

# Torte and Tautology

The Fourth International Conference on Terminology and Knowledge Engineering (TKE 96) in Vienna provided a snapshot of the fast-moving terminology industry. So what is the present picture?

The Technical University in Vienna hosted TKE 96, as a reminder of Austria's commitment to terminology development and harmonisation which, in recent history, stretches back a quarter of a century to the founding of Infoterm in 1971 but which, as the Austrian government representatives pointed out, at the same time, echoes the progress made in the 19th century towards multi-lingualism in the former Austro-Hungarian Empire.

TKE 96 provided the opportunity for computer scientists, terminologists and users from around the world to come together under the same roof and the ensuing mix of disciplines and viewpoints made for a very stimulating few days.

The tone of the opening speeches gave the delegates a sense of the high stakes to be played for in this field. The European Commission Director of Informatics, *Eduard Brackeniens*, outlined the financial and administrative challenges the European Union faces with 11 official languages at present, and the daunting prospect of these increasing to 20 in the foreseeable future. Nevertheless, he stressed the liberating role multi-lingualism has in expanding people's ability to make full use of their potential, with the ringing phrase 'diversity = democracy'.

For the Austrian government, *Mr Kneucker* and *Mr Kögerler* emphasised the importance of swiftly developing standards in terminology which could foster effective and inclusive free trade for the benefit of all.

During the two day conference 51 papers were arranged in nine sessions, of which the three largest charted terminology's relationship with the Philosophy of Science, Language Engineering and Knowledge Data Management. It was clear that enormous technical progress had been made in all areas in the three years since the last TKE conference, especially with regard to work strategies, use of the World Wide Web and development of language technology.

## Work strategies

Some intimation of what the future may hold came in presentations which showed the way a number of universities are now structuring their research so that it runs in tandem with the needs of industry. Managing research as a consortium in this way allows attention to be focused on the needs of the end user or customer and the product can be made available in

a realistic time frame, while the university benefits from investment in skills and equipment. Once the project is shown to be viable, it can be honed efficiently in a real time situation.

There are drawbacks to this approach. As *Robert de Beaugrande* of the Institut für Anglistik at Vienna University pointed out, the impulse to preserve competitive advantage can be dangerous if it leads to greater exclusion, where results are not shared and work is duplicated. However, now that time and efficiency imperatives are so strong, such consortia seem to be an increasingly popular way of delivering results.

*Eva Dauphin* for Aero-spatiale and *Khurshid Ahmad* for the University of Surrey gave the results of the Transterm project which took place in conjunction with GSI-Erlä and Electricité de France. Electricité de France was also represented by *Elie Naulleau* who spoke on the subject of tagging term components with semantic information. Transterm is a toolbox for recycling industrial terminology which consists of corpus management tools, using a Virtual Corpus Manager developed by *Lee Gillam* called SystemQuirk, morphological analysis tools and a terminology management system.

*Jörg Schütz* gave the results of the University of Saarbrücken's project with BMW and Mercedes to create an on-line multilingual technical information service for the car industry. This combines artificial intelligence, language and Internet technologies to give an integrated and intelligent information brokering system.

After depicting the changes which the South African National Translation Service has undergone in recent years and giving an insight into Afrikaans terminology development, *Milde Jordaan-Weiss* brought news of the University of Pretoria's work with EpiUse in producing machine translation software for a variety of other indigenous South African languages, such as Swahili, Setswana, Sepedi, Zulu and Xhosa plus six European language pairs.

*Walther von Hahn* presented DB-MAT, a joint venture in improving German/Bulgarian knowledge

Gary  
Goldfinch



*Eduard Brackeniens*

representation in Machine Aided Translation between the University of Hamburg and the Bulgarian Academy of Sciences, sponsored by the Volkswagen Foundation.

Other consortia with a more uniformly industrial base included the Finnish MT development company Kielikone, in partnership with Nokia and represented by *Harri Arnola* and the work of *Peter Weingartz* of Media Online in Kempen in the development of an ISO which will permit standardised data exchange between suppliers and customers of specific design objects or technical catalogues. This work is the result of a number of major German industrial companies pooling resources.

### Use of the Internet

A number of papers here focused attention on how the Internet is being used in a dual capacity. The first of these is the use of Web sites for the capture and monitoring of terminology data. The second is the use of the Internet for on-line delivery of machine translation. At the moment these remain distinct, but it is clear that technology is converging on the information superhighway.

Work on capturing and monitoring data is currently progressing for both database management and information retrieval. Under the first use comes the work of the University of Rennes 2 who, as *André Le Meur* explained, have set up the Balneo computer network to attest to French neologisms, and to manage, update and exchange them. Also active in this field are *Erich Schweighofer* and *Dieter Scheithauer* of Vienna University in researching legal terminology on the Internet and *Christine Jacquin* and *Maurice Liscouet* of l'Institut de Recherche en Informatique de Nantes in pursuing terminology extraction from text corpora available on the Internet.

Work in mapping and retrieval techniques for terminology information is mainly focused on visualisation. *Christian Galinski* of Infoterm, Vienna in his discussion of semiotics, mentioned the role alphanumeric, graphic and pictorial nonverbal forms may have in representing the information content of the future. *Heiner Benking* of the Research Institute for Applied Knowledge Processing in Ulm proposed a data organisation scaffolding as a development from existing 'universal' ordering systems to improve access to the increasing amount of terminology held in specific domains.

As for the use of the Internet to deliver machine translation, this contains a variety of work which was elaborated during the EAMT workshops immediately following TKE96.

### Development of language technology

This encompasses the development of language technology to embrace languages besides English

and especially what *Magnar Brekke* from the Norwegian Business School at the University of Bergen calls 'lesser-used living languages'. As one might expect, the development spread in this area is immense. A great deal of progress has already been made in some European languages such as French and German, while in languages like Welsh, Catalan and others such as Norwegian, (See *Fergus Geraghty's* article in this issue), the work is just beginning. *Magnar Brekke* and *Khurshid Ahmad's* paper on the viability of a corpus-based approach to terminology collection (possibly using the Internet) in these languages seemed to point to a promising future.

Japanese is the leader among the East Asian languages. Speaking for the in-house team at Ricoh Ltd in Japan, *Masumi Narita* discussed the *tanyukuman* or 'language assistant', an English to Japanese word translation device for use on a normal fax/copy machine, which can, for example, annotate an English copy text with Japanese word translations. To achieve this, it contains an Optical Character Reader, a Language Processor and a Layout Controller. Within the constraints of a dictionary less than 2Mb in size, if desired the *tanyukuman* can even supply the translations in a workbook style format, graded by difficulty.

On the other hand, *Reza Mansuri* of Potsdam University indicated the great strides which had to be made in Iran, where terminology work began almost from scratch after the 1979 revolution, following many years of Anglicisation. Handicapped by lack of funds and the problems involved in adapting the cursive script to an on-line environment, a telephone hot-line for neologisms has now been established. All these speakers expanded by example on what *Eduard Brackeniers* had said earlier about the benefits of own language development to knowledge dissemination and democracy.

To sum up, TKE96 was characterised by an eclectic range of papers and opinions and by an equally broad range of activities. Thanks are due to the Austrian Government and the Institut Français for their respective receptions, and to the TKE96 organisers for preparing such fertile ground.

### ■ Exeter change

The Dictionary Research Centre at the University of Exeter, UK, which plays a leading role in European lexicography, became part of the university's School of English and American Studies in August. The School also took in at the same time the Department of Applied Linguistics.