Ranking of remaining alternative target language sentences using translation probabilities from the word dictionary and a bigram

target language model derived from the training corpus.

Steps 3-4, filtering and reordering, can be repeated any number of times, until a satisfactory result is reached.

The T4F processing steps require a number of resources, they are:

Dictionaries	word form lexicon, supertag lexicon, English supertag lexicon, Swedish supertag lexicon
Rules	supertag rules, filter rules, order rules, expansion rules
Ranking	statistical translation model bigram language model

Data for dictionaries and statistical ranking are automatically derived from the aligned training corpus while the rules are handcrafted. The following section describes the contents of the dictionaries and how they were extracted from the corpus.

3.2 Lexicons

The dictionaries required by the T4F system are created *directly* from word-alignments of supertagged texts. Table 3 contains the dictionaries with a sample dictionary entry created from the alignment of *used-används* in the sentence pair below.

E: When a connection file is [used]
S: När en anslutningsfil [används]

Dictionary	Sample entry
Wordlinks	E word – S word (used – används)
Englex	E word – E supertag (used – pos:en-act:pass-fin:inf-tns:pres)
Swelex	S word – S supertag (används – pos:v-aux:no-fin:fin-mod:no-tmp:pres-voice:pass)
Superlinks	E supertag – S supertag (pos:en-act:pass-fin:inf-tns:pres – pos:v-aux:no-fin:fin-mod:no-tmp:pres-voice:pass)

Table 3. T4F dictionaries and their contents.

Alignments are made using two alignment tools, I*Link and I*Trix (Ahrenberg et al. 2003, Merkel et al. 2003).

I*Link is used to manually link corresponding segments in parallel texts. I*Link was used to align the first part of the training corpus to create high-quality basic dictionaries.

I*Trix is an automatic alignment system. Data from manual alignment with I*Link can also be utilized to improve automatic alignment with I*Trix.

3.3 Evaluation

The system was trained on sentences of length 2-20 words from Access XP help files. The automatic evaluation measures BLEU (Papineni et al., 2001) and NEVA (Forsbom, 2003) were applied during development of T4F to evaluate translation quality.

In the evaluations, we compare T4F with a fairly sophisticated statistical baseline system, built for the purpose, that includes tokenization, a simple target word look-up and the bigram-based statistical ranking procedure used in T4F to resolve the final translation. The baseline system uses the same lexical resources as T4F but does not apply the supertag constraints during transfer.

Part 1: Basic system

To create the core system, 600 sentence pairs from the help texts were manually aligned and used to create basic dictionaries for translation. The rule bases (supertag-, filter-, order- and expansion rules) were developed with these sentences in mind. Comparing the T4F system with the baseline system on translations of this part of the corpus gave the evaluation scores in Table 4.

System	BLEU score
T4F	0.71
Baseline	0.62

Table 4. BLEU scores for T4F and baseline system on first part of training corpus.

Part 2: Alignment test

The effect of word alignment accuracy on translation quality was measured in a test where 400 sentence pairs were aligned using three different setups, two automatic and one manual alignment. Expanding the basic dictionaries, created from the first part of the corpus, with lexicon data derived from the different alignments resulted in three sets of dictionaries. We name the sets, *Auto1*, *Auto2* and *Manual*, where the quality of the alignment increases for each set. The dictionaries in T4F are extracted directly from the alignments, which means that the quality of the alignment decides the quality of the lexicon. Figure 1 presents the BLEU scores for translation of the 400 source sentences with the three dictionary sets, comparing T4F system performance with the baseline system.

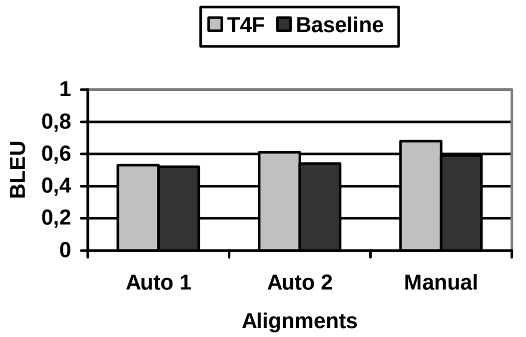

Figure 1. Alignment test

The results indicate that T4F performs better than the baseline as the lexicon quality improves. When the alignment is at its worst (*Auto1*), T4F is only slightly better than the baseline, but as alignment quality improves (*Manual*) the gap between T4F and baseline scores reaches that of the core system.

The alignment test shows that, with more accurate alignment, the difference between baseline system and T4F is accentuated. This implies that the supertag constraints and application of rule-bases (filter- and order rules) in T4F need sufficiently accurate dictionaries to have a positive effect on translation. The following example will illustrate this point by comparing the results and intermediary steps of translating a sentence with T4F using resources from *Auto1* and *Manual*:

E: This action can provide a visual indication that the macro is running

S: Den här instruktionen är ett visuellt tecken på att makrot pågår

T4F-Auto1: Den här instruktionen kan provide ett visuellt tecken som det makrot pågår
BLEU = 0.3928

T4F-Manual: Den här instruktionen är ett visuellt tecken på att makrot pågår
BLEU = 1.0000

The translation S is the reference translation. The translation produced from *Auto1* dictionaries has a rather low BLEU score (0.3928), while the high-quality dictionaries in *Manual* produce a perfect translation, which receives the BLEU score 1.

With the first set, *Auto1*, the words in the phrase “the macro” have the following target alternatives after transfer.

the → {alla, de, en, ingen, sökningen, den, det, posten, NULL, vilka, vilken}
macro → {verifieringsmakron, makro, makron, makrot, makrots}

After application of the rule bases, before the final ranking step, the following alternatives remain,

the → {alla, de, sökningen, det, posten, NULL, vilka}
macro → {verifieringsmakron, makro, makron, makrot, makrots}

The only filter rule used above was:

RM pos:det-gen:utr IF Pos[h]=gen:neu

This rule removes incongruent determiners, i.e. determiners with gender *utrum* that modify heads of gender *neutrum*. The fact that all alternative translations of “macro” have gender *neutrum*, allows us to remove determiners of a different gender below this position. Filter rules are generally stated as above, identifying token:tag-pairs that can be removed from a position depending on context.

Because of inadequacies in the automatic alignment, the lexicon will contain errors. In this example, we lack common syntactical features for the position of “the”. This position contains two nouns and therefore we cannot refer to the alternatives in this position as determiners. In effect, the filtering stops here, since we can not, for example, make use of the rule “Remove indefinite nouns above definite determiners”. The remaining alternatives must be handled by the statistical ranking, which produces the phrase ”det makrot”.

Lexical transfer of “the macro” with the manually aligned resources in *Manual* generates fewer target word alternatives:

the → {de, den, det, NULL, vilken}
macro → {makro, makron, makrot, i+ett+makro, makrots}

Several filter rules are applicable here. After removing incongruent determiner “vilken”, the first position consists of definite determiners, which allows removal of indefinite nouns in the second position according to the rule:

RM pos:n-def:ind IF Pos[d]=pos:det-def:def

When all filter rules have been applied, only two alternative translation remains:

the → NULL
macro → {makrot, makrots}

The two remaining combinations “NULL makrot” and “NULL makrots” are passed through the ranking module, which decides on the correct translation “NULL makrot”.

These translation examples show that in the current T4F system, even a moderate number of errors in the dictionary data have a strong negative effect on the applicability of filter rules.

Part 3: Effects of shifts

Using the same sample as in Tables 1 and 2, a test was made to see the effects of the occurrence of different types of shifts in the reference translation. The results are shown in Table 5.

Category	T4F	Baseline
One-to-one mappings	0,70	0,66
Additions only	0,57	0,55
Inversions only	0,65	0,45
Additions and inversions	0,48	0,42

Table 5. Comparison of T4F and the baseline system on a sample drawn from training data. The scores used are BLEU on the full sample.

As expected, translation quality gets worse the more shifts the reference translation includes. The most interesting result is that the greatest difference in performance between T4F and the baseline system occurs with reference translations that include inversions but no additions. This means that the order rules and expansion rules clearly gives T4F an advantage, when they can be applied.

Part 4: Full system on test data

The complete dictionary resources were created from automatic alignment of the remaining training data (4382 sentence pairs). The final system was used to translate 277 unseen test sentences. The evaluation result on the test corpus is presented in Table 6.

System	BLEU score
T4F	0.31
Baseline	0.30

Table 6. BLEU scores for T4F and baseline system on test corpus.

As can be seen, T4F and baseline translations score about equal when using dictionary resources derived from automatic alignments. The relatively low BLEU score on test data (≈ 0.30) compared to translation of comparable training data (≈ 0.50) can be attributed to:

- Lexical gaps: New words in the test data are missing in the T4F dictionary since they did not appear in the training corpus.
- Missing supertag links: Source words may have translations in the word dictionary, but a similar context (supertag) was not present in training data, and so, there are no matching contexts in the supertag dictionary.
- Statistical model: Lexical gaps and new words means that new unseen bigrams are introduced, which may result in inaccurate ranking of translations.

3.4 Dependencies in transfer

Many constructions can be translated in different ways. For example, an English *of*-genitive construction can be translated into Swedish by an isomorphic construction, or by an *s*-genitive, as in the following examples:

E: the contents of the file
S1: innehållet i filen
S2: filens innehåll

E: the borders of controls
S1: kantlinjerna på kontrollerna
S2: kontrollernas kantlinjer

As can be seen, the first alternative (S1) is isomorphic with the English original, while the second involves a reordering. It should also be noted that different forms are used in the two Swedish constructions.

A problem for a direct system such as T4F is that the dependencies in transfer are not captured. If we restrict the attention to the alternatives S1 and S2 only, and disregard the definite article, we have three binary choices to make[3]:

contents → {innehållet, innehåll}
of → {i, NULL}
file → {filen, filens*}

Without any extra machinery this gives us eight different possibilities to consider. While the ranking process could no doubt do some of the job of eliminating the six impossible combinations, we would prefer a solution that never considered them.

A similar problem concerns the choice between an active and a passive construction, as in the following example:

E: the groups you create do not affect source data
S: källdata påverkas inte av grupperna som du skapar

The translation produced by T4F for this source sentence is the active "Grupperna som du skapar påverkar inte källdata", which is fine in itself, but not what the translator produced. In fact, this case is similar to the previous one in that there are three main tokens involved:

affect → {påverka, påverkas+av}
groups → {grupperna, grupperna*}
source data → {källdata, källdata*}

Now, if we want the system to be able to produce all sensible alternatives in cases such as these, we would need access to rules that handle the dependencies in some way. This is the kind of case for which syntactic transfer rules are especially useful. In particular, as the passive and active constructions would constitute exclusive alternatives, they would not be mixed in transfer, but belong to separate paths.

To accomplish something similar within the limitations of T4F, we must first of all allow the creation of different paths on the target side. This means that we would split the set of alternatives for a given position into subsets, and create two separate schemas to work on, one for each subset. The split should occur only when the alternatives for the position carry features that are mutually exclusive, such as active and passive. Similarly, the NULL translation of the genitive use of *of* could be assigned a feature that sets it apart from other prepositions, and this feature could be used to split the set of alternative translations for *of*. This is not sufficient, however. Once the split has been made, these features must be used to identify the tokens whose translations depend on it, so that they are treated the right way. In essence this means adding filtering and ordering rules that refer to this particular feature and whatever syntactic information is relevant

4 Conclusions

We have provided a restricted and preliminary analysis of the text type of on-line help texts in terms of the occurrence of translation shifts. Slightly less than half of the sentences in the chosen sample have been translated in a one-to-one fashion and a large part of the remaining sentences with the aid of a single problematic structural shift. This indicates that a direct machine translation system could go a long way to automate the translation from English to Swedish of this text type.

We sketched the architecture of a direct, corpus-based English-Swedish MT system called T4F, which has been evaluated on the chosen text type. The evaluation shows that T4F has a much better performance than a statistical baseline system on the training data. In particular, the order

[3] The asterisk on an alternative is used here for expository purposes to indicate that this form cannot be used without being transposed to a different position from that of its source word.

and filter rules improve performance considerably when changes of word order are necessary. However, the performance of T4F is highly dependent on accurate alignment. With automatic alignment, T4F does not perform any better than the baseline system. Also, performance on test data is markedly worse than on training data. The difference can to some extent be explained by lexical gaps, but not completely. Thus, there is ample room for improvements. One line of work will investigate how filtering and ordering can be made more effective by allowing the search space to be split, when mutually exclusive syntactic constructions are competing alternatives for the same subset of source tokens. This, in essence, amounts to using the equivalents of syntactic transfer rules when the situation demands it. Another important line of work will investigate how the alignment process can be improved.

Acknowledgement

This research has been supported by VINNOVA (The Swedish Agency for Innovation Systems) as part of the project Corpus-based Machine Translation, KOMA, contract no. 2002-02407.

References

Lars Ahrenberg, Magnus Merkel and Michael Petterstedt. 2003. Interactive Word Alignment for Language Engineering. Project note at The 10th Conference of the European Chapter of the Association for Computational Linguistics April 12-17, 2003 Agro Hotel, Budapest, Hungary (EACL-2003).

Helge Dyvik. 1995. Exploiting structural similarities in machine translation. *Computers in the Humanities 28.*

Eva Forsbom. 2003. Training a Super Model Look-Alike: Featuring Edit Distance, N-Gram Occurrence, and One Reference Translation. I Proceedings of the Workshop on Machine Translation Evaluation: Towards Systemizing MT Evaluation, held in conjunction with MT SUMMIT IX, s. 29-36. New Orleans, Louisiana, USA, 27 september.

Magnus Merkel , Michael Petterstedt and Lars Ahrenberg. 2003. Interactive Word Alignment for Corpus Linguistics. Proceedings of Corpus Linguistics 2003. UCREL Technical Papers No 16, Part 1, pp. 533-542

Sergei Nirenburg, Harold Somers and Yorick Wilks. 2003. *Readings in Machine Translation*. Cambridge, Mass. The MIT Press.

Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jung Zhu. 2001. Bleu: a method for automatic evaluation of machine translation, IBM research report rc22176 (w0109- 022). Technical report, IBM Research Division, Thomas, J. Watson Research Center.

Pasi Tapanainen and Timo Järvinen. 1997. A non-projective dependency parser. Presented at ANLP-97 ACL, Washington, D.C.

Bernard Vauquois. 1976. Automatic translation - A survey of different approaches. Coling-76, Ottawa. (quoted from Nirenburg et al., 2003).

Per Weijnitz, Anna Sågvall Hein, Eva Forsbom, Ebba Gustavii, Eva Pettersson, Jörg Tiedemann. 2004. The machine translation system MATS – past, present and future. This volume.