The 2nd International T3L Conference: Tradumatics, Translation Technologies & Localisation

9th International Conference on Translation, Department of Translation, Interpreting and East Asian Studies

Universitat Autònoma de Barcelona

10th-11th October 2016

Departament de Traducció
i d'Interpretació i d'Estudis
de l'Àsia Oriental

UAB
Universitat Autònoma de Barcelona

TABLE OF CONTENTS

This document contains the abstracts for the conference. You can look for more information here:

http://jornades.uab.cat/t3lconference/

NEW USES OF MACHINE TRANSLATION IN THE TRANSLATION WORKSTATION

Mikel L. Forcada,

Universitat d'Alacant

http://www.dlsi.ua.es/~mlf/

Machine translation (MT) has now been with us for a long time. Over the years, translators have experimented with MT, and many have adopted it as a resource for their work. The incorporation of MT into the professional translation environment has depended largely on the possibilities offered by the various translation editing and management (TEM) systems, also known as computer-assisted translation (CAT) tools. The speaker will present emerging uses of MT in TEM systems, such as interactive MT (at the subsegment level), MT-based repairs of partial translation memory matches and estimates of the quality of MT at the word level in partial matches.

PANORAMA 2020: WHICH WILL BE THE MOST SOUGHT-AFTER PROFESSIONAL PROFILES IN THE TRANSLATION MARKETPLACE?

Olga Blasco

The Rosetta Foundation

https://www.therosettafoundation.org/

In response to the global paradigm shift over the last decade, there have been rapid technological changes to achieve total connectivity, maximum automation, high productivity and non-stop service to clients. Small, medium and large translation companies are adopting innovative commercial and operational strategies to position themselves in an increasingly competitive market. What professional profiles do they need to make it possible now and in the future, and why?

THE QUALITY OF POST-EDITED MACHINE TRANSLATION: WHAT'S IT TO YOU?

Van Egdom Gys-Walt

Zuyd University of Applied Sciences

https://www.peterlang.com/view/product/23511

Machine translation and post-editing are often subjected to stern and lofty criticism. The community of translators seem to speak with a unified voice when airing the complaint that, with time, the profession will be endangered or, at least, radically transformed by both phenomena, limiting her/his role to that of a post-editor or even to that of an all-round linguistic or intercultural expert, and that, as a consequence of this transformation, quality standards will no longer be upheld.

Instead of easing or allaying the fears of the community, translation studies scholars have contributed to those fears by comparing the quality of post-edited machine translations to target texts made 'from scratch'. Although the qualitative superiority is established or corroborated time and again, these studies alert us to the fact that the loss of quality is probably offset by other factors (e.g. an increase in productivity)

Beyond the theoretical ken of the aforementioned studies lies a terrain that is. fraught with perilous obstacles, the 'hidden' domain of quality. The assumption that quality is a stable concept or that quality is only defined by the profession has left many scholars us mired in error for decades. These kneejerk assumptions are reflected in the aforementioned studies. In a recent study, we have felt the need to venture a leap into the great unknown by defining quality as perceived quality (per definitionem).

In our practice-oriented research, the students of the Zuyd simulated translation bureau were asked to produce four PE-versions of two English source texts. Each version was related to a degree of editing. The eight post-edited machine translations were distributed to LSP's as well as to end-users, along with a survey form. With our pilot-study, we have hoped to gain demonstrably deeper understanding of the quality standards of both respondent groups, with a view to developing a rationale for what is commonly called the 'fitness for purpose' of the PE text.

POST-EDITING SUBOPTIMAL MACHINE TRANSLATION: WHY, WHEN AND HOW?

Nora Aramberri

University of the Basque Country (UPV/EHU)

http://www.ehu.eus/ehusfera/ixa/

It is well known that machine translation (MT) systems do not always provide high-quality output. This does not necessarily mean that translators cannot benefit from it coupled with post-editing. Nevertheless, little attention has been devoted to analysing what happens in these scenarios. In this line, this presentation aims to start a discussion on aspects relating to MT quality and the post-editing task, specifically, potential advantages, changes in the post-editing approach and training.

MACHINE TRANSLATION AND POST-EDITING METRICS – MYTH AND FACTS

Blanca Vidal

Lucy Software and Services

www.Lucysoftware.com

The main -and still unresolved- issue with MT in combination with post-editing (MT+PE) is the projection of the post-editors productivity to define a price tag to the job to be done.

Over the decades a lot of PE guidelines, workshops and trainings defined skills that post-editors should have and criteria they have to take care of when post-editing. And we saw different approaches to measure the productivity of post-editors in comparison to normal translation.

IBM Germany gave us the possibility to measure exactly the post-editing and translation productivity of many translators for every segment they worked on. So we were able to see the translators "ground productivity" (incl. the use of TM matches) in relation to the post-editing productivity.

In our presentation we will show some figures which allow us to de-mystify and uncover some information, empty promises and rumors circulating in the translation industry.

A Description of Post-editing, from Translation Studies to Machine Learning

Félix do Carmo and Belinda Maia

Faculdade de Letras da Universidade do Porto

https://sigarra.up.pt/flup

Since Holmes' map of the discipline, there has been a space in Translation Studies for research on the translation process. However, when Statistical Machine Translation (SMT) appeared, the description of the translation process in Translation Studies seemed to be overshadowed by the efficiency of algorithms and mathematical techniques to suggest matches and rebuild near functional sentences from big bilingual data. Although it is considered a quality requirement, post-editing is regarded as a very simple process by translation companies and their clients, who require no specification of techniques, but simply usable results.

In this paper, we move away from the concepts of the translation and the post-editing processes as seen by the most current Translation Studies research, in order to integrate the perspective of post-editing in the SMT literature. Inspired by this, we suggest that, in order to distinguish it from Translation and Revision, Post-editing of SMT can be best defined by its technical procedures. Finally, this view of Post-editing leads to the suggestion that the integration of Machine Learning into translation tools should be adjusted in order to support the work of post-editors more efficiently. This paper is based on the experience of one of the authors in managing and doing research in translation and post-editing projects.

MACHINE TRANSLATION: A SECRET LOVER TURNED INTO THE LIFE PARTNER

Diego Bartolomé

tauyou language technology

www.tauyou.com

No matter if we like it or not, Machine Translation (MT) is here to stay. Certainly, there are several fields where MT (nor CAT tools) will never be used because the process of creating a translated text requires a significant creativity. However, it is currently used in many technical fields, and also for verticals such as e-commerce where in many cases a machine translated sentence plus a light post-editing is good enough for the end client. However, don't worry, top quality translations will never disappear! But some other types of translation will be replaced by Big Data processing and natural language processing ...

In this presentation, we will show why MT might not be well-suited for certain fields, but also the good translations produced by MT in fields such as e-commerce or hospitality with Big and Small Data. Then, the optimum process to deal with post-editing jobs from the freelance translator point of view will be described, discussing key aspects such as the discount or how to deal with complaints. Finally, we will cover how tools in the market deal with machine translation post-editing, and some products specifically designed for the task.

In the end, we will show that if MT helps freelance translators increase their profits or have more free time, and also opens new business opportunities, the optimum option we all have is to embrace change and love MT.

METHODOLOGY OF MT POST-EDITORS TRAINING. PRACTICAL EXPERIENCE WITH PEMT AS A PART OF TRANSLATION STUDIES AT CONSTANTINE THE PHILOSOPHER UNIVERSITY IN NITRA.

Jakub Absolon

Constantine the Philosopher University in Nitra

www.UKF.sk

In my presentation I would like to explain why we understand there is a need for specific methodology for PEMT training, based on experience both with professional translators and students.

Current students are more than familiar with statistical based machine translation and therefore it should be reflected in methodology we use. They are “natural-born Googlers” and that is why they should understand how MT can help them but also how it could threaten them. Our main interest is to find out which are the most important competencies for translators using PEMT as a natural technique of translation process and what are the differences (if any) in comparison to a traditional translator. The second goal is to find out which of the competencies of PE are trainable, to what extend and what is the best methodology for training.

EL TRADUCTOR COMO USUARIO AVANZADO DE INFORMÁTICA: HABILIDADES PARA LA TRADUCCIÓN AUTOMÁTICA

Amparo Alcina

Universitat Jaume I

http://tecnolettra.uji.es/es/

La era digital y globalizada en que estamos inmersos nos hace pensar que hoy en día cualquier persona es usuaria nata de la informática y se "lleva bien" con ella. Sin embargo, en el aula de traducción podemos observar numerosas deficiencias de conocimientos y habilidades informáticas de los estudiantes de traducción en el manejo de las diversas tecnologías. La traducción automática también se presenta en forma intuitiva para el gran público, pero el manejo que se exige al traductor de esta herramienta va mucho más allá de esta primera intuición. En este trabajo pretendemos caracterizar cuáles son las luces y sombras de los conocimientos, habilidades y actitudes en que deben formarse los traductores para el manejo de esta herramienta.

WORKSHOP: WHEN MY (TRANSLATION) MEMORY IS NOT ENOUGH: A PRACTICAL WORKSHOP ON HOW TO CREATE TRANSLATION MEMORIES FROM THE WEB AND USE THEM TOGETHER WITH MACHINE TRANSLATION.

Gema Ramírez

Promsit

www.prompsit.com

Translation memories are great: they prevent translating the same sentence twice. But, where to find them when you have no previous translations? The answer can be out there: Internet is plenty of multilingual content that can be seen as a big and powerful TM. Until now, only big players and their IT teams have been able to explote this mine, but in this workshop we will show you how you can have this power at just some clicks with a new tool: Bicrawler. We will then work on combining translation memories with machine translation to get the best of current technologies without suffering them.

Prompsit will be running this workshop. Prompsit is a translation technology company with a strong focus on empowering translators with useful tools (www.prompsit.com).

WORKSHOP: THE TAUYOU MACHINE TRANSLATION PLATFORM

Diego Bartolomé

tauyou language technology

www.tauyou.com

Are you tired of using Google or Microsoft Machine Translation (MT) engines where you do not have confidentiality and control? Would you like to know how to build customized MT engines to double your productivity? How can you further explore the possibilities MT offers? In this workshop, we will describe the basics of the tauyou MT platform, to then show the optimum process to achieve significant gain in productivity. Translation will be faster and better after it.

MTradumàtica

Ramon Piqué

Adrià Martín

Universitat Autònoma de Barcelona

www.tradumatica.net

La qualitat de la Traducció Automàtica (TA) es pot millorar mitjançant l'entrenament de motors estadístics a partir de corpus bilingües i monolingües.

El traductor està cridat a tenir un paper rellevant en aquest punt i entenem que cal contemplar aquests processos en la formació especialitzada del traductor.

L'MTradumàtica és una eina per a entrenar motors de TA estadístics (TAE) a partir de corpus. Es tracta d'una interfície gràfica basada en web per a Moses. L'usuari pot carregar bitextos, per a crear models de traducció, i montextos per a crear models de llengua.

Amb l'MTradumàtica aquest procés se simplifica per mitjà d'una interfície intuïtiva i facilita l'aproximació per part dels professionals de la traducció a la gestió d'aquests processos.

MTradumàtica (English version)

The quality of machine translation (MT) can be improved by training statistical engines from monolingual and bilingual corpora.

The translator is called to play an important role in this environment and it is fundamental to include these processes in the training of specialized translators.

The MTradumàtica is a tool to train statistical MT engines (SMT) from different corpora. It is a web-based GUI for Moses. The user can load bi-texts to create translation models, and monolingual texts to create language models.

With MTradumàtica this process is simplified through an intuitive and easy interface and it facilitates translation professionals the management of these training processes.

EXPLOITATION OF A COMPARABLE CORPUS OF USER-GENERATED CONTENT FOR POST-EDITING STRATEGIES

Miguel Ángel Candel-Mora

Universitat Politècnica de València

http://www.upv.es/entidades/DLA

With the Web 2.0 and the active participation of users, online consumer-generated reviews have become a clear reference in purchasing decision-making processes, and on occasions sometimes exceed the reliability and authenticity of advertising campaigns from manufacturers.

User reviews may appear in different formats and with different structures (Vázquez, 2014): as evaluations of a product, in the form of dialogue in a forum, or in the case of this study, as unstructured free text for the evaluation of a tourism product (hotel reviews), in which, in addition to specific linguistic features, other aspects such as the reviewers' reliability, naturalness, or credibility play a key role.

Hotel reviews in particular have been studied extensively (Schemmann, 2011) and research has clearly identified its structure and revealed that opinions are not only conveyed through linguistic resources, but there are other genre-specific features such as intertextuality, the profile of the reviewer, and paralinguistic elements that contribute to the reliability and credibility of consumer reviews (Pollach, 2006; Vásquez, 2012).

From the point of view of post-editing, and keeping in mind the purpose of the target text (O'Brien, 2005), work to improve machine translation would basically consist in repairing and accommodating the MT output - in varying degrees - to the appropriate target language linguistic conventions.

Thus, the research question that motivates this paper is that in the case of user-generated reviews in the domain of tourism, the message is not only transmitted through linguistic resources but there are other elements or textual artifacts that should be taken into consideration in the post-editing strategy, in addition to relevant grammar and stylistic PE guidelines (Babych, 2014; Vilar et al., 2006).

Several studies have already confirmed that there are no universal guidelines for post-editing (Allen, 2003; TAUS, 2010), and each genre requires specific quality rating scales. Thus, this work highlights the need to pay special attention to the textual conventions during any post-editing strategy in addition to identifying error patterns common to most

post-editing guidelines. More specifically, the objective of this work is to compare textual characteristics of user reviews originally written in English and in Spanish from data derived from a corpus-based approach analysis that serve as standards in MT output post-editing tasks.

Furthermore, this objective is further expanded with two subobjectives:

1- to extract MT-related data to elaborate training material for future pos-editors, and

2- to reuse the translation knowledge extracted to add bilingual dictionaries to the parallel corpus of the MT system to improve its performance.

MACHINE TRANSLATION IN SPANISH LSPS

Olga Torres-Hostench

Marisa Presas

Pilar Cid

Universitat Autònoma de Barcelona

projecta.tradumatica.net

In this presentation we will discuss with the audience the results of a report carried out on the use of Machine Translation (MT) and MT Post-editing (PE) in Spanish Translation Service Providers in 2015. The report is part of a research project called ProjecTA, funded by the Spanish Ministry of Economy and Competitiveness (Ref. FFI2013-46041-R) and aimed at analysing the flow of MT+PE work in the professional translation sector in Spain. Quantitative data were collected through an online survey, which was reviewed by experts in Statistics and TSP representatives. The survey was sent individually to 189 Spanish TSPs during January and February 2015 and 55 surveys received were valid, which corresponds to the 29.5% of the sample. The survey consisted in 17 questions, distributed in three areas: (a) basic identification information from the TSP (location, number of employees, annual turnover, etc.); (b) Services and specialties offered by the TSP, most common source and target languages, types of customers, etc.; (c) use of MT and PE in the company (use of MT; reasons not to use MT; percentage of MT use in the company projects; type of MT system; ownership of the MT system; translators' acceptance of PE work; percentage of PE use in the company projects). . The quantitative results show that 47.3% of Spanish language service providers use MT and that 45.5% of these use MT in only 10% of their total projects. Furthermore, qualitative data were also obtained via telephone interviews, personal interviews with experts and a focus group session in second quarter 2015. The qualitative results reveal that the decision to implement an MT system depends on multiple factors: the business's economic capacity, technological capacity and the knowledge and attitude of the business's human resources.

On the correctness of machine translation: an English-Italian case study

Giovanni Cerasani

Binari Sonori

www.binarisonori.it

In order to understand the present capability of machine translation, I drew inspiration from the essay "On the correctness of machine translation: A machine translation post-editing task" to make a study of five texts, different from each other in both difficulty and subject. They were analysed between May and July 2015 through the Google Translate (statistical) and SYSTRANet (hybrid) systems, as my intention was to verify the reliability of free MT, which is more likely to be used by common web users.

Unlike the quoted study, my analysis on the English-Italian MT output was carried out with the source texts constantly at hand. Therefore I adopted different evaluation methods from the 1-to-5 scale originally used for both meaning and grammar. Since most of the sentences were long and syntactically complex, and since I wanted to focus on the recognition of a "core meaning", I chose a 1-to-4 scale in order to evaluate the semantic accuracy:

- 1 – The core meaning is obscure or distorted;

- 2 – The core meaning is not very clear;

- 3 – The core meaning is clear;

- 4 – The whole sentence is clear.

As far as grammar is concerned, in my opinion neither the quality nor the quantity of errors can be mathematically established. Not only do the various error types not weigh the same, but a single type may not be equally difficult to understand and correct, according to the context. That's the reason why I chose a 1-to-3 scale:

- 1 – The average error quantity and quality requires massive post-editing;

- 2 – The average error quantity and quality requires light post-editing;

- 3 – The sentence is correct.

We can extract some interesting data from the 376 sentences analysed (188 for each system):

• 184 sentences (49%) are completely clear. 28 of these (7%) are grammatically correct;

• 148 sentences (39%) are ungrammatical, and 41 of these (11%) are obscure.

The most common semantic error is the incorrect translation of polysemous words. English terms may occasionally be kept: sometimes they are out-of-context anglicisms generally accepted in the Italian language today; some others they are non-standard English words which are, thus, not identified by the MT systems themselves. Multiword expressions may also be tricky and, as a consequence of the individual translation of the lexemes, the overall meaning is often lost.

As for grammar, it may be the case that verbs are not translated in the most suitable tense or mood: inconsistencies can sometimes be found in a single sentence. Phrasal verbs are a source of misunderstandings both from a semantic (i.e. the "get" phrasal verbs) and a syntactic perspective (when prepositions are taken separately). Word order is mostly retained: it is certainly positive with SVO and brief sentences, but it is not when the latter are longer and have a more complex syntax. Gender and number agreement are not always guaranteed.

LOS ERRORES DE LA TRADUCCIÓN AUTOMÁTICA CHINO > ESPAÑOL

Hong Zhang

Universitat Autònoma de Barcelona

www.uab.cat

El presente trabajo tiene como prioridad principal ver los tipos de errores más habituales en el uso de la traducción automática chino>español con el objetivo de enseñar esos errores durante el aprendizaje de español como lengua extranjera. La traducción automática está acaparando un gran interés científico e investigador en los últimos años. Con el desarrollo tecnológico y el acelerado ritmo de vida cotidiana se ha producido un aumento de demanda en el uso de traducciones automáticas por parte de los estudiantes chinos que aprenden el idioma español. Sin embargo, el resultado de los sistemas de la traducción automática a menudo no otorga sentido ni coherencia a las palabras, y genera muchos errores. Los estudiantes deberían aprender a contrarrestar de manera correcta los errores de la traducción automática. Muchos autores han identificado los principales errores producidos por la traducción automática (Hutchins y Somers, 1992; Sitman y Piñol, 1999; Mercedes, 2002; Giménez, 2009; Dorr, 1993; Arnold, 2003; Bennett, 1993; Yang y Lange, 2003; Lindop y Tsujii, 1991; Tsujii y Fujita, 1991; y Whitelock, 1992, etc.).

Mi presente trabajo consiste en el análisis crítico y contrastivo de los errores que se cometen en la traducción automática de chino al español, en el uso de los materiales de prensa china. La prensa, por sus características tales como la periodicidad, la novedad, el interés y la objetividad es un tipo de texto muy traducido con traducción automática. El propósito de esta investigación es hacer una lista de los errores comunes de la traducción automática en prensa. Siguiendo una metodología descriptiva, se corregirán las traducciones automáticas siguiendo las recomendaciones de la industria para posedición. Se marcará los errores en diferentes colores y cada color representará un tipo de error. Por medio de esa posedición, se podrá elaborar una lista preliminar de los errores más habituales en traducción automática de chino al español ya sean errores semánticos, léxicos, sintácticos, morfológicos, ortográficos y otros errores del español. Se quiere observar si los errores de traducción automática son de todo tipo o si se producen un tipo de errores más que otros.

En el futuro la autora empleará los tipos de error identificados en este trabajo para elaborar materiales didácticos para aprender español con actividades de traducción automática. Actualmente no existen investigaciones sobre el uso de la traducción automática como herramienta para aprender español por lo que esta investigación tiene un gran sentido a nivel académico.

LA POSTEDICIÓN COMO NUEVA ESPECIALIZACIÓN TRANSDISCIPLINAR EN LA ERA DE LA INFORMACIÓN GLOBALIZADA

Coral Díez Carbajo

Universidad de Salamanca

www.usal.es

La post-edición se presenta como una solución para hacer frente a la cantidad de textos que requieren traducción hoy en día. Por lo general, hasta ahora, se ha concebido como una especialización para traductores. No obstante, como advierten investigaciones recientes (Ruano 2013, Bielsa 2005), en la era de la información, la traducción está integrada hoy en día en numerosas prácticas de redacción y procesamiento textual que permiten la circulación de información y noticias en la era global. Por esta razón, la formación en post-edición puede ser interesante para otros perfiles profesionales, como los periodistas, que, como han mostrado autores como Hernández Guerrero (2009) o Schäffner (2012), asumen operaciones de traducción en su labor diaria en prácticas como la transedición. Por ello, y basándonos en la experiencia personal, consideramos fundamental idear acciones formativas en post-edición destinadas tanto a futuros traductores como a periodistas que permitan a los estudiantes familiarizarse con los procedimientos de la post-edición para completar esta tarea de manera satisfactoria. En este sentido, en esta ponencia presentamos un curso destinado a estudiantes de traducción y periodismo con un plan de estudios organizado por semanas y concebido desde la perspectiva de las universidades de artes liberales estadounidenses. La importancia de la enseñanza de este curso va más allá del aprendizaje que puedan extraer los estudiantes que reciban formación sobre cómo post-editar correctamente y cómo obtener resultados de post-edición de mayor calidad. Así, creemos que esta propuesta permite contribuir a explorar la utilidad de la post-edición en ámbitos transdisciplinares muy diversos.

TRADUCCIÓ AUTOMÀTICA I LLENGÜES MINORITZADES: EL CAS DEL SARD

Gianfranco Fronteddu

Università degli Studi di Cagliari

www.unica.it

La Traducció Automàtica (TA) ha esdevingut els darrers anys una eina utilitzada per professionals, investigadors i la societat en general. Malgrat que actualment els esforços en recerca i desenvolupament se centren en la millora de la qualitat, la TA en l'estat actual de la tecnologia encara pot suposar molts avenços per a tot un seguit de llengües que habitualment no tenen representació en aquests nivells. És el cas de la combinació lingüística que volem presentar, italià-sard.

El projecte que presentem neix d'una col·laboració entre universitats (la Universitat Autònoma de Barcelona i la Universitat d'Alacant) i l'empresa Prompsit, amb el finançament de Google per mitjà de la iniciativa Google Summer of Code que, cada any, permet a estudiants de grau i de postgrau el desenvolupament de projectes de codi obert durant l'estiu.

La creació d'un traductor automàtic que incorpori la llengua sarda, per les característiques d'aquesta llengua, té diverses implicacions. En primer lloc, pel fet que es tracta d'una llengua en procés d'estandardització, els recursos tant lingüístics (obres de referència) com tecnològics (corpus, correctors) són escassos. En segon lloc, aquesta escassedat, especialment de textos publicats seguint la norma estàndard, fa que calgui decantar-se per un sistema de Traducció Automàtica Basada en Regles (TABR). Apertium, inicialment un projecte públic anomenat OpenTrad, és especialment indicat per a la traducció entre parells d'idiomes de la mateixa família, i es basa en diccionaris i regles de transferència escrits amb llenguatge d'etiquetes. A més, el fet que es tracta d'un projecte lliure, disposa d'una comunitat molt nombrosa que ofereix suport als nous usuaris.

El nostre projecte demostra (i reivindica) un cop més que el perfil del traductor és vàlid per a formar part d'equips de treball per a la creació de motors de TA. En aquesta comunicació, il·lustrarem les diverses fases de creació del motor, les dificultats amb què ens hem trobat i la manera com les hem superades.

PRODUCTION-READY MACHINE TRANSLATION ENGINES WITH KANTANMT

Laura Casanellas Luri

KantanMT

www.kantanmt.com

Learn how to build a MT engine for a production environment in less than two hours

KantanMT is a leading provider of cloud-based Statistical Machine Translation (SMT), offering a global client base cost-effective access to the latest SMT technologies. The KantanMT.com platform is an intuitive and easy to navigate application, allowing members to build and manage SMT engines within the Amazon Web Services (AWS) cloud. The performance of SMT engines can be rapidly improved using KantanMT's unique data analysis and data visualisation technologies (KantanBuildAnalytics™), and SMT project management is simplified using KantanAnalytics™ - a SMT fuzzy-match technology used to plan, cost and schedule projects. Other features of KantanMT.com include: - PEX automatic post editing - GENTRY parsing technology - Instant Segment Re-training (KantanISR™). KantanMT seamlessly integrates with most Computer Aided Translation (CAT) tools and web applications, offering users instant access to on-demand Machine Translation. Members have reported an average increase in productivity of 60% after introducing Kantan Machine Translation into translation workflows.

New to the Kantan platform is KantanLQR™, a language quality evaluation tool that automates the process of Human Evaluation and offers a number of industry standard KPIs to choose from. Also available, is the new KantanFleet™, a set of pre-built engines in a number of verticals and language pairs.

During this workshop we aim to build a Custom Machine Translation engine through the KantanMT platform. Each attendee will be provided with access to the platform, as well as training and testing data to be used during the exercise.

We will create an engine, learn how to check the engine's initial quality using KantanBuildAnalytics, together we will identify methods to fix issues found in the training data through Pre-processing, Rejects Report and Gap Analysis features. We will learn how to keep track and compare different iterations of the engine using KantanTimeline™. Finally, we will translate a file using the recently created engine and analyse the output through Kantan's quality estimation report.

In the second part of the session, we will use KantanLQR to set up a Human Evaluation project, and finally we will review the output using the editing environment with the aim of improving the engine further to production ready status.

POST-EDITING IN A FEEDBACK FRIENDLY MT ENVIRONMENT

Blanca Vidal

Lucy Software and Services

http://www.lucysoftware.com/espanol

Post-Editing in a feedback friendly MT environment ...and other advantages of deep Rule-based Machine Translation Systems

The rule-based MT (RBMT) systems have often been marked as outdated and not innovative. But in reality, a deep RBMT system such as Lucy LT can yield very good results in many professional MT environments, especially for language pairs where SMT still fails to find solutions to handle language specific issues, and when not enough bi-lingual training material for the statistical training is available.

Moreover, the Lucy LT system can be quickly adapted to the customer's needs.

In the workshop will show you

• the issues post-editors will see with MT output coming from a deep RBMT system,

• how easily the output quality could be improved with feedback from the post-editor.

Due to the interactive character of the workshop, we would recommend to limit the number of participant to max 15 persons.

Índice de ponentes