Translating and the Computer 26 Conference

Abstracts

Day One: 18 November 2004 (08.30-17.00)

Chair of Day One: Professor Tony Hartley

Day One Presentations:

The Status of the Free-Lance Translator in the 21st Century: Greatness or Misery, Prestige or Mediocrity, Globalization Hero or Victim? *Yves Champollion, France*

Globalization means different things to different people. For translators, globalization is the age when people talk to people without borders and through more and more diverse media. The internet makes communication instant, free, ever-present, free-flowing - but the one thing computers can't do is translate. This is one of Globalization's bottlenecks, which places the translator in a crucial situation. The increase of translation is a technical challenge. It is also a moral issue. Translators are what the Greek language called hermeneuts : they don't only convey information, they serve as bridges, go-betweens, almost ambassadors between people of different cultures. The translator is not just another trader who exploits the lack of understanding among people to make a living, he has to remain a person who offers his diplomatic skills, his vast culture, to make people understand each other, work together. The time has come for the creation of a world guild of translators. The following paper wishes to offer a historical and moral perspective for this endeavour.

Modern MT Systems and the Myth of Human Translation: Real World Status Quo. *Richard Jelinek, The PetaMem Group, Germany/Czech Republic*

This paper objects to the current consensus that MT systems are generally inferior to human translation in terms of translation quality. In our opinion, this belief is erroneous for many reasons, the both most important being a lack of formalism in comparison methods and a certain supineness to recover from past experience. As a side effect, this paper will provide evidence for a much more favourable judgement of the performance of contemporary MT systems. We will present and discuss known methods of evaluation, give real world examples of both machine and human translation and finally suggest a universal formal evaluation method to handle both human, as well as machine translation output in a comparable fashion.

Machine Translatability and Post Editing Effort: How do they relate? *Sharon O'Brien Dublin City University, Ireland*

Now it is 2004 and we still hear reports about the ever-growing demand for translation. According to an IDC survey of the globalisation, internationalisation and localisation market, the annual growth rate is 16.3%. The localisation and translation services part of this market are growing at an annual rate of 14.6% (Van der Meer: 2003). Clearly, this level of growth increases the need for faster translation throughput. In addition, there is a growing demand for "gist" translation , or translation for informational purposes only. Controlled Language works by reducing

the occurrence of so-called "translatability indicators" (Bernth 1999a, 1999b, Bernth and Gdaniec 2002) in a text. By reducing the number of translatability indicators in the text, we are, hopefully, also reducing the post-editing effort. This paper will explore this assumption in more detail by asking the question: what is the correlation between specific translatability indicators (in English) and post-editing effort (for a German target text)?

Controlled Translation as a New Translation Scenario – Training the Future User. *Celia Rico and Enrique Torrejon, Universidad Europea de Madrid, Spain*

Focusing on the translation needs of the automotive industry which were discussed during the 2002 TOPTEC Symposium as reported by Sharon O´Brien, it was agreed that three main areas need to be carefully catered for: effective terminology management, increasing use of hybrid automation models comprising TM and MT, and rising demand of translators specialized in the automotive domain. We developed a training programme as a result and this paper specifically addresses the needs highlighted with examples of student sessions illustrating the use of controlled language rules to pre-edit texts from service manuals, the changes to improve MT output, the use of post-editing rules of Spanish MT output as well as the difficulties encountered. Finally, the paper concludes with a proposal for a formal training program specifically suited for freelance and agency translators planning to work for the automotive industry. The program reviews the background knowledge required for this technical field and has as its backbone the use of ENGSPAN and other MT software (Systran, Promt XT) in the controlled translation process of automotive information.

Assessing Controlled Language rules: Can they improve performance of commercial Machine Translation systems? *Johann Roturier, Dublin City University, Ireland*

This paper presents an evaluation of a set of Controlled Language rules selected within the framework of a feasibility study of translation automation. The main objective of the project is to improve the machine translatability of an English corpus of a global Internet security technology company, Symantec. The corpus contains alert notifications stored in XML format that are generated from an SQL database. Due to the time-critical nature of this type of communication, which needs to be promptly distributed in a number of languages, MT presents itself as a prospective candidate.

Bringing together writers and translators – XML based content management of product documentation at Autodesk. *Mirko Plitt, Autodesk, Switzerland*

As a leading CAD software company, Autodesk produces each year tens of thousands of pages of printed manuals and online help which are translated into up to twenty languages. Until recently, as in many other companies, the content creation and translation processes consisted to a large extent of disconnected, time-consuming manual steps, causing high cost. Based on this analysis, Autodesk decided to integrate the product documentation processes into one single Content Management System which spans the content lifecycle from initial creation through update and translation to publishing. Autodesk's CMS project started in early 2003 as a shared initiative across its authoring teams and the localisation team. In a first phase, criteria for the system were defined, followed by a thorough vendor selection. As no commercial product seemed to cover the entire set of criteria (vendors tend to be

strong on either authoring or translation, but not both) the chosen vendor had to first develop the missing features. A successful production pilot was carried out during the first months of 2004, and the system went to production in June of this year.

Multilingual document management and workflow in the European institutions
Dr Joanna Drugan, Centre for Translation Studies, University of Leeds

In the year when ten new accession countries have joined the European Union, the institutions are jointly reviewing effective multilingual document management strategies and workflow. The integration of a range of translation software and tools is also under discussion, not least because a new Call for Tenders for the supply of a translation memory tool for all EU institutions is in process. What problems relating to data management, integration of translation technology and workflow have been identified by one of the world's largest suppliers of multilingual translation? What solutions have been suggested and finally selected, and what was the rationale behind such decisions? How successfully have such solutions been implemented within and across the EU institutions? And what new challenges are anticipated by those managing workflow in the EU? This paper will address these questions by presenting original findings from a recent six-week research placement during which the author was based in the English Unit of the Directorate General for Translation of the European Commission, with additional access to translation tools specialists, members of the committee on translation workflow, translators from both pre- and post-accession member states, Heads of Units and indeed other institutions, including the Parliament, Court of Auditors and Translation Centre, both in Luxembourg and Brussels. Research took the form of tailored interviews, work shadowing, performing translation tasks using in-house tools and observing training in translation tools. Issues relating to translation workflow and data management were thus investigated across different languages and institutions, resulting in a comprehensive, up-to-date and comparative overview of effective translation management strategies within an expanding multilingual non-profit organisation. Research findings will be of interest to a wide audience including translators, managers, trainers and academic researchers. The effect of latest developments in translation software and user aspects of translation software will be examined via a practical problem-solving approach.

Development and Fielding of the Phraselator ® Phrase Translation System
Ace J. Sarich, Marine Accoustics Inc., USA

Over the past five years, VoxTec, a division of Marine Acoustics, Inc., teaming with SRI International, has developed and refined and fielded the Phraselator®, a handheld one-way voice-to-voice phrase-based language translator. The Defense Advanced Research Projects Agency (DARPA) and a DARPA Small Business Innovative Research (SBIR) grant have funded the research and development and initial production. VoxTec has evolved the concept from a PC based system to a handheld device. After the 9/11 terrorist attack on the World Trade Center, Phraselator development was accelerated. Initial deployment of the prototype Phraselators to U. S. Military forces operating in Afghanistan in support of Operation Enduring Freedom began in March, 2002. Lessons learned from Afghanistan were folded into the redesigned Phraselator P2. To date, over 2,000 Phraselators P2s have been manufactured and delivered to users around the world.

Day Two: 19 November 2004 (08.30-17.30)

Chair of Day Two: Daniel Grasmick

Day Two Presentations:

Multilingualism in an enlarged Union - Challenges and opportunities for the European Commission's translation service.
Klaus Ahrend, Acting Head of Unit, European Commission, Directorate-General for Translation

Rule-based versus statistics-based MT
Dr Gregor Thurmair, Linguatec

Parallel concordancing and translation. *Michael Barlow, University of Auckland, New Zealand*

In this presentation, the use of the parallel concordancer, ParaConc, a Windows program designed to analyse parallel (translated) texts, will be demonstrated. Michael will also discuss some advantages and drawbacks of this kind of program for different users. The successful searching and analysis of parallel texts depends on the presence of aligned text segments in each language corpus (and, of course, on the availability of parallel corpora). The alignment utility in ParaConc is semi-automatic. When files are loaded, the user enters information about the format of the files either through reference to SGML tags or via specifications of patterns. The user specifies the form of headings and the form of paragraphs. ParaConc uses the information to align the documents at this level and the user can make adjustments by merging/splitting units, as appropriate.

Searchable Translation Memories. *Chris Callison-Burch, Linear B Ltd., UK*

Linear Bs' searchable translation memories allow a translator to type in a phrase and retrieve a ranked list of possible translations for that phrase, which is ordered based on the likelihood of the translations. The searchable translation memories use translation models similar to those used in statistical machine translation. In this paper, we first describe the technical details of how the TMs are indexed and how translations are assigned probabilities, and then evaluate a searchable TM using precision and recall metrics.

Dynamical Visualisation of Nested Correspondences. *Christophe Chenon, GETA, France*

In this presentation, Christophe will introduce a new method of highlighting correspondences in segments of text that the translator will be able to use quickly – this new method is called TransTree and relies on XML.

Who's afraid of CAT - Redefining the boundaries of Translation. *Janet Carter-Sigglow, Forschungszentrum Jülich, Germany*

On the basis of examples drawn from our working environment in the largest interdisciplinary research establishment in Europe, this presentation intends to raise

the consciousness of translators concerning the exploitable potential of their linguistic and electronic skills. Illustrations will be given of how electronic tools extend translators' core competences and enable them to offer new enhanced services. By the same token, translators are encouraged to take the initiative and make software developers and dictionary publishers aware of their needs so that they cease to be passive consumers of translation technology.

Problems with CAT tools in translations into Central and Eastern European Languages. *Andrew and Jurek Nedoma, Lido-Lang Technical Translations, Poland*

After a short presentation of the level of usage of CAT tools in new EU member countries and candidate countries, the core part of this paper focuses on problems related to the use of CAT tools for translations into Central and Eastern European Languages. Analysed problems are classified into several groups: alphabet-related, grammar rules-related, linguistic problems, technical problems. The presentation is based on Lido-Lang's 3-year experience in the use of CAT tools. Therefore, all problems discussed present real situations, real questions or real answers to requirements that may rise and do rise during processing of translation projects in the languages mentioned above.

Localising Nations, saving languages: moving from Unicode to Language Engineering. *Pat Hall, Open University, UK*

It is believed that all peoples would benefit from the use of computers and access to the Internet, a belief that was reflected in the World Summit on the Information Society in Geneva at the end of 2003. But computers, and now the Internet, are dominated by western nations working in English and other major languages of the developed world. If computers and the internet are to be widely used, clearly it should take place using the language of that community, just as all other activities do. Using Unicode is a first step, and it has been proposed the Unicode should be funded to provide encodings for all known writing systems. In this paper I will spell out what is required to move beyond Unicode to the language technologies that underpin the use of Unicode and software localisation, looking at developments in South Asia. The countries of the region are moving towards using computers and the Internet in their own local languages.

The Certified Localisation Professional. *Reinhard Schäler, Localisation Research Centre, University of Limerick, Ireland*

The Institute of Localisation Professionals (TILP) was established in 2002 as a non-profit organisation and merged in 2003 with the US-based Professional Association for Localization (PAL). TILP's objective is to develop professional practices in localisation globally. TILP is owned by its individual members. It coordinates a number of regional chapters in Europe, North America, Latin America and Asia. This presentation will outline the CLP programme and is aimed at course providers interested in offering TILP accredited courses, employers planning to make CLP certification a requirement for future employees, and individual professionals planning to develop their professional career.

Aslib is pleased that TILP has this year become one of the supporting organisations of this Conference. For more information on TILP, please visit www.localisation.ie