

Trados: Ten Years On

Trados may be small — currently just five people — but the Stuttgart company enjoys a well earned reputation in the linguistic engineering community for being both innovative and highly productive software developers.

Now entering its second decade, Trados has built up a solid technological base and a good market position: its MultiTerm package is the leading terminology management package at the moment. But success has not come to Trados overnight. Rather, since 1988, when Trados began developing its software, it has slowly but steadily built up a customer base. “1992 was really the turning point” says Jochen Hummel, who co-founded the company with Iko zu Knyphausen in 1984 to supply translation services to IBM. “1992 was the year we released MultiTerm for Windows and our DOS Translator’s Workbench package” He adds that it was also the year in which contacts that Trados had cultivated resulted in some sizable site licenses for the company’s software. Trados has also thrived on the growing popularity of local area networks. Whereas it takes considerable discipline for individual translators to develop a standalone terminology database, a shared termbase is such a natural application for a network that it is almost inconceivable for a high-tech organization to be without one. Naturally, technology is just half the equation; there is also the question of content. Growing amounts of data, including source texts in electronic form, previously translated materials, and specialized terminology databases, also contribute to making termbases and repetition managers (translation memory) increasingly valuable components of the multilingual documentation production process.

With a growing market comes increased competition and Trados cannot afford to stand still. Correspondingly, the company is in the process of launching an array of new products, including the long-awaited introduction of the Window’s version of the Trados Translator’s Workbench, MultiTerm Pro, MultiTerm Lite, and MultiTerm Dictionary, a package for compiling read-only termbases. Earlier this year, the company started shipping TAlign, a utility for automatically generating translation memory files from existing translations for use with the Workbench package. TAlign was written by Matthias Heyn, a computational linguistic from the University of Stuttgart who after serving as a consultant for Trados for several years has joined the company as its third partner. Based on the research done at AT&T’S Bell Labs by Ken Church and Bill Gale in the statistical analysis of text, TAlign is to our knowledge the first commercial bilingual text alignment package to reach the market.

Whereas the DOS version of the Trados Workbench was a standalone package with an integrated editor, the new Windows version can be used with standard Windows wordprocessing packages via the Windows DDE interface. Trados supplies DDE macros for integrating its software with Word, WordPerfect 6.0, and Ami Pro. Like the DOS version, the Windows Workbench includes MultiTerm, enabling users to scan source texts for both exact and approximate matches of terms as well as phrases. Hummel, who was largely responsible for the development of the Trados Workbench, says that any file size limitations found in the DOS version are no longer an issue. Trados has incorporated fuzzy matching algorithms licensed from a German University which make it possible to search the translation memory database with blinding speed, even across translation memory databases tens of megabytes in size. These algorithms, which are based on neural network techniques, are language independent and they work in conjunction with the compound matching algorithms newly developed by Heyn. “In German, Dutch, Danish, and Swedish, you find a lot of compound nouns, particularly in technical texts, where sixty to eighty percent of nouns can be compounded” says Heyn. “However, often you won’t find the compounds listed as such in your termbase, or they

are listed inconsistently. If your search routine can only handle single words, you're lost!" Heyn's compound recognition routines can handle up to four-part compound split genitives and separable verb prefixes.

Hummel is thoroughly convinced of the soundness of Trados architecture, arguing that conventional relational databases are not necessarily well suited for the storage of linguistic data. One way in which he likes to demonstrate the speed of the Trados search routine is by making a search with a leading wildcard (an asterisk). It takes only a few seconds longer to perform than normal searches. "Don't try that with relational database" he says. Not only is searching very quick in the new package but importing and loading translation memories is impressively fast. The fuzzy matching scheme has also been incorporated in the most recent version of MultiTerm, MultiTerm Pro.

As with all such repetition managers, the Workbench translation memory is most effective on suitable material. Not surprisingly, this is primarily highly repetitive technical documentation. Freelance translators or teams of translators who handle a variety of diverse materials will find it less effective. Hummel is keenly aware of this and also demonstrates using the fuzzy search function as a kind of bilingual concordance tool, in which a translation memory file is used not for processing repetitions but as an online reference to see how a given term has been used in previous translations. This corpus browsing function is one of the most attractive features of both the Windows Workbench and MultiTerm Pro, partly because the response times are excellent.

Naturally, the Trados Workbench translation memories can be shared among multiple users. Hummel says the maximum number of concurrent users depends on a number of factors (server configuration, network speed, etc.) but in practice is likely to exceed the number of translators working on the same text. The Workbench can be configured with two translation memory files active; this would comprise typically foreground and background files, whereby the foreground file is specific to the project and updated instantaneously and the background file is read-only and is updated manually with more general text.

Designing an add-on for an existing wordprocessor may seem less work than starting from scratch but looks are deceptive. Knyphausen and Hummel have had numerous problems tackling the file formats for packages like Word and WordPerfect. Says Knyphausen, "we preserve the formatting of the original text, including display attributes. That means in essence that our programs are like little wordprocessors in themselves" With the complexity of today's programs, that means a lot of file format detective work, and the situation is not made easier by rampant bugs in the early releases of the new versions of these packages. Hummel says the most fundamental problem is the lack of a standard document interchange format. "All of the programs implement formatting slightly differently. Even RTF isn't always reliable" he laments. "After all these years, we still don't have useful standards in this field" Like countless software developers, Trados spends a depressingly large proportion of its time wrestling with proprietary file formats; like its colleagues, it is essentially having to rediscover the wheel each time, devoting time that would be far better dedicated to improving the functionality of their programs.

Another new addition to the Trados lineup is MultiTerm Dictionary, a package for creating read-only terminology lists. This was developed to meet the needs of several of Trados's larger customers, who wanted to be able to distribute electronic terminology lists within their organizations. MultiTerm Dictionary incorporates compression technology licensed from AND Software, a Rotterdam company which has made a specialty of data compression for electronic publishing purposes. Trados is the first company to license AND's compression library, CompLex. Thanks to CompLex, MultiTerm Dictionary achieves an optimal balance of speed and memory requirements.

MultiTerm Dictionary consists of two parts: a dictionary compiler which incorporates the CompLex engine and an

access program, which, aside from the lack of editing functions, largely resembles elder sibling, MultiTerm. MultiTerm Dictionary is highly configurable, providing the termbase compiler with a large latitude of control over its onscreen personality. It also offers “drag-and-drop” pasting of terms into any Windows wordprocessor.

While MultiTerm Pro and MultiTerm Dictionary address the needs of large users, Trados is also catering to the needs of freelance translators with its MultiTerm Lite. Lite is similar in functionality to Pro, only it lacks the latter’s support for networks and graphics and its termbases are limited to eight thousand entries in size.

On a consultancy basis, Trados also does data conversions for its customers, and earlier this summer, Knyphausen was readying a termbase for the European Parliament. Called Euterpe, the termbase includes some fifty thousand entries in the nine working languages of the EU, and Knyphausen was faced with a delicate political consideration in the otherwise seemingly innocuous interface of the program. Onscreen, MultiTerm Dictionary uses a small flag to denote the language of each given entry; it is a small but effective visual clue for navigating a termbase, especially one with nine languages.

However, for the socio-linguistically sensitive Eurocrats, national flags are a contentious symbol for a given language, having the natives of Belgium, Luxembourg, Ireland, and—shortly—Austria to contend with. A solution came in the form of making it possible to substitute text tags for the flag icons, a more politically correct solution in the most literal sense of the phrase, albeit a more prosaic one.

Hummel and Knyphausen acknowledge that it has occasionally been difficult in the past to convince potential of the long-term viability of the company. Part of that may have been due to the relative youthfulness of the two—they founded the company in their early twenties—as well as their lack of ‘formal’ credentials in a country which dotes on titles. Eschewing such pretensions, Knyphausen and Hummel much prefer their software to be judged purely on its technical merits but inevitably potential customers also take other factors into consideration, such as the long-term stability of a given supplier. However, Knyphausen points out that the size of a company is not necessarily a guarantee of its commitment to a given product. “IBM could drop TM/2 and it would hardly make a ripple within the company” he says. “But our products are our only means of income” This point of view is only bolstered by the recent demise of Keyterm. Years of putting in appearances at CeBIT and other trade shows have reinforced Trados’s presence and image of stability. allowing the company has built up a what is by now an impressive customer list.

Among companies and organizations which have settled on the Trados software are the Swiss PTT and the European Parliament, the latter of which is probably the largest Trados user, upwards of a thousand users (including freelance translators). Another big Trados customer is Oracle Europe, which settled on the Workbench as its standard platform last fall and has since required its translation service vendors, including Softrans (Dublin) and Donnelley Language Services (Amsterdam) to work with the package. On the far side of the Atlantic, Trados’s US subsidiary, MCB Systems (San Diego) has also been booking some substantial successes, most notably the International Monetary Fund, which recently installed a fifty-user MultiTerm system.

One factor much in favor of the Trados software is that it is written by people who know how translators work. As many people in the translation community will know, Trados’s other personality was that of “INK Deutschland” for many years, and indeed Hummel and Knyphausen originally established Trados to supply translation services to IBM. In the early days, they too worked with the INK TextTools, one of the first translation packages for the PC. INK eventually stopped further development of the package, giving Hummel and Knyphausen the incentive to develop a follow-up to it, and that became MultiTerm. As the software development side slowly evolved, the translation business

provided some welcome financial support as well as a useful testbed. However, with the acquisition of INK by RR Donnelley and the subsequent reorganization of the INK network, Hummel and Knyphausen closed down the translation side of things to concentrate entirely on software development. To complete the transition, the company recently relocated from an office in downtown Stuttgart to a leafy residential neighborhood where they occupy the top half of a handsome pre-war villa discretely wired for Ethernet. One of the many attractions of the new quarters is a roomy top floor which Trados will use for training purposes.

Trados has employed contract programmers in the past but Hummel and Knyphausen, together with their new partner, Matthias Heyn, prefer to rely primarily on their own forces. "It can be extremely difficult to motivate people to go the last measure in software development" says Knyphausen. "When the interesting problems have been solved and it is just a matter of testing and debugging and refining the interface" However, during the past year, they hired a crack object-oriented programming consultant to assist them make the difficult transition to an object-oriented software architecture. "We never could have done that alone" says Knyphausen of the grueling experience. "There were times when we were ready to tear our hair out" But both of them say the result was worth the effort; with now a nearly complete foundation of base class modules, the Trados software will be easier to maintain and improve in the future. While object-oriented programming does not achieve anything that can't be done with traditional methods, Hummel points out that certain enhancements, notably extending the linguistic coverage of modules are facilitated by this approach. In principle, the Trados software supports upwards of twenty languages, including national variants, but in practice 'support' ranges from just being able to indicate it in a menu to full morphological analysis, for which Trados obviously supports fewer languages. If a customer wants better support for Hungarian, and is willing to pay for it, Trados can easily implement it thanks to the well designed modular architecture of the software.

Up until recently, the users of Trados software who called the company for technical support more often than not got either Hummel or Knyphausen themselves directly on the line, but now Trados has a full-time person just for handling user support and training, freeing up the three developers to concentrate on programming. "Taking support calls certainly sharpens your commitment to good product design" notes Hummel wryly. "After the fifth or sixth caller comments on the same anomaly, you sure won't forget to change it in the next release"

For Hummel and Knyphausen, developing commercial end-user products is an exacting craft, requiring a rare combination of imagination, technical skills, practical knowledge, and sheer stubborn perseverance, things that can be only successfully cultivated among a small team. That makes them all the more sceptical about big projects for which dozens of programmers are deployed. "Big teams of programmers are simply not efficient" states Hummel unequivocally. "There is far too much communications overhead. Moreover, you can't simply throw more programmers at a project which is running behind" he says. "that's counter-productive" Granted, though, being small does have its limitations. Trados does not have the resources for extensive 'engineering' endeavors, such as porting its software to different hardware platforms, for example. Nor can it embark upon ambitious efforts to develop the linguistic resources needed for deeper levels of syntactic and semantic processing. Both Hummel, Knyphausen, and Heyn are keenly aware of this. For this reason they are eagerly on the lookout for technologies which they can license, such as the neural network search algorithms and the data compression schemes which have been incorporated in the latest generation of their software. Hummel expresses the hope that eventually the large sums of money the European Union puts into NLP R&D will also result in public domain materials, such as lexicons a immense value. Hummel sees this as a healthy division of labor; universities and research institutes developing the underlying linguistic processing modules and basic resources, while companies like Trados supply the user-interface development skills to create commercial products. Neither is the company adverse to the possibility of co-marketing arrangements, although at the moment none are in place outside that with its North American distributor, MCB Systems. Over the years, Trados's

marketing has been low-keyed but certainly sufficient to keep the modestly scaled operation going.

Inevitably, comparisons arise between the Trados Workbench and the new EuroLang Optimizer package, and it is undoubtedly daunting for a company the size of Trados to be suddenly faced with a competitor many times its size. Not only that, IBM is entering the fray as well, with a modestly priced Windows version of TM/2. But as Jochen Hummel points out, each of these packages reflects distinct philosophies in their approach to computer-aided translation, and offer different solutions for different problems. Optimizer, at least on paper, promises support for a variety of systems and is based on an ambitious client/server architecture designed primarily for the management of large translation jobs. The Trados Workbench, on the other hand, offers more refined linguistic analysis and has been carefully engineered to increase the productivity of single translators and small workgroups. Hummel sums it up as the difference between a hierarchical and a collective approach.

Looking towards the future, Hummel envisages being able to offer incremental advancements of source analysis in their software, raising the translator's productivity step-by-step. Naturally, there will be challenges in achieving this across all twenty or so languages the company supports, but Hummel suggests that there might be techniques which they can employ, such as statistical approaches, whereby the need for large amounts of linguistic information are partially circumvented. Also on the horizon is the move to Unicode, possible on the basis of the Gamma Unicode Server.

Well positioned as market leader in a small but lucrative niche, Trados nonetheless faces increasing pressure from both below, as mass-market oriented developers set their sites on the large foreign language and English-as-a-second-language market, and above, as established players, such as IBM and Siemens, look for ways of exploiting the sizable NLP R&D activities within their organizations. Against the background of a rash of consolidations in the software industry and the disturbing sight of some of yesterday's most successful software companies now faltering, Trados supplies compelling evidence that enterprise and innovation are still very much alive and well among smaller players.