

## **Kielikone MT Takes User-Friendliness Seriously**

Kaarina Hyvönen and Harri Arnola<sup>1</sup>  
Kielikone Ltd.  
P.O. Box 126  
00211 Helsinki, FINLAND  
kaarina@kielikone.fi, hamola@cc.hut.fi  
www.kielikone.fi

### **1. Introduction**

Kielikone MT stands for innovative machine translation R&D which has produced a theoretically and commercially successful MT product, the TranSmart® Finnish-English machine translation system. A prestigious jury indirectly praised also the theoretical quality of the system when it elected TranSmart the best Finnish software product of the year 1996. Hundreds of daily users in three major Finnish companies (a steel company, an information technology company, and a telecommunications company) are proof of the commercial value of TranSmart.

Critical and theoretical success may be sweet, but the ultimate judge of any artificial system is the user. Therefore, Kielikone MT has from the outset been particularly sensitive to the expressed or anticipated wishes of end users. This has meant, first of all, creating a thorough customization process through which the machine translation system is adapted to the information technology environment and language conventions of a new customer company. However, also individual users expect the system to adapt smoothly to their needs. In this paper we mention company level customization only in passing and focus mainly on the features we have designed with the end user in mind: a user-friendly interface, the ability to preserve document formatting, the facilitation of post-editing, and document-specific dictionaries.

It should be clear that, like all data processing systems, TranSmart is subject to continuous improvement. The features discussed below represent only the innovations that have already been implemented and are present in the shipping version. They are, by no means, our last word on the subject of user-friendliness: we will continue to improve TranSmart in close cooperation with our users.

### **2. User-Friendly Interface**

It cannot be emphasized strongly enough how important the quality of a user interface is for the usability of a data processing system. The development of an MT system requires good innovations and hard work. It is only human that the designers of a complete system should feel proud of their achievement and want to raise the system into a central position in a workflow. However, from the user's viewpoint, an MT system is just one utility function among many. It should be an uncomplicated,

---

<sup>1</sup> Formerly Harri Jäppinen.

inconspicuous, and reliable tool which may be called to perform its duty only when there is a need for it.

We must confess here in parenthesis that we came to know this the hard way. Initially, we designed an MT workstation version which runs on Unix workstations (Jäppinen, Hartonen, Kulikov, Nykänen and Ylä-Rotiala, 1993), complete with a text editor. Many attractive features were designed which would have made using the system and particularly post-editing rough translations very easy - if the users had used it!

The hard reality was that none of our pilot user companies were willing to reorganize their workflow so that the MT workstations would be utilized in the most efficient manner possible nor were the end users willing to produce and edit documents in one environment and then jump to another one for translating and post-editing them.

Back to the drawing board. In Finland companies commonly use PC LAN's, and MS Word is the most popular word processing program. Therefore, we decided to integrate the TranSmart system with MS Word and let the users call the MT function in the most natural and unoffending way, from their familiar word processor. And by integrate we mean total integration - in the first PC version of TranSmart there is no TranSmart user interface in the strict sense of the word. Currently we install the TranSmart system on an NT server in the LAN of the customer, and for the end users TranSmart is just a new pull-down menu or a toolbar in their MS Word interface. This is what the default toolbar looks like:



Figure 1: The default toolbar for TranSmart

As is customary in current Windows versions, the toolbar is a customizable item. The TranSmart end user may create a TranSmart toolbar which corresponds to his or her needs. The figure above shows the default toolbar which contains the five essential TranSmart functions: translate the document, retrieve the translation, show/hide translation tags as colors, monitor translation progress, and help.

Such complete integration, painful as it may have been to our developer's egos, was exactly what was needed. It brought TranSmart in close contact with the writing process and made a special learning process for a new user practically superfluous. TranSmart became a computer application which is very easy to introduce to new users and different workflows.

### **3. Ability to Preserve Document Formatting**

On the road to a satisfied user who really takes full advantage of the software, a friendly user-interface is only the first step. One of the numerous issues standing between an MT application and a happy end user is document formatting, a very sensitive issue. In the digital world, written texts appear in a host of various formats. One extreme are texts which are produced fast and which have almost no formatting

save the linear ordering of words. E-mail messages are good and voluminous representatives of such "format-free" texts. The other extreme are texts which incorporate figures and tables, and in which the layout of the text is richly formatted. The latter may have been produced with such devotion and intensity that should a single misplaced delimiter or misspelled word be found in the text after its release, the author would feel great anguish. Printed financial statements of large corporations represent this text type well.

Ideally, an MT system should be able to handle reliably all formats of texts and produce translations in exactly the same format as the originals. However, the practical world does not permit idealism to prevail in MT. It is obvious that the less formatting a text contains the easier it is for an MT system to produce an intact translation. And conversely: the richer the formatting of the document, the greater are the risks that an MT system will wreak havoc. At any rate the user must tolerate a certain degree of format infidelity in the rough translation as word and phrase lengths are likely to change during translation and therefore the translation will occupy at least a slightly different space than the original document for example on paper.

TranSmart uses Rich Text Format (RTF) to tackle the problem of preserving formatting. The RTF format is able to preserve most of the formatting made with MS Word. TranSmart converts an original Word document into the RTF format before translation, preserves the RTF coding throughout the translation process, and converts the translated document back into Word format. As a rule, formatting is preserved quite well. For example, below is a piece of formatted text and its unedited rough translation by TranSmart. The text is an excerpt from a press release published by the Finnish steel company Rautaruukki Oy and it concerns air pollution. As the focus here is on document formatting, the translation was produced using general dictionaries only. Thus the quality of the rough translation in this first example is not the best possible. For better translation quality, see later examples in which the dictionary setup is larger.

Finnish source text:

Pölyn tiedetään yleisesti huonontavan ilman laatua ja likaavan ympäristöä. Terveydellistä haittaa ihmiselle ja myrkyllisiä vaikutuksia eliöstölle aiheuttavat varsinaisesti hienojakoinen pöly ja pölyn mahdollisesti sisältämät haitalliset aineet, kuten *raskasmetallit* ja *polysykliset aromaattiset hiilivedyt* (PAH).

Raahen terästehtaalla (RTt) pölyä muodostuu ja irtaantuu esimerkiksi liikenteestä, materiaalien käsittelystä ja erilaisista prosessitoiminnoista. Tehtaalle on rakennettu yli 60 erilaista pölyjen talteenotto laitosta. Tällaisia ovat *sähkö-* ja *tekstiilisuodattimet*, *syklonit* ja *pesurit*. Vuonna 1996 pölypäästöjä vähennettiin peruskorjaamalla sintraamon multisyklonit ja lisäämällä talteenotetun pölyn käyttöä sintraamolla. Vuonna 1997 valmistuu terässulaton pölyn keräys- ja puhdistusjärjestelmä. Lisäksi on tarkoitus tehostaa sintraamon pölynpoistoa, rakentaa pellettivaunujen purkauspaikan pölynpoistojärjestelmä, lisätä masuunin

sähkösuotimen puhdistustehoa ja uusia masuunin rikinpoistolaitteistot.

Metallipitoisuudet ovat pysyneet samalla tasolla useita vuosia. Raahen ilmanlaadulle on tunnusomaista korkea rautapitoisuus. Sinkkipitoisuus on myös hieman kohonnut. Muuten raskasmetallipitoisuudet ovat lähellä tausta-arvoja.

	mg / m <sup>3</sup>
lyijy	0,007 - 64
kadmium	0,003 - 1,1
rauta	0,62 - 4 160
sinkki	0,03 - 460
mangaani	0,01 - 16,7
kupari	0,029 - 12
arseeni	0,007 - 1,9

Taulukko 1. Metallipitoisuuksien tausta-arvoja ilmassa

English rough translation by TranSmart:

It is known that dust lowers the quality of air generally and dirties the environment. The health risk to the human being and the poisonous effects on the population are in fact caused by fine dust and by the harmful substances possibly contained by the dust, such as *by heavy metals* and *by polycyclic aromatic hydrocarbons (PAH)*.

At the steel mill of Raahe (RTt) dust forms and is emitted for example by traffic, the handling of materials and by the different process operations. The more than 60 different collection plants of dust have been built for the factory. Such are *electricity* and *textile filters*, *cyclones* and *pesun* -. In 1996 the dust emissions were reduced by renovating the multisykloni of the sintraamo and by increasing the use of *talteenottaa* dust with the sintraamo. In 1997 the collecting system and cleaning system of the dust of the terässulatto is completed. Furthermore, it is intended to intensify the dust removal of the sintraamo, to build the *pölynpoistojärjestelmä* of the *purkauspaikka* of *pellettivaunu*, to increase the cleaning effect of the electricity filter of the blast furnace and to renew the rikinpoistolaitteisto of the blast furnace.

The metal contents have stayed at the same level for several years. A high iron content is characteristic of the air quality of Raahe. The zinc content also has risen a little. Otherwise the heavy metal contents are near the background levels.

	mg / m <sup>3</sup>
lead	0,007 - 64
cadmium	0,003 - 1,1
iron	0,62 - 4 160
zinc	0,03 - 460
manganese	0,01 - 16,7
copper	0,029 - 12
arsenic	0,007 - 1,9

Table 1. Background levels of metal contents in air

In this case, document formatting, such as indentation, different text formats and the table, is preserved quite nicely. Yet, as explained above, formatting is not always preserved intact. The RTF conversion itself is not error-free and does cause occasional errors in translations. The most common formatting problems are related to different phrase and sentence lengths in source and target languages. A special problem is presented by what we call "pseudo formatting". By this we mean formatting that looks like real formatting but that has, in fact, been created (often extremely clumsily and with astonishing patience and imagination) using tabulators or space characters rather than real Word formatting functions. Pseudo formatting cannot be detected by TranSmart and will not machine translate well. Here is an example of what would happen to the indented chapter in the text above if the formatting had been created using the tabulator and by placing hard returns at the end of each line.

At the steel mill of Raahe (RTt) dust forms and is emitted for example by the traffic from handling of materials and from different process operations. To the factory one is built the more than 60 different collection plants of dust. Such are *electricity* - and *textile filters, cyclones* and *pesuri*. During year 1996 dust emissions it was reduced by renovating the multisykloni of the sintraamo and by adding use of talteenottoa dust with the sintraamo. In 1997 it is completed collecting system and cleaning system of the dust of the terassulatto. Furthermore, there is a purpose to intensify the dust removal of the sintraamo, to build the purkauspaikka of pellettivaunu pölynpoistojärjestelmä, to increase the cleaning effect of the electricity filter of the blast furnace and to renew the rikinpoistolaitteisto of the blast furnace.

The effects are quite disastrous, both for form and for content. Pseudo formatting is, admittedly, a provocative example. Nevertheless, even ignoring pseudo formatting, there is still much to be done. Today's increasingly sophisticated word processing tools present a real challenge for formatting preservation. However, in view of what used to happen to most formatting before we came up with RTF conversion, we feel that we have come a long way. TranSmart is able to retain most of the formatting with satisfactory fidelity - an ability that has proved to be of immense importance to end users.

#### 4. Facilitating Post-Editing

The cold fact is that in terms of translation correctness, there is and never will be a perfect MT system. No matter how easy to use, all of them make mistakes and users

have to check results, correct errors, retranslate, and post-edit the rough translation in order to get the translations right, however, as unacceptable as translation errors are, they are not completely unheard of even in human translations. It may also be argued that only rarely if ever does a human translator produce a final polished translation at one go. Rather he or she edits and rewrites incomplete versions of a translation - often many, many times.

Undoubtedly, errors caused by a machine are different from those produced by humans: some of the mechanical errors mutilate syntax, and some of the rough translations may even be incomprehensible. Human translations, from the first version to the final one, are never that bad in quality. The syntax is usually approximately correct and the sentences are comprehensible. So while the process of post-editing is common to both human and machine translation, the type of post-editing required is different. In human translation post-editing usually aims at making the message more precise or better in style.

Since editing is quite normal in human translation, MT systems are more attractive if they offer easy and convenient post-editing functions. If an MT system had intelligent editing functions tuned specifically for the efficient correction of the various error types, users might not feel that the post-editing of machine translations was an arduous task at all. The TranSmart system still has a long way to go towards highly sophisticated post-editing functions. Some modest steps have, however, already been taken towards that goal. They are related to lexical phenomena and focus on two things: reducing the amount of translations which might require post-editing, and making it easy to spot translations that TranSmart predicts might need post-editing.

#### **4.1 Reducing the Amount of Material Requiring Post-Editing**

By far the best way to facilitate post-editing would seem to be to reduce the amount of translations requiring post-editing. This is easier said than done, and, as all MT developers know, usually involves slow, expensive and unending lexical work. However, sometimes a language gives developers a break. When translating from Finnish into English, compound nouns have turned out to be a case in point. Finnish is pregnant with compound nouns. While English opts for a collocation to express a nominal concept, say, data base, Finnish uses a compound noun: *tietokanta* (*tieto+kanta*). Commonly the defining part (the leftmost part) of a compound noun is either in the nominative case, as in the example just given, or in the genitive case as in *kirkonkello* (*kirkon* is a genitive form of the noun *kirkko*) for a church bell. Compound nouns are very common in Finnish texts, particularly in texts dealing with special fields.

Perhaps consequently, the formation of a compound noun often is a productive process. For example, a factory is *tehdas* in Finnish. In Finnish the generic name for almost any kind of factory is easily produced by attaching the name of the things produced in the factory to the noun *tehdas*. This means that the compound is highly productive. The Finnish words for shoe factory and car factory are *kenkätehdas* and *autotehdas*, respectively (a shoe is *kenkä* and a car is *auto*). Should somebody begin to produce coat sleeves in an industrial process, he or she would have *hihatehdas* (a

sleeve is *hiha*). Such a compound does not exist in any of the dictionaries of Finnish language because nobody has so far established such an enterprise. The point is that it is impossible to list all possible Finnish compounds in a dictionary. Since new concepts and hence compound nouns are formed almost daily in fast advancing new fields, new concepts cause a substantial lexical problem for MT in Finnish. Naturally, they also present the editor of a rough translation with quite a task.

The TranSmart system implements the following solution which, paradoxically, takes advantage the productivity of Finnish compound nouns. If the base form of a noun is not found in the system dictionaries, TranSmart tries to recognize it as a two-part compound. If TranSmart is able to recognize both parts, it produces a productive translation by combining the translation equivalents of the parts and tags the suggested translation. Should a piece of text include the noun *hihatehdas*, TranSmart would produce the default translation *sleeve factory* and mark the suggestion with a translation tag. This makes it much easier for the editor of the rough translation to deal with productive compound nouns. First, because it reduces the number of translations which might need heavy post-editing (such as consulting dictionaries and rewriting) since the productive translations suggested by TranSmart are often usable. Second, because, thanks to translation tags, it makes productive nouns easy to spot and check in the rough translation.

## 4.2 Translation Tags

The basic idea behind translation tags is simple. It is the same as the basic idea of facilitating post-editing: what is known to benefit from checking should be made easy to find and edit. In addition to productively translated compound nouns (discussed above), TranSmart adds translation tags to three types of translations: words not recognized as Finnish words, words for which there are no translation equivalents in TranSmart's dictionaries, and culture bound translations.

The most common type of phenomenon that a post-editor runs across in a rough translation produced by TranSmart are words for which no translation equivalents are found in TranSmart's dictionaries, and words not recognized as Finnish words. New proper names are a good example of the latter. Proper names are especially worth marking in a rough translation because while some proper names need to be translated, some must not be translated. For instance, proper names of people do not usually have translations but geographical names often do. London is *Lontoo* in Finnish but New York is *New York*.

In a rough translation TranSmart reproduces unrecognized and thus unanalyzed words (e.g. new proper names) verbatim, and analyzed words which lack translation equivalents in their basic form. It also adds translation tags to them to make it easy for the post-editor to locate them. The post-editor can then either just replace the words with a correct translation or, preferably, correct the situation by adding a word to one of TranSmart's dictionaries. In addition to facilitating the post-editing process, tagging unanalyzable words and words which lack translation equivalents is useful from the point of view of dictionary maintenance. It provides a quick and efficient way to browse for potential new lexical entries.

The third and final type of translation with translation tags, culture bound translation equivalents, are a familiar phenomenon for anyone who has had to communicate between two cultures. Culture bound expressions are expressions whose meaning is so closely related to their cultural background that it is impossible to translate them directly from one language to another. For example the school systems in Finland and Britain are so different from each other that for many school levels there are no direct translation equivalents. The same holds true for example for the army and the legal system. A human translator usually handles difficult culture bound expressions by explaining what is meant. In MT explaining translations is not an option. An MT system has to make do with one translation equivalent and TranSmart is no exception. Hence, TranSmart tags culture bound translation equivalents to make sure that the post-editor pays special attention to these expressions.

All translation tags are implicit and they are made explicit only if a user wishes so. The user can have the translation tags displayed either as colors, which is very convenient on screen, or as underlining which is a good choice for documents which will be printed. Each phenomenon has its own color and type of underlining and the user can make the tags visible and invisible with a simple toggle command in the TranSmart menu or toolbar. The tags are created using Word's hidden text features which is why it is especially important that once the translation is polished the user is able to delete the tags completely. This way no extra information remains in the document. An excerpt from the Rautaruukki text used in the previous example demonstrates translations tagged with underlining. Notice that the translation is again produced with general dictionaries only which means that many domain specific words are not translated. In further examples we shall see an improvement in this respect.

Finnish source text:

Raahan terästehtaalla (RTt) pölyä muodostuu ja irtaantuu esimerkiksi liikenteestä, materiaalien käsittelystä ja erilaisista prosessitoiminnoista. Tehtaalle on rakennettu yli 60 erilaista pölyjen talteenottolaitosta. Tällaisia ovat sähkö- ja tekstiilisuodattimet, syklonit ja pesurit. Vuonna 1996 pölypäästöjä vähennettiin peruskorjaamalla sintraamon multisyklonit ja lisäämällä talteenotetun pölyn käyttöä sintraamolla. Vuonna 1997 valmistuu terässulaton pölyn keräys- ja puhdistusjärjestelmä. Lisäksi on tarkoitus tehostaa sintraamon pölynpoistoa, rakentaa pellettivaunujen purkauspaikan pölynpoistojärjestelmä, lisätä masuunin sähkösuotimen puhdistustehoa ja uusia masuunin rikinpoistolaitteistot.

English rough translation by TranSmart:

At the steel mill of Raahe (RTt) dust forms and is emitted for example by traffic, the handling of materials and by the different process operations. The more than 60 different collection plants of dust have been built for the factory. Such are electricity and textile filters, cyclones and pesuri -. In 1996 the dust emissions were reduced by renovating the multisykloni of the sintraame and by increasing the use of



~~talteenottaa~~ dust with the ~~sintraamo~~. In 1997 the collecting system and cleaning system of the dust of the ~~terässulatto~~ is completed. Furthermore, it is intended to intensify the dust removal of the ~~sintraamo~~. to build the ~~pölynpoistojärjestelmä~~ of the ~~purkauspaikka~~ of ~~pellettivaunu~~. to increase the cleaning effect of the electricity filter of the blast furnace and to renew the ~~rikinpoistolaitteisto~~ of the blast furnace.

Explanations for the tags are the following. Finnish words which TranSmart's analysis of Finnish word forms has not been able to recognize are marked with dotted underlining. Finnish words for which TranSmart does not know the translation equivalent are marked with ~~striketrough~~. Productively translated compounds nouns are marked with double underlining. This example does not contain any culture bound expressions which are tagged with ~~dotted underlining and striketrough~~.

So far in our discussion of efforts aimed at making TranSmart as user-friendly as possible we have concentrated on measures that are defined by reduction: integrating the user interface completely into a word processor, minimizing the amount of changes in document formatting during translation, and making post-editing as easy as possible by decreasing the number of post-edited items and the energy spent in locating and checking them. For a final point, let us focus on the opposite, expansion. What better example than dictionaries?

## 5. Customer-Specific and Document-Specific Dictionaries

One of the main problems in MT is to choose the correct word senses and their correct translations. Words often have multiple senses and the different senses of a word require different translations. To take a simple example from Finnish, the noun *tavu* translates into English either as *byte* or as *syllable*, depending on whether the text deals with information technology or linguistic or literary matters.

Some people have claimed that because of the word sense problem, MT has to perform deep semantic analysis in order to make the senses clear. In our opinion, deep semantic analysis would call for too high a price for something that can be solved at least partially by other means. Deep semantic analysis would bring about difficult theoretical problems and their implementations would be computationally expensive. Furthermore, deep semantic analysis would greatly increase the cost of customizing and maintaining the system. Therefore, Kielikone MT has opted for a different route for reducing the word sense problem. Our answer is a powerful and flexible dictionary setup which is based on customer-specific dictionaries and enhanced by document-specific dictionaries.

### 5.1. Customer-Specific Dictionaries

Due to its flexible system architecture TranSmart is able to employ several dictionaries simultaneously. The dictionaries are accessed in a specific precedence order. Typically, a TranSmart configuration contains three types of dictionaries: customer-specific dictionaries, domain-specific dictionaries and general dictionaries.

In practice, both customer-specific and domain-specific dictionaries are created for each customer during a company level customization project. A TranSmart user company usually has one customer-specific dictionary which contains general company terminology (such as product names and administrative terminology), and several different domain-specific dictionaries (for example a steel company might need an IT dictionary, a financial dictionary, and several dictionaries pertaining to the different phases of their steelmaking process). The end user may choose the desired dictionary setup for each translation with just a few mouse clicks before translating the document with TranSmart. The system searches for a translation equivalent first in the customer-specific dictionary, then in the domain-specific dictionaries, and finally in the general dictionaries. The figure below illustrates the setup. Note that while a company usually has several domain-specific dictionaries, only two at a time are used in a translation. The last dictionaries to be used, the general dictionaries, are, of course, not optional for the user: they are used in all translations.

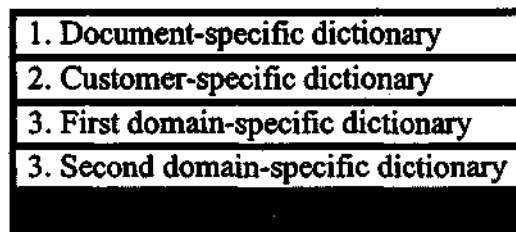


Figure 2: The TranSmart dictionaries in runtime order

The beauty of customer-specific dictionaries is twofold. First, such a dictionary setup greatly reduces the word sense problem. For example, if a customer's texts deal with information technology, the domain-specific dictionary translates *tavu* with *byte*. The other word sense, a syllable, would be included in the general dictionary.

Second, customer-specific dictionaries work wonders as far as user-friendliness is concerned. Customer-specific dictionaries help to both customize and homogenize terminology for all users in a user company. The company is able to disseminate its official terminology very effectively. With the help of carefully compiled and accurate domain-specific dictionaries, the company can brush up its communications by helping its employees to write using precise terminology which is up to date. Furthermore, the user company has full control over its TranSmart dictionaries. During the customization project, we train a few people in their staff to compile and maintain the customer-specific dictionaries. This ensures consistent and riskless dictionary maintenance. The following example shows the already familiar Rautaruukki text excerpt now translated with TranSmart using Rautaruukki company dictionaries to illustrate the effect of a comprehensive and well-constructed company dictionary. In this example the rough translation quality is very close to optimal. There are no tagged translations left in the rough translation.

At Raahe Steel Works (RSW) dust forms and is emitted for example by traffic, the handling of the materials and by the different process operations. More than 60 different dust collection plants have been built for the works. *The bag fitters , cyclones and scrubbers are such* . In 1996 the dust

emissions were reduced by renovating the multicyclones of the sintering plant and by increasing the use of recovered dust at the sintering plant. In 1997 the dust collection system and purification system of the Steel Plant is finished. Moreover it is intended to intensify the dust removal of the sintering plant, to build the dust extraction system of the discharge station of the pellet wagons, to increase the cleaning efficiency of the electric precipitator of the blast furnace and to renew the desulphurizing equipment of the blast furnace.

Company dictionaries are obviously advantageous from the point of view of the end user. Ideally, he or she has valid terminology already inserted in his or her rough translation. However, centralized dictionary maintenance may sometimes be something of a nuisance to the user who is trying to produce a translation, due yesterday, of a 100-page document in which the noun *client* appears umpteen times when the translation equivalent he or she would have preferred is *customer*. This is one of the reasons why we created document-specific dictionaries.

## 5.2 Document-Specific Dictionaries

Individual users in a company often have their own wishes for translation equivalents which override the company-specific choices described above. For example, a user may use colloquial or slang words not recorded in company dictionaries. Often e-mail messages between colleagues use informal style. A company may not want to include such words in the company-specific dictionaries because their usage and translations are not official. A user may also want to add a word to TranSmart dictionaries if his or her piece of text discusses matters that the company-specific dictionary does not cover. Finally, the user may want to correct a word choice because the translation equivalent given in the TranSmart dictionaries is not suitable for his or her purposes (cf. *client* and *customer* above).

What these users have in common is that they want to modify the TranSmart dictionary immediately. For them to contact the dictionary maintenance people, inform them of their special need, wait for the dictionary maintenance people to update the company dictionaries and to release the update, as safe and productive a process as it is, is far too slow and thus no solution. To accommodate these lexical adjustments of provisory nature, TranSmart supports document-specific dictionaries.

The document-specific dictionary is a fast and easy way to choose translation equivalents for a given document. The user activates the document-specific dictionary with a command in the TranSmart menu, types in words and their translation equivalents in a dialog box using a very simple syntax, retranslates the document, and gets a better rough translation. The document-specific dictionary is fully reusable in later translations of the document and it can be edited at will.

As TranSmart searches for translation equivalents first in the document-specific dictionary, it is possible to override all other dictionaries with the document-specific dictionary. This makes it an effective but also potentially dangerous tool. We have had to take precautions to prevent lexical corruption for example by allowing only certain parts-of-speech in the document-specific dictionary. Moreover, the word

choices are used only in the translations of that specific document. Should the user end up adding the same word frequently into different document-specific dictionaries, he or she is encouraged to report it to the dictionary maintenance people who may decide to add it into the company dictionaries for the benefit of all users.

Let us look at this with the help of an example. We have already seen how the Rautaruukki sample text translates when correctly and incorrectly formatted, and when translated with Rautaruukki company dictionaries. Let us now suppose that the user is working with the rough translation produced using Rautaruukki dictionaries. He or she is not satisfied with the translation equivalent *dust* for the Finnish noun *pöly* in the last translation and wants to have *pöly* translated with the translation equivalent *powder dust*. He or she simply creates a document-specific dictionary, types in an entry for *pöly* assigning the *noun powder dust* as the translation equivalent, and translates the text again. The figure below shows the document-specific dictionary he or she would have created.

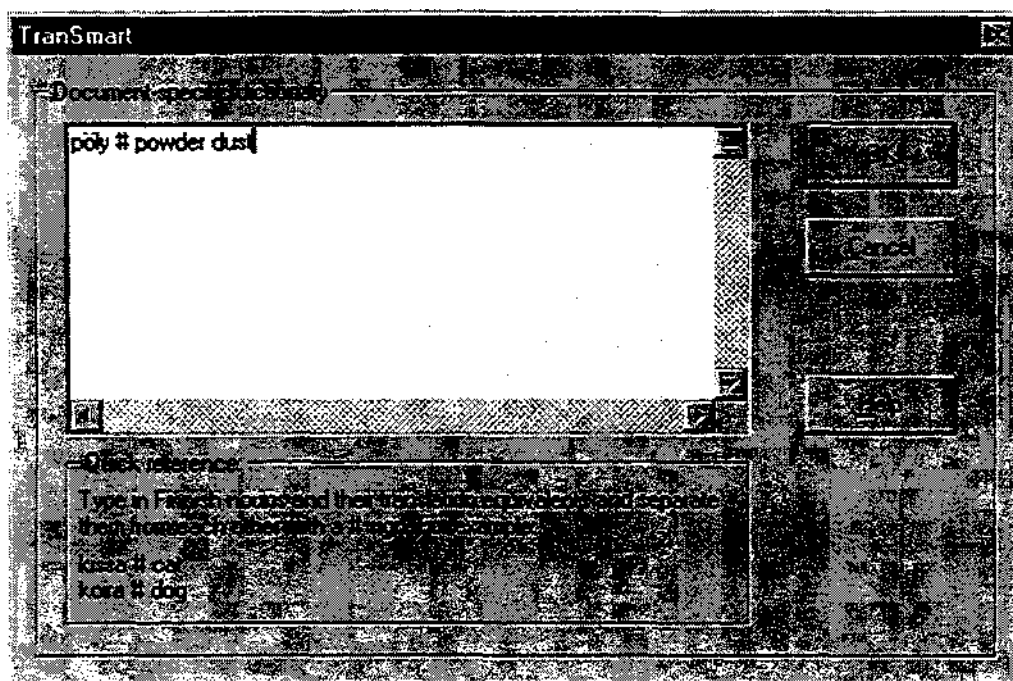


Figure 3: The document-specific dictionary dialog box

After retranslating the text with TranSmart, the user would get the following rough translation. Like the previous example, this translation represents close to optimal rough translation quality producible with TranSmart.

At Raahe Steel Works (RSW) powder dust forms and is emitted for example by traffic, the handling of the materials and by the different process operations. More than 60 different dust collection plants have been built for the works. *The bag filters , cyclones and scrubbers are such .* In 1996 the dust emissions were reduced by renovating the multicyclones of the sintering plant and by increasing the use of the recovered powder dust at the sintering plant. In 1997 the dust collection system and purification system of the

Steel Plant is finished. Moreover it is intended to intensify the dust removal of the sintering plant, to build the dust extraction system of the discharge station of the pellet wagons, to increase the cleaning efficiency of the electric precipitator of the blast furnace and to renew the desulphurizing equipment of the blast furnace.

It is safe to say that the document-specific dictionary is an extremely user-friendly feature. It lets the individual end user make lexical adjustments in an efficient but safe way, and at the same time indirectly supports centralized dictionary maintenance.

## **Conclusion**

An MT system is a complex and challenging computer application. The challenges lie not only in technical or linguistic R&D issues, but also increasingly in questions relating to usability. These can be roughly divided into two categories: general usability issues and MT-specific usability issues. General usability issues incorporate such things as the graphical user interface. Nowadays, well designed and ergonomical user interfaces are a necessity that many users already know to expect. MT-specific usability issues stem partly from everyday use of MT, in other words, how people use an MT system and how they would like to use it, and partly from general characteristics of MT, such as translation quality.

For us, usability translates into practice as user-friendliness. We have striven to address user needs and wishes from the very beginning of Kielikone MT. We began developing our MT technology more than a decade ago. Over those years, the concept of user-friendliness has evolved radically. We started out in command-line format on Unix workstations, with one type of elementary translation tags which had to be removed by hand in an ASCII text editor. Today, we have achieved an integrated user interface in a de facto standard word processing program, good quality formatting preservation, effective reduction of translations which need post-editing together with many features that facilitate post-editing, as well as highly but safely customizable dictionaries both for company and individual use. We call it a good start.

## **Acknowledgements**

Building a commercial MT system is a group effort; it calls for several different kinds of talent. Although only two names appear on the front page of this paper, many others have contributed to Kielikone MT and the TranSmart system. In particular, we want to mention the following current employees of Kielikone Ltd. who have been involved in the later phases of TranSmart R&D: Susanna Auhtola, Kimmo Huovila, Jukka-Pekka Juntunen, Terhi Ronkainen, Tuomas Tammi, and Kari Tuomainen. The Technology Development Center and the Finnish National Fund for Research and Development (Sitra) have supported our R&D work financially. Appropriately enough, Kielikone MT has also greatly benefited from user feedback. We wish to point out that constructive comments and suggestions by Nokia Telecommunications

Oy, Rautaruukki Oy, and Trantex Oy have contributed substantially to the user-friendliness of TranSmart.<sup>2</sup>

## References

- Arnola, H., Hyvönen, K., Juntunen, J-P., Linnanvirta, T. and Suoranta, P. 1996. "Kielikone Finnish-English MT System TranSmart in Practical Use" in *Proceedings of Translating and the Computer 18*, London.
- Arnola, H., Hyvönen, K., Juntunen, J-P., Linnanvirta, T. and Suoranta, P. 1996. "Kielikone Machine Translation Technology and Its Perspective on the Economics of MT" in *Proceedings of the MT Workshop at the TKE*, Vienna.
- Arnola, H., Hyvönen, K. and Linnanvirta, T. 1997. "Kielikone Machine Translation How We Did It", *elsnews*, 6.2, pp. 4-5.
- Jäppinen, H., Hartonen, K., Kulikov, L., Nykänen, A. and Ylä-Rotiala, A. 1993. "KIELIKONE Machine Translation Workstation", in S. Nirenburg (ed.) *Progress in Machine Translation*, IOS Press, Amsterdam, pp.173-184.
- Jäppinen, H., Kulikov, L., and Ylä-Rotiala, A. 1991. "KIELIKONE Machine Translation Workstation" in *Proceedings of MT Summit III*, Washington D.C.

---

<sup>2</sup> TranSmart is a registered trademark of Kielikone Ltd. MS Word is a registered trademark of the Microsoft Corporation.