[Translating and the Computer 23: papers from the Aslib conference held on 29 & 30 November 2001]

TopTrans: Interactive Machine Translation

S. Schachtl¹ and A. Fraser-Tillmann²

1 Senior engineer at Siemens Corporate Technology (CT), Department for Human Computer Interaction (IC5) Otto-Hahn-Ring 6, 81739 München, Germany Stefanie.Schachtl@mchp.siemens.de

2 Translation manager at Siemens Business Services (SBS), Training and Services (TS), Sprachendienst (SD), Hofmannstraße 51, 81359 München, Germany Audrey.Fraser-Tillmann@mchh.siemens.de

ABSTRACT

The paper will report on the interactive machine translation system TopTrans, that has recently been developed at Siemens Corporate Technology and is currently used for translating the user guides for Siemens mobile phones and the Gigaset telephones from German to English. This article points out the specific characteristics of interactive machine translation implemented in TopTrans (controlled language, interactive disambiguation, terminology support, stylistic selection etc.), describes the supporting software (translation memories, text filter and editor) and gives a summary of experiences in practical use with regard to the efficiency of using TopTrans in the current application. A demonstration of the system will be given in the talk.

1. INTRODUCTION

Within our department for Human Computer Interaction at Siemens Corporate Technology (CT IC 5) we offer a new translation service to our partners at SBS Language Services. Using our interactive machine translation system TopTrans we provide translations of user documentation from German to English. Over a period of more than a year now, this partnership proved to be mutually beneficial: As we were able to significantly reduce the time spent on producing high quality translations by employing the TopTrans system, Language Services found a reliable and fast provider and we at CT IC 5 were able to thoroughly check our translation software and ascertain its benefit.

The Toptrans system has developed over a period of three years, starting in 1996, combining the knowledge sources that had accumulated from previous work on knowledge based text understanding and speech translation¹. In addition to creating the engine for the machine translation component itself we implemented a set of features we thought might be especially useful for producing high quality machine translations.

¹See for example: Block, H. U. and Schachtl, S. (1992) or Block, H. U. (1993). For further development on speech translation see: Block, H. U., Gehrke, M. and Schachtl, S. (2000).

The TopTrans system as we use it today is equipped with:

- ► a lexical database with approx. 84.000 word pairs,
- ▶ software for comfortable on-/off-line acquisition of missing lexical entries,
- software for the development of document(ation) specific terminology,
- ► an engine for full machine translation from German to English
- ► text filters for Framemaker and HTML documents
- a graphical user interface that is specifically designed to support the task of interactive machine translation and
- specific software for the development and maintenance of our translation memories.

The TopTrans system does not incorporate any knowledge from domain modelling, this sort of knowledge is replaced by effectively supporting interactive user input (see the paragraph on 'syntactic disambiguation' below).

2. INTERACTIVE MACHINE TRANSLATION WITH TOPTRANS

In the following paragraphs we emphasise the special features implemented in order to achieve high quality output from the automated translation process:

Lexical acquisition

The lexical database of the Toptrans system can be updated by the user during the translation process or in a separate process. In order to put a new word into the lexicon, the user only needs to enter the typical usage of the word in German and in English. A short dialog with the system ensues to complete the information. The inflected forms of the lexical entries are generated automatically and stored in the lexical databases. The management of the lexicon (such as deleting or modifying entries) must be carried out in a separate process on these databases. The following gives an illustration of the lexical acquisition dialog:

User input:	Der A schreibt ein Wort ganz neu.	The A rewrites the complete word.
System:		Please enter infinitive:
User input:		rewrite
Result:	vmorph(ha, [ein Wort ganz neu	vmorph(eDel,rewrite,rewrote,rewritten),
	'schreiben ein Wort ganz neu'(Ev,A,POS, SPECS)	'rewrite@@ the complete word'(Ev,A,POS,SPECS)

Example 1: TopTrans - Lexical acquisition

The lexical acquisition tool of TopTrans is designed in such a way that the user may comfortably acquire lexical items that consist of several words and wordings with idiomatic usage (cf. Example 1, above). In our experience it is this kind of

information which is especially needed to achieve stylistically and terminologically correct translations. For the same reasons we decided against including an automatic word analysis in the TopTrans engine. Experience showed that translating nominal compounds automatically by splitting them up and translating them noun by noun nearly always failed to produce a terminologically correct translation. Therefore all compounds have to be entered in the lexicon as new words².

Controlled input

TopTrans machine translation does not operate robustly. If the engine cannot completely analyse a translation unit in the source language it does not translate it. Unlike conventional machine translation systems TopTrans has no tools for shallow (i.e. phrasal or robust) translation. In order to prevent the unit from appearing in German in the English text the TopTrans user has to rephrase it in the TopTrans editor, so that the TopTrans analysis accepts it (for information on the extent of required rephrasing, see the paragraph on 'Controlled German' below).

Original:	Jedem sein eigenes Mobilteil!	Original:	Jedem sein eigenes Mobilteil!
CDG:		CDG:	
Translation:	Jedem sein eigenes Mobilteil!	Translation:	Everyone has their own handset.



Syntactic disambiguation

A very important feature for achieving correct machine translations is to get all the grammatical relations within the translation unit straight. Unfortunately the engine's grammatical analysis alone cannot master this task, as natural language is bound to be highly ambiguous. For example, the German unit "*Das BIOS startet das Betriebssystem*' can mean both: "*The BIOS starts the operating system*," and "*The operating system starts the BIOS*". If the machine fails to hit on the correct interpretation of an ambiguous unit, post-editing the translation can be most troublesome.

In order to deal with this problem TopTrans creates a list with all interpretations of the unit that are grammatically possible before translating it. The user is presented with this list in the TopTrans editor, where the crucial parts are highlighted and the correct interpretation can then be selected by mouse click:

² This proved to be a highly strategic decision which cut off an entire application domain from the possible applications for TopTrans: The ratio potential in the translation of documents that are characterised by new and changing vocabulary such as newsletters can not be increased by the use of TopTrans. Too much effort would need to be put into terminological research for updating the lexicon which would not be offset by the effectiveness of the machine translation process.

Original:	z.B. können unvollständige Rufnummern mit einem "?" gekennzeichnet sein.
CDG:	z.B. können unvollständige Rufnummern <mark>mit</mark> einem "?" <mark>gekennzeichnet</mark> sein.
TopTrans:	mit verbindet "?" und Rufnummern (mit ",?" refers to 'Rufnummern')
	mit verbindet <u>"?" und gekennzeichnet</u> ('mit "?" refers to 'gekennzeichnet')
Translation:	e.g. incomplete phone numbers may be indicated by a"?".

Example 3: TopTrans - Syntactic disambiguation

The system saves those answers in an internal representation and uses them as additional information for the ensuing automatic transfer process. This information is also stored in a special translation memory (TopTrans Semantic Memory) for further use.

Lexical disambiguation

Picking out the correct translation for a specific word in a given context can be a highly demanding task. Notwithstanding all efforts spent on terminology by editorial boards and language services among others there will always remain a substantial amount of ambiguous words even within technical documentation. In order to tackle this task we developed a cluster of tools in TopTrans that are designed to support the interactive translation process in this respect:

For each document the user can output a list with all words unknown to the system and check the new translations with Language Services (if there is enough time).

If the document is terminologically very sensitive the user may also check all occurrences of nouns, verbs, adverbs and adjectives and their translations offered by the system before the actual translation process starts (We have, however, never used this feature as it is very time consuming).

The entries in the Toptrans lexical database can be supplemented with information concerning specific terminology lists. This information can be organised hierarchically so that the system will automatically select the translation that is most appropriate for a given document (Although in theory a very attractive feature, this kind of automatic preselection was soon overtaken by using the TopTranslator's Terminology', described below).

In the actual translation process the user is presented with a list of all possible translations for every ambiguous word that occurs in the text (if available, the translation is enriched with a German comment quoting German synonyms). The user can now select the required translation from this list: This step of disambiguation immediately follows the syntactic disambiguation.

Original:	z.B. können unvollständige Rufnummern mit einem "?" gekennzeichnet sein.	
CDG:	z.B. können unvollständige Rufnummern mit einem "?" gekennzeichnet sein.	
TopTrans:	"Rufnummer" ==> : "call number"	
	"Rufnummer" ==> : "telephone number"	
	"Rufnummer" ==> : "phone number"	
	"Rufnummer" ==> : "number"	
Translation:	e.g. incomplete phone numbers may be indicated by a"?".	

Example 4: TopTrans - Lexical disambiguation

If the correct translation is not listed it can be acquired on-line (see the paragraph on lexical acquisition above). Information about which translation was selected in which context is also stored in the TopTrans Semantic Memory.

To shortcut the task of lexical disambiguation we implemented the TopTranslator's Terminology' feature, which became one of the most useful tools of the system. It works as follows: In the step of lexical disambiguation the user may select a specific translation by double clicking the desired reading. This fixes the translation setting for the remainder of the document. If more than one translation for the word is needed, the user can also select a subset of the complete list and will only be presented with this reduced list henceforth. These settings can be stored after the translation process is finished. They constitute a reliable knowledge source for documents of the same type of documentation and can be profitably used in checking terminology with language services.

Depth of disambiguation

The user can set how strictly the steps of disambiguation are to be carried out in the translation process. There are four levels:

At level 0, the system lists all interpretations that are syntactically conceivable, even those interpretations which in fact the grammar would be able to eliminate correctly by its own means (e.g. whether "sich" in "er öffnet sich" (it opens) is an obligatory reflexive pronoun).

At level 1, the user enables the predefined generalizations in the grammar which are (nearly) always correct so that those questions need not be answered any more.

At level 2, generalisations which are relevant for the syntactic structure and which have been stored from former translation processes (TopTrans Semantic Memory, see below) are automatically accepted. This reduces the effort in answering the questions put by the system significantly. However, in order to achieve a satisfying result the TopTrans Semantic Memory must have already been well trained with similar documentation.

At level 3, generalisations on lexical disambiguation stored in the Semantic Memory are also accepted automatically. This feature is subordinate to the settings in TopTranslator's Terminology, described above).

At level 4, TopTrans can be operated like a full machine translation system, although not robustly. The user does not have to answer any questions on disambiguation and TopTrans translates all grammatically correct phrases automatically.

Generation of the target language

The TopTrans engine incorporates separate analysis, transfer and generation modules. All units that have been analysed successfully by the grammar and for which disambiguation has been carried out by the user are assigned a lexical semantic representation. The transfer module of TopTrans then automatically maps the source language representation to a semantic representation of the unit in the target language:

```
[jeder_pron(B,1,C),pronominal(B)=yes,number(B)=sg,
    ein_indef(D,3,E),modif(D, F),
'Mobilteil_norm'(F,4,G),number(F)=sg,
haben(H,B,F,2,I),
    voice(H)=active,mod(H) = [ind],tempus(H)= pres,status (H) = fin,
fullstop(H,5,J),stype(H)=prop,sigma(H,full_clause)].
```

TopTrans lexical semantic representation: English

Example 5: TopTrans - Semantic representation for source and target language

These representations contain sufficient information for the English grammar of TopTrans to generate the desired English translation of the unit. Using a separate generation component for the target language ensures that the system output always conforms to the rules of the target language grammar as have been implemented in TopTrans.

Stylistic selection

A good translation is characterised by the fact that its style is adapted to the topic and complies with the native speaker's feeling for language. Correct word order is especially important for the stylistic characteristics of the clause (functional perspective). Whereas certain word order regularities such as SPO (Subject Predicate Object) constitute a syntactic feature of English grammar, it is not specified in the grammar whether e.g. a subordinate clause is to precede or to follow the main clause or whether an adverbial phrase is fronted or not. However, the native speaker has strong preferences as to what is the correct word order in the given context. As it is next to impossible to sufficiently generalise these word order idiosyncrasies (the word order of the source language being no help) we designed the Toptrans generation component so that after the machine translation process, the user is presented with all word order variations for a given unit that are grammatically possible. Again the user has to select the most fitting translation from this list.

Original:	z.B. können unvollständige Rufnummern mit einem "?" gekennzeichnet sein.
CDG:	z.B. können unvollständige Rufnummern mit einem "?" gekennzeichnet sein.
TopTrans:	e.g. incomplete phone numbers may be indicated by a"?". incomplete phone numbers may be indicated by a"?" for example.
Translation:	e.g. incomplete phone numbers may be indicated by a"?".

Example 6: TopTrans - Stylistic selection

3. SUPPORTING TOOLS

In the course of the project we also designed and optimised software to support the interactive translation process. The chief results are shortly outlined in this chapter.

TopTrans translation memories

Every document that has been translated using TopTrans is automatically in a format that is adapted to feed a Translation Memory database: The TopTrans editor (see below) splits the text to be translated into individual translation units. When the translation process is finished, each unit of the document is stored in its original, rephrased and translated form. Although the TopTrans architecture provides an interface to commercially available translation memories we only use our own memories at the moment which meet the rudimentary translation memory functions without fuzzy match.

In addition, TopTrans provides a memory in which the semantic relations among clausal parts such as predicate - object and PP - attachment can be stored (see the paragraph on 'syntactic disambiguation' above). When this memory is enabled it helps reducing the questions that are put by the system in the disambiguation phase considerably. The effect of this memory is that it generalises across different syntactic constructions. For example, if somewhere in the documentation the clause "Unvollständige Rufnummern sind mit "?" gekennzeichnet' (incomplete phone numbers are indicated by "?") has been translated, the system stores the information that "mit"?" (by "?") is most likely to refer to "kennzeichnen" (indicate) wherever this relation is grammatically possible. This means that an ambiguous clause like "Deshalb wurden unvollständige Rufnummern mit "?" gekennzeichnet' will be automatically translated as "Therefore incomplete phone numbers have been indicated by "?" and not as "Therefore incomplete phone numbers with "?" have been marked'. Actually, machine translation in combination with TopTrans Semantic Memory can be seen as a sort of intelligent fuzzy match.

TopTrans text filter and editor

The GUI of the TopTrans editor follows examples of editors known from commercially available translation memories. The user is always presented with two units of the text at one time (the unit to be currently translated and its preceding unit), where each unit appears in its three manifestations: the original, the rephrased and the translated clause (whenever the relevant steps have been carried out), see Figure 1 below.

Datei Bearbeiten	Bearbeiler: Steff Einheiten, 64 Seiten: 000 0.36 0.61 0.06 0.11 zuletzt. Steff 26.3.2001 14:11
Stop	Fræftern 518 ihr Gigaset zu einer schnurichen reiefonanlagetse
Such V Such	Arweitern Sie Ihr Gigaset zu einer schnurlosen Telefonanlage:##
Aodineni, 🤤 🖓	
Verwendung:	
ligenes 🕺 🕅 🔬 🚲 🖓	
zurück vor	
zurück vor	ledem sein eigenas Mobilteilę
zuruck vor	
ZURŪCK VOR	
24/994 1. 1 Post 200	I KIT PALL US LOOK SET SUL (DOORASANA) WANDARD LITZMAN LITELSEF SCHUNG 30
zurück vor	NT PN H1 QL E OH Backganaig Verginden, Teren LE SE T Hille 30
zurück vor Prüfen	<pre>CB-Jeder hat sein eigenes Mobilteil#</pre>
	Har F. H. R. F. A. Britshind, Andmin, Man, P. A. Man, Man, Man, Man, Man, Man, Man, Man,
Prüfen	Har F. H. R. F. A. Britshind, Andmin, Man, P. A. Man, Man, Man, Man, Man, Man, Man, Man,
Prijfen Disambiguleren Obersetzen	Har F. H. R. F. A. Britshind, Andmin, Man, P. A. Man, Man, Man, Man, Man, Man, Man, Man,
Prifén Disambiguieren Überseizen	<pre>CB>Jeder hat sein eigenes Mobilteil#</pre>
Priden Disambignieren Obersetzen Vergessen	<pre>CB>Jeder hat sein eigenes Mobilteil#</pre>
Prüfen Disambiguleren Ubersetzen Vergessen U Ubers U Ubers U Ubers	<pre>CB>Jeder hat sein eigenes Mobilteil#</pre>
Prilén Disambignieren Oberseizen Ungessen UKA RE EE Wortste	<pre> <</br></br></br></br></br></br></br></br></br></br></br></br></br></br></br></br></br></br></br></pre>
Prilén Disambigateren Obersetzen Vergessen G Ubers UK E E E Wortiste	<pre> <</br></br></br></br></br></br></pre>
Prixen Disambiguieren Disambiguieren Ubersetzen Vergessen UA PE EE Vortiste Com* U	<pre>CB>Jeder hat sein eigenes Mobilteil# Everyone has their own handset# Everyehe has their own handset#</pre>
Prilén Disambigateren Obersetzen Vergessen G Ubers UK E E E Wortiste	<pre> <</br></br></br></br></br></br></br></br></br></br></br></br></br></pre>

Figure 1: TopTrans editor

The criteria according to which the TopTrans text filter splits the document into translation units are punctuation as well as layout information. The user can interactively combine or split units by pressing the relevant soft keys in the editor. The newly created units now form a single unit for the translation process and the proper layout will be reconstructed afterwards. Parentheses are always translated separately from the text surrounding them. After they have been translated they will be inserted in the target language either next to the translation of the word where they were located in German or at the end of the unit.

A special feature of the documentation we have to deal with is that graphics or variables are embedded in units like normal nouns. In this case a special symbol (<XX>) is displayed which is treated by the engine like a normal noun:

Original:	Neue Nummer mit <xx></xx> bestätigen.
CDG:	Neue Nummer mit <xx></xx> bestätigen!
Translation:	Confirm new number with <xx></xx>

Example	7: TopTrans	s - Text variables
---------	-------------	--------------------

Formatting information which is relevant for translation such as bold, italic etc. and font types is visualised in the editor. TopTrans treats this kind of formatting information as part of the lexical information of the word it characterises. Like other lexical information it migrates through the whole translation process and is represented in the target language at the corresponding word. This means that TopTrans translates this kind of formatting information synonymously and not according to position. The relevant font information is reconstructed according to its position in the source language only in cases where just part of the word is

highlighted (e.g. "Multifrequency"). If this fails the user will be informed of the fact there might be something wrong with the fonts and make the necessary changes in the editor.

General formatting information such as tab stops, markers or cross references is represented as '#' characters in the editor and will automatically be inserted in the translated clause according to their position in the original clause. These characters cannot be deleted. The markers for conditional text in Framemaker for example belong to this group. When TopTrans is used to translate documents with conditional text the user is always presented with the whole text, the conditions being reconstructed afterwards by the textfilter.

The TopTrans editor specifically supports the various steps of interactive translation by providing the following soft keys:

If this key is pressed the engine analyses the German translation unit and marks the result according to the colour metaphor red (not analysable) yellow (ambiguous) - green (unambiguous). Unknown words are listed in the interaction window below the units. They can be entered on-line.

Disambiguieren: If this key is pressed, a selection menu with the possible interpretations of the unit is displayed in the interaction window. The relevant parts of the unit are highlighted.

If this key is pressed the unit, is automatically translated into English and a selection menu with the stylistic variations of the translation is displayed in the interaction window.

: If this key is pressed the unit is automatically copied into the translation in its original form. This is useful for units that need not be translated, e.g. parts in the document that have already been localised etc.

The translation provided by Toptrans can be post-edited on-line in the translation window of the active unit (typical case: inserting or deleting the definite determiner). Instead of the soft keys there are also keyboard macros available. Usually the steps of interactive translation are not performed unit by unit but for (longer parts of) the whole document. For this purpose, the editor offers the menu commands 'Alles Prüfen' (check all), 'Alles Disambiguieren' (disambiguate all) and 'Alles übersetzen' (translate all). The TopTrans editor also provides status information on the document (how many pages does the document have in total and how many pages have already been checked, disambiguated, translated etc.), soft keys and search functions for simple navigation in the document and menu commands for enabling/disabling the translation memories and terminologies for the relevant documentation. There is also a soft key for lexicon look up.

When the translation process is finished the translated as well as the rephrased document can be output. The textfilter will reconstruct the original layout (up to now we have reliable filters for Framemaker and HTML). It is also possible to output an aligned version of the document without formatting information, in which the original, rephrased and translated units are listed one below the other:

Orig: CDG: Transl:	<xx> <xx> <xx></xx></xx></xx>
Orig: CDG: Transl:	Auswählen und bestätigen. Auswahlen und bestatigen. Select the menu item and confirm.
Orig:	Bisher gültige System-PIN (4-stellig; Lieferzustand 0000) eingeben.
CDG:	Bisher gültige System-PIN (4-stellig; Lieferzustand 0000) eingeben.
Transl:	Enter the currently valid system PIN (4 digits; factory setting: 0000) .

Example 8: TopTrans - Aligned output

4. EXPERIENCE IN PRACTICAL USE

Starting from March 2000 we have been using TopTrans to provide English translations of customer documentation for Siemens Gigaset and mobile phones in co-operation with our partners at SBS language services. We have translated about 750 pages of Framemaker documents using the system. The following table shows the development in the translation process over the year by giving the figures of the actual effort we invested for single documents. **Roman numbers** identify the documentation, alphabetic letters indicate the individual documents. The figures for **Effort MT** include the time it took the TopTrans user to proof-read the translated document's Contents and Index in Framemaker. **Effort QA** indicates the time needed for Quality Assurance (layout & language) of our translations by our partners:

Document	Period	Pages	Effort MT	Effort QA
Document la	June 2000	85	56 hours	10 hours
Document IIa	December 2000	70	48 hours	10 hours
Document IIb	January 2001	114	40 hours	6 hours
Document Ib	January 2001	90	48 hours	5 hours
Document III	February 2001	40	12 hours	4 hours
Document Ic	February 2001	72	24 hours	19 hours
Document IV	February 2001	13	6 hours	4 hours
Document Id	March 2001	110	40 hours	10 hours
Document le	August 2001	87	15 hours	2 hours

Table 1: Overview: TopTrans in practical use

As can be seen from the figures time savings increase considerably when several documents out of one documentation are translated. The better the terminological knowledge sources and the semantic memory, the greater the benefit of using TopTrans in translating documentation.

Efforts that concern the maintenance of the system itself are not included in this

table. To a small degree, the development of the system still accompanies the current work on translation, in order to incorporate further improvements that have been identified during operational use.

Controlled German

When we started the conceptual development of TopTrans in 1996, we designed a sketch for a Controlled Siemens Documentary German (CSDG) and proposed and discussed it first in our editorial boards, then in public³. This sketch was intended to serve as groundwork for an interactive machine translation system that followed examples from the English-speaking area⁴.

Necessary revisions to the original concept emerged when the TopTrans system had been piloted in co-operation with Language Services in summer 99. It became apparent that rephrasing the original German text according to the stipulations of CSDG took significantly more time than it would take to answer the additional questions the system puts when the language is not controlled. We therefore decided to free the system from all those restrictions on controlled German that we had put there in order to enforce less ambiguous input⁵. TopTrans German grammar now essentially describes all syntactic constructions that are permitted in normal standard language. Syntactic constructions that involve very complex ellipses such as gapping constructions must be rephrased by the user.

The engine of TopTrans is no black box. This means that it can easily be adapted to special requirements resulting from linguistic idiosyncrasies of new documentation. For example if the new documentation is characterised by the fact that it uses special syntactic expressions to express imperatives which are not mentioned in standard grammar the TopTrans grammar will be enriched with the appropriate rules describing the new phenomenon.

'Best of' mode of operation

With growing experience, different modes of operating the system have evolved that strongly depend on the status of the document:

When the lexicon and the semantic memory of the system are in good shape because it has already been used to translate other documents from the same documentation, the most effective procedure is the following: The entire document is automatically checked in batch mode and mapped against the Translation Memories. The TopTrans user now carries out the step of rephrasing the unanalysable units, updating the lexicon and simultaneously re-checks the rephrased units. When all units have passed TopTrans syntactic analysis the disambiguation phase follows. Enabling TopTranslator's Terminology and the Semantic Memory of the relevant documentation helps to reduce the necessary effort for this considerably. When the whole document has been made unambiguous for the system, the actual machine translation can be carried out in batch mode. The last interactive steps include selecting the correct word order variations and a

³ See: Schachtl, S. (1996a), Schachtl, S. (1996b). and Lehrndorfer, A. and Schachtl, S.

^{(1998).} ⁴ See for example: Farrington, G. (1996), Godden, L. and Means, L. (1996) or Nyberg, *E.* H.

⁵ For example the system now accepts all syntactically possible variations of expressing a condition in German (i.e. subordinate with finite verb in front position ("blinkt die Anzeige", "if the display flashes") or in end position with wenn or falls in complementizer position ("wenn/falls die Anzeige blinkt', "if the display flashes").

machine assisted overall check on formatting information. Finally the translated document is output, proof-read once by the Toptrans user and then sent to Language Services.

Starting on new documentation with TopTrans requires some effort in preparation for the actual translation to run smoothly. Ideally a few related documents and their approved translations are translated using TopTrans. During this preparation phase the lexicon will be adequately updated, TopTranslator's Terminology will be created and TopTrans Semantic Memory for the documentation will be fed.

If there is no time for thorough preparation, the TopTrans user must proceed with the actual translation of the document as if translating it manually, i.e. translating unit by unit from top to bottom. By continuously updating the supporting modules (lexicon, terminology and translation memories) during this work, the translation process becomes considerably faster as work progresses.

5. CONCLUSIONS

In using TopTrans in the way described here we violated one of the basic laws in translating technical documentation, which states that one is to translate **into** the mother tongue. Using TopTrans within our department, we have always translated **from** the mother tongue into the foreign language. However, we were able to demonstrate that a slightly extended quality assurance by a native speaker suffices for texts translated with a well trained TopTrans system to satisfy the highest quality requirements⁶.

The reason why we chose this way of operation was based solely on organisational considerations. There are no conceptual barriers against TopTrans being used by native English speakers. The rephrasings of the German text that are necessary do not put increased demands on what translators usually have to deal with and answering the questions in the disambiguation phase is everyday labour. But, it might increase the efficiency of using TopTrans when the lexical update and stylistic selection is carried out by a native speaker.

References

Block, H. U. and Schachtl, S. (1992). Trace & Unification Grammar. In Proceedings of the 14th International Conference on Computational Linguistics (COLING-92).

Block, H. U. (1993). Compiling Trace and Unification Grammar for Parsing and Generation. In Strzalkowsky, T., (ed.), *Reversible Grammar in Natural Language Processing.* Boston, Dordrecht, London: Kluwer. 155-174.

Block, H. U., Gehrke, M. and Schachtl, S. (2000). Adapting a Large Scale MT System for Spoken Language. In Wahlster, W. (ed.). *Verbmobil: Foundations of Speech to Speech Translation.* Berlin,

⁶ Quoting from our last QA comment: "In general, the manual was in good shape. I changed some syntactical oddities and there were a few mistakes (typos and repeats of words) that I straightened out, but no major surgery was required. The machine did a nice job!"

Heidelberg. Springer. 394-410.

Farrington, G. (1996). AECMA Simplified English: an Overview of the International Aircraft Maintenance. In *Proceedings of the First International Workshop on Controlled Language Applications (CLAW)*. Leuven.1-21.

Godden, L. and Means, L. (1996). The Controlled automotive Service Language (CASL) Project. In *Proceedings of the First International Workshop on Controlled Language Applications (CLAW).* Leuven. 106-114.

Lehrndorfer, A. and Schachtl, S. (1998). Controlled Siemens Documentary German and Toptrans. In *Technical Communicators Forum 3-98.*

Nyberg, E. H. and Mitamura, T. (1996). Controlled Language and Knowledge-Based Machine Translation: Principles and Practice. In *Proceedings of the First International Workshop on Controlled Language Applications (CLAW).* Leuven. 74-83.

Schachtl, S. (1996a). Kontrolliertes Deutsch. Anforderungen, Randbedingungen und Entwurf. In *tekom Herbsttagung. Zusammenfassung der Referate.* Stuttgart. 78-80.

Schachtl, S. (1996b). Requirements for Controlled German in Industrial Applications. In *Proceedings of the First International Workshop on Controlled Language Applications (CLAW)*. Leuven. 143-149.