Translating and the Computer 36

AsLing

27-28 November, 2014
One Birdcage Walk, London

Abstracts

Filling in the gaps: what we need from subsegment TM recall

Kevin Flanagan
Swansea University

Alongside increasing use of Machine Translation (MT) in translator workflows, Translation Memory (TM) continues to be a valuable tool providing complementary functionality, and is a technology that has evolved in recent years, in particular with developments around subsegment recall that attempt to leverage more content from TM data than segment-level fuzzy matching. But how fit-for-purpose is subsegment recall functionality, and how do current Computer-Assisted Translation (CAT) tool implementations differ? This paper presents results from the first survey of translators to gauge their expectations of subsegment recall functionality, cross-referenced with a novel typology for describing subsegment recall implementations. Next, performance statistics are given from an extensive series of tests of four leading CAT tools whose implementations approach those expectations. Finally, a novel implementation of subsegment recall, 'Lift', is presented (integrated into SDL Trados Studio 2014), based on subsegment alignment and with no minimum TM size requirement or need for an 'extraction' step, recalling fragments and identifying their translations within the segment even with only a single TM occurrence and without losing the context of the match. A technical description explains why it produces better performance statistics for the same series of tests and in turn meets translator expectations more closely.

Almost fifty years after the (first?) ALPAC report

Gábor Prószéky
MorphoLogic & Pázmány University,
Budapest, Hungary

The presentation will provide a historical flashback of Machine Translation (MT) by reviewing the significant milestones in its development and will reflect as to what the future has in store. To start with, the early developments after the Second World War II will be outlined. Next, the presentation will elaborate that the addition of morphological, syntactic and semantic knowledge did not lead to expected improvements which in turn, triggered the ALPAC report in 1966.

In the 1990s that landscape significantly changed due to the emergence of large amount of language data (corpora) which offered new opportunities for the rise and deployment of Statistical Machine Translation (SMT). SMT has been recently enhanced by the incorporation of morphological, syntactic and semantic information but the results are still not as good as expected. The presentation will review these recent developments and will reflect as to what are the options for the EU decision makers given that high quality MT is still a desideratum...

Beyond prescription: What empirical studies are telling us about localization crowdsourcing

Miguel A. Jiménez Crespo
Rutgers University

Translation crowdsourcing represents a new and quickly evolving phenomenon that has attracted the attention of industry experts and scholars alike. During recent years the industry has released a number of publications, mainly case studies and best-practice reports, while academic disciplines such as Computational Linguistics and Translation Studies (TS) have primarily focused on empirical studies. This paper attempts to compare and critically analyze research produced from both perspectives and locate these different approaches within the wider cycle of applied and theoretical/descriptive research. The findings of empirical studies on volunteer motivation and quality in TS will be contrasted with the best practices in the industry. This analysis will show a potential avenue to engage both perspectives to collaborate towards closing the existing research gap.

Using Cross-Language Information Retrieval and Statistical Language Modelling in Example-Based Machine Translation

Nasredine Semmar
CEA, LIST
Vision and Content Engineering Laboratory
Gif-sur-Yvette, France

Othman Zennaki
CEA, LIST
Vision and Content Engineering Laboratory
Gif-sur-Yvette, France

Meriama Laib
CEA, LIST
Vision and Content Engineering Laboratory
Gif-sur-Yvette, France

In this paper, we present the CEA LIST Example-Based Machine Translation (EBMT) prototype which uses a hybrid approach combining cross-language information retrieval and statistical language modelling. This approach consists, on the one hand, in indexing a database of sentences in the target language and considering each sentence to translate as a query to that database, and on the other hand, in evaluating sentences returned by a cross-language search engine against a statistical language model of the target language in order to obtain the n-best list of translations. The English-French EBMT prototype has been compared to the state-of-the-art Statistical Machine Translation system Moses and experimental results show that the proposed approach performs best on specialized domains.

Representing intra- and interlingual terminological variation in a new type of translation resource: a prototype proposal

Koen Kerremans
Vrije Universiteit Brussel

In this study, terminological variation pertains to the different ways in which specialised knowledge is expressed in written discourse by means of terminological designations. Choices regarding the use of term variants in source texts (i.e. intralingual variation) as well as the different translations of these variants in target texts (i.e. interlingual variation) are determined by a complex interplay of contextual factors of several kinds. For translators, it is therefore important to know the different language options (i.e. variants) that are available when translating terms and to know in which situational contexts certain options are more likely to be used.

To this end, translators often consult bi- or multilingual translation resources (e.g. terminological databases) to find solutions to certain translation problems. Different possibilities are offered in terminological databases to represent and visualise intra- and interlingual variants. In conventional terminology bases, terms in several languages usually appear on concept-oriented term records. This particular way of structuring and visualising terminological data has its roots in prescriptive terminology in which terms are merely viewed as 'labels' assigned to clearly delineated concepts (Picht and Draskau 1985). In ontologically-underpinned terminological knowledge bases or TKBs, terminological data tend to be represented in networks comprised of conceptual and semantic relations (Kerremans et al. 2008; Faber 2011; Durán Muñoz 2012; Peruzzo 2013). As opposed to traditional ways of representing terminological data (e.g. on the basis of alphabetically sorted lists, tables or matrices), such networks allow for a flexible and dynamic visualisation of data that may be connected to one another in several ways.

The aim of this article is to reflect on how visualisations of terms, variants and their translations in networks can be improved by taking into account the contextual constraints of the texts in which they appear. To this end, a novel type of translation resource has been developed, resulting from a semi-automatic method for identifying intralingual variants and their translations in texts.

A prototype visualisation of this resource will be presented in which terms, variants and their translations appear as a contextually-conditioned network of 'language options'. The proposed model derives from the Hallidayan premise that each language option or choice acquires its meaning against the background of other choices which could have been made. The choices are perceived as *functional*: i.e. they can be motivated against the backdrop of a complex set of contextual conditions (Eggins 2004). Changing these contextual conditions causes direct changes in the network of terminological options that are shown to the user.

iCompileCorpora: A Web-based Application to Semi-automatically Compile Multilingual Comparable Corpora

Hernani Costa
University of Malaga

Gloria Corpas Pastor
University of Malaga

Miriam Seghiri
University of Malaga

This article presents an ongoing project that aims to design and develop a robust and agile web-based application capable of semi-automatically compiling monolingual and multilingual comparable corpora, which we named iCompileCorpora. The dimensions that comprise iCompileCorpora can be represented in a layered model comprising a manual, a semi-automatic and a Cross-Language Information Retrieval (CLIR) layer. This design option will not only permit to increase the flexibility of the compilation process, but also to hierarchically extend the manual layer features to the semi-automatic web-based layer and then to the semi-automatic CLIR layer. The manual layer presents the option of compiling monolingual or multilingual corpora. It will allow the manual upload of documents from a local or remote directory onto the platform. The second layer will permit the exploitation of either monolingual or multilingual corpora mined from the Internet. As nowadays there is an increasing demand for systems that can somehow cross the language boundaries by retrieving information of various languages with just one query, the third layer aims to answer this demand by taking advantage of CLIR techniques to find relevant information written in a language different from the one semi-automatically retrieved by the methodology used in the previous layer.

Terminology Management Revisited

Nizar Ghoula
University of Geneva
and The Olanto Foundation

Jaques Guyot
The Olanto Foundation

Gilles Falquet
University of Geneva
and The Olanto
Foundation

Large repositories publishing and sharing terminological, ontological and linguistic resources are available to support the development and use of translation. However, despite the availability of language resources within online repositories, some natural languages associations cannot be found (rare languages or non-common combinations, etc.). Consequently, multiple tools for composing linguistic and terminological resources offer the possibility to create missing language associations. These generated resources need to be validated in order to be effectively used. Manually checking these resources is a tedious task and in some cases hardly possible due to the large amount of entities and associations to go through or due to the lack of expertise in both languages. To solve this matter and generate sound and safe content, tools are needed to

automatically validate and filter associations that make no sense. Hence, a validation tool is based itself on external resources such as parallel corpora which need to be either collected or created and filtered. To solve these matters we propose a set of tools that generate new terminological resources (**myTerm**) and a filter them using a parallel corpus generated by another tool (**myPREP**). We describe our methodology for terminology management and we describe its implementation within an original framework.

Rule-based Automatic Post-processing of SMT Output to Reduce Human Post-editing Effort

Victoria Porro
Université de Genève
FTI/TIM

Johanna Gerlach
Université de Genève
FTI/TIM

Pierrette Bouillon
Université de Genève
FTI/TIM

Violeta Seretan
Université de Genève
FTI/TIM

To enhance sharing of knowledge across the language barrier, the ACCEPT project focuses on improving machine translation of user-generated content by investigating pre- and post-editing strategies. Within this context, we have developed automatic monolingual post-editing rules for French, aimed at correcting frequent errors automatically. The rules were developed using the Acrolinx^{IQ} technology, which relies on shallow linguistic analysis. In this paper, we present an evaluation of these rules, considering their impact on the readability of MT output and their usefulness for subsequent manual post-editing. Results show that the readability of a high proportion of the data is indeed improved when automatic post-editing rules are applied. Their usefulness is confirmed by the fact that a large share of the edits brought about by the rules are in fact kept by human post-editors. Moreover, results reveal that edits which improve readability are not necessarily the same as those preserved by post-editors in the final output, hence the importance of considering both readability and post-editing effort in the evaluation of post-editing strategies.

Quality Assurance process in translation

Jerzy Czopik

Quality does not start, when the translation is finished. To deliver a high quality product a well designed process is necessary. In the best case it starts already before the translation is assigned to a translator.

The session will start with defining quality and looking at the measures to achieve it. After having done that, we have a good starting point to talk about checking quality.

Quality of a translation cannot be achieved by using tools like CAT or QA-tools. These tools can only provide some help, but cannot replace the human. Nevertheless good quality can be improved, if the tools are used properly. But only then – improper use will cause a lot of misunderstandings and problems.

We shall thus talk about quality checking, focused on the target language. Tools like SDL Trados Studio, MemoQ or Xbench allow you to configure the QA-checking modules, but in quite different ways. Here not only the knowledge of the tool, but also some understanding of the target language is necessary. Best case QA-checking should be done by people understanding both source and target language. Unfortunately very often this process is done by project managers, who typically cannot have command of as many languages as the languages of the projects they manage.

During the session I would like to show why understanding target is also necessary when doing QA-checking.

Intelligent Translation Memory Matching and Retrieval Metric Exploiting Linguistic Technology

Rohit Gupta	**Hanna Bechara**	**Constantin Orasan**
University of Wolverhampton	University of Wolverhampton	University of Wolverhampton

Translation Memories (TM) help translators in their task by retrieving previously translated sentences and editing fuzzy matches when no exact match is found by the system. Current TM systems use simple edit-distance or some variation of it, which largely relies on the surface form of the sentences and does not necessarily reflect the semantic similarity of segments as judged by humans. In this paper, we propose an intelligent metric to compute the fuzzy match score, which is inspired by similarity and entailment techniques developed in Natural Language Processing.

Improving fuzzy matching through syntactic knowledge

Tom Vanallemeersch
University of Leuven

Vincent Vandeghinste
University of Leuven

Fuzzy matching in translation memories (TM) is mostly string-based in current CAT tools. These tools look for TM sentences highly similar to an input sentence, using edit distance to detect the differences between sentences. Current CAT tools use limited or no linguistic knowledge in this procedure. In the recently started SCATE project, which aims at improving translators' efficiency, we apply syntactic fuzzy matching in order to detect abstract similarities and to increase the number of fuzzy matches. We parse TM sentences in order to create hierarchical structures identifying constituents and/or dependencies. We calculate TER (Translation Error Rate) between an existing human translation of an input sentence and the translation of its fuzzy match in TM. This allows us to assess the usefulness of syntactic matching with respect to string-based matching. First results hint at the potential of syntactic matching to lower TER rates for sentences with a low match score in a string-based setting.

Integrating Machine Translation (MT) in the Higher Education of Translators and Technical Writers

Marion Wittkowsky
Flensburg University of Applied Sciences

This paper describes how MT is integrated into a course project for translation and technical writing students. The course project is based on the idea of combining controlled language and a pre-editing step in order to achieve an effective way to prepare contemporary technical documentation for rule-based machine translation (RBMT). I will explain what I mean by "contemporary" within the context of technical documentation and why this attribute plays an important role within the decision-making process to integrate CL, pre-editing and MT in the course project, which also includes practical exercises for the students. In addition, the reason why RBMT is the MT method chosen within the context of multilingual text production is explained.

Top-down or bottom-up: what do industry approaches to translation quality mean for effective integration of standards and tools?

Joanna Drugan
University of East Anglia

The diverse approaches to translation quality in the industry can be grouped in two broad camps: top-down and bottom-up. The author has recently published a decade-long study of the language services (Quality in Professional Translation, Bloomsbury, 2013). Research for the study covered translation providers from individual freelance translators working at home, to large-scale institutions including the European Union Directorate-General for Translation, commercial translation companies and divisions, and not-for-profit translation groups.

Within the two broad 'top-down' and 'bottom-up' camps, a range of further sub-models was identified and catalogued (e.g. 'minimalist' or 'experience-dependent'). The shared distinctive features of each sub-group were described, with a particular focus on their use of technologies.

These different approaches have significant implications for, first, the integration of industry standards on quality, and, second, the efficient harnessing of technology throughout the translation workflow.

This contribution explains the range of industry approaches to translation quality then asks how these map on to successful integration of standards, and features of the leading tools which are designed to support or enhance quality.

Are standards and technologies inevitably experienced as an imposition by translators and others involved in the translation process? Significantly, no straightforward link was found between a 'top-down' or 'bottom-up' approach to assessing or improving translation quality and effective use of tools or standards. Instead, positive practice was identified across a range of approaches.

The discussion outlines some painless ways these developments are being channelled to improve quality, or more frequently, to maintain it while meeting tighter deadlines. Some models existed beyond, or were partially integrated in, 'professional' translation (e.g. pro bono translators, and volunteer Open Source localizers).

What lessons can we learn from enthusiasts in such communities, who sometimes adopt or create approaches voluntarily?

Getting the best out of a mixed bag

Terence Lewis
MITI

This paper discusses the development and implementation of an approach to the combination of Rule Based Machine Translation, Statistical Machine Translation and Translation Memory tecnologies. The machine translation system itself draws upon translation memories and both syntactically and statistically generated phrase tables, unresolved sentences being fed to a Rules Engine. The output of the process is a TMX file containing a varying mixture of TM-generated and MT-generated sentences. The author has designed this workflow using his own language engineering tools written in Java.

Terminology finding in the Sketch Engine: an Evaluation

Adam Kilgarriff
Lexical Computing Ltd.

The Sketch Engine is a leading corpus query tool, in use for lexicography at OUP, CUP, Collins and Le Robert, and at national language institutes of eight countries, and for teaching and research in many universities. Its distinctive feature is the 'word sketch' a one page, automatic, corpus, derived summary of a word's grammatical and collocational behaviour.

Very large corpora and word sketches are available for sixty languages.

A number of tools and resources have recently been added with translators and terminologists in mind. The resources are parallel corpora: EUROPARL-7 and the various datasets available in the OPUS collection. The tools are bilingual word sketches and the term finder. These have been reported on in previous Asling/Aslib conferences.

One remaining task is to make the Sketch Engine functions conveniently accessible to translators and terminologists. We have recently done this via IntelliWebSearch, a tool which lets the user highlight text in the environment they are working in, which could be a CAT tool or Microsoft Word, and, with a key sequence, query their preferred database. So now the key sequence can take the translator or terminologist to a browser window showing the word sketch, or parallel concordance, or any of a number of other reports, for the expression they are working on.

Machine Translation Quality Estimation Adapted to the Translation Workflow

Sabine Hunsicker
euroscript Deutschland GmbH

Alexandru Ceausu
euroscript Luxembourg S.à r.l.

The varying quality of machine translation (MT) poses a problem for language service providers (LSPs) which want to use MT to make the translation production process more efficient. In this user study we describe the MT confidence score we developed. It predicts the quality of a segment translated by MT and it is fully integrated into the translation workflow.

The Dos and Don'ts of XML document localization

Andrzej Zydroń
CTO XTM International

XML is now ubiquitous: from Microsoft Office to XHTML and Web Services it is at the core of electronic data communications. The separation of form and content, which is inherent within the concept of XML, makes XML documents easier to localize than those created with traditional proprietary text processing or composition systems.

Nevertheless, decisions made during the creation of the XML structure and authoring of documents can have a significant effect on the ease with which the source language text can be localized. For example, the inappropriate use of syntactical tools can have a profound effect on translatability and cost. It may even require complete re-authoring of documents in order to make them translatable.

This presentation highlights the potential pitfalls in XML document design regarding ease of translation and provides concrete guidance on how to avoid them.

AutoLearn<Word>

Kurt Eberle
Lingenio GmbH

AutoLearn<word> extracts new translation relations for words and multiword expressions (MWE) of any category from bilingual texts of any size in high quality and prepares the information found as a conventional dictionary entry - with morpho-syntactic and semantic classifications and contextual use conditions.

The function uses Lingenio's MT-system and analysis components as knowledge source, integrates its results into these and, by this bootstrapping approach, adapts dictionary and MT to the needs of the customer. Manual intervention is restricted to a very reduced number of difficult cases and can be carried out easily in an ergonomic graphical user interface, without need of effortful training. This is enabled by the underlying MT-architecture with rule-based core and additional statistical features.

The use conditions connected to the new dictionary entries are derived from the local representation the considered word or expression is part of in the considered reference(s). They restrict the corresponding translation to similar cases so that interferences with other translations in the dictionary are avoided.

A basic version of the function is already available in the current version of Lingenio's *translate*.

Quality Evaluation Today: the Dynamic Quality Framework

A. Görög
TAUS

Translation quality is one of the key topics in the translation industry today. In 2011, TAUS developed the Dynamic Quality Framework (DQF) in an attempt to standardize translation quality evaluation. In this paper, we will describe common approaches to translation quality and introduce the TAUS framework for QE. We will show that the development of this framework, initiated by the industry, was necessary to fill the gap between theory and practice. In We will give a short summary of the survey on quality evaluation and DQF that was conducted in the summer of 2014 among users of the DQF tools. Finally, we will suggest some ways academia and industry could and should collaborate with each other in the field of quality evaluation in the future.

Far from the Maddening Crowd: Integrating Collaborative Translation Technologies into Healthcare Services in the Developing World

Erin Lyons
University of Maryland

Crowdsourced and collaborative translation technologies have been at the centre of a heated debate in the translation industry in recent years, as questions have been raised regarding labour practices, the widespread integration of machine translation (MT) as well as concerns regarding quality and professional practices. However, despite the criticism of this emergent technology, the union of collaborative translation platforms and mobile communication technology has bridged a knowledge, resource and communication gap in the developing world, allowing healthcare and medical services to be re-imagined to reach a previously unimaginable community – often instantaneously. The rich data network supplied by mobile phones, when combined with automated data integration, can now be merged with translation services to contribute to initiatives, such as slowing the spread of malaria or stopping stock-outs of life-saving drugs at local clinics. We will take a closer look of the role of translation (machine translation, human translation and controlled language) in some of the leading crowdsourced translation applications, how translators bridge the gap between algorithm and on-the-ground communication and the implications for the development of "lite", mobile-ready versions of CAT tools and TMs.

Is Machine Translation Ready for Literature?

Antonio Toral
Dublin City University

Andy Way
Dublin City University

Given the current maturity of Machine Translation (MT), demonstrated by its growing adoption by industry (where it is mainly used to assist with the translation of technical documentation), we believe now is the time to assess the extent to which MT is useful to assist with translating literary text. Our empirical methodology relies on the fact that the applicability of MT to a given type of text can be assessed by analysing parallel corpora of that particular type and measuring (i) the degree of freedom of the translations (how literal are the translations) and (ii) the narrowness of the domain (how specific or general that text is). Hence, we tackle the problem of measuring the *translatability* of literary text by comparing the degree of freedom of translation and domain narrowness for such texts to texts in two other domains which have been widely studied in the area of MT: technical documentation and news. Moreover, we present a

pilot study on MT for literary text where we translate a novel between two Romance languages. The automatic evaluation results (66.2 BLEU points and 23.2 TER points) would be considered, in an industrial setting, as extremely useful for assisting human translation.

A Tool for Building Multilingual Voice Questionnaires

Alejandro Armando Université de Genève
Pierrette Bouillon Université de Genève
Manny Rayner Université de Genève
Nikos Tsourakis Université de Genève

We describe a prototype platform for creating multilingual voice questionnaires. Content is defined using a simple form-based language with units for questions, question-groups and answers; questionnaire definitions are compiled into efficient speech recognition packages and tables, and the resulting applications can be deployed over the web on both desktop and mobile platforms. We sketch our initial questionnaire application, which is designed for gathering information related to availability of anti-malaria measures in sub-Saharan Africa. It contains 114 question-groups and 218 questions.

Translating implicit elements in RBMT

Irina Burukina
ABBYY
RSUH, Computational Linguistics department

The present paper addresses MT of asymmetrical linguistic markers, in particular zero possessives. English <-> Russian MT was chosen as an example; however, obtained results can be applied to other language pairs (English – German / Spanish/Norwegian etc.).Overt pronouns are required to mark possessive relations in English. On the contrary, in Russian implicit possessives are regularly used, thus making it very important to analyze them properly, not only for MT but also for other NLP tasks such as NER, Fact extraction, etc. However, concerning modern NLP systems the task remains practically unsolved. The paper examines how modern English <-> Russian MT systems process implicit possessives and explores main problems that exist concerning the issue. As no SB approach can process IP constructions properly, linguistic rules need to be developed for their analysis and synthesis;the main properties of IPs are analyzed to that end. Finally, several rules to apply to RB or model-based MT are introduced that help to increase translation accuracy.

The present research is based on ABBYY Compreno © multilanguage NLP technologies that include MT module.

Losses and Gains in Computer-Assisted Translation: Some Remarks on Online Translation of English to Malay

Mozhgan Ghassemiazghandi — University Sains Malaysia

Tengku Sepora Tengku Mahadi — University Sains Malaysia

The article begins with a concise study of the significance of the translation technology in modern life as well as machine and computer-assisted translation. It then describes the technology accessible to translators and examines the losses and gains of the used tools in computer-assisted translation such as the electronic dictionaries, and specifically Google translate. The paper studies the influence of the online dictionaries on the professional translator with a view to determining the extent translation done using such dictionaries can be accurate. Loss in machine translation is inevitable due to the differences between English and Malay as they are entirely two different languages and not-related language pairs for translation. The online dictionary and translation software cannot replace the human translator and guarantee high-quality translations, despite their efficiency and outlooks. Online dictionaries and other translation means accelerate and facilitate the translation process only by minimizing the expected time for translation. Combination of electronic technologies with comprehensive knowledge of the translator and translation theory may result in a high-quality translation. Translation software and programs nonetheless, will not replace humans even in the future. As mentioned, the main aim of the paper is to investigate the new technologies in machine translation tools to discover the losses and gains in translation of English to Malay by using online dictionaries. Machine translations employing online dictionaries are compared with the translation done by a human translator to analyze the probable errors in machine-translated texts.

Affective Impact of the use of Technology on Employed Language Specialists: An Exploratory Qualitative Study

AnneMarie Taravella
Université de Sherbrooke

A well-established fact in the information systems literature is the importance of human aspects of technology use. In our doctoral research, we look into the emotional effort that employed language specialists have to put in their daily work, in the light of an increased use of language technology tools (LTT) by language service providers. In 2011 and 2012, we conducted qualitative studies to understand how LTT were perceived by language specialists. We observed

translators and other language specialists at work and conducted 12 in-depth interviews. We noticed that respondents often mentioned affective constructs, such as stress or anxiety, even when not prompted to describe their affective state. We then reanalyzed our transcripts and written notes in search for answers to the following specific question: "What affective variables do language specialists spontaneously mention when asked to describe their use of LTT?" Using content analysis, we found that respondents often mention some form of occupational stress, or relief of occupational stress, along with other affective variables, in relation with the use of LTT. We argue that emotional well-being and stress relief should be measured and serve as a guide for the design and implementation of language technology tools.

Solving Terminology Problems More Quickly with 'IntelliWebSearch (Almost) Unlimited'

Michael Farrell
IULM University, Milan

Michael Farrell received several descriptions of university courses to translate from Italian into English in early 2005. The syllabuses boiled down to a list of topics and laws of mathematics and physics: not many complex sentences, but a great deal of terminology which needed translating and double checking with the utmost care and attention.

To do this, he found himself repeatedly copying terms to his PC clipboard, opening his browser, opening the most appropriate on-line resources, pasting terms into search boxes, setting search parameters, clicking search buttons, analysing results, copying the best solutions back to the clipboard, returning to the translation environment and pasting the terms found into the text.

He quickly realized that he needed to find a way to semi-automate the terminology search process in order to complete the translation in a reasonable time and for his own sanity. He immediately started looking around for a tool, but surprisingly there seemed to be nothing similar to what he needed on the market. Having already created some simple macros with a free scripting language called AutoHotkey, he set about writing something that would do the trick.

The first simple macro he knocked out gradually grew and developed until it became a fully fledged software tool: IntelliWebSearch. After speaking to several colleagues about it, he was persuaded to share his work and put together a small group of volunteer beta- testers. After a few weeks of testing on various Windows systems, he released the tool as freeware towards the end of 2005.

At the beginning of his workshop, Michael Farrell will explain what prompted him to create the tool and how he went about it. He will then go on to describe its use and its limitations, and

show how it can save translators and terminologists a lot of time with a live demonstration, connectivity permitting.

The workshop will conclude with a presentation revealing for the first time in public some of the features of a new version which is currently being developed under the code name "IntelliWebSearch (Almost) Unlimited" (pre-alpha at the time of writing).

The workshop is aimed at professional translators, interpreters and terminologists in all fields, especially those interested in increasing efficiency through the use of technology without lowering quality standards.

Twitter Crowd Translation — Design and Objectives

Eduard Šubert

Czech Technical University
in Prague

Ondřej Bojar

Czech Technical University
in Prague

This paper describes our project to support translation of streaming texts on social networks, in particular Twitter. Since machine translation of this type of content is still almost unusable, we rely on volunteers to provide and score the translations. The translations will serve as a testbed and development data for our MT systems tuned for this domain. The project thus serves multiple purposes: From the users' point of view, we would like to provide a smooth access to timely information in foreign languages. From our translators' point of view, we want to provide them with interesting content and material to improve their language skills. Finally, we admit that our project is still primarily a research exercise: As MT researchers, we are interested in learning to handle the specific challenges that this type of content brings. We hope to acquire an interesting collection of data for MT development and to gradually improve our MT processing pipeline for this type of text.

SMT for restricted sublanguage in CAT tool context at the European Parliament

Najeh Hajlaoui
European Parliament
Luxembourg

This paper shows that it is possible to efficiently develop Statistical Machine Translation (SMT) systems that are useful for a specific type of sublanguage in real context of use even when excluding the exact Translation Memory (TM) matches from the test set in order to be integrated in CAT "Computer Aided Translation" tools. It means that the included part is quite different from the existing translations and consequently harder to translate even for an SMT system trained on the same translation data.

Because we believe on the proximity of sublanguages even though it is still hard to practically define the sublanguage notion, we are proposing on the framework of the MT@EP project at the Directorate General for Translation (DG TRAD) of the European Parliament (EP) to develop SMT systems specific for each EP Parliamentary Committee optimised for restricted sublanguages and constrained by the EP's particular translation requirements.

Sublanguage-specific systems provide better results than generic systems for EP domains showing a very significant quality improvement (5-25% of BLEU score), mainly due to the EP context specificity and to the proximity of sublanguages. This approach is also satisfactory for pairs of under-resourced languages, such as the Slavic families and German.

Task-based Teaching of Computer-aided Translation in a Progressive Manner

Jessica Xiangyu Liu
University of Hong Kong

The teaching of computer-aided translation is commonplace in academic institutions in recent years. More research has been done and more works have been published in this area. While much has been written on the theoretical and conceptual aspects of computer-aided translation and the contents of the course, little has been done on its practical aspect.

This workshop will present the classroom practice modules of the Introduction of Computer-aided Translation, an MA course at the Department of Translation of The Chinese University of

Hong Kong. The author will discuss how to teach the use of computer-aided translation systems in a progressive manner through demonstrations and classroom practice, and from basic functions to advanced operations.

This workshop will also present some pedagogical reflections on the teaching of computer-aided translation systems.

It is hoped that it will lead to a rethinking of the way of computer-aided translation systems should be taught.

What it intends to propose is to bring the learning of translation technology closer to the real world through systematic training, thus responding to the changing professional requirements that translators face in their workplace.

Chairs

João Esteves-Ferreira

João Esteves-Ferreira graduated in Arts, Business Administration and Terminology.
He qualified as a Sworn Translator in Switzerland (1977) and as a Conference Interpreter (1983).
He has held several posts in Swiss professional translation organisations, culminating with the Presidencies of ASTTI (Swiss Association of Translators, terminologists and Interpreters) and ASTJ (Swiss Association of Sworn-in Translators).
João served as Council Member of the Fédération internationale des Traducteurs (FIT) from 1996 to 2005 and as Chairman of FIT Europe 2005-2008. He was the Founder and first Chairman of FIT Translation Tools and Technology Committee (2000-2005).
In 2000, he founded tradulex, the International Association for Quality Translation, which he has chaired since its inception.
He is also President of ASLING, the International Association for Advancement in Language Technology and Co-Chair of the TC 36 Conference.
His current activities, besides translating and interpreting, are the coordination of tradulex and the training of professional colleagues.
João has published a great number of papers on legal translation, translation technology and professional issues.

Juliet Macan

Juliet Macan was born in Malta, educated in England, where she studied Sociology and Psychology at Leeds University.
After working for fifteen years as a freelance translator in Italy, specialising in medicine, pharmacology and plant pathology, in 1991 she joined Intracoop, as a senior language consultant.
In 1994 she had her first encounter with a CAT tools: IBM Translation Manager, Trados and DejaVu.
In 1999 she became Translation Tools Manager of the new company Ic.doc, with a strong emphasis on technology. She was responsible for ensuring optimum use of translation technology by the company, training in-house staff and external freelancers, advising customers, evaluating new projects, problem-solving and

testing of new tools such as SLDX, Multitrans, across, Memo-Q, SDL Studio, XTM and ONTRAM.
She also investigated the Quality assurance functions of these tools in comparison with the stand-alone QA tools such as Error Spy, QA Distiller and XBench, overseeing the introduction of Quality Assurance procedures within the company. Engaged as a consultant in the new company Arancho Doc, setup at the beginning of 2011.
She has lectured at Bologna and Palermo universities, given presentations at numerous conferences throughout Europe and held workshops on translation tools and QA procedures and technology for language technology specialists, project managers and translators. An expert in translation technology and the problems related to new authoring methods, she provides consultancy services and training for international companies.
She is Vice president of AsLing (Association internationale pour la promotion des technologies linguistiques), a not for profit association set up in Geneva in 2014 to promote the development, knowledge and use of translation technology in the academic sector, large international organisations and industry.
Currently, she is lead chair of the 36th Translating and the Computer Conference in London.

Ruslan Mitkov

Prof. Dr. Ruslan Mitkov has been working in Natural Language Processing (NLP), Computational Linguistics, Corpus Linguistics, Machine Translation, Translation Technology and related areas since the early 1980s.
His research output was highlighted as being internationally leading in the last UK Research Assessment Exercise (RAE 2008). Whereas Prof. Mitkov is best known for his seminal contributions to the areas of anaphora resolution and automatic generation of multiple-choice tests, his extensively cited research (more than 210 publications including 9 books, 25 journal articles and 25 book chapters) also covers topics such as machine translation, natural language generation, automatic summarisation, computer-aided language processing, centering, translation memory, evaluation, corpus annotation, bilingual term extraction, automatic identification of cognates and false friends, NLP-driven corpus-based study of translation universals and text simplification.
Prof. Mitkov is author of the monograph Anaphora resolution (Longman) and sole Editor of The Oxford Handbook of Computational Linguistics (Oxford University Press) which has been hailed as the most successful Oxford Handbook. Current prestigious projects include his role as Executive Editor of the

Journal of Natural Language Engineering (Cambridge University Press), Editor-in-Chief of the Natural Language Processing book series of John Benjamins publishers, and Consulting Editor of Oxford University Press publications in Computational Linguistics. He is also working on the forthcoming Oxford Dictionary of Computational Linguistics (co-authored with Patrick Hanks) and the forthcoming second, substantially revised edition of the Oxford Handbook of Computational Linguistics.

Prof. Mitkov has been invited as a keynote speaker at a number of international conferences.

He has acted as Programme Chair of various international conferences on Natural Language Processing (NLP), Machine Translation, Translation Technology, Translation Studies, Corpus Linguistics and Anaphora Resolution.

He is asked on a regular basis to review for leading international funding bodies and organisations and to act as a referee for applications for Professorships both in North America and Europe. Ruslan Mitkov is regularly asked to review for leading journals, publishers and conferences and serve as a member of Programme Committees or Editorial Boards.

Prof. Mitkov has been an external examiner of many doctoral theses and curricula in the UK and abroad, including Master's programmes related to NLP, Translation and Translation Technology.

Dr. Mitkov has considerable external funding to his credit (more than £ 15,000,000) and is currently acting as Principal Investigator of several large projects, some of which are funded by UK research councils, by the EC as well as by companies and users from the UK and USA.

Ruslan Mitkov received his MSc from the Humboldt University in Berlin, his PhD from the Technical University in Dresden and worked as a Research Professor at the Institute of Mathematics, Bulgarian Academy of Sciences, Sofia.

Dr. Mitkov is Professor of Computational Linguistics and Language Engineering at the School of Humanities, Languages and Social Sciences at the University of Wolverhampton which he joined in 1995 and where he set up the Research Group in Computational Linguistics. His Research Group has emerged as an internationally leading unit in applied Natural Language Processing (the output of the Research Group was classed by RAE 2008 as internationally leading, internationally excellent and internationally recognised) and members of the group have won awards in different NLP/shared-task competitions. In addition to being Head of the Research Group in Computational Linguistics, Prof. Mitkov is also Director of the Research Institute in Information and Language Processing. The Research Institute consists of the Research Group in Computational Linguistics and the Research Group in Statistical Cybermetrics, which is another top performer in the recent RAE.

Ruslan Mitkov is Vice President of ASLING, an international Association for promoting Language Technology.

In recognition of his outstanding professional/research

achievements, Prof. Mitkov was awarded the title of Doctor Honoris Causa at Plovdiv University in November 2011. In November 2014 Dr. Mitkov will be conferred the Profesor Honoris Causa at Veliko Tarnovo University.

Olaf-Michael Stefanov

Olaf-Michael Stefanov is an IT professional with a strong focus on multilingualism. During 36 years on staff at the United Nations he managed various information-technology related areas, the last being Library and Linguistic Support for Vienna headquarters, which included reference and terminology support for the editorial, translation and interpretation sections.

Having introduced the first completely web-based multilingual terminology database handling Arabic, Chinese, Cyrillic and Latin scripts for input, query, and output, VINTARS, he presented it at Translating and the Computer - 20, in 1998. He also introduced digital dictation and voice-recognition into the translation workflow of several international and multinational organizations.

Although retired from the UN he continues to serve in the site administration and management of JIAMCATT, an information exchange among governmental and intergovernmental language professionals, serves as co-moderator of the JIAMCATT Working Group on Standardization and Interoperability and has implemented multilingual Web 2.0 and CMS tools for JIAMCATT.

He is also active in Tiki, a leading open source CMS, wiki and Groupware tool and was active in drafting the ITS 2.0 (Internationalization Tag Set) standard under the aegis of the World Wide Web Consortium (W3C) in 2013.

He has been a member of the Programme Committee of FEISGILTT since 2012.

He is co-founder, Vice President and coordinateur of ASLING, the International Association for Language Technology which took over the Translating and the Computer conference series from ASLIB in 2014.

Having served as co-chair of conferences in this series from 2000 and as lead chair in 2013, he is again a chair in 2014.

Based in Vienna, Olaf-Michael is actively engaged worldwide in a variety of multilingual projects and conferences

David Chambers

David Chambers recently retired, worked for over 20 years with the World Intellectual Property Organization (WIPO), a specialised agency of the United Nations in Geneva, initially as Head of the Patent Translation Section and subsequently as Head of the Language Division responsible for the Organization's translation and interpretation activities.

This followed an earlier career as translator, interpreter, and head of translation sections in the private sector. He has always shown keen interest in the development of computer assisted translation tools, and developed their practical application for the specific type of translation work at WIPO.

In 2000, he presented a paper at the 22nd annual Translating and the Computer Conference (TC22) on *Automatic Bilingual Terminology Extraction – a Practical Approach*, outlining work done at WIPO in this context in the field of patent translation.

Since retirement, he continues to take an active interest in computer systems and translation technologies.

A regular participant at the Conference since the mid 1990's, he has been a member of the Conference Programme Committee since 2011, acted as session Chair in 2011 and 2012 and will again chair at the 2014 Conference.

He holds a BA (Hons) in Applied Language Studies and is a Member of the Chartered Institute of Linguists.

Speakers

Alejandro Armando

Alejandro Armando is a certified translator graduated from the School of Languages, Universidad Nacional de Córdoba, Argentina.
He works as a freelance translator for a number of international organizations and private sector clients.
He is currently finishing a Master of Arts in Translation at the University of Geneva and conducting research on automatic translation and speech recognition.

Hanna Bechara

Hanna Bechara previously studied at Eberhard Karls University of Tubingen, where she earned her Bachelor's degree in Computational Linguistics and Computer Science in 2010.
After completing an internship at Dublin City University, she started research on statistical post--editing for machine translation and was awarded her Masters of Science (by Research) from Dublin City University in 2013.
She joined the EXPERT project as ESR 12 on January, 6th 2014, at the University of Wolverhampton.
Her research centres around evaluation methods of different machine translation systems, specifically the hybrid systems proposed by the EXPERT projects.
This work is under the supervision of Dr. Constantin Orasan and Prof. Ruslan Mitkov.

Ondřej Bojar

Ondřej Bojar graduated in computer science in 2003 and received his Ph.D. in computational linguistics in 2008 at the Faculty of Mathematics and Physics, Charles University in Prague.
He now works as an assistant professor at the faculty. His main research interest is machine translation.
He participated in the Johns Hopkins University Summer Engineering Workshop in 2006 as a member of the Moses team. Since then, he is regularly taking part in WMT shared translation tasks mainly with systems based on Moses and adapted for English-to-Czech translation.
He was the main local organizer of MT Marathons 2009 and 2013 held in Prague.

Irina Burukina

Irina Burukina completed her Diploma with Honours (BA equivalent) in June 2012 at the St. Petersburg State University, Department of Linguistics where she majored in Mathematical Linguistics; her final qualifying paper: "Integration of multiword expressions in the RussNet thesaurus structure".
Then, in June 2014, she graduated from the Russian State University for the Humanities, Institute for Linguistics; with a Diploma with Honours (MA equivalent). She majored in Computational Linguistics; her final qualifying paper: "Syntax of implicit possessives in Russian. Implicit possessives recognition and translation".
Irina gained other research experience working 2010 and 2012 as a Research assistant in the RussNet project: automatic extraction of collocations in Russian (semantic and statistical approaches), and 2013 as a Research assistant in the General Internet Corpus of Russian project.
Her working experience:
June 2012 – present. Linguist, Lexical Semantics Group, Technology Development Department; ABBYY Headquarters (Linguistic Software Company), Moscow, Russia.
June 2011 – May 2012. Linguist, Tree Syntax Group, Technology Development Department. ABBYY Headquarters (Linguistic Software Company), Moscow, Russia.
Russian is Irina's native language. She has proficiency in English, advanced reading skills in German and basic communication skills in Spanish.
Her Computer skills: proficient user; Java, Python, R.

Hernani Costa

Hernani Costa is currently a Marie Curie Early Stage Researcher in the Department of Translation and Interpreting at the Faculty of Philosophy and Humanities, University of Malaga, Spain. His main research interest lies in the Computational Linguistics and Artificial Intelligence areas, especially its practical application in the fields of Translation Technologies, Natural Language Processing, Information Extraction and Information Retrieval. He is also interested in (or has worked on) a number of other topics such as Recommender Systems, Multiagent Systems, Affective Computing, amongst others.

Hernani completed his BSc and MSc on Informatics Engineering in the Bologna model at the Department of Informatics Engineering of the University of Coimbra (UC) in 2010.

It was during his Master degree that he started his research activities. In particular, he developed a system capable of acquiring semantic knowledge from any kind of Portuguese text. Besides that, he also analysed the benefits of applying similarity distributional metrics, based on the occurrence of words in documents, on the system outputs. In the same academic year, he applied for a research grant to work in the "Automatic Construction of Ontologies in Portuguese" project, where he explored popular distributional similarity measures with the purpose of quantify relational triples in an automatic fashion.Furthermore, in September 2011, he was invited by LAP LAMBERT Academic Publishing to publish his MSc thesis. The book was published on October 2011 named "Automatic Extraction and Validation of Lexical Ontologies from Text: Creating Lexical Ontologies from text". In a total of six years, he studied at the aforementioned institution, developing skills on the field of Computer Science, except in the academic year of 2007/2008 where he integrated the Erasmus program for a year at the University of Vigo, Spain. During this period, besides starting to acquire skills in the Natural Language Processing area, he developed his skills in interpersonal relations (meeting Erasmus students from other cultures and languages), teamwork, research, organisation and autonomy which enabled him to develop writing and speaking skills in Spanish, as well as in English. In October 2010, he applied for a scholarship and, between December 2010 and August 2013, he worked on the project "Forms of Selective Attention in Intelligent Transportation Systems", at the Cognitive and Media Systems (CMS) group, at the Department of Informatics Engineering of the University of Coimbra. In this project, a two-parted agent architecture was implemented, with an agent responsible for gathering Points of Interest (POIs) from a location-based service, and a set of Personal Assistant Agents (PAAs) collecting information about the context and the intentions of its respective user. In each PAA were embedded a set of Machine

Learning algorithms, with the purpose of ascertaining how wellsuited these classifiers are for filtering irrelevant POIs, in a completely automatic fashion. During the Autumn 2011, he also developed an online service for browsing Portuguese semantic relations for the Linguateca project. It is also important to mention that he has three years of experience on teaching area. In the first two years he taught at Lousã Professional School (2010/2011 and 2011/2012) and in the academic year 2012/2013 he taught at Coimbra Institute of Engineering (ISEC). As always, he is highly motivated to find new challenges that defy his competences and skills in computer science field. That is why he enrolled the doctoral program in September 2013 at the Department of Translation and Interpreting, at the Faculty of Philosophy and Humanities of the University of Malaga, Spain.

Jerzy Czopik

Jerzy Czopik was born and grew up in Cracow, where he started to study mechanical engineering.
In 1986 he moved with his wife to Dortmund/Germany. Here he finished the mechanical engineering study and started a translators and interpreters career in 1990.
Jerzy is approved trainer for SDL Trados Studio and MultiTerm.
Together with his wife he is certified by LICS according to EN 15038. He also acts as LICS auditor for this standard.
In 2011 he published a manual on SDL Trados Studio in Polish. He is also very active on various lists and forums, helping with SDL Studio problems.

Joanna Drugan

Dr Joanna Drugan is Senior Lecturer in Applied Translation Studies at the University of East Anglia, UK.
Her main research interests include translation quality, translation ethics and translation technologies.
Her most recent book is *Quality in Professional Translation* (Bloomsbury, 2013).
She is currently researching real-world ethical challenges when professional translators and interpreters are not available, particularly in healthcare and social work, and ways in which training and technology might support professionals and service users faced with such challenges.

Jo holds an MA (Hons) and PhD in French from the University of Glasgow, Scotland.
She previously worked at Reading University and Leeds University, where she was a founder member of the Centre for Translation Studies and ran the MA Applied Translation Studies for over a decade.
She was awarded a National Teaching Fellowship and became a member of the Higher Education Academy in 2008.
She has served as a member of the Peer Review Council for the Arts and Humanities Research Council since 2012 and was selected as a founding member of the Publication Integrity and Ethics Council in 2013.
Since joining UEA in 2012, Jo has led specialist Masters modules in translation technologies, translation as a profession, and research methods, and an undergraduate module on translation and globalisation.
She is Director of Graduate Studies for the School.

Kurt Eberle

Kurt Eberle received his dissertation and habilitation in linguistics and computational linguistics from the university of Stuttgart in 1991 and 2004.
He holds master degrees in Romance Languages and Mathematics received from the universities of Freiburg and Heidelberg in 1983 and 1987.
From 1987 until 1997 he was involved in various NLP projects at the university of Stuttgart and at IBM research in Heidelberg. In 1997 he joined the MT group there where he was responsible for the development of German-French.
In 1999 he was one of the co-founders of Lingenio GmbH (named Linguatec Entwicklung & Services at that time).
Since 2007 Kurt Eberle is associate professor at the University of Heidelberg and since 2009 general manager of the company.
He has published approx. 50 articles and monographs in the fields of MT, syntax and semantics and designed and managed a number of innovative products in the fields of MT and dictionaries at IBM and at linguatec Development&Services/Lingenio.

Michael Farrell

Michael Farrell is primarily a freelance technical translator, but is also an untenured lecturer in computer tools for translation and interpreting at the IULM University (Milan, Italy).
He is an Atril Certified Training Partner and the author of "A Tinkerer's Guide to Structured Query Language in Déjà Vu X".
He is also the developer of the freeware terminology search tool IntelliWebSearch and a qualified member of the Italian Association of Translators and Interpreters (AITI).

Kevin Flanagan

Kevin Flanagan is pursuing a full-time PhD at Swansea University.
He spent many years as a software developer, using language skills with French clients, prior to starting work as a freelance technical translator.
Having used a number of translation memory (TM) systems in that capacity, he developed a prototype TM system providing more effective sub-sentential recall.
His PhD research focusses on extending and refining the system, formalising the theoretical principles and delivering a production-ready implementation.

Johanna Gerlach

Johanna Gerlach started working as a Research and Teaching Assistant at the Translation Technologies Department of the University of Geneva in 2008.
She contributed to the MedSLT and CALL-SLT projects, developing linguistic resources for German.
Currently, she is involved in the ACCEPT European project, investigating pre- and post-editing technologies for user generated content.
In 2012, she began working on her PhD thesis, which focuses on the development and evaluation of pre-editing rules for French forum content.

Nizar Ghoula

Nizar Ghoula is a PhD candidate in Information Systems at the University of Geneva.
His fields of interest include the semantic web, knowledge representation and ontology alignment.
He joined the OLANTO team in order to collaborate on the development and enhancement of many CAT tools.
He is also a project manager within the Executive department of the University of Geneva to design a solution for executive programs management.

Attila Görög

Attila Görög has been involved in various national and international projects on language technology in the past 10 years.
He has a solid background in Quality Evaluation, Post-Editing and Terminology Management.
Attila is interested in globalization issues and projects involving CAT tools.
His webinars and workshops discuss hot topics in the translation industry with aim of making participants future proof.
As a product manager at TAUS, he is responsible for the TAUS Evaluation platform also referred to as the Dynamic Quality Framework or DQF.

Jacques Guyot

Jacques Guyot is a senior computer scientist with over 20 years of experience in turning break-through technologies into professional solutions.
He is also a researcher at the University of Geneva.
He holds a PhD in Computer Sciences from the University of Geneva.

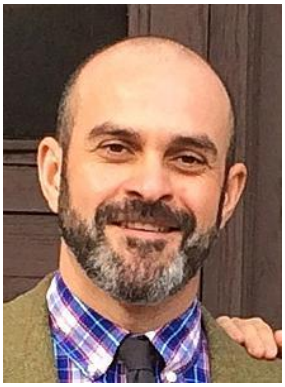
Najeh Hajlaoui

Najeh Hajlaoui received his PhD in computer science from Joseph Fourier University (Grenoble, France) in 2008 on Multilingualization of ecommerce systems handling spontaneous utterances in natural language.
In 2002 he received his MS in information systems at Joseph Fourier University, and his Joint European Diploma MATIS (Management and Technology of Information Systems).
He is currently Senior Researcher and Project Manager for Machine Translation at the European Parliament in Luxembourg (since August 2013).
Before joining the Idiap Research Institute in December 2011, he has been a Research Fellow at the University of Wolverhampton (UK) in 2011, a Postdoctoral Researcher at Orange Labs (Lannion, France) in 2010, and an Associate Lecturer at Jean Monnet University (Saint- Étienne, France) from 2007 to 2009.

Sabine Hunsicker

Sabine Hunsicker is a computational linguist with experience in statistical natural language processing with a focus on machine translation.
Sabine completed her M.Sc. in computational linguistics at Saarland University in 2009. Her thesis concerned example-based machine translation enhanced with statistical methods.
Before euroscript, she worked at the German Research Center for Artificial Intelligence (DFKI) in Saarbrücken. Her research topics were hybrid machine translation and she was involved with the Euromatrix Plus and ACCURAT research projects with a strong focus on integrating linguistic analysis into the SMT workflow.
Her areas of expertise include linguistic analysis, data mining as well as information retrieval.

Miguel A. Jiménez Crespo



Miguel A. Jiménez-Crespo is an Assistant Professor in the Department of Spanish and Portuguese at Rutgers University, where he directs the MA program in Spanish Translation and Interpreting.
He holds a PhD in Translation and Interpreting Studies from the University of Granada, Spain.
His research focuses on the intersection of translation theory, translation technology, digital technologies, corpus-based translation studies and translation training.
He is the author of *Translation and Web Localization* published by Routledge in 2013, and has published extensively in the top international journals in the discipline of Translation Studies.

Koen Kerremans

Koen Kerremans obtained his Master's degree in Germanic Philology (Dutch-English) at Universiteit Antwerpen in 2001, his Master's degree in Language Sciences - with a major in computational linguistics - at Universiteit Gent in 2002 and his PhD degree in Applied Linguistics at Vrije Universiteit Brussel in 2014 (Title of his dissertation: 'Terminological variation in multilingual Europe. The case of English environmental terminology translated into Dutch and French').
His research interests pertain to applied linguistics, language technologies, ontologies, terminology (variation) and translation studies.
He currently holds a position as post-doctoral researcher and teaching assistant at the department of Applied Linguistics (Faculty of Arts and Philosophy) of Vrije Universiteit Brussel (VUB) where he teaches applied linguistics, terminology and several Dutch language courses.
He is a member of VUB's research group 'Centrum voor Vaktaal en Communicatie' (Centre for Special Language Studies and Communication).

Adam Kilgarriff

Adam Kilgarriff is Director of Lexical Computing Ltd.
He has led the development of the Sketch Engine, a leading tool for corpus research used for dictionary-making at Oxford University Press, Cambridge University Press, HarperCollins, Le Robert and elsewhere.
His scientific interests lie at the intersection of computational linguistics, corpus linguistics, and dictionary-making.
Following a PhD on "Polysemy" from Sussex University, he worked at Longman Dictionaries, Oxford University Press, and the University of Brighton prior to starting the company in 2003.
He is a Visiting Research Fellow at the University of Leeds.
He has been an Expert Witness in a number of legal cases involving trademarks.
He is active in moves to make the web available as a linguists' corpus and was the founding chair of ACL-SIGWAC (Association for Computational Linguistics Special Interest Group on Web as Corpus).
He has been chair of the ACL-SIG on the lexicon and Board member of EURALEX (European Association for Lexicography).

Terence Lewis

Terence Lewis, MITI, entered the world of translation as a young brother in an Italian religious order, when he was entrusted with the task of translating some of the founder's speeches into English. His religious studies also called for a knowledge of Latin, Greek and Hebrew.
After some years in South Africa and Brazil, he severed his ties with the Catholic Church and returned to the UK where he worked as a translator, lexicographer (Harrap's English-Brazilian Portuguese Business Dictionary) and playwright.
As an external translator for Unesco he translated texts ranging from Mongolian cultural legislation to a book by a minor French existentialist.
At the age of 50 he taught himself to program and wrote a Dutch-English machine translation application which has been used to translate documentation for some of the largest engineering projects in Dutch history.
For the past 15 years he has devoted himself to the study and development of translation technology.

Jessica Xiangyu Liu

Jessica Xiangyu Liu is a research postgraduate at Department of Translation, The Chinese University of Hong Kong, Hong Kong.
She is in her first year of the MPhil programme in Translation.
She has a strong interest in the teaching computer-assisted translation.systems, and the hybrid of machine translation and computer-assisted translation.
Before conducting her study, she worked as a research assistant at Centre for Translation Technology, CUHK (2010-2013) where she was engaged in the research and training of computer-assisted translation software.

Erin Lyons

Erin M. Lyons is a full-time French to English and Italian to English translator, medical writer and consultant and the Owner and President of BiomedNouvelle.
She is also an Adjunct Professor of Translation in the Master's Degree programme at the University of Maryland (USA).
Her primary areas of focus include clinical research, pharmaceuticals, medical devices, and cosmetic products.
Furthermore, she continues to work on the development of BabelNouvelle®, a mobile-based translation technology employing crowdsourcing and machine translation to facilitate medical services in the developing world.
Ms Lyons has split her professional career between Europe and the U.S.
She has a BA in Romance Languages and Literature from the University of Chicago and an MA in Italian and French Translation from the Monterey Institute of International Studies.

Ms Lyons frequently presents her work to private groups, universities and at international conferences, including at ATA Annual Conferences, the World Congresses of the International Federation of Translators and the International and France ProZ.com Conferences.

Tengku Sepora Tengku Mahadi

Tengku Sepora Tengku Mahadi (Associate Professor Dr.) is Dean of the school of Languages, Literacies and Translation at the University Sains, Malaysia.
She lectures in translation theories and practice and supervises research in Translation Studies at MA and PhD levels.
She is the author of *Text-wise in Translation* (2006), and co-author of *Corpora in Translation: A Practical Guide* (2010).

Victoria Porro

Victoria Porro holds a bachelor and a master in Translation Studies and Translation Technologies.
She joined the Translation Technologies Department of the University of Geneva as a Research and Teaching Assistant in June 2012.
Since then, she devotes most of her time to the EU-funded ACCEPT project she participates in and she is currently designing a PhD project in post-editing and machine translation.
She is most interested in opening new lines of research in post-editing and advocates for the recognition of post-editing as a highly skilled activity.

Gábor Prószéky

Gábor Prószéky is a professor and vice dean of the Faculty of Information Technology and Bionics at the Pázmány Péter Catholic University (Budapest) and CEO of MorphoLogic, a leading Hungarian language technology company.
He is in charge of the MTA-PPKE Hungarian Language Technology Group co-financed by the Hungarian Academy of Sciences and the Pázmány University.
He graduated at the ELTE University both in software engineering and in general & applied linguistics. He holds a PhD (1994) in computational linguistics. In 2005 he received the title of Doctor of the Hungarian Academy of Sciences.

Since his university years, he has been involved in more than thirty R&D projects in human language technologies (HLT), and computational and theoretical aspects of humanities.
His research interest covers various aspects of computational analysis of highly inflectional languages, intelligent dictionaries and machine translation.
Aside of more than 140 scientific publications mainly on HLT, he is the author of three comprehensive books on human language technologies.
Among others, he was a Board Member of the European Language Resources Association and the President of the Lexicographical Committee of the Hungarian Academy of Sciences (2006-2012). Since 2013 he has been the president of the Association of Hungarian Applied Linguists, and in 2014 he became the president of the Council of Social Sciences and Humanities of the Hungarian Scientific Research Fund.
In 1991, with software engineer colleagues working on human language technology applications, he founded MorphoLogic, the first language industry company in Hungary. Since then, MorphoLogic's various applications have been licensed by Microsoft, IBM, Xerox, among others. In 1999, MorphoLogic won the ***IST Prize*** of the European Commission.
Gábor Prószéky received Hungary's highest award, the ***Széchenyi Prize***, for his activities in 2000. Among others, he also received the ***Kalmár Award*** of the John von Neumann Computer Society (1995), ***IT Manager of the Year*** (2002), ***Award for the Hungarian IT*** (2005),***Special Prize to the IT Lecturer of the Year*** (2009) and ***Dennis Gabor Award*** (2010).

Emmanuel (Manny) Rayner

Emmanuel (Manny) Rayner is a Collaborateur Scientifique (senior non-teaching research post) at the University of Geneva's Multilingual Processing Group, and has previously held positions at SRI International (1991-1999) and NASA Ames Research Center (1999-2005).
Over the last 20 years, his research has focused on construction of speech-enabled dialogue systems.
He has played key roles in several major projects in this area, including SRI's Spoken Language Translator, NASA's Clarissa and the Open Source Regulus platform.
At the University of Geneva, he has worked primarily with speech translation and CALL.
He has more than 100 refereed publications in speech technology, computational linguistics, machine translation and artificial intelligence.

Nasradine Semmar

Nasredine Semmar obtained his Ph.D. in 1995 at University of Paris-Sud (France) in 1995 on Multimedia software localization.
He worked from 1996 to 2000 as an R&D engineer in Lionbridge Technologies – Bowne Global Solutions, he designed and implemented tools for Computer Aided Translation and he participated in delivering the multilingual version of MS Windows 2000.
He then joined SAP - Business Objects where he worked from 2000 to 2002 as an expert in software internationalization and localization.
Since 2002, he has been working as a research scientist at the Vision and Content Engineering Laboratory (LVIC) where he has implemented the treatment of Arabic in the CEA LIST NLP platform for a cross-language search engine and he has developed several tools for sentence and word alignment from parallel corpora.
He is the convenor of the work group “Multilingual information representation” of the ISO/TC37/SC4 and he participates as an expert in promoting the MLIF standard (Multilingual information framework).

Eduard Šubert

Eduard Šubert studies informatics and mathematics at Czech Technical University in Prague.
He was introduced to computational linguistics in course at the university.
Among other interests, Eduard is responsible for the creation of science popularizing video content of his faculty's YouTube channel.
Additionally, he works on a computer simulation of lens polishing process for the Academy of Sciences of the Czech Republic.

AnneMarie Taravella

AnneMarie Taravella, cert. tr. (OTTIAQ) is a doctoral student and part-time faculty at Université de Sherbrooke, in Québec (Canada), as well as a member of Ordre des traducteurs, terminologues et interprètes du Québec (OTTIAQ).
She holds a BA in Translation and a MA in Translation Studies, both from Montreal-based Concordia University. She is also a graduate from Université de Paris-IX Dauphine, France, in Management Science.
AnneMarie is now pursuing a Doctorate in Business Administration (DBA) at the Faculty of Administration of Université de Sherbrooke, under the supervision of Alain O. Villeneuve, DBA.
Her research interests are translation work organization, adoption of information technologies in organizations, workplace well-being and positive organizational scholarship.
Her doctoral research focuses on the variation of language specialists' affective states in the workplace.
Her research is supported by the Canadian Social Sciences and Humanities Research Council.

Antonio Toral

Antonio Toral (Dr.), Research Fellow at Dublin City University (DCU).
Obtained his MSc in Computer Science in 2004 and PhD in Computational Linguistics in 2009 from the Universitat d'Alacant (Spain).
He worked as a researcher in CNR-ILC (Italy) from 2007 to 2009, involved in the EU-FP7 projects KYOTO and FLaReNet.
He joined DCU in 2010 where he has been working as a postdoctoral researcher to date, in the EU-FP7 projects Abu-MaTran (coordinator), QTLaunchPad, PANACEA and CoSyne.
He has published more than 70 peer-reviewed papers, has served in the scientific committee of international conferences and workshops and has reviewed papers for three indexed journals of the field.
He has also organised evaluation tasks at the SemEval and EVALITA forums.

Tom Vanallemeersch

Tom Vanallemeersch is a researcher at KU Leuven, Centre for Computational Linguistics.
He has been working in the language technology sector for twenty years, both in academia and industry. His activities mainly involve translation memories, machine translation and alignment of bilingual resources.
He performed work in these fields at Xplanation, LNE International, Lessius University College, Systran (project in collaboration with Centre for Computational Linguistics) and the European Commission.
Other types of language technology he dealt with are multilingual dictionary processing (University of Liège), text-to-speech (Lernout and Hauspie), text mining (Temis) and terminology extraction (coordinator of project at Dutch Language Union).

Marion Wittkowsky

Marion Wittkowsky is a lecturer in the Department of International Technical Communication at the Flensburg University of Applied Sciences in Germany since 2007.
She teaches courses in technical writing, technical translation, and applied computer linguistics.
Prior to her position at the University she worked for ten years as a technical translator, project manager and finally as a business unit manager at a language service provider.
A major focus of her translation work was post-editing the machine translation of SAP release notes.

Angelika Zerfass

Angelika Zerfass, a consultant, trainer and provider of technical support for users of translation tools is located in Bonn, Germany.
After her studies (degree in translation for Chinese and Japanese plus Computational Linguistics) she came to the translation industry first as a training specialist for a tools provider in 1997 and has been working freelance since 2000.
Her main focus being the translation technologies, she is traveling the world on a mission to make translation tools understandable.

Andrzej Zydroń

CTO @ XTM International, **Andrzej Zydroń** is one of the leading IT experts on Localization and related Open Standards. Zydroń sits/has sat on, the following Open Standard Technical Committees:

- LISA OSCAR GMX
- LISA OSCAR xml:tm
- LISA OSCAR TBX
- W3C ITS
- OASIS XLIFF
- OASIS Translation Web Services
- OASIS DITA Translation
- OASIS OAXAL
- ETSI LIS
- DITA Localization
- Interoperability Now!
- Linport

Andrzej has been responsible for the architecture of the essential word and character count GMX-V (Global Information Management Metrics eXchange) standard, as well as the revolutionary xml:tm (XML based text memory) standard which will change the way in which we view and use translation memory.
Andrzej is also chair of the OASIS OAXAL (Open Architecture for XML Authoring and Localization) reference architecture technical committee which provides an automated environment for authoring and localization based on Open Standards.
He has worked in IT since 1976 and has been responsible for major successful projects at Xerox, SDL, Oxford University Press, Ford of Europe, DocZone and Lingo24 in the fields of document imaging, dictionary systems and localization.

Andrzej is currently working on new advances in localization technology based on XML and linguistic methodology.

Highlights of his career include:
1. The design and architecture of the European Patent Office patent data capture system for Xerox Business Services.
2. Writing a system for the automated optimal typographical formatting of generically encoded tables (1989).
3. The design and architecture of the Xerox Language Services XTM translation memory system.
4. Writing the XML and SGML filters for SDL International's SDLX Translation Suite.
5. Assisting the Oxford University Press, the British Council and Oxford University in work on the New Dictionary of the National Biography.
6. Design and architecture of Ford's revolutionary CMS Localization system and workflow.
7. Technical Architect of XTM International's revolutionary Cloud based CAT and translation workflow system: XTM.

Specific areas of specialization:
1. Advanced automated localization workflow
2. Author memory
3. Controlled authoring
4. Advanced Translation memory systems
5. Terminology extraction
6. Terminology Management
7. Translation Related Web Services
8. XML based systems
9. Web 2.0 Translation related technology