KIELIKONE FINNISH-ENGLISH MT SYSTEM *TRANSMART* IN PRACTICAL USE

Arnola[1], Harri; Hyvönen, Kaarina; Juntunen, Jukka-Pekka; Linnanvirta, Tim;
and Suoranta, Petteri
Kielikone Ltd.
P.O. Box 126, 00211 Helsinki, FINLAND
e-mail: kkoy@kielikone.fi
WWW: http://www.kielikone.fi

This paper discusses certain practical requirements for machine translation products. An innovative generic machine translation technology (KMTech) is presented and a commercial machine translation system (TranSmart) which is based on that technology is introduced. The current version of TranSmart translates from Finnish into English. The paper evaluates TranSmart informally vis-a-vis the practical features discussed in the beginning of the paper.

1. THE PRACTICAL REQUIREMENTS FOR AN MT SYSTEM

Building an MT system is a strenuous task. It requires both theoretical insight and a great amount of practical work to compile the lexicons and design the translation rules for the system. In the field of commercial MT applications there have been some unfortunate failures in which attention was focused on theoretical aspects and possible user requirements were ignored. As a result, nothing workable came out from the efforts. In this paper we describe the Kielikone Machine Translation technology which has taken practical user requirements seriously early on. User requirements were considered as sign posts for the theoretical game.

What, then, are the practical features a user might appreciate in an MT system? Below we list four requirements in random order.

Efficiency. There are certain limit requirements for the translation speed of an MT system. One cannot give a precise figure since the speed requirement depends on the function and set-up of an MT system. Yet, one might say that if the translation of a single sentence takes several seconds, the applicability of the system is quite limited; if it takes dozens of seconds to translate a sentence, it is next to impossible to find any application for the system.

Quality. Translation quality is probably the most important practical parameter for an MT system. Unfortunately it is very difficult to measure translation quality in an objective manner. Nevertheless, for a given task there is a quality level which makes a difference in the sense that if a system produces quality beyond the level it is easier to find customers, and if the system produces quality which is below the level its chances

---

[1] Formerly Jäppinen.

of commercial success are very slim. Such a quality level cannot be described precisely and it depends on the type of intended use of the MT system.

Adaptation to workflow. An MT system translates sentences from one language to another. Translation is seldom an isolated process, but an integrated function in a large process of the information processing workflow. A case in point is technical documentation. Technical documents are created, edited, transmitted, stored, accessed, translated, again edited and so forth in a complex cyclical process. In such an environment translation is but one step on a long and bumpy road. Furthermore, documents are frequently dressed in a format which is easy to read, and different text editors have different formatting codes. Thus, it is important for an MT system to have such a structure that it can be smoothly adapted to a customer's information flow and document formats.

User friendliness. Since there are no fully automatic high quality MT systems, an MT system often has to take into account that a human user wants to review translated documents and to postedit them when needed. If a system supports postediting work well, slips in quality are more readily forgiven. If, on the other hand, error correction is a difficult task, even minor errors may cause irritation. Great attention should be paid to the needs of the user.

When designing its MT system TranSmart, Kielikone Ltd. has taken the practical requirements mentioned above very seriously. Potential users have participated in the development work from the onset.

Next, we discuss briefly the basic technology of Kielikone MT and the TranSmart system which is an application of that base technology. After the introduction of the system we return to the practical issues mentioned above and review how we have succeeded in them.

2. KIELIKONE BASE MT TECHNOLOGY

The strong industrial growth in the 20th century is at least partly due to economy of scale. To produce commodities it is often wise business practice to invest - even heavily - in the production plant. Increased automation lowers the marginal costs of the products, and there is a break even point in sales where initial costs are paid off and increased sales start bringing great profits.

The software engineering viewpoint of Kielikone MT technology borrows the basic tenet if not spirit of economy of scale. With considerable initial effort we designed a generic (language independent) virtual machine, an abstract **MT Engine,** which is the basic processing module of an MT system (Figure 1). The abstract MT Engine becomes a real functioning virtual machine when rules are written for a given task and a proper strategy is chosen for it. Such an MT Engine instance transforms linguistic trees for a specific translation subtask.

An **MT system,** then, has an extremely simple but flexible linear architecture. Translation (rather, transfer) is a chain of MT Engine applications (Figure 2). Each

engine application performs a certain subtask, contributing to the whole task of translating sentences of one language to another. The number of MT Engine applications used depends on how the translation problem is divided into subproblems. The architecture can be changed at any time by defining new subtasks and inserting new MT Engine applications in appropriate places in the chain. We call this technology **KMTech** (short for Kielikone Machine Translation Technology).

In terms of economy of scale, KMTech pays off in that the same MT Engine software can be used for each subtask required in a particular MT system. It pays off in other ways as well. The design of an MT system for another language pair requires less investment. Kielikone is willing to license KMTech to other MT system developers.

KMTech has an impact on efficiency, too. Assume that an MT system applies the MT Engine nine times (Figure 4). Any decrease in processing time due to optimizing of the code of the MT Engine means nine-fold savings in the processing time of the MT system.

KMTech does not commit the designer of an MT system to any specific linguistic theory. Yet, in parallel with the software design we have found the **dependency theory** of sentence structure particularly fitting for machine translation. The MT Engine is especially well suited for the transformation of dependency trees. We do not want to enter linguistic discussion here. Suffice it to imply the flavor of the dependency theory by exhibiting the dependency structure of the sentence *This is a simple tree* (Figure 3 a) and contrast it with a constituent structure of the same sentence (Figure 3 b).

A dependency structure does not indicate the configuration of a sentence like a constituent structure does. Rather, a dependency structure makes explicit the annotated modifying relations between the words in a sentence. Each word, except one, modifies one word, its regent. Hence the structure of a sentence is topologically a tree. Modifying (dependency) relations often have semantic content which is why a dependency structure comes closer to the logical structure of a sentence than a constituent structure.

In terms of machine translation techniques, KMTech applies the transfer approach. Translating a sentence means analyzing the sentence into its dependency structure and transforming it into a meaning-preserving target dependency tree. Then the target tree is transformed into a linear string of word forms according to the grammar of the target language.

3. TRANSMART FINNISH-ENGLISH MT SYSTEM

Kielikone's TranSmart Finnish-English system is a fully implemented MT system which relies on KMTech (Jäppinen et ai., 2; Jäppinen et ai., 3). Its basic architecture is shown in Figure 4. The shaded squares represent MT Engine applications. The analysis has two parts: the morphological analysis of word forms and the dependency analysis of sentence structures. As the figure shows, transfer activates the MT Engine seven times (lexical transfer three times) and synthesis two times.

4. PRACTICAL REQUIREMENTS REVISITED

After this introduction of KMTech and TranSmart we review the practical requirements discussed in the beginning of this paper to see how TranSmart satisfies them.

4.1 Efficiency

For an input sentence in Finnish TranSmart produces a morphological analysis of the word forms, parses the sentence, and executes the MT Engine nine times, seven times for transforming the Finnish dependency tree into its English equivalent and two times for English sentence synthesis (word forms and a linear string of word forms) (Figure 4). TranSmart translates several sentences per second on average on a UNIX platform. The precise speed depends on the type of the hardware. The parser produces only one tree, and only one tree is progressively transformed into a target tree. The time spent translating a sentence is a linear function of the number of words in the sentence.

We find the efficiency of TranSmart satisfactory for most conceivable purposes. The translation speed will accelerate automatically as processors become faster. We would like to generalize these figures by stating that an MT system built using KMTech will translate a few sentences per second, presuming that the parser of the system does not have exponential time behavior.

4.2 Quality

Translation quality is extremely difficult to measure accurately and objectively. Too often an MT system vendor claims that a system is, say, 87 % correct or accurate. In our opinion, such precise figures, even if given as a range of two integers, are completely misleading. There is no single integer number or even a range of them that would accurately tell something about such a fuzzy notion as translation quality.

Serious efforts have been made to evaluate MT systems using statistical methods combined with human judgment. One might, for example, use a statistically significant amount of data and rate the comprehensibility of target sentences on the one hand and their accuracy on the other, using, say, a scale of 1 to 7 as judged by human evaluators. Such results undoubtedly correlate with translation quality, but the results are not very objective. One person ranks translations differently from another. Furthermore, such an evaluation method is very strenuous and cannot be used often. One should also note that the quality of an MT system varies from one text type to another.

Kielikone has developed a simple evaluation method which does not attempt to establish absolute quality figures but tries to tell something about the maturity of the

system vis-a-vis different text types. We call this evaluation method **window evaluation.**

In window evaluation a representative text sample is translated twice with the system. The first translation is run without any corrections or tuning. Let us call this the **rough rough** translation of the text. After the first round the shortcomings of the translation are analyzed. Errors in the system are corrected and missing lexical or structural rules are added. After the corrections the text is translated again to produce what we call the **polished rough** translation. No postediting is performed. The customer gets both translations plus numerical data about the corrections that had been made. Using such data he or she should be able to determine 1) the current quality of the system vis-a-vis his or her text type without any customer-specific tuning, 2) the best possible quality the system can deliver, and 3) how far the system is from the final quality vis-a-vis this particular text type.

An example visualizes the method. Table 1 shows the text of a news item printed in a major Finnish newspaper. The original text was 56 sentences long. To save space, the table shows only the first 30 sentences. The text has not been pre-edited in any way. Notice that sentence number 7 has serious parsing errors and its translations are therefore incorrect (marked with an asterisk). In the rough rough column the words for which no translation equivalents were found in the system lexicons are printed in bold face. If the missing word is a bipartite compound and translation equivalents for each part were found in the bilingual lexicons, TranSmart performs a part-for-part translation as a default. These productively translated compound words are marked in italics in the rough rough column. In the polished rough column typography indicates the corrections as follows: lexical corrections are italicized and structural corrections (that is, corrections made in general translation rules) are in bold face. The polished rough translations have not been post-edited in any way.

TABLE 1: Example text in window evaluation

#	Source	Rough rough target	Polished rough target
1	**AURINKOENERGIAN VOIMALLA LIIKKU-VASTA VENEESTÄ PUUTTUU MOOTTORIN SÄKSÄTYS**	THE **SÄKSÄTYS** OF THE MOTOR IS MISSING FROM THE BOAT WHICH MOVES WITH THE POWER OF SOLAR ENERGY	THE *CLATTER* OF THE MOTOR IS MISSING FROM THE BOAT WHICH MOVES WITH THE POWER OF SOLAR ENERGY
2	**Kolmen vuoden kehittelyllä yhdistettiin suomalainen puuvene ja aurinkoenergia**	With the three year development the Finnish wooden boat and solar energy were connected	With the three year development the Finnish wooden boat and solar energy were *combined*
3	Tavallisen soutuveneen perässä on pikkuinen moottori.	There is a little motor behind an ordinary rowing-boat.	There is a little motor behind an ordinary rowing-boat.
4	Vene liukuu vedessä kuin unelma, tasaisesti, äänettömästi ja saasteettomasti.	The boat slides in water like a dream, evenly, silently and **saasteettomasti.**	The boat slides in water like a dream, evenly, silently and *without pollution.*
5	Sähkömoottori saa voimansa auringosta.	The electric motor gets its power from the sun.	The electric motor gets its power from the sun.
6	Suomalaisen soutuveneen on työstänyt aurinkoveneeksi	The Finnish rowing-boat has been worked to be a *sun boat* by	The Finnish rowing-boat has

#			
7*	Juha Nyman Särkisalosta. Hanketta on ollut kehittämässä myös Albican-verkostoyritys sekä tukemassa Nesteen aurinkoenergiaan keskittynyt tutkimusosasto.	Juha Nyman from Särkisalo. * The project has also been developed by the Albican *network company* and has been supporting the *research department* which has concentrated on the solar energy of Neste.	been modified into a *solar boat* by Juha Nyman from Särkisalo. * The project has also been developed by the Albican network company and has been supporting the research department which has concentrated on the solar energy of Neste.
8	"Tämä on kolmen vuoden kehittelyn tulos.	"This is the result of the three year development.	"This is the result of the three year development.
9	Hanke on edennyt pikkuhiljaa ja kevättalvella idea toteutettiin käytännössä.	The project has proceeded little by little and in late winter the idea was carried out in practice.	The project has proceeded little by little and in late winter the idea was carried out in practice.
10	Pääajatus oli yhdistää vanha suomalainen puuvene ja aurinkoenergia", selittää Nyman.	The main idea was to connect an old Finnish wooden boat and solar energy", Nyman explains.	The main idea was to *combine* an old Finnish wooden boat and solar energy", Nyman explains.
11	**Paneeli piiloon penkin alle**	**Panel to hiding place under bench**	**Panel to hiding place under bench**
12	Teknisesti aurinkovene vaikuttaa sangen yksinkertaiselta.	The *sun boat* seems very simple technically.	The *solar boat* seems very simple technically.
13	Aurinkopaneeli lataa akun, josta sähköinen perämoottori saa voimansa.	The solar panel loads the accumulator from which the electric outboard motor gets its power.	The solar panel loads the accumulator from which the electric outboard motor gets its power.
14	Paneelin voi jättää näkyviin tai työntää piiloon istuimen alle.	The panel can be left in sight or can be pushed to the hiding place under the seat.	The panel can be left in sight or can be *hidden* under the seat.
15	Vene saavuttaa noin kymmenen kilometrin tuntinopeuden ja yksi lataus riittää reilun tunnin ajomatkaan täydellä vauhdilla.	The boat will reach about the ten kilometre speed per hour and one charging is enough for the good hour drive on a full speed.	The boat will reach *the speed of about ten kilometres per hour* and one charging is enough for the good hour drive *at* full *speed.*
16	Nymanin aurinkoveneverstaaseen on kantautunut uutisia, joiden mukaan Ruotsiin on syntymässä hänelle kilpailija.	News according to which the competitor being born to Sweden to him have been carried in Nyman's *sun boat workshop.*	*News* according to which the competitor *is being born in* Sweden to him has *reached* Nyman's *solar boat workshop.*
17	Muista pohjoismaisista kilpailijoista ei ole tietoa.	There is no information about other Nordic competitors.	There is no information about other Nordic competitors.
18	"Ruotsin TV4 esitteli taannoin ruotsalaisyrityksen valmistaman aurinkoveneen.	"Sweden's TV4 demonstrated recently the *sun boat* made by the Swedish company.	"Sweden's TV4 demonstrated recently the *solar boat* made by the Swedish company.
19	Sen sähköratkaisut vaikuttivat tosin kovin alkeellisilta.	However, its electric solutions seemed very elementary.	However, its electric solutions seemed very elementary.
20	Myös Saksassa päin on tiettävästi muutama tämänsuuntainen hanke", sanoo Nyman.	There probably are a few such projects also somewhere in Germany", Nyman says.	**Also somewhere in Germany there probably are a few such projects",** Nyman says.
21	**Tavoitteena Keski-Eurooppa**	**As objective Central Europe**	**As objective Central Europe**
22	Aurinkovene on ollut esillä muutamilla venemessuilla.	A few *boat exhibitions* have had a *sun boat* up for discussion.	A *solar boat* has been *shown at* a few boat *exhibitions.*
23	Suomalaiset messuvieraat	The Finnish *exhibition guests*	The Finnish exhibition guests

	ovat jonkin verran vieroksu-neet Nymanin venettä.	have shunned Nyman's boat a little.	have shunned Nyman's boat a little.
24	"Perussuomalainen soutuveneen ostaja on aika konservatiivinen.	"The *basic Finnish* rowing-boat buyer is quite conservative.	"The basic Finnish rowing-boat buyer is quite conservative.
25	Yksi kommentti on kuulunut, että veneemme on digitaa-leilla pilattu puuvene."	One comment has been that our boat is a wooden boat that has been spoiled with **digitaali."**	One comment has been that our boat is a wooden boat that has been spoiled with *digitals."*
26	Sen sijaan keskieurooppalaiset ovat olleet kovasti innostu-neita.	Instead the Central Europeans have been very excited.	Instead the Central Europeans have been very excited.
27	"Etenkin saksalaiset ovat olleet ihastuneita aurinko-veneen ympäristöystävälli-syyteen."	"Especially the Germans have been attracted to the kindness to the environment of the *sun boat."*	"Especially the Germans have been attracted to the kindness to the environment of the *solar boat"*
28	Nykyisellään Nymanin vene ei kuitenkaan kelpaa Keski-Eurooppaan.	As it is now Nyman's boat does not suit to Central Europe, however.	As it is now Nyman's *boat is not suitable,* however, *for* Central Europe.
29	Siellä päin veneitä säilytetään pääasiassa telillä, koska vedessä olevia venepaikkoja on niukasti.	There the boats are retained mainly on the spindle because there are scantly *boat places* in water.	There the boats are *kept* mainly *on* the spindle because there are scantly *places for boats* in water.
30	"Jos nykyisen veneen nostaa aina vesireissun jälkeen telille, puu kuivuu ja seuraa-vassa vesillelaskussa vene täyttyy vedellä." ...(cont.)	"If it lifts the present boat always after the **vesireissu** to the spindle, the tree will dry and in the following launch the boat will become Ml of water." ...(cont.)	"If **one lifts the present boat** always after the *boat trip* to the spindle, the tree will dry and in the following launch the boat will become full of water." ...(cont.)

The polished rough translation in Table 1 does not yet represent the final say of TranSmart, since certain phenomena (such as the proper assignment of articles or the proper ordering of adverbials) have not been handled conclusively yet. There are several such errors in the polished rough translation. In addition to this, sentence number 30 has a fatal word sense error. In Finnish the word *puu* means tree, wood, or timber, depending on the context. In the translation the wrong equivalent *tree* is chosen *(the tree will dry...* should read *wood will dry...).* This error cannot be corrected by any general method and the error is therefore left as it is.

The text called for 18 lexical corrections and the following two structural (general) corrections: in sentence number 20 the surface ordering of adverbials was corrected and in number 30 the translation of a passive Finnish sentence was corrected so that the formal subject *one* is used instead of the pronoun *it.* Both corrections are general.

The distance between the rough rough and the polished rough translation is shown in Table 2. The distance is a real number between 0 and 1. It is calculated simply by dividing the number of sentences requiring corrections by the total number of sentences. Figure 5 shows the distances graphically. The evaluation tells that the general translation rules cover this text type well (the distance is 0,07), but the lexical distance is quite long - 0,5 - which means that every second sentence requires lexical correction. The reader has to decide for him/herself how good the absolute quality of the rough rough and polished rough translation is.

TABLE 2: Quality distance of the example text

Correction type	Domain-specific	Generic	Combined
New or corrected words	7	11	18
New or corrected structural rules	0	2	2
Lexical distance	0.23	0 33	0 5
Structural distance	0	0 07	0 07

4.3 Adaptation to workflow

In industrial environments translation is only rarely an isolated function. Document processing at a customer site often sets various adaptation requirements for an MT system if machine translation is to be seamlessly integrated into the whole workflow. TranSmart can be adapted to customer needs in several different ways.

4.3.1 Integration to information processing systems

The first application of TranSmart was a UNIX workstation version (Jäppinen et ai., 2; Jäppinen et ai., 3). This solution aims at making postediting as easy as possible. However, the workstation solution has some drawbacks. A customer may have several employees who need translations only occasionally. A UNIX workstation is too costly a solution for such a user configuration. We created, therefore, also a client/server solution of TranSmart in which the employees of an organization can send documents to TranSmart server via the organization's internal e-mail system. There is also a client/server solution which integrates an external word processing program (Microsoft Word) with a TranSmart server. All these system configurations are in use at our customer sites.

4.3.2 Integration to document formats

A document produced using a modern word processing program often contains a large amount of non-linguistic formatting information. When implementing a machine translation system preserving such data is of crucial importance for two reasons:

1. The graphical form of the translated document is generally expected to resemble the source document as closely as possible.

2. An MT system often achieves better translation quality using this extralinguistic information.

Usually the author of a source language document has put considerable effort into formatting. If the document is converted and translated in raw ASCII format, all this work must be re-done, often at a high cost.

It is important to realize that while the machine translation process cannot possibly preserve **all** formatting and layout properties, it is well worth the effort to try to preserve as much as possible. In Finnish-English machine translation (or even hand translation), pagination is often lost because the translation is typically longer than the original. But it is still much easier to re-paginate a poorly paginated translation which is in Word format than to make a Word document out of a plain ASCII file.

One of the most important areas in which a machine translation system can benefit from formatting data is the detection of sentence boundaries (rather, translation unit boundaries). Like most practical machine translation systems, TranSmart translates the source language text one sentence (or one translation unit) at a time, and to function properly it should determine the sentence boundaries correctly. Any algorithm for detecting sentence boundaries in natural language texts must involve heuristics and is therefore prone to occasional failure.

The formatting information of a word processor document makes the problem of detecting sentence boundaries simpler. For example, unaided, a whole table may be translated as if it were one long sentence. A much better translation result will be obtained if the table is pre-edited manually by inserting sentence boundary tags between the elements. Obviously, such manual pre-editing is costly, slow and error-prone.

The TranSmart MT system is able to preserve and utilize formatting information which is expressed in the RTF (Rich Text Format) document formatting language. Most word processing programs can read and write documents in RTF format and preserve most of the formatting information. TranSmart includes a RTF preconverter and a RTF postconverter, which preserve the RTF formatting information

- by separating, before translation, the formatting information from the natural language information and tagging the formatting information as non-translatable,

- by inserting sentence-boundary tags into the text in places where the RTF formatting indicates that a translation unit cannot continue (for example between a heading and a paragraph, and between the table cells), and

- by reconstructing, after translation, the formatting information which has been separated and tagged as non-translatable before translation.

There are additional benefits in using word processor format as the output format of an MT system. Typographical data can be used to facilitate the postediting process. For example, TranSmart uses ~~overstriking~~ to highlight a word for which no translation equivalent was found in the system lexicons, and <u>double underlining</u> to highlight a bipartite compound word for which no translation equivalent was found in the lexicons but which was translated by concatenating the translations of the parts of the compound (Figure 6). The user may choose if and how these warnings are indicated in the translation.

4.3.3 Integration of previous translations

It is typical of technical documentation that once the core document has been translated, only parts of it need modifying. When an MT system translates technical documents it is therefore important that the system is able to store, recognize and retrieve previous high-quality translations and translates anew only sentences with no previous translations available. A solution to this adaptation problem is a translation memory.

There is another reason to use translation memories. Fully automatic machine translation has quality limits. One way to overcome such limits is to access and adapt high quality human translations (or corrected MT translations) stored in a translation memory. Obviously, the extent to which translation quality is improved in such a hybrid system depends first and foremost on the amount of recorded data. Since in practice an exactly matching sentence is only rarely found in the memory, the added value of a translation memory depends also on how intelligently the translations of nearly matching sentences can be used.

Translation memories usually perform fuzzy matches on **strings** of words. TranSmart offers a linguistically intelligent translation memory which stores the dependency **trees** of source sentences. Matching trees rather than strings offers several benefits. To name one, dependency trees neutralize variations in constituent order in a natural way thus facilitating the adaptation of near-matches. Relying on dependency trees also makes it possible to further normalize the sentences before matching. Redundant or semantically irrelevant nodes can be removed. Also, a thesaurus or a synonym lexicon can easily be introduced into the matching process.

The closest matching translation found in the translation memory database may not be a perfect match. This means that as such the translation cannot be used for the source sentence. For conventional translation memories this normally means also that the close but not perfectly matching translation is left unused. The translation memory of TranSmart features a module for accommodating small lexical and attributical differences so that the translation can still be used (Juntunen, 1). For instance, if we are translating the sentence *"The man found the cats"* and the only matching translation in the translation memory happens to be *"The man found the dog",* TranSmart's intelligent translation memory is able to find out that the only non-matching word is a noun which is the Object in the sentence and since the dependency trees of the two sentences are identical from this point of view, it suggests the sentence *"The man found the cats"* as a match and even replaces the singular noun *"dog"* with the required plural noun *"cats"*.

4.4 User friendliness

4.4.1 Bilingual editor

The first version of TranSmart, the workstation application, was designed to make the user interface convenient for postediting purposes. Thus we developed a bilingual text editor which has two synchronized windows side by side, one displaying the source text, the other showing the translation (Figure 7).

The TranSmart editor utilizes linguistic information about the text. In addition to the usual text editing functions (insertion, deletion, searching/replacing etc.), the editor offers some linguistically intelligent editing functions:

- *Alignment of source and target language sentences and words.* When the user clicks a word on the target window, the editor underlines the corresponding source sentence and **highlights** the corresponding source language word.
- *Special function keys* for quickly correcting errors particular to Finnish-English translations such as English articles (Finnish has no articles and this poses problems in MT).
- *Target language synonym lexicon.* The user may select a target language word and ask for possible synonyms for it. For example, in the Finnish-English system the synonym lexicon displays *large* and *great* as possible synonyms for *big.* A selected synonym replaces the original translation.
- *Translation equivalent lexicon.* The user may select a target language word and ask for alternative translations for it. The system offers alternative translations for the corresponding source word. For example, the Finnish-English system displays *make* and *cook* as translation equivalents for the verb *boil*, if the source word is *keittää* (as in *keittää kahvia => make coffee, keittää ruokaa => cook food, keittää vettä => boil water).* Again, a selected alternative automatically replaces the old translation in the target text in a correctly inflected form.

4.4.2 Interfacing with an external word processor

Many potential users of machine translation systems are accustomed to some specific editor or word processing program and are reluctant to learn to use new text editors. There is therefore an alternative user interface in TranSmart. It integrates the MT system with PC-based text editors. The first implementation integrates TranSmart with Microsoft Word under Microsoft Windows.

5. TRANSMART IN PRACTICAL USE

Kielikone MT has been under development for several years. The project was initially supported by the Sitra Foundation, and later on, in the product development phase, by

the Technology Development Center of the Ministry of Trade and Industry. From the outset there has been strong user participation in the work. The companies listed below have participated in the development work both financially and concretely. Although systems have already been sold to customers outside this consortium, Kielikone has so far mainly concentrated on catering to the needs of the members of the consortium.

5.1 The consortium

Nokia Telecommunications Oy, net sales over FIM 10 billion, a subsidiary of the Nokia Group, is one of the pilot customers who have participated in the development work from the beginning. The original intention of the company was to use TranSmart as a workstation to support customer documentation. In this domain the translation memory feature plays an important role as documents often have common sections. During the current year (1996) a new application has spontaneously surfaced at Nokia. The company has plenty of employees whose native tongue is not Finnish. Occasionally they face documents which are available only in Finnish. Thanks to the TranSmart server version installed in the company network they are now able to obtain quick rough translations which often satisfy their needs (Nuutila, 4).

Rautaruukki Oy, a Finnish steel company, net sales over FIM 9 billion, is another pilot customer. The TranSmart system is installed as a server in the company net, and employees can access it through the internal electronic mail system.

Trantex Ltd., a Finnish translation services company, net sales over FIM 28 million, is a third pilot customer. The company specializes in localizing software products but produces translation services in other fields as well. Text types and domains tend to vary greatly in a translation services company and since MT requires domain-specific lexical tuning, the benefits of MT are not so clear in this area.

5.2 MT services

Jointly with Trantex Ltd., Kielikone has established a translation services company, Transwise Oy, to offer machine translation services fast and at a reasonable price. This service function has two goals. The first and paramount goal is to offer customers a new form of attractively priced translation services. The second, subsidiary goal is to collect texts from different sources for purposes of system tuning. Since the texts are used also for tuning, translation speed has not yet been as high as the Kielikone MT technology would ordinarily allow. Transwise Ltd. intends to install the TranSmart system in the WWW in the near future and start offering fast on-line translation services.

REFERENCES

1. Juntunen, J-P., Intelligent translation memory. M.Sc. Thesis, Helsinki University of Technology, 1996 (in Finnish).

2. Jäppinen, H., Kulikov, L., and Ylä-Rotiala, A., KIELIKONE Machine Translation Workstation. Proc. MT Summit III, Washington, DC, 1991.

3. Jäppinen, H., Hartonen, K., Kulikov, L., Nykänen, A., and Ylä-Rotiala, A., KIELIKONE Machine Translation Workstation. In Nirenburg, S. (Ed.), Progress in Machine Translation 10 S Press, 1993.

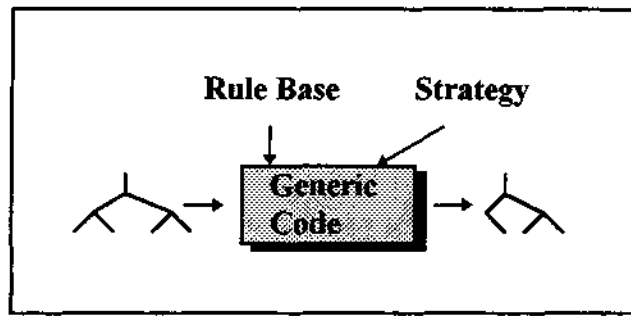4. Nuutila, P., Roughlate service for in-house customers. Translating and the Computer 18, London, 1996.
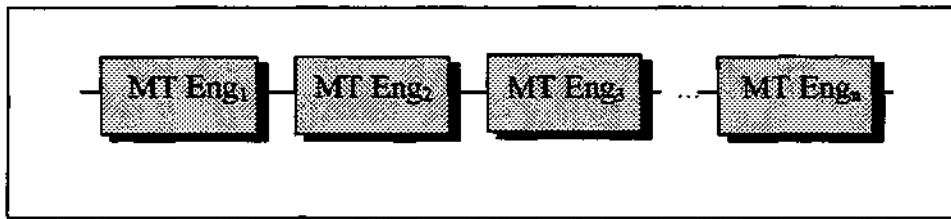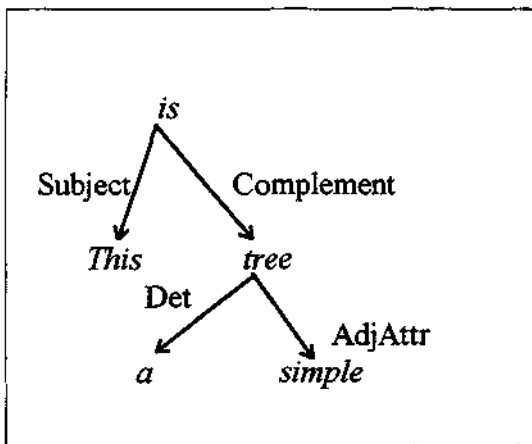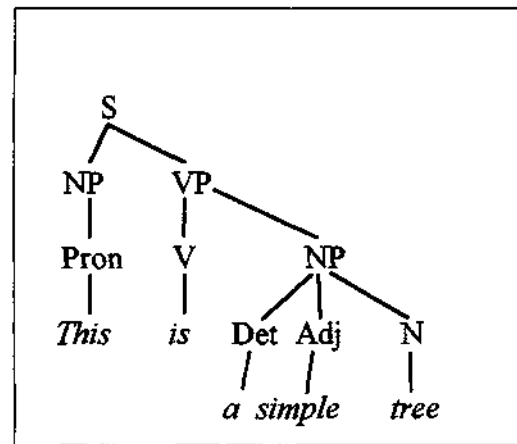
Figure 1: MT Engine

Figure 2: Linear transfer architecture

a) Dependency tree   b) Constituent tree

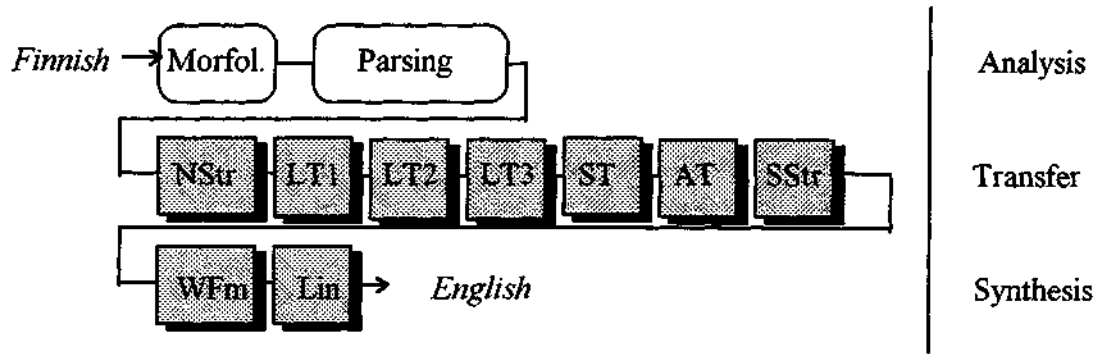Figure 3: Dependency and constituent structures of sentence

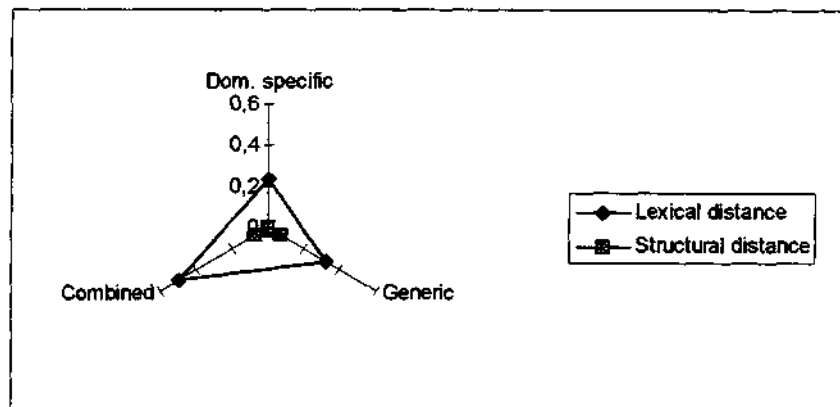Figure 4: Architecture of the TranSmart Finnish-English system

Figure 5: The distances in the window evaluation

Hillary Clinton tapaa Suomen-vierailullaan naisvaikuttajia

Yhdysvaltain presidentin puoliso käy myös Ahtisaaren vieraana

Rouva **Hillary Clinton** saapuu tänään tiistaina vierailulle Suomeen. Yhdysvaltain presidentin puoliso tulee Helsinkiin illalla Virosta.

Hän tapaa keskiviikkona presidentti **Martti Ahtisaaren** ja rouva **Eeva Ahtisaaren.** Ahtisaaret taijoavat vieraalle lounaan Mäntyniemessä.

Aamupäivällä Hillary Clinton keskustelee suomalaisten naisvaikuttajien kanssa naisten asemasta. Tapaamiseen on kutsuttu ainakin ulkoministeri **Tarja Halonen** (sd), eduskunnan puhemies **Riitta Uosukainen** (kok), **puolustusministeri Anneli Taina** (kok), veroministeri **Arja Alho** (sd) ja Helsingin kaupunginjohtaja **Eva-Riitta Siitonen** (kok).

Rouva Clintonin päätökseen tulla Suomeen vaikutti naisten äänioikeuden juhlavuosi. Suomalaiset naiset saivat 90 vuotta sitten ensimmäisenä Euroopassa äänioikeuden.

Suurlähettiläsparin ystävä

Presidentti **Bill Clinton** ja rouva Hillary Clinton ovat hyviä ystäviä Yhdysvaltain suurlähettilään **Derek Shearerin** ja tämän puolison **Ruth Goldwayn** kanssa. Ystäväpariskunnan tapaaminen on eräs syy siihen, että Hillary Clinton pistäytyy Suomessa.

~~Hillary~~ Clinton meets influential women on its Finland visit

The spouse of the president of the United States also visits Ahtisaari

Mrs. ~~**Hillary**~~ **Clinton** will arrive on the visit in Finland today on Tuesday. The spouse of the president of the United States will come from Estonia to Helsinki in the evening.

He will meet president **Martti Ahtisaari** and Mrs. **Eeva Ahtisaari on Wednesday** . Ahtisaaris entertain a guest to lunch in Mäntyniemi.

In the morning ~~Hillary~~ Clinton will discuss the womens' position with Finnish influential women. At least ##Minister for Foreign Affairs## **Tarja Halonen** (~~sd~~), speaker of the parliament **Riitta ~~Uosukainen~~** (~~kek~~)**, ##Minister of Defence## Anneli Taina** (~~kek~~)**,** tax minister **Arja Alho** (~~sd~~) and mayor of Helsinki **Eva-Riitta Siitonen** (~~kek~~) have been invited for the meeting.

The jubilee year of the womens' franchise affected Mis. Clinton's decision to come to Finland. The Finnish women got the franchise 90 years ago as the first in Europe.

Ambassador pair friend

President **Bill Clinton** and Mrs. ~~Hillary~~ Clinton are good friends with the ambassador of the United States, **Derek Shearer** and with this spouse, **Ruth Goldway** The meeting of the friend couple is a reason for the fact that ~~Hillary~~ Clinton drops in Finland

Figure 6: An example of a document with generated warning information

Source Text File: uutinen.se Mode: AS FOLLOW

File   Edit   Help

Uusimman murhasyytteen mukaan mies oli
ensin riidellyt mökin omistajan kanssa.
Käsirysyn jälkeen syytetty oli lähtenyt pois,
mutta hän oli palannut muutaman tunnin
kuluttua kostamaan. Hän oli syyttäjän
mukaan kaatanut mökin lattialle paloöljyä,
sytyttänyt sen tuleen ja estänyt tukevassa
humalassa ollutta mökin omistajaa
pääsemästä pois palavasta mökistä. Syytetty
on kiistänyt murhasyytteet. Kuulusteluissa
hän on kertonut yksityiskohtaisesti
tapahtumista, mutta kaksi viikkoa sitten
pidetyssä istunnossa hän perui puheensa.
Hän oli omien sanojensa mukaan kertonut
poliisille valheita päästäkseen
mahdollisimman pian vapaalle jalalle. Kaikki
syytteet perustuivat pääosin miehen omaan
kertomukseen. Avustaja antoi oikeudelle
todistuksen hoidosta ja korosti, ettei mies
ollut lomilla tai muillakaan vapailla palon
aikana. Avustajan mielestä miehen muutkin
tunnustukset joutuvat kokonaan uuteen
valoon, jos tuoreinta syytettä koskeva
kertomus todetaan valheeksi. **juttu** käsittelyä
jatketaan parin viikon kuluttua.

Target Text File: uutinen.te Mode: AS

File   Edit   Dictionary   Functions   Help

According to the newest murder charge the
man had first quarrelled with the owner of
the cottage. After brawl the accused had left
but he had gone back after a few hours to
revenge. According to the prosecution, he
had poured paraffin on the floor of the
cottage, had set fire to it and had prevented
the heavily drunken owner of the cottage
from getting out of the burning cottage. The
accused has denied the murder charges. In
the questioning he has in detail talked about
the events but in the session that was held
two weeks ago he took back his words. He
had told the police lies according to his own
words to be released as soon as possible.
All charges were based on mainly the man's
own story. The helper gave the court the
certificate of the care and emphasised that
the man was not during the fire on holiday or
on other leave. In the helper's opinion, the
man's other confessions will fall in a totally
new light, too, if the story concerning the
freshest charge is found to be a lie. The
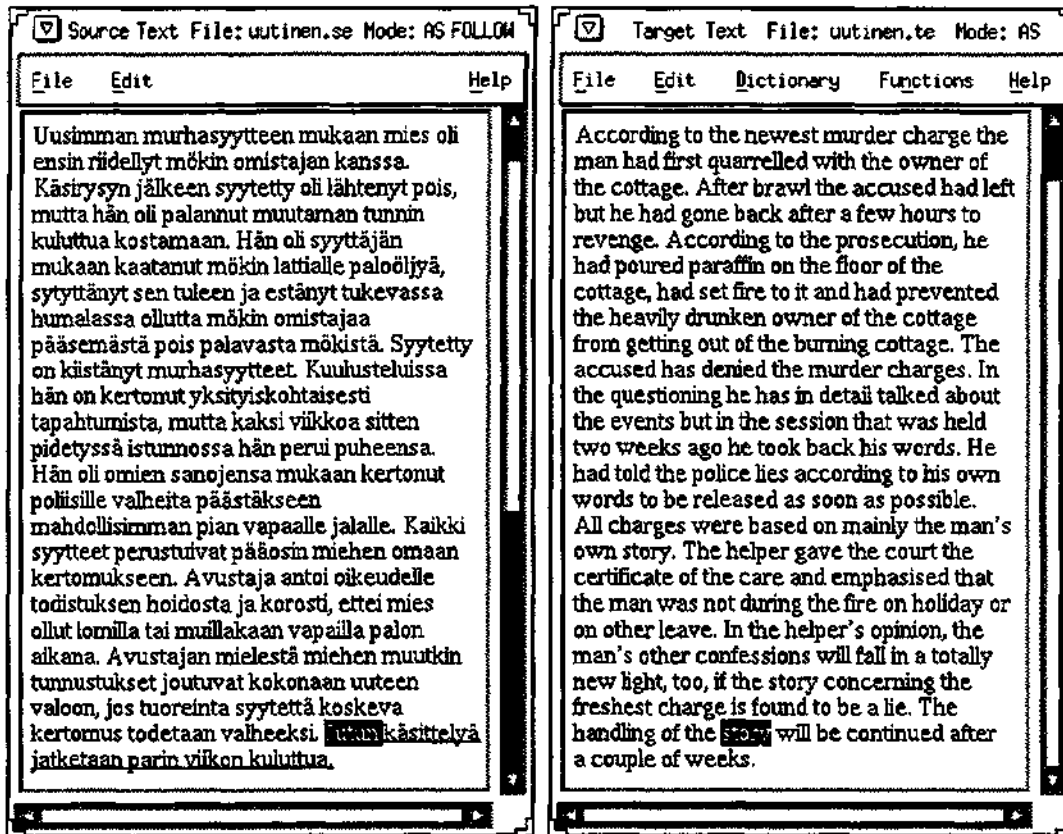handling of the **story** will be continued after
a couple of weeks.

Figure 7: User interface of the bilingual TranSmart editor