This document is part of the Project "Machine Translation Enhanced Computer Assisted Translation (MateCat)", funded by the 7th Framework Programme of the European Commission through grant agreement no.: 287688.

Machine Translation Enhanced
Computer Assisted Translation

D6.4 - Third report on dissemination and exploitation

Authors: Alessandro Cattelan and Manuela Speranza

Dissemination Level: Public

Date: November 16th, 2014

Grant agreement no.	287688
Project acronym	MateCat
Project full title	Machine Translation Enhanced Computer Assisted Translation
Funding scheme	Collaborative project
Coordinator	Marcello Federico (FBK)
Start date, duration	November 1st 2011, 36 months
Dissemination level	Public
Contractual date of delivery	October 31st, 2014
Actual date of delivery	November 15th 2014
Deliverable number	D6.4
Deliverable title	Third report on dissemination and exploitation
Type	Report
Status and version	Final
Number of pages	25
Contributing partners	TRANSLATED, FBK
WP leader	FBK
Task leader	TRANSLATED
Authors	Alessandro Cattelan and Manuela Speranza
Reviewers	Frédéric Blain
EC project officer	Alexandra Wesolowska
The partners in MateCat are:	Fondazione Bruno Kessler (FBK), Italy
	Université Le Mans (LE MANS), France
	The University of Edinburgh (UEDIN)
	Translated S.r.l. (TRANSLATED)

For copies of reports, updates on project activities and other MateCat-related information, contact:

FBK
MateCat
Manuela Speranza manspera@fbk.eu
Povo - Via Sommarive 18 Phone: +39 0461 314 521
I-38123 Trento, Italy Fax: +39 0461 314 591

Copies of reports and other material can also be accessed via http://www.matecat.com

Executive Summary

This document reports on the dissemination and exploitation activities carried out during the third year of the project. On the one hand, we focused our activities on disseminating information about the research and technological advancements to the research community. On the other, we worked hard to set the ground for the commercial exploitation of the product developed during the three years of the EC-funded project.

In particular, we focused on creating an online presence and on attracting the first groups of early adopters from the translation industry, both professional translators and language service providers. With that aim in mind, we organized online competitions, user group events, online webinars and took part in a number of conferences to actually get to talk to our users.

This document also presents the list of the activities carried out and the documentation and marketing material created to support the dissemination and exploitation efforts.

Table of Contents

1 Introduction

MateCat aims at improving the integration of machine translation (MT) and human translation within the computer aided translation (CAT) framework. The ultimate goal of the project is to improve productivity of professional translators and to enhance their work experience with MT in a professional setting. Large part of the communication activities therefore focuses on promoting the technology among professional translators and language service providers. At the same time, the research partners used all available channels to disseminate information about the research and technological advances to the research community.
This deliverable describes the activities in both areas. In terms of dissemination, it reports on the activities carried out by the research partners to present the research results through the publication of technical papers in the top scientific venues and journals, and presentations at scientific conferences. In terms of exploitation of the commercial product, the activities focused on creating an online presence and developing the brand through the redesign of the website, the creation of promotional material such as videos and gadgets, the organization of contests and the active use of the MateCat technology by the commercial partner for most of the translation projects for its customers. All of these activities resulted in the adoption of MateCat by over 3,000 professional users and the translation of over 50 million words with it.

2 Scientific dissemination

The following scientific papers have been published/presented during the third year of the project (they are all published under an open license and available through the MateCat website):

- Bernardo Magnini, Marco Baroni, Marcello Federico, and Roberto Navigli. Recent Advancements in Human Language Technology in Italy. Intelligenza artificiale 7, 2013, p. 91-100.
- Amin Farajian, Nicola Bertoldi, and Marcello Federico. Online Word Alignment for Online Adaptive Machine Translation. EACL Workshop on Humans and Computer-assisted Translation, Gothenburg, Sweden, 26 April 2014, p. 84–92.
- Christian Buck, Kenneth Heafield, and Bas van Ooyen. N-gram Counts and Language Models from the Common Crawl. 9th Language Resources and Evaluation Conference (LREC 2014), Reykjavik, Iceland, 26-31 May 2014, p. 3579-3584.

- Marco Turchi and Matteo Negri. Automatic Annotation of Machine Translation Datasets with Binary Quality Judgements. 9th Language Resources and Evaluation Conference (LREC 2014), Reykjavik, Iceland, 26-31 May 2014, p.1788-1792.
- Marco Turchi, Antonios Anastasopoulos, José G. C. de Souza, and Matteo Negri. Adaptive Quality Estimation for Machine Translation. 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), Baltimore, USA, 22-27 June 2014, p. 710-720.
- José G. C. de Souza, Jesus González-Rubio, Christian Buck, Marco Turchi, and Matteo Negri. FBK-UPV-UEdin participation in the WMT14 Quality Estimation shared-task. 9th ACL Workshop on Statistical Machine Translation (WMT 2014), Baltimore, USA, 26-27 June 2014, p. 322-328.
- Mihael Arcan, Claudio Giuliano, Marco Turchi, and Paul Buitelaar. Identification of Bilingual Terms from Monolingual Documents for Statistical Machine Translation. 4th COLING International Workshop on Computational Terminology (Computerm), Dublin, Ireland, 23 August 2014, p. 22-31.
- Marcello Federico, Nicola Bertoldi, Mauro Cettolo, Matteo Negri, Marco Turchi, Marco Trombetti, Alessandro Cattelan, Antonio Farina, Domenico Lupinetti, Andrea Martines, Alberto Massidda, Holger Schwenk, Loic Barrault, Frederic Blain, Philipp Koehn, Christian Buck, and Ulrich Germann. The MateCat Tool. 25th International Conference on Computational Linguistics (COLING 2014), Dublin, Ireland, 23-29 August 2014, p. 129-132.
- Christian Girardi, Luisa Bentivogli, Mohammad Amin Farajian, Marcello Federico. MT-EQuAl: a Toolkit for Human Assessment of Machine Translation Output. 25th International Conference on Computational Linguistics (COLING 2014), Dublin, Ireland, 23-29 August 2014, p. 120-123.
- José G. C. de Souza, Marco Turchi, and Matteo Negri. Machine Translation Quality Estimation Across Domains. 25th International Conference on Computational Linguistics (COLING 2014), Dublin, Ireland, 23-29 August 2014, p. 409-420.
- Prashant Mathur, Mauro Cettolo, Marcello Federico, and José G. C. de Souza. Online Multi-User Adaptive Statistical Machine Translation. 11th biennial conference of the Association for Machine Translation in the Americas (AMTA 2014), Vancouver, Canada, 22-26 October 2014, p. 152-165.
- Mihael Arcan, Marco Turchi, Sara Tonelli, Paul Buitelaar. Enhancing Statistical Machine Translation with Bilingual Terminology in a CAT Environment. 11th biennial

conference of the Association for Machine Translation in the Americas (AMTA 2014), Vancouver, Canada, 22-26 October 2014, p. 54-68.

- Mauro Cettolo, Nicola Bertoldi, and Marcello Federico. The Repetition Rate of Text as a Predictor of the Effectiveness of Machine Translation Adaptation. 11th biennial conference of the Association for Machine Translation in the Americas (AMTA 2014), Vancouver, Canada, 22-26 October 2014, p. 166-179.
- José G. C. de Souza, Marco Turchi, and Matteo Negri. Towards a Combination of Online and Multitask Learning for MT Quality Estimation: a Preliminary Study. AMTA 2014 Workshop on Interactive and Adaptive Machine Translation (IAMT 2014), Vancouver, Canada, 22 October 2014, p. 9-19.
- Prashant Mathur, and Mauro Cettolo. Optimized MT Online Learning in Computer Assisted Translation. AMTA 2014 Workshop on Interactive and Adaptive Machine Translation (IAMT 2014), Vancouver, Canada, 22 October 2014, p. 32-41.
- Ulrich Germann. Dynamic Phrase Tables for Statistical Machine Translation in an Interactive Post-editing Scenario. AMTA 2014 Workshop on Interactive and Adaptive Machine Translation (IAMT 2014), Vancouver, Canada, 22 October 2014, p. 20–31.
- Marcello Federico, Matteo Negri, Luisa Bentivogli, Marco Turchi. Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models. Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), Doha, Qatar, 25-29 October 2014, p. 1643-1653.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), Doha, Qatar, 25-29 October 2014, p. 1724-1734.
- Mauro Cettolo, Nicola Bertoldi, Marcello Federico, Holger Schwenk, Loïc Barrault, and Christophe Servan. Translation Project Adaptation for MT-Enhanced Computer Assisted Translation. Machine Translation Journal, Volume 28, Issue 2 (2014), p. 127-150.

The following scientific papers have been accepted for publication during the third year of the project:

- Marco Turchi, Matteo Negri, and Marcello Federico. Data-driven Annotation of Binary MT Quality Estimation Corpora Based on Human Post-editions. Machine Translation Journal. To appear.

- Nicola Bertoldi, Patrick Simianer, Mauro Cettolo, Katharina Waeschle, Marcello Federico, and Stefan Riezler. Online Adaptation to Post-Edits for Phrase-Based Statistical Machine Translation. Machine Translation Journal, Special issue on Post-editing. To appear.
- José G. C. de Souza, Marco Turchi, Matteo Negri, Antonios Anastasopoulos. Online and Multitask learning for Machine Translation Quality Estimation in Real-world scenarios. Italian Computational Linguistics Conference (CLIC-it 2014), Pisa, Italy, 9-10 December 2014. To appear.
- Mauro Cettolo, Nicola Bertoldi, and Marcello Federico. Adattamento al Progetto dei Modelli di Traduzione Automatica nella Traduzione Assistita. Italian Computational Linguistics Conference (CLIC-it 2014), Pisa, Italy, 9-10 December 2014. To appear.
- Frédéric Blain, Amir Hazem, Fethi Bougares, Loic Barrault, and Holger Schwenk. Project Adaptation over Several Days. 2nd Translation in Transition Conference. Mainz, Germany, 29-30 January 2015. To appear.

Besides the conferences/workshops where the above scientific papers have been presented, MateCat has also been promoted as described below.

Conferences/workshops:

- Attendance and discussions with industrial users to present MateCat by LIUM at the GALA Annual Conference, 23-26 March 2014, Istanbul (Turkey).
- Invited talk "New frontiers for Machine Translation Quality Estimation: from the lab to the industry" by FBK at TAUS Quality Evaluation Summit, 4 June 2014, Dublin (Ireland).
- Exhibit “MateCat : un outil d'aide à la traduction” by LEMANS at JEP 2014 (Journées d’Études sur la Parole), 26 June 2014, Le Mans (France).
- Demo “MT-EQuAl: a Toolkit for Manual Assessment of Machine Translation Output” by FBK at the 25th International Conference on Computational Linguistics (COLING 2014), 26 August 2014, Dublin (Ireland).
- Demo “The Matecat Tool” by FBK, Translated, LEMANS and UEDIN at the 25th International Conference on Computational Linguistics (COLING 2014), 26 August 2014, Dublin (Ireland).
- Lab “Humans in the Loop for MT Improvement: a Hands-on Experience with Manual Error Annotation” by FBK at the 9th Machine Translation Marathon (MTM 2014), 9 September 2014, Trento (Italy).

- Keynote "CAT Tool" by Translated at the 9th Machine Translation Marathon (MTM 2014), 12 September 2014, Trento (Italy).
- Workshop organization: "Interactive and Adaptive Machine Translation (IAMT 2014)", workshop co-organized by FBK and UEDIN (in collaboration with the CASMACAT project) at AMTA 2014, 22 October 2014, Vancouver (Canada).
- Invited talk "User-adaptive machine translation in the MateCat tool" by FBK at the AMTA Workshop on Interactive and Adaptive Machine Translation (IAMT 2014), 22 October 2014, Vancouver (Canada).
- Showcase "The MateCat Tool" by FBK and Translated at the 11th biennial conference of the Association for Machine Translation in the Americas (AMTA 2014), 24 October 2014, Vancouver (Canada).
- Booth by Translated at the TAUS Annual Conference 2014, 27-28 October 2014, Vancouver (Canada).
- Panel "MT Adoption and Usage" by Translated at the TAUS Annual Conference 2014, 28 October 2014, Vancouver (Canada).
- Exhibit "MateCat" by Translated at the Localization World Conference and Exhibits, 30-31 October 2014, Vancouver (Canada).

Other events:

- Lecture "La traduzione automatica a supporto dei traduttori", held by FBK on 12-13 December 2013 at Univesity of Sassari (Italy).
- Course "Traduzione automatica", held by FBK on November 2013 - January 2014 at University Institute ISIT Trento, (Italy).
- One-day tutorial for the EU translators organized by the DGT held by Translated and FBK on 16 January 2014 in Luxembourg.
- Exhibition "MateCat: traduzione automatica a supporto dei traduttori" held by FBK on 8 May 2014 at Festa dell'Europa 2014, FBK, Trento (Italy).
- Stages for students in translation studies, 12-24 May 2014, at FBK Trento (Italy).
- Student summer internship by Lisa Beinborn (University of Darmstadt, Germany) "Large-Scale Distributed Language Models", 30 June - 22 August 2014, at FBK Trento (Italy).
- Student summer internship by Daniel Torregrosa Rivero (University of Alicante, Spain) "Error Tagging", 16 June - 22 September 2014, at FBK Trento (Italy).

- Student summer internship by Marion Weller (University of Stuttgart, Germany) "Learning from Human Post-Editions", 28 June - 27 September 2014, at FBK Trento (Italy).
- Demo by LIUM on 11-15 September 2014 at Le Mans exhibition.
- Talk "MT Quality Estimation and Evaluation for Professional Translators", held by FBK on 17 September 2014 at eBay, San Jose, California (USA).
- Talk "Offline and Online Adaptation Methods in Statistical Machine Translation from Human Post-editions", held by FBK on 18 September 2014 at eBay, San Jose, California (USA).
- User Group Event, organised by Translated and FBK on 28 October 2014 at Pan Pacific Vancouver Hotel, Vancouver (Canada).

3 Commercial exploitation

3.1 Brand identity

We worked on defining the brand identity for MateCat, an iterative and ongoing process which is helping us position correctly the tool on the market to meet specific user needs.

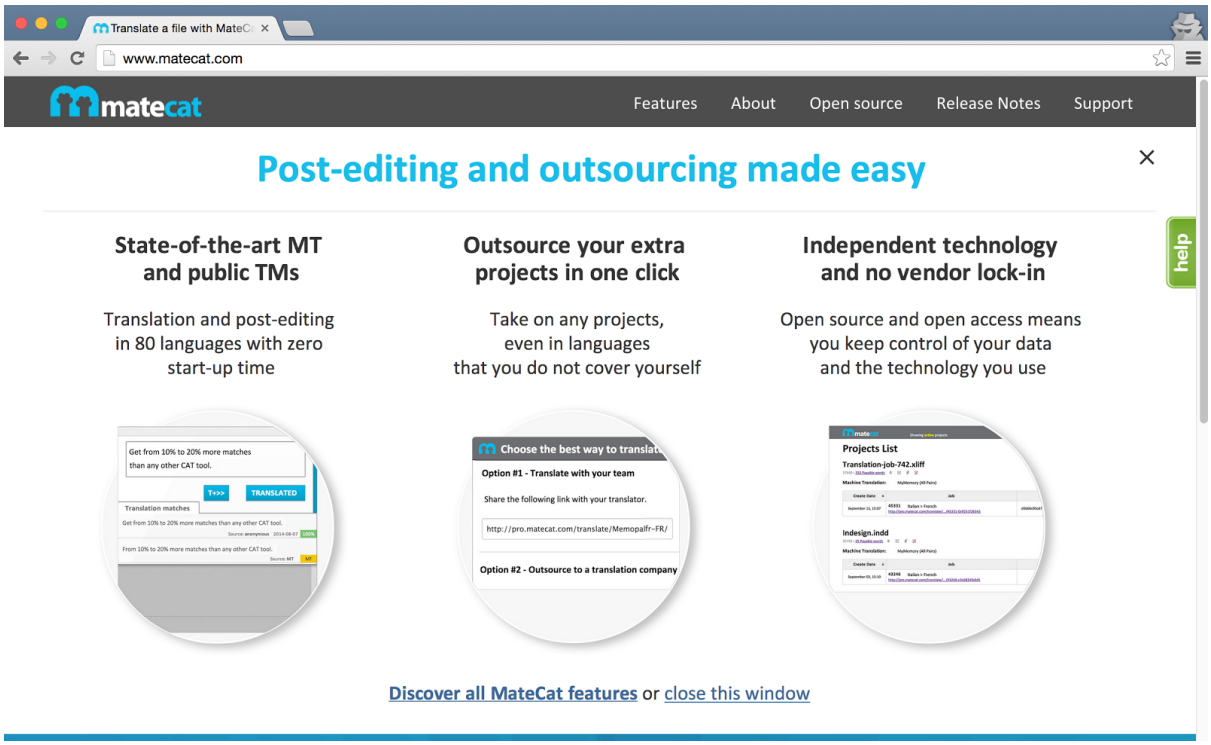


Fig. 1: Tagline and key features on the Matecat website.

The first activity revolved around the definition of the tagline and value proposition based on the feedback received from the early adopters and on the business model for the commercial product. We settled for the tagline "Post-editing and outsourcing made easy" which matches users' expectations in the first stage of the product release.
Another key feature that we stress on the MateCat website and tool is the fact that "MateCat gives you more matches than any other CAT tool" thanks to the integration of the best machine translation engines with the largest collaborative translation memory in the world.

The mascot

As part of the definition of the brand identity, we also created a mascot that accompanies the logo in most of our communication activities. It is a smart and friendly-looking blue cat designed to help users to engage with the tool and immediately recognise it. The mascot appears on a number of pages of the website and the tool, taking on different roles to either draw attention to a certain element or just to illustrate the content of the pages.
The cat is intended to become a sign of recognition for the MateCat tool itself.

Fig. 2: The MateCat mascot at Localization World in Vancouver (Oct. 2014)

3.2 Website

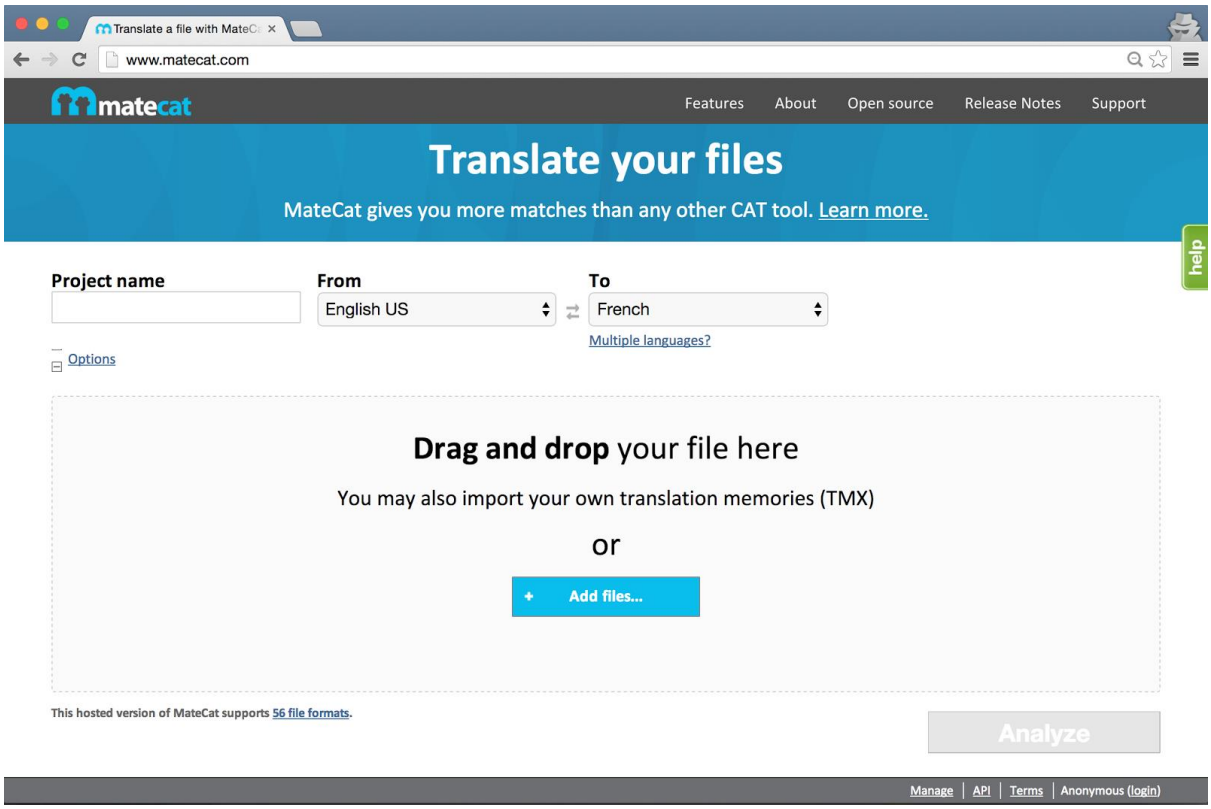


Fig. 3: MateCat website: http://www.matecat.com/

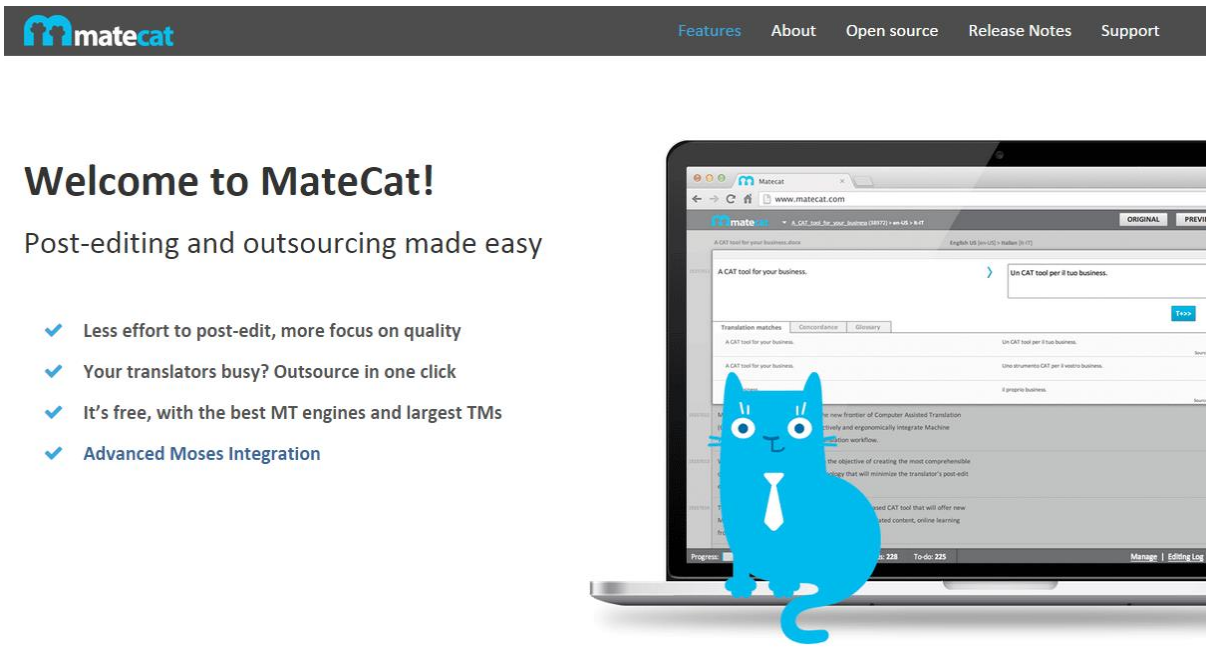


Fig. 4: MateCat website, *Features* page.

Starting in July 2014, we worked on the redesign of the website to make more explicit the commercial side of the technology we developed. The new website is now built around the MateCat tool and users landing on the home page http://www.matecat.com can start using the tool immediately. Browsing through the sections of the website, users are presented with all the key information on the product, the user manual and the scientific reports and publications. There is a specific section dedicated to the open source release which provides information on how to install the software and where to download the source code.

In less than a month since the new website went live (from October 16, 2014, until November 7th, 2014), we registered 3.071 visits by users who viewed altogether 17.121 pages and spent, on average, 2 minutes and 25 seconds on a single page. The most visited sections were: *Features*, *Open source*, *Support* and *About*.

3.3 User manual

With a constantly growing user base, the need of a user manual became ever more clear. We documented all the functionalities of MateCat in an ebook available for download from our website:

http://www.matecat.com/wp-content/uploads/2014/10/MateCatUserManualandInstallation Guide_v3.pdf

The same content is available on the website in the *Support* section which also provides a specific search function that returns multiple results when users search for a specific topic.

The content of the user manual and support section is constantly updated with information on the latest releases. As of today, the content is only available in English. However, a Chinese version of the user manual should be available at the beginning of 2015 thanks to the efforts of enthusiastic users who decided to voluntarily translate the manual into their native language.

3.4 Videos

We created three videos, two for promotional reasons and one as a training resource.

The promotional videos were created in collaboration with a professional agency and are focused on promoting the advantages of MateCat as compared to other CAT tools (content reuse, machine translation integration and outsourcing). They were released soon after the Translation World Cup (see below) as a marketing aid to promote MateCat to people who visited the Translation World Cup pages or took part in it. They are now published in the MateCat YouTube channel. The first one is a longer video just over 2 minutes long, while the second is much shorter and intended as a teaser for MateCat.

Note that the cat used in these videos has been updated and improved over the following months to get to a more appealing and effective mascot.

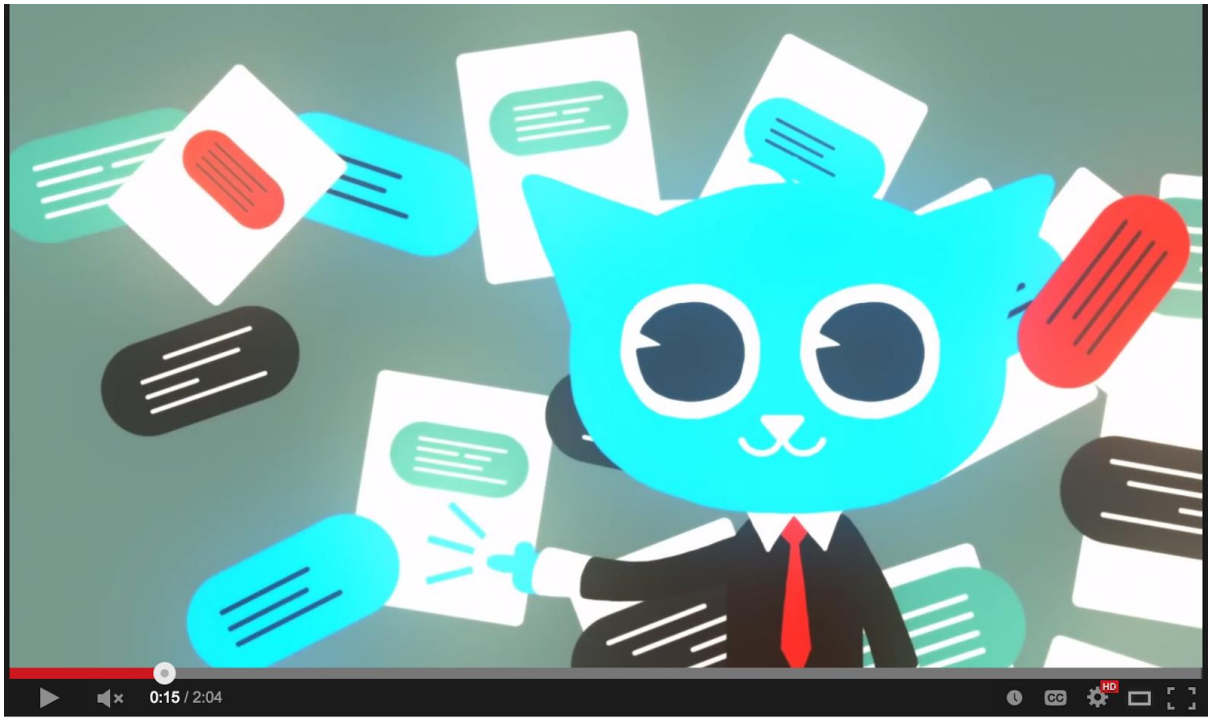


Fig. 5: Longer promotional video (http://youtu.be/4MXW35duI3c).

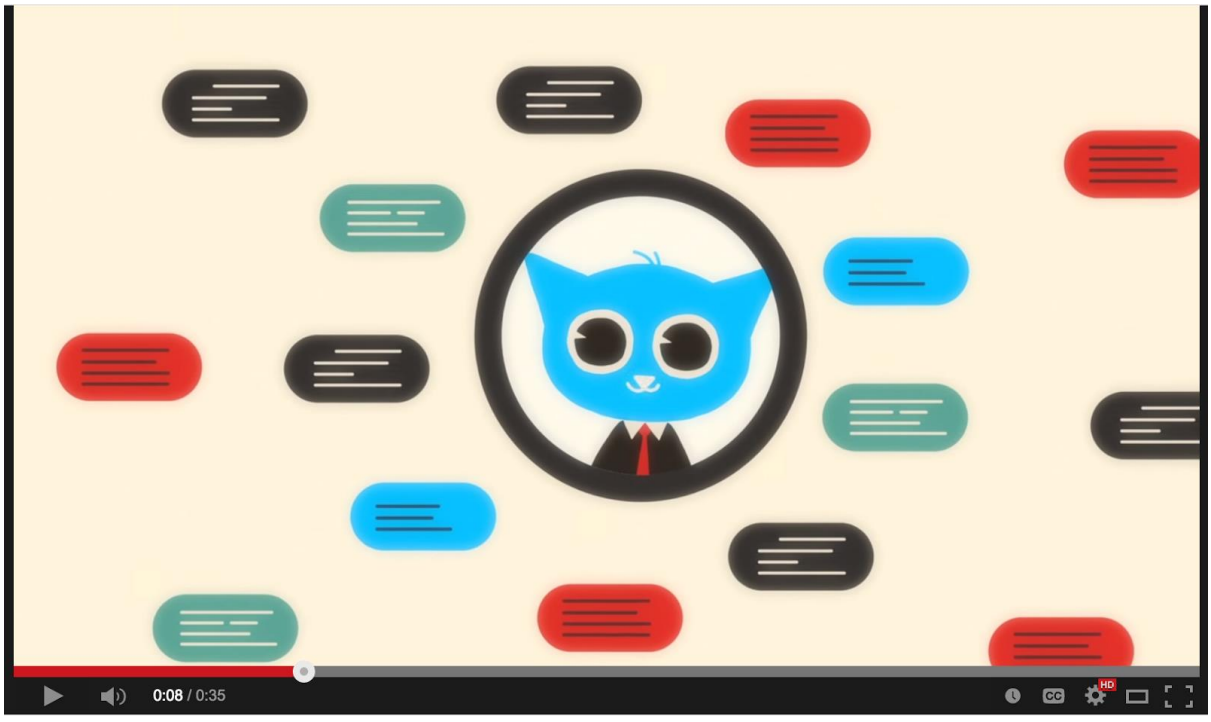


Fig. 6: Shorter promotional video (http://youtu.be/a0BRq70LA3M).

Finally, we created and published a video tutorial to show users how to get started translating with MateCat.

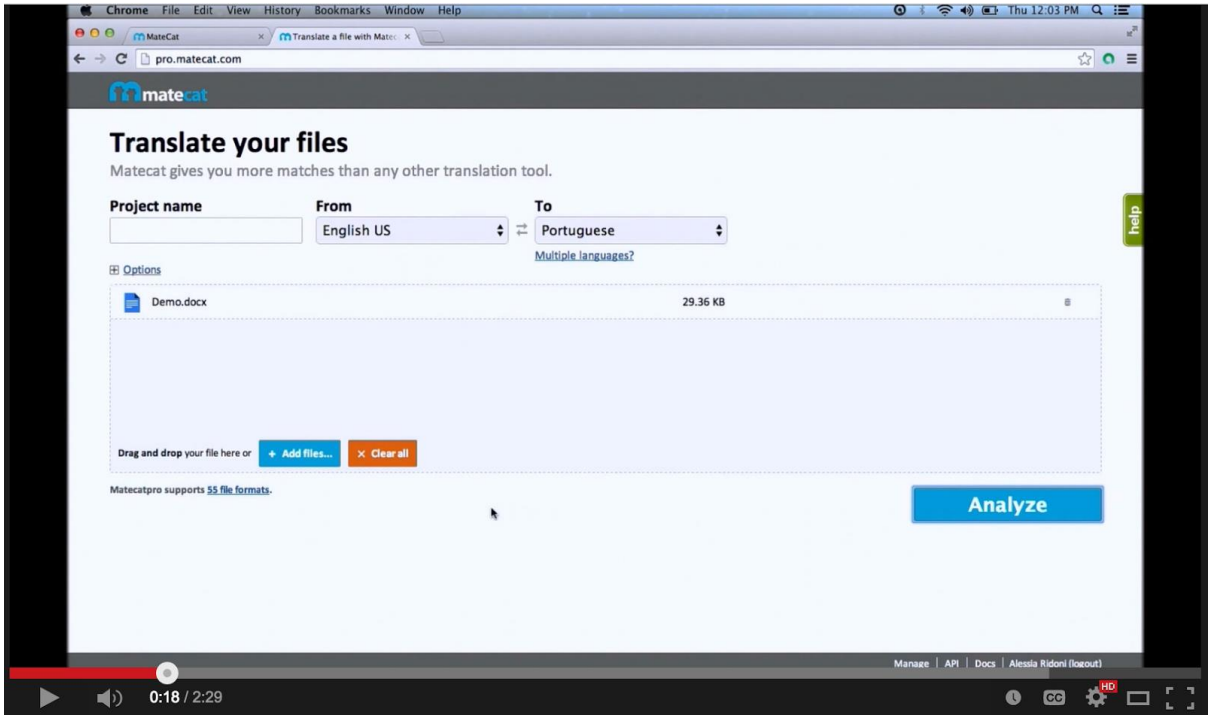


Fig. 7 : Video tutorial (http://youtu.be/LQSnBQWEZeU).

3.5 Translation World Cup

Translated srl organized a contest in collaboration with Proz.com, the largest translation community online, with the goal to promote the tool among professional translators. The contest was a translation from English to any target language of a short text about the FIFA World Cup hosted at http://worldcup.matecat.com; to win the contest, translators had to share their work on social networks and ask their friends and contacts to vote by sharing the translation. The translator who collected the most shares won the contest. Overall, 3,124 people took part in the contest and 1.429 people completed their translation and shared their work on social network.

Thanks to this campaign, word about MateCat spread virally on social networks and the contest generated 130,955 pageviews in little more than two weeks (June 12th - 27th, 2014). In addition, during the same period the MateCat website registered 6,916 pageviews. We counted 44,221 unique visitors for the Translation World Cup and 3,222 for the MateCat website, that is an increase of 401.15% over the previous month. Almost 40% of the users that visited the MateCat website came directly from the Translation World Cup website.

The contest also boosted the request for trying out MateCat which, at the time, was in private beta. In the weeks when we ran the contest, 797 users registered for the private beta, while the total amount of registered users for the previous two years was 1,422 people.

Fig. 8: Translators who won the contest for their language pair, were sent a medal.

3.6 Online presence

Together with the macro activities described above, we worked on building a presence for MateCat in other online websites and social networks. The main goal is to connect with key influencers and potential users from the translation industry.

Wikipedia

We published a page on Wikipedia[1] which provides a brief and complete description of the tool, including technical details, a description of the project and the members, enriched with references. It was created to consolidate the MateCat brand.

Twitter

As of November 7th, 2014, @MateCat[2] had posted 340 tweets and had 448 followers. Through the activity on Twitter we spread the word about MateCat and get in touch with some key people of the language industry.

Facebook

[1] http://en.wikipedia.org/wiki/MateCat
[2] https://twitter.com/matecat

MateCat created a closed group as well as a public page on Facebook. The private group was designed to connect and discuss with the beta testers of MateCat while the public page is a branding tool. As of November 7th, 2014, the private beta group had 188 members while the MateCat page, created on May 26th, 2014, has 449 fans.

Google+

A Google+ page was also created for MateCat to share all the relevant news and events about the tool. This page is aimed to complete the digital presence of MateCat on the most used social networks and has a role in the search engine optimization (SEO) strategy for the MateCat website.

Linkedin

MateCat has also a company page on LinkedIn to promote the software. It has been used only to find new members for the team and to promote MateCat events.

YouTube

We set up a channel on YouTube to upload the videos we created to promote MateCat. This channel is also intended to host a series of video tutorials to help translators and project managers to discover how MateCat can increase their productivity. The YouTube channel and the other social network accounts will continue to be used even after the conclusion of the EC-funded project to promote the adoption of the MateCat technology.

3.7 Newsletter

During the third year of the project, we published the third issue of the MateCat newsletter to report on the latest activities and results obtained.

MateCat Newsletter 03 covered the following issues (full articles available in Annex I):

- Translation world cup: 3,000 translators used MateCat for the first time
- MateCat is in private beta: 1,500 translators are already using it
- From North America to the Middle East: MateCat's research activities reach around the world
- Twice is nice: Summer internships at FBK

The third issue of the newsletter was sent to 1,835 readers registered via the MateCat website and was opened by 26.2% of the recipients (average for email newsletters in the same industry is 14.9%); over 5.9% of the readers clicked on the articles' links to visit the MateCat website (average for the industry is 1.9%).

We also circulated other messages (a dozen in total) to send further information about the release of MateCat and the MateCat Translation World Cup, which were opened by between 40 and 75 percent of the recipients.

3.8 MateCat Launch Party

Members of the consortium launched publicly MateCat during an event organized specifically for new and existing users in Vancouver on October 28th, 2014. Approximately 60 people attended the event during which Marcello Federico, Marco Trombetti and Alessandro Cattelan presented the goals of the project and gave a demo of the CAT tool and of the MT online learning functionality.

The most important part of the event was the networking after the presentations during which the members of the MateCat team gave one-to-one presentations of MateCat to most of the attendees, collecting feedback on the project and on specific user needs.

As a follow up to the event, Translated srl has been contacting all attendees to further discuss how MateCat can be integrated in the processes of the language service providers or companies they work with.

Fig. 9: The MateCat team at the MateCat launch party in Vancouver.

4 Conclusions and future exploitation plans

During the project and especially over the last year, the technology developed by the MateCat consortium has been used extensively by Translated srl for most of its translation projects. Indeed, Translated srl entirely replaced previous CAT tools it had been using with the Mate-

Cat tool. This process resulted in extensive feedback being collected from account managers, project managers and translators which, ultimately, influenced the development process and allowed the consortium to develop an enterprise level, open source CAT tool.

The software, and especially the MateCat tool, is now ready to be used and deployed by other language service providers and corporations. The hosted version available at www.matecat.com is comparable to most commercial CAT tools in terms of functionalities and file formats support, and is superior to most in terms of interoperability since it exposes APIs which allow to integrate MateCat in translation workflow management systems and to easily connect to custom MT servers.

After the conclusion of the European project, development and commercialization of MateCat will not stop and will, in fact, enter the most active phase with the promotion of the technology to language service providers and corporations.

The exploitation plan foresees the promotion of the technology as free and open source to facilitate its adoption by other companies. The interest shown during the events in Vancouver and the feedback collected seem to highlight the importance of keeping the software free and open source since it opens up opportunities for integration in other platforms, such as integrated development environments (IDE), translation mobile apps, translation workflow systems, etc.

Translated srl will continue developing and promoting the software using an innovative business model intended to disrupt the current translation technology sector: offering for free a technology that up to now costs thousands of dollars. Instead of charging the users for using the technology, Translated srl will sell its translation services inside the MateCat platform by creating the easiest tool for language service providers to outsource their translation projects. For this model, Translated srl will leverage previous technologies it developed, such as T-Rank, an algorithm which can automatically select the best available translators for any project, and its database of over 100,000 professional and vetted translators.

Every year, language service providers outsource to other language service providers translation services worth over 5 billion dollars. Positioning the MateCat tool as a free software to carry out those translations will allow MateCat to intercept part of the outsourced translations in that segment of the market.

Annex I – Newsletter 03

Translation World Cup: 3,000 translators used MateCat for the first time

The MateCat team is composed of people from different countries of the world and it was therefore natural as the recent World Cup approached, that our spirit of competition was rekindled. Thus the Translation World Cup was born, an international translation competition that we organised in collaboration with Proz to allow professionals in our industry to get to know MateCat.

More than three thousand people were involved in the Translation World Cup. To win you had to translate a story about the World Cup oracles and submit your work to a vote by social network users. This mechanism provided the translators and MateCat a great deal of visibility as the translations published generated more than 130,000 page views.

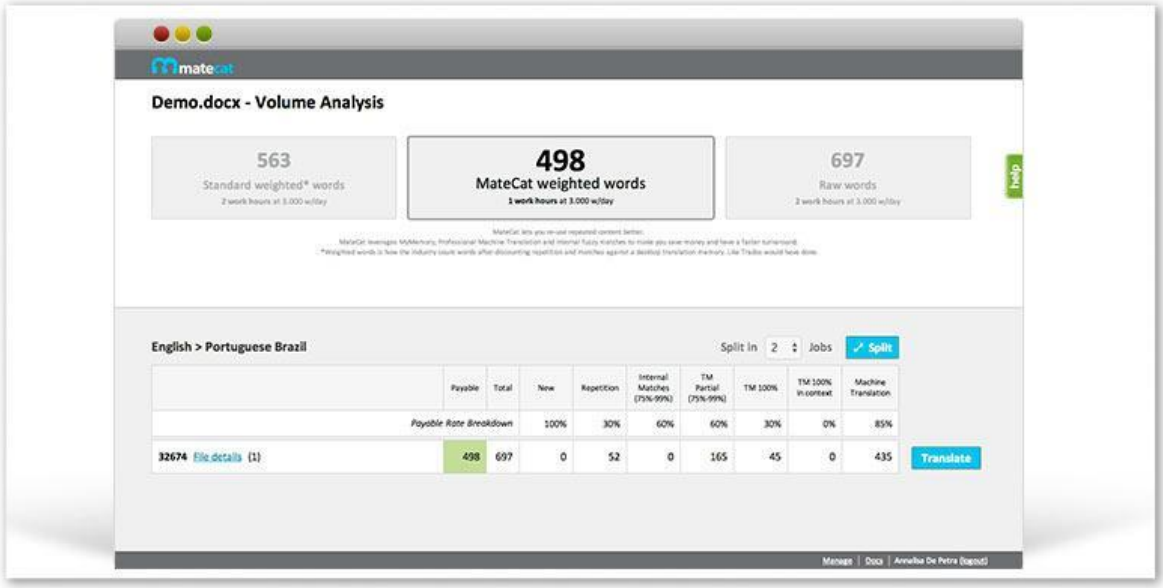
The translators were enthusiastic and we received a lot of feedback, most of it positive. The names and contact details of the winners were sent to the 49,000 agencies registered on Proz. Some of these, along with more than 900 professionals who participated in the contest, asked to be admitted to the upcoming private beta of MateCat.

This experience has made us aware that our software can also be used in ways other than that for which it was designed, i.e. supporting the work of translators and language service providers. For example, we can organise collaborative online translation contests and we are interested in exploring these directions. Furthermore, since the MateCat code is open source,

anyone can use it for creating platforms to translate and disseminate news and important documents quickly in cases such as humanitarian crises.

MateCat is in private beta: 1,500 translators are already using it

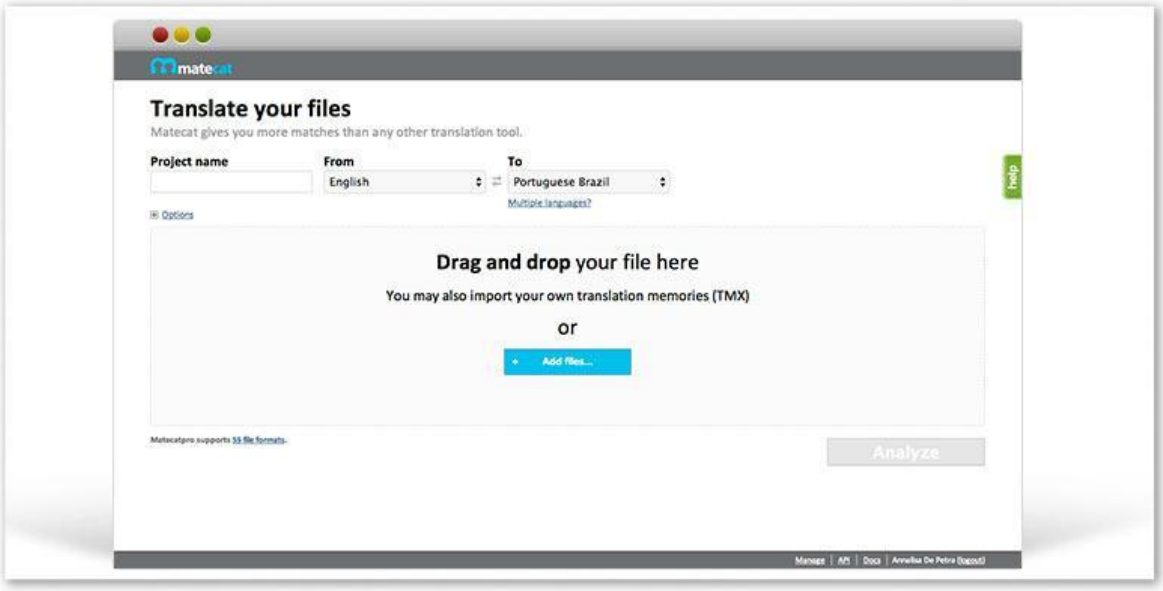
Around the middle of July, we launched the private beta of MateCat. For the first time people outside our team were able to use the CAT tool we're working on. We invited the first 1,500 translators who registered on the site to use MateCat and we will soon be adding more people (you can ask to participate here). We suggested that early users who want to be more involved in the project join a closed group on Facebook where they can share and discuss their feedback, any bugs and features not yet implemented.



The findings of the first MateCat users have been extremely positive. The translators relied on the software immediately and typically, after a first test, began using it regularly for their work, or at least for part of it, according to the statistics of use.

This activity is helping us to further improve the CAT tool. Every day we receive emails from translators who ask us for additional features, many of which are already in our road map, or from users who tell us about little bugs that need fixing.

One of the features the translators seem to appreciate most is not having to worry about managing translation memories. In fact, with MateCat, they don't need to continually update their TMs or do manual backups: memories are always available and updated in real-time anywhere there is a browser and an Internet connection.



Translators are also happy to use MateCat because it provides everything they need in a single work environment: document conversion, free tips from machine translation, possibility to split and merge documents and, above all, many more matches than any other translation software is able to provide them.

Our goal, with the private beta Facebook group, is to lay the foundations for a MateCat Users Group in which translators and language service providers can help each other and share useful resources for their work. At the same time, this is helping us to finalise the last details before the launch of the public beta.

From North America to the Middle East: MateCat's research activities reach around the world

Dissemination of the research conducted in MateCat is going on very successfully with a series of new scientific publications. More specifically, a paper on adaptive quality estimation was presented at the top international conference ACL 2014, the 52nd Annual Meeting of the Association for Computational Linguistics, which took place in Baltimore. MateCat researchers also participated in other international conferences (LREC – the Language Resources and Evaluation Conference – in Reykjavik, and COLING – the International Conference on Computational Linguistics – in Dublin) and workshops (the 9th Workshop on Statistical Machine Translation at ACL, the 4th International Workshop on Computational Terminology at COLING 2014, and the EACL Workshop on Humans and Computer-assisted Translation), with a total of nine scientific papers covering different topics in the field of machine translation and computer assisted translation.

MateCat is also advancing significantly in terms of dissemination in scientific journals. Two articles appeared respectively in the Prague Bulletin of Mathematical Linguistics and in Intelligenza Artificiale, the official journal of the Italian Association for Artificial Intelligence, while three more will appear soon in the journal Machine Translation.
This autumn we will further expand the reach of MateCat by presenting five scientific articles, two at EMNLP 2014 (the Conference on Empirical Methods in Natural Language Processing that will to be held in Qatar) and three at AMTA 2014 (the Eleventh Biennial Conference of the Association for Machine Translation in the Americas that will be held in Canada), where we will also propose a tutorial on Matecat, a workshop on Interactive and Adaptive Machine Translation in a joint effort with the CasmaCat project, and a technology showcase.

Twice is nice: Summer internships at FBK

For the second time since the start of the project, MateCat hosted resourceful students for a summer internship at FBK. Daniel, Marion and Lisa are all PhD students and they are spending 6 to 10 weeks in Trento.

Lisa Beinborn, a PhD student at the Technical University Darmstadt, has already completed her project on Large-Scale Distributed Language Models. She investigated the application of distributed and scalable database architectures to build servers for large-scale language models able to supply n-gram probabilities to multiple machine translation and speech translation engines. The project implied programming work and experimental investigation with very large corpora and with FBK's computing cluster.

Daniel Torregrosa, a PhD student at the University of Alicante, is still conducting his work on Error Tagging. By generating high-quality error tags over the target and the post-edition, in fact, it is possible to achieve different goals, such as providing visualization tools that help researchers identify the pitfalls of their methods, showing human post-editors the common errors that appear while using SMT, training statistical algorithms that automatically identify and fix common errors, or even feeding back the SMT system for improving future translations.

The main objective of Daniel's internship is to explore the current state of the art in error tagging, and improve it either by refitting the existing tools to exploit the new source information, or by developing new methods that use the new information with more success.

Marion Weller, a PhD student at the University of Stuttgart, is focusing on Learning from Human Post-Editions. More specifically, she is working on building an automatic post-editing system with correction rules derived from corrections on machine translation output provided by human post-editors. One major challenge in this project is that of sparse resources: as they learn from human post-editions, only a small amount of training data is available, particularly when comparing to the amount of parallel data typically used to train SMT systems. To maximally exploit the small data set, they combine different forms of information, ranging from specific (full surface forms) to generalized (e.g. tags), into model. Another focus in this project is the evaluation of the impact of different types of correction rules applied by the post-editing system.