

A Fresh Look at MT

An aerial photograph of an industrial park or factory complex. The image shows a dense arrangement of large, rectangular industrial buildings with light-colored roofs, interspersed with green spaces and parking lots. The perspective is from a high angle, looking down on the facility. The sky is overcast and hazy, creating a soft, diffused light across the scene.

A professional translator reviews the Personal Translator machine translation application.

by Angela Moisl

When Language International asked me to evaluate the latest version of Personal Translator, a machine translation application, I told them I had no experience with machine translation whatsoever. The only qualifications I could bring to the table were that of being a professional translator and a German native speaker. It seems that this was precisely what they were looking for. Users of earlier versions of the application or of other machine translation tools looking for information on how this application compares to earlier releases or competitors therefore will be disappointed. Professional translators hoping for a good laugh about the inevitable mistranslations produced by a machine translation application will also be disappointed, although not completely. As tempting as it is, it is ultimately unfair to pitch the quality a human translator can produce against that of a machine translation tool, which is confined to dictionaries and algorithms to extract and produce "meaning".

Installing Personal Translator

Personal Translator (PT) is a product of Linguatex, a company founded in 1992 in Germany and specializing in the development and marketing of language technologies with the emphasis on automatic translation and speech recognition. PT 2002 is the latest release of their automatic translation application, released May 2002. It comes in three versions: PT 2002 Home, PT 2002 Office and PT 2002 Office Plus, each with increased features and capabilities—and price. The Office Plus version (EUR 249,—) German-English was the version I tested. PT 2002 is also available for the language pair French-English, with a French user interface.

The installation of the application, with system requirements that should be more than easily met by any fairly up-to-date computer configuration, went without a hitch. I opted for the default application configuration set by the system. After installation I would be able to use the application with a German user interface, translate either within PT or Microsoft Word (97, 2000 or 2002), access PT from within any other Windows application (through PT Direct), translate HTML pages with PC Web and have the computer read words and sentences to me in case I needed help with pronunciation. As a deviation from the default setup I also added an English user interface. Later I realized that through this installation procedure I had also installed a system dic-

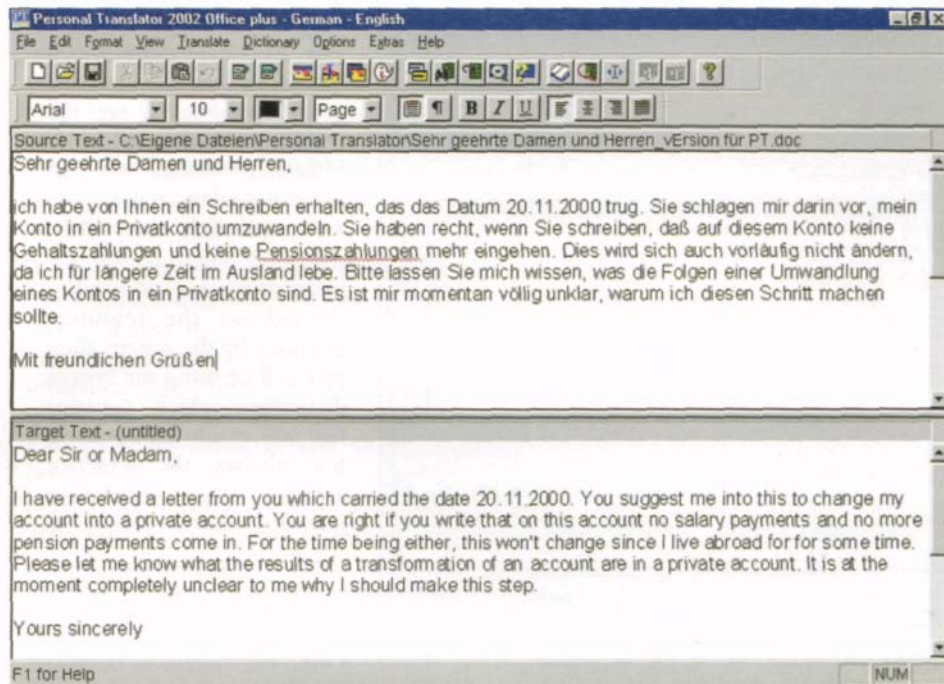


Figure 1

tionary, an empty user dictionary and a translation memory filled with a few segments. But more about that later.

After launching Personal Translator I found a typical Windows user interface with a menu bar and toolbar and the screen divided horizontally into two fields (see Figure 1). The message was clear: The source text in the upper window, translations in the lower. "Just copy a source text to the upper window or open it there and press the Translate button", you might be tempted to think. But working with the

The translation process consists of a preparatory phase, the actual translation phase, and in case you are aiming for translations of publishable quality, also a post-translation phase.

PT is not quite that easy and the translation process more complex, as a careful reading of the manual revealed. At least, if one aims for the best possible translation quality.

Getting ready for translation

The translation process consists of a preparatory phase, the actual translation

phase, and in case you are aiming for translations of publishable quality, also a post-translation phase. The preparatory phase and the post-translation phase is where the human factor comes in, while the actual translation phase is completely left to the application. Watch it while it works and you will see a sentence of the source text being copied to the target window and then run through the five sub-processes that make up the translation: segmentation with the sentence as the basic segment, lexical research, syntactical analysis, lexical transfer and structural transfer. Once processed, the sentence will appear in the target language in the target window.

Before you can relish the fruits of the computer's labors you will have to perform a few actions and make a few crucial decisions that will influence the quality of the target text. First of all, you need to search the text for unknown words or terms and add those to the dictionary, which is not necessarily an easy task as I will later explain. Setting or confirming the translation direction detected by the system, however, is simple. The same applies to excluding text from translation. The translation direction-specific options in the Translation Options menu are more difficult to define. After all, the user must already have some idea of the source text to decide if an impersonal request must be translated by an imperative (German to English) or how the system should deal with the pronoun "you" (English to German).

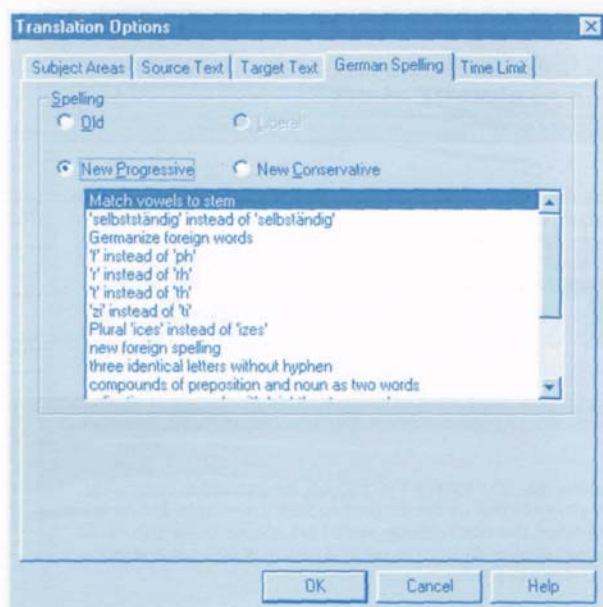


Figure 2

Many options are available for setting the spelling for German texts, including the Swiss variant of the language. A German text can use old, liberal, new progressive and new conservative spelling and for German as a target language the user can even tailor his own spelling from a list of new-spelling features (see Figure 2). For English, the choice is between American and British English, both for input and output of texts.

tions relevant to that subject area. Also determine which dictionary should overrule other dictionaries, set the sentence archive options and influence from which archive earlier translations are extracted, at which matching value. If you are just starting out and need to depend on the resources provided by the system, then you will be using the system dictionary, which contains 620,000 words. Except for a few phrases, the sentence archive, i.e. translation memory, will be empty at this point.

Translating

Once the system is configured, the actual translation can start. In choosing my source texts, I took Linguattec at their word. According to their product brochures, the type of source text does not matter that much, from the odd Web site to the e-mail to a friend in the U.S., from business letters to entire books, the Personal Translator will provide faster and better translations. Later I learnt that Linguattec themselves used mostly computer-related texts to test their application. Unaware of this at that point, I

I finally came up with was an arbitrary collection of small texts, ranging from the more general to the more technical, from private to business texts, in both translation directions.

In order to provide an idea of the capabilities of Personal Translator, I will present one of the more general texts, chosen rather randomly since the results were similar for all the translations. All of them contained words and sentences which survived the transitional process surprisingly well, and others which ended up in awkward or unintelligible translations.

One of the texts subjected to translation was taken from an English book explaining to foreigners and the interested Dutchman alike the peculiarities and habits of the Dutch (*The Undutchables* by Colin White & Laurie Boucke), a useful guide for me when I moved to the Netherlands. The passage I chose seemed not particularly complicated to me and since Linguattec claims that PT "significantly improved translation quality for texts with complex structures", I thought it fair to submit it to the application. See Table 1 for the source text and its translation by Personal Translator.

Anyone familiar with the Dutch culture will confirm that this is what many Dutch birthday parties look like. Anyone familiar with the German language can tell at first

| Source text | Personal Translator |
|---|--|
| ¹ Birthday parties begin around 8 pm and are held at the home of the birthday boy/girl. | ¹ Geburtstagsfeiern beginnen etwa 8 p.m. und werden im Hause der Geburtstagsjungen /Mädchens gehalten. |
| ² Be sure to bring flowers and a gift if you want to be invited again. | ² Achten Sie darauf, Blumen und ein Geschenk zu bringen, wenn Sie wieder eingeladen werden wollen. |
| ³ The event somewhat resembles an open house. ⁴ After entering, you will be ushered into the living room which, for the occasion, will resemble a doctor's waiting room, with chairs arranged in a circle. ⁵ On them will be seated an array of relatives interspersed with the odd friend and neighbour. ⁶ The relatives will welcome you to what at first appears to be a group therapy session, with all the appeal thereof. | ³ Das Ereignis ähnelt ein wenig einem offenen Haus. ⁴ Nach dem Betreten werden Sie ins Wohnzimmer geführt, das für den Anlass dem Wartezimmer eines Doktors mit in einem Kreis arrangierten Stühlen ähnelt. ⁵ Auf ihnen sitzt ein Aufgebot an Verwandten fügte mit dem merkwürdigen Freund und Nachbar ein. ⁶ Die Verwandten begrüßen Sie dazu, was zuerst eine Gruppentherapie-sitzung davon mit dem ganzen Appell zu sein scheint. |
| ⁷ The welcoming ceremony consists of walking around the room and shaking hands with each person. ⁸ For some unknown reason, the relatives will extend their congratulations to you, the guest -- then mutter their name unintelligibly. | ⁷ Die Zeremonie zur Begrüßung besteht daraus, um die Raum- und Zitterhände herum zu gehen, mit jeder Person. ⁸ Aus irgendeinem unbekanntem Grund fordern die Verwandten ihre herzlichen Glückwünsche zu Ihnen, den Gast - murmeln ihren Namen dann unverständlich. |

Table 1

But there are more decisions to be made. Choose a certain subject area from a long list and the application will choose transla-

ransacked my personal library and searched through my computer and the internet on a quest for sample texts. What

reading that the translation shows some lexical, syntactical and grammatical problems and would not yet be suited for

publication. However, Personal Translator is not completely unaware of this. After translation, the user can activate a log function that shows those terms and sentences that were not recognized by the system.

**Still, the translation application
wins it from the human,
especially if only informative
translations are needed.**

Personal Translator warned about a translation error in sentence 5 (see Table 1): *On them will be seated an array of relatives interspersed with the odd friend of neighbour.* The application did not succeed in recognizing the syntactical function of the term *interspersed with* that serves to further qualify the *array of relatives* (which I find very suitable translated as *ein Aufgebot an Verwandten*). While Personal Translator is in general capable of recognizing this syntactical construction (A sentence like *He entered the house built of stone* was correctly translated with *Er betrat das aus Stein gebaute Haus.*), my guess is that the translation might have failed on account of the lexical information available for the term *intersperse* (see Figure 3).

Although the translation of other sentences also resulted in syntactical awkwardness (#6) and one completely wrong target sentence (#7), no warning was issued by the application. It had been able to fit the constructions of these sentences into its algorithms, but of course could not recognize that the result would distort the meaning or at least make it difficult to grasp. The translation of sentence 7, by the way, is one of the more delightful mistranslations I came across: The source sentence "*The welcoming ceremony consists of walking around the room and shaking hands with each person.*" was translated by PT as "*Die Zeremonie zur Begrüßung besteht daraus, um die Raum- und Zitterhände herum zu gehen, mit jeder Person.*" Back-translated by myself this reads something like: "*The welcoming ceremony consists of walking with each person around the spatial and trembling hands.*"

This mistranslation is particularly interesting as the system's idioms dictionary not only contains a correct translation of the term *to shake hands* but also some cultural information on this habit. Since the dictionary had been activated prior to

translation, the only possible explanation for the error is that PT interpreted the words *room and shaking hands* as two compound nouns *room hands* and *shaking hands* and contracted them into the German *Raum- und Zitterhände*, which in principle I find quite a feat in a machine translation application.

Another interesting mistranslation, this one from another text, is a purely lexical one: *Free-lance translation* was translated as *Gratislanzenübersetzung*, which means something like "translations of lances or spears one gets for free". Although it is easy to laugh about it, one should not overlook the fact that the system was capable of recognizing that these three terms need to be turned into one German compound noun and that this conversion resulted in a grammatically correct word. The error is easily explained by searching the dictionary. The term *free-lance* misses from the system dictionary on which above translation is based, but it does contain the words *free* and *lance*. Since Personal Translator is capable of recognizing and translating compound words, the above translation was a logical one ... for a computer.

Other lexical problems in the birthday text become understandable after checking the dictionary. Take the non-existing German expression *Glückwünsche zu jemandem fordern* (translation for *to extend congratulations*) as an example. As you can see in the screenshot from the dictionary (Figure 4),

**PT interpreted the words room
and shaking hands as two
compound nouns room hands and
shaking hands and contracted
them into the German Raum-
und Zitterhände, which in
principle I find quite a feat in a
machine translation application.**

the system chose the translation *fordern* on account of the condition which specifies that the word must be related to a "living being". Some reference in the sentence to a living being (relatives?) made the system choose this translation. Mind that the correct translation was also included in the dictionary under 2.4. However, the word *congratulation* is not mentioned in the condition.

I will refrain from a further analysis of the translation. It is clear that Personal Translator is capable of coping with certain complicated grammatical and syntactical structures (see for example sentence 4 in the birthday text), while at the same time it will produce texts with shortcomings. On Linguatrec's Web site ([**Since Personal Translator is
capable of recognizing and
translating compound words,
the above translation was a
logical one ... for a computer.**](http://www.lin-</p></div><div data-bbox=)

[guatrec.de](http://www.linguatrec.de)) and in the PT online help text, ample and reasonable warnings are issued as to the limits of machine translation. One factor that is mentioned is the suitability of the source text for machine translation. Of course, sentences of limited length and with simple sentence construction will lead to better results. Another factor of importance, demonstrated by the examples given previously, is the quality of the dictionary.

Fine-tuning the system- dictionary and sentence archive

As mentioned earlier, Personal Translator ships with an extensive system dictionary that contains 620,000 entries (in the Office Plus version). These entries include what I would like to call "semantic clues" allowing the system to make a choice if a word has more than one translation. A subject area or condition might be such a clue. According to Linguatrec, PT is even able to analyze these clues across sentences. In addition, grammatical and syntactical information, e.g. on prepositions, type of objects, etc. can also be stored with every word in the dictionary.

The system dictionary provided with the application is read-only, which seems a very reasonable precaution to me. If a user wants to add new entries or edit existing ones, then these will be stored in the user dictionary. Terms can be looked up in as many as 16 dictionaries accessible at the same time, user-defined, imported from lists or bought from Linguatrec, but only the one currently defined as user dictionary can be edited. By typing in a translation and choosing among the basic grammatical forms noun, proper noun, verb, adjective and adverb, one gets a stan-

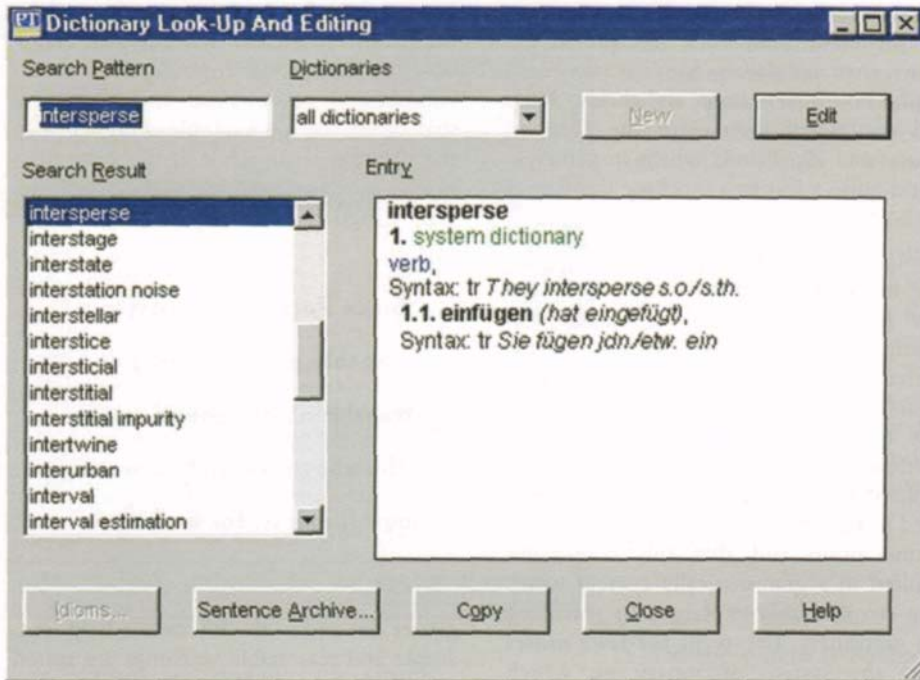


Figure 3

standard dictionary entry that then needs to be adapted and fine-tuned by what is called “extended coding”.

I found it fairly easy to add new translations and manage the extended coding options for nouns, proper nouns, adjectives and adverbs. For example, after having defined *free-lance* as an adjective and as an adverb to be translated with *freiberuflich*, no more *Gratislanzen* appeared in my translations. However, to my shame I must admit that despite several attempts I never managed to define a verb and in particular its syntactical information in such a way that

it resulted in a satisfying translation. My impression is that it requires a person with a good grasp of the grammar and a good knowledge of both languages to update the dictionary.

Personal Translator also contains another feature that helps the user to get the most out of a translation, i.e. the sentence archive or the translation memory. Every time a sentence is translated, the source sentence and its translation are stored in an archive. Depending on the archive settings, Personal Translator will use an already existing translation if it encounters the same

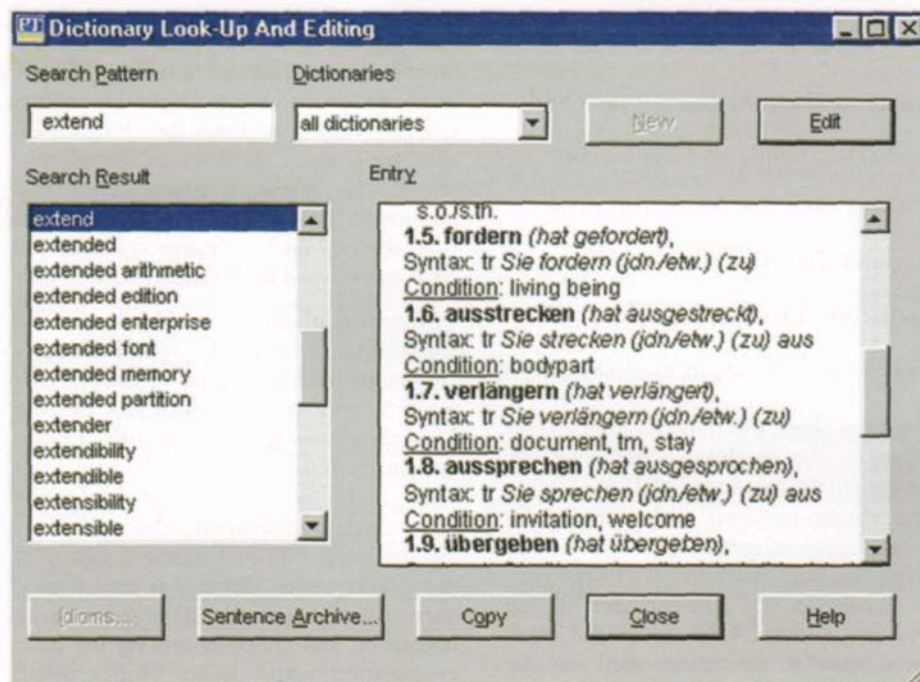


Figure 4

or a similar source sentence in the translation it is working on.

While each and any sentence translated by the application is added to the automatic archive, edited translations will be entered into what is called the manual archive. Linguatrec advises to always activate the archiving function when the target text is edited in the post-translation phase, so that the final sentences are added to the manual archive. I personally found the sentence-by-sentence archiving option rather awkward to work with. Source sentence and target sentence are presented in a dialog box in such a way that they are rather difficult to tell apart and editing text right in this window is not possible (see Figure 5). This is a point that is definitely worth some more attention from Linguatrec’s developers.

However, Personal Translator’s sentence archive offers features comparable with dedicated translation memory applications. A text can be analyzed before translation to see how many sentences will be pre-translated using the archive. The color coding of the target text will provide information on the matching value for the individual sentences. Archived sentences can be stored with additional information (attributes), e.g. on who did the translation. And users can also add previous translations to the archive by placing source and translation in the respective windows, in a process that other applications call *alignment*. Just like the dictionaries, sentence archives can be easily searched for translations and edited, imported and exported, and prioritized for use in the translation.

After working my way through almost all features and functions of the application it became clear to me that Personal Translator is not a system in which all you have to do is push the Translate button. Of course, you can restrict yourself to this, but then you have to accept the inadequacies of the resulting translation. If one is aiming for more quality, then the system needs to be brought and kept alive by adding other dictionaries, by maintaining and updating the existing ones, and by continually adding translations to the manual archive.

Who benefits most?

When I was asked to evaluate Personal Translator, one of my first questions had been: Who is it for? “Professional translators”, was the answer that I was given. And it is true—to get the most out of this system and to produce translations of publishable quality, a person with a good knowledge of the source and the target

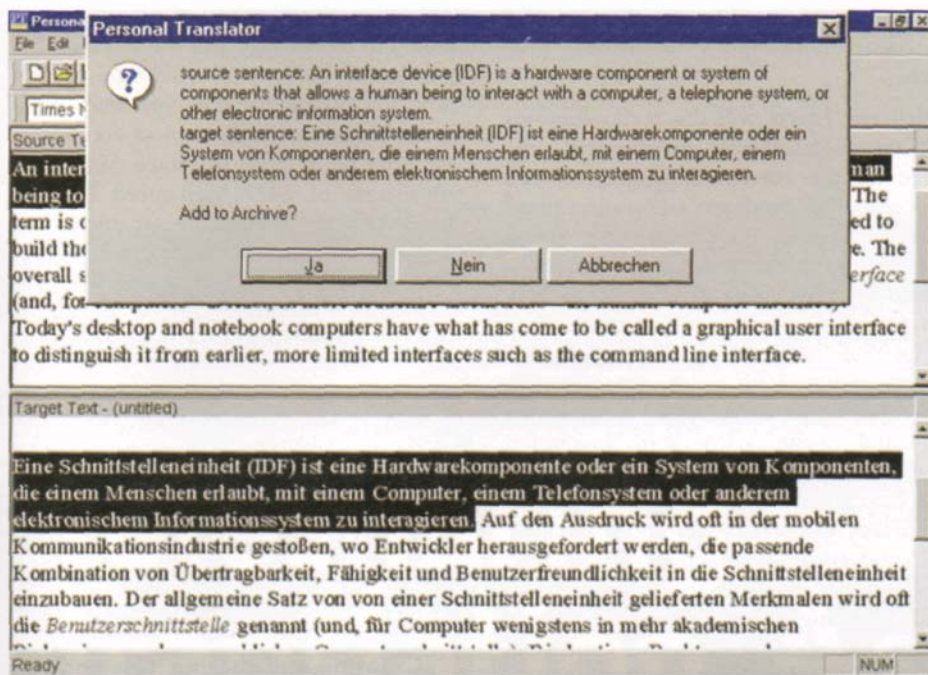


Figure 5

language and of the mechanics of translation is needed.

But there are also many features that seem to have the non-professional user in mind, one who is primarily interested in information retrieval, where grammatical and syntactical shortcomings are acceptable. Take, for example, PT Web, which enables users to translate Web pages directly in Internet Explorer or in Netscape Navigator, while preserving the layout of the original page. Or take PT Direct. If you have a text in front of you, no matter in which Win-

Within these restraints and within a contained system of repetitive texts, this cost-efficient, easy-to-use and easy-to-learn versatile system will surely have its merits.

dows application, and need to know the translation of a sentence, PT Direct will provide "die kleine Übersetzung zwischendurch", as Linguattec so nicely phrases it. Another example is the pronunciation feature, which reads dictionary entries to you or the full source and target sentences while you are translating. Nice as these features are, I think that for the professional translator they are more gimmicks than important tools.

The sentence archive, however, seems to have been added with the professional translator in mind. Reuse of earlier translations is a feature of interest to a professional who works with recurring texts. Next to reuse, a professional translator's considerations in choosing such a system might be speed and quality.

As to speed, PT is definitely faster than the human translator in the actual translation phase. Even if a human translator would not have to go through the cyclic process of analysis, understanding, researching and recreation of text to come to a first draft, typing alone of a target text would take longer. Of course, in order to be able to compare the speed of a human and automatic translator, the time needed for pre- and postproduction has to be taken into consideration too. Still, the translation application wins it from the human, especially if only informative translations are needed. And over time the speed will increase with a growing dictionary and a growing sentence archive.

Linguattec had a usability study performed by the German Fraunhofer Institut according to which use of the Personal Translator resulted in average time savings of 41%. The study involved 40 candidates with basic computer knowledge and varying knowledge of English who were asked to translate and write business letters with and without the use of PT (see Linguattec's Web site for details and results of this study).

As to quality, this study brought about a result that I found rather telling: While users with a low or medium knowledge of

English produced target texts of better quality when using the Personal Translator, users with above-average knowledge of English, a category to which professional translators belong, produced texts of lesser quality when working with the automatic translation application. Apparently, the word-for-word translation that they edited to generate the final version restricted them in the use of their full idiomatic language creation capabilities.

The answer to my question might therefore be somewhere in the middle. Translations, also for use in a professional and business environment, are produced by people with a variety of trainings and backgrounds. By professional translators I mean those highly-trained individuals who are capable of dealing with complex texts which they can recreate according to the needs of their clients in another language. This group of professionals, and I am thinking in particular of free-lance translators, will have to compare the qualities and pro and cons of using a application such as Personal Translator with those of translation memory software.

But then there is also a wide group of people employed in business and industry with its ever growing need for international communication that is not trained for translation as such, but needs to produce and understand texts in a foreign language. They are probably the ones who can benefit most from Personal Translator Office Plus. However, it seems very important to me that these users of PT are aware of the fact that no machine translation application produces publishable quality at the first go. Within these restraints and within a contained system of repetitive texts, this cost-efficient, easy-to-use and easy-to-learn versatile system will surely have its merits.

Angela Moisl is German language specialist for Lionbridge in Amsterdam. She can be contacted at angela_moisl@lionbridge.com.