PROJECT FINAL REPORT

Grant Agreement number: 287688

Project acronym: MATECAT

Project title: Machine Translation Enhanced Computer Assisted Translation

Funding Scheme: Collaborative project

Period covered: from November 1st 2011 to October 31st 2014

Name of the scientific representative of the project's co-ordinator[1], Title and Organisation: Marcello Federico, Head of research unit, FBK

Tel: +39 0461 314552

Fax: +39 0461314591

E-mail: federico@fbk.eu

Project website address: http://www.matecat.com/

[1] Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

4.1 Final publishable summary report

4.1.1 Executive Summary

Worldwide demand of translation services has dramatically accelerated in the last decades, as an effect of the market globalization and the growth of the Information Society. Human translation provides the best quality but is in general time consuming and expensive. Machine translation, on the other hand, is fast and cheap but far from publication quality. Computer assisted translation (CAT) tools are currently the dominant technology in the translation and localization market, and those including machine translation (MT) engines are on the increase. In fact, empirical studies conducted with professional translators have reported significant productivity and even quality gains when translators *post-edit* MT suggestions rather than translate from scratch.
The MateCat project leveraged the growing interest towards the integration of human and machine translation by developing new research directions for machine translation technology, aiming at enhancing the productivity but also the user experience of translators. The project team facing this ambitious goal included three academic research labs, particularly active on MT open source software, and a very dynamic and technology-driven translation company.
After 3 year, MateCat has succeeded in most of its challenges. It has created a new web-based CAT technology, initially developed to field test advanced MT functions but now also ready for the market[2]. MateCat developed adaptive MT technology that improves translation quality by learning almost instantly from human corrections. It developed more accurate quality estimation technology, to rate the utility of MT suggestions before they are shown to the user. It developed automatic support of terminology to supply less expert translators with technical terms specific to the documents they are working on.
By running extensive experiments with professional translators working with the MateCat Tool, the project also opened new directions in the way MT technology can be field tested and productivity as well as translation quality improvements can be reliably measured. Finally, all scientific and technological results of the MateCat project have been made available to the wide community, both under form of publications and open source software.

4.1.2 Summary Description

Worldwide demand of translation services has dramatically accelerated in the last decades, as an effect of the market globalization and the growth of the Information Society. Human translation provides the best quality but is in general time consuming and expensive. Machine translation, on the other hand, is fast and cheap but far from publication quality. Computer assisted translation (CAT) tools are currently the dominant technology in the translation and localization market, and those including machine translation (MT) engines are on the increase. In fact, empirical studies conducted with professional translators have reported significant productivity and even quality gains when translators *post-edit* MT suggestions rather than translate from scratch. However, MT research has been focusing for long time on providing ready to use translations rather than suggestions useful to human translators. Filling this gap was the goal of the MateCat project. In other words, MateCat addressed the scientific and technical challenge of *integrating human and machine translation* in ways to enhance the productivity but also the user experience of translators.

Computer-assisted translation (CAT) is a broad term, which identifies specific tools and software used by language professionals to increase the productivity and improve the quality of their work. This definition covers software as diverse as specialised text editors, spell checkers, grammar checkers, terminology databases, dictionaries, translation memories, electronic dictionaries, etc. Main components of any CAT tool are the editor where users read the source text and input their

[2] The MateCat Tool is accessible online at http://www.matecat.com

translations and the translation memory (TM), a database used to store translated texts. Texts are generally broken down into minimal units, called segments, which usually consist of sentences or paragraphs. When a translator opens one such segment, the CAT tool scans the database looking for pre-translated segments matching the source text. A list of matches is retrieved and ranked according to the similarity with the source text. The segments found in the TM which match the source text exactly, are called exact matches or 100% matches and usually do not require any intervention from the translator. Fuzzy matches are those that only partially match the source text and are assigned a different score expressing the percentage of similarity.

Machine translation is nowadays dominated by the so-called statistical approach, in which the translation process is expressed as a search problem that computes an optimal sequence of rule to apply. Translation rules are automatically extracted from a large parallel corpus and a probabilistic model over the translation rules that is build and optimised to best fit the data. According to the employed probabilistic model, the sequence of rules may generate linear or hierarchical structures. Progress in MT research has quickly found a way to the marketplace. Most prominently, perhaps, is the large-scale effort by Google to make translation for many language pairs available online, which uses standard statistical MT methods paired with massive computer clusters. Google's MT targets the problem of Web page translation, while commercial offerings of companies such as Language Weaver in the United States and Systran in Europe offer specialized statistical and hybrid systems to individual clients. Note that the academic research community, including the research groups participating in the MateCat project, is at the forefront of extending the state of the art in the field. Moreover, two of the research partners are co-developers of the Moses toolkit, which is nowadays the most popular open source statistical MT software in the world, and of popular open source tools such as Irstlm.

Most recent work in machine translation is motivated by the application scenario of fully automatic translation of text, typically triggered by a user who wants to understand the meaning of a foreign text. At the same time, the professional translation industry has started to incorporate new translation technology into their workflow. While the use of translation memories is widespread, the deployment of MT is more sporadic and has encountered more resistance by human translators.

MateCat aimed at improving the state of the art by investigating new research issues related to the integration of MT into the CAT working process and advancing statistical MT technology along three main directions:

- *Self-tuning MT*, i.e. to train MT for specific domains once before starting a translation project or periodically during its life;
- *User adaptive MT*, i.e. to quickly adapt MT from user feedback to instantly spot lexical errors and user preferences to avoid her correcting the same errors over and over;
- *Informative MT*, i.e. to supply more information to enhance users' productivity and work experience, such as quality estimation and terminology support.

MateCat aimed to develop a new web-based CAT tool integrating the novel MT functionalities implemented in the project. The tool offers all features of a state-of-the-art CAT tool -- e.g. support of file formats, translation memory, terminology database, concordancer -- support interoperability with different MT engines and TMs through open APIs, and offers logging functions in order to permit measuring user productivity. Progress in the project was cyclically measured with lab and field tests respectively addressing performance of single MT components and the impact of the new MT functions on user productivity. User productivity was measured according to two standard key performance indicators: (i) *post-edit effort*, roughly the percentage of the MT suggestions that have been corrected; and (ii) *time to edit*, namely the translation speed expressed either in seconds per word or words per hour. Field tests were run with the MateCat Tool, on real translation projects, and with professional translators. Lab tasks covered four translation directions targeting English from, respectively, German, French, Spanish and Italian. Translation project were taken from two well data supplied linguistic domains: *legal*, including documents from the European Commission, and *information technology*, including software documentation.

The final goals of the project were to significantly enhance user productivity, thanks to self-tuning, user-adaptive and informative MT. For this purpose, suitable evaluation protocols were defined in order to run experiments with the MateCat tool under contrastive post-editing conditions. As reference baseline a strong domain adapted MT system was assumed, trained with conventional methods on exactly the same data as the enhanced MT system before starting the translation project.

MateCat built on well-established open source and proprietary technology developed by the project partners, such as Moses and MyMemory, which had been already adopted worldwide among research and industry. Thanks to the leadership role of the project partners, the new technical advances pursued by MateCat permitted to further improve the scientific and technological leadership of Europe in the translation sector and to foster a widely accepted vision and roadmap about the application of statistical MT.

MateCat's results were released in open source and made available to a variety of organizations operating in the translation sector, including European SMEs offering translation services, translation offices of large public bodies and IT companies, and educational institutions. The involvement of a strong international User Group in the project as well as the pursued dissemination activities have fostered a closer dialogue and partnership between research and industry, improved understanding of user and market requirements, and triggered innovation and technology uptake within and outside the project.

Dissemination activities were conducted along the whole duration of the project. Partners regularly presented their work at established academic and industrial venues, such as workshops, conferences, journals, and magazines. Research partners of MateCat also participated in international evaluations measuring the competitiveness of the machine translation technology developed within the project against the international state of the art. All project results were released in open software and documented to encourage their use as well as the uptake of the technical challenges addressed in the project. A User Group was set-up including many user and industrial representatives who were kept informed about the project through a newsletter. A final event presenting the project results was organized at the end of the project, too.

4.1.3 Main S&T results/foregrounds

MateCat identified a series of scientific and technological challenges for the seamless integration of Machine Translation (MT) in the computer-aided translation (CAT) human translation workflow.

In this final report, we first outline the final version of the MateCat tool as resulting from the implementation spanning the three years of the project. Then, we summarize the main scientific results of the project, grouped according to the three research challenges described above: self-tuning MT, user-adaptive MT, and informative MT. Finally, we provide the general outcomes of the three field tests that were designed to measure the utility and usability of the developed MateCat Tool prototypes, as well as performance of the single MT components.

4.1.3.1 The MateCat Tool

The objective of MateCat is to improve the integration of MT and human translation within the CAT framework. The integration of suggestions from an MT engine as a complement to translation memory (TM) matches is motivated by recent studies which have shown that post-editing MT suggestions can substantially improve the productivity of professional translators. MT research in MateCat has converged into a brand-new CAT software, which is both an enterprise level translation workbench (currently used by several hundreds of professional translators) as well as an advanced research platform for integrating new MT functions, running post-editing experiments and measuring user productivity. The MateCat tool is a web-based CAT tool providing translators with a professional work environment, integrating TMs, terminology bases, concordancer, and MT. The MateCat tool combines features of the most advanced systems (either commercial, like the popular SDL Trados Workbench,[3] or free, like OmegaT[4]) with new functionalities. These include: i) an advanced API for the Moses toolkit,[5] customizable to languages and domains, ii) ease of use through a clean and intuitive web interface that enables the collaboration of multiple users on the same project, iii) concordancer, terminology databases and support for customizable quality estimation components and iv) advanced logging functionalities.

The tool is completely developed as open source software, is distributed under the LGPL open source license, and has been already successfully deployed for business, research and education. Probably today the MateCat tool represents the best available open source platform for investigating, integrating, and evaluating under realistic conditions the impact of new MT technology on human post-editing. Indeed, after the three-years development within the MateCat project, the MateCat tool has reached a high level of maturity and solidity, which makes it suitable even for professional use. In fact, Translated S.r.l., the commercial partner of the consortium, has switched its entire production to MateCat, translating over 35 million words with it. In the first few months after the switch, over 3,000 translators have started using the MateCat tool for professional work.

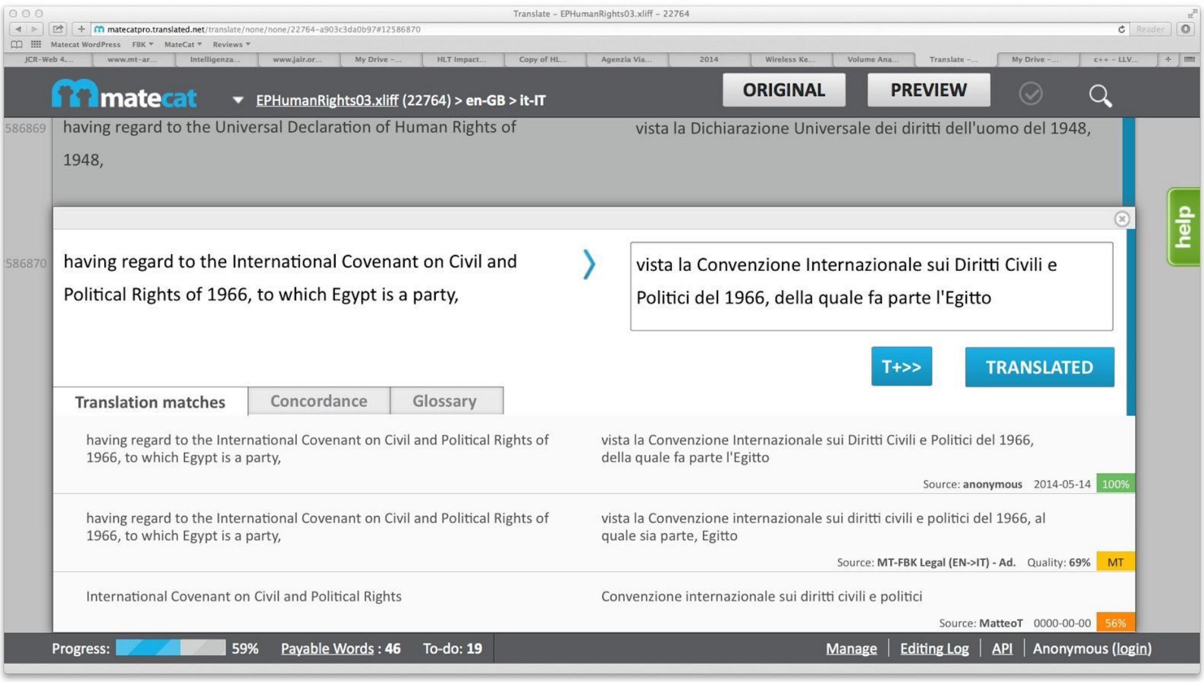

Figure 1: The MateCat tool editing page.

---

[3] http://www.translationzone.com/

[4] http://www.omegat.org/

[5] http://www.statmt.org/moses/

4.1.3.1.1 A Short Tour

The MateCat tool runs as a web-server accessible through Chrome, Firefox and Safari. The CAT web-server connects with other services via open APIs: the TM server MyMemory,[6] the commercial Google Translate (GT) MT server, and a list of Moses-based servers specified in a configuration file. While MyMemory's and GT's servers are always running and available, customized Moses servers have to be first installed and set-up. Communication with the Moses servers extends the GT API in order to support self-tuning, user-adaptive and informative MT functions. The natively supported document format of MateCat tool is XLIFF,[7] although its configuration file makes it possible to specify external file converters. The tool supports Unicode (UTF-8) encoding, including non Latin alphabets and right-to-left languages, and handles texts embedding mark-up tags.

The tool is intended both for individual translators and managers of translation projects involving one or more translators. A translation project starts by uploading one or more documents and specifying the desired translation direction. Then the user can optionally select a MT engine from an available list and/or a new or existing private TM in MyMemory, by specifying its private key. Notice that the public MyMemory TM and the GT MT services are assumed by default. The successive step is the volume analysis of the document, which reports statistics about the words to be actually translated based on the coverage provided by the TM. At this stage, long documents can be also split into smaller portions to be for instance assigned to different translators or translated at different times. The following step starts the actual translation process by opening the editing window. All source segments of the document and their corresponding target segments are arranged side-by-side on the screen. By selecting one segment, an editing pane opens (Figure 1) including an editable field that is initialized with the best available suggestion or with the last post-edit. Translation hints are shown right below together with their origin (MT or TM). Their ranking is based on the TM match score or the MT confidence score. MT hints with no confidence score are assigned a default score. Tag consistency is automatically checked during translation and warnings are possibly shown in the editing window. An interesting feature of the MateCat tool is that its URL page, which also includes the currently edited segment, uniquely identifies each translation project. This permits more users to simultaneously access and work on the same project. Moreover, to support simultaneous teamwork on the same project, translators can mark the status (draft, translated, approved, rejected) of each segment with a corresponding colour (see Figure 1, right blue bar). The user interface is enriched with a search-and-replace function, a progress report at the bottom of the page, and several shortcut commands for the skilled users. Finally, the tool embeds a concordance tool to search for terms in the TM, and a glossary where each user can upload, query and update her terminology base. Users with a Google account can access a project management page which permits them to manage all their projects, including storage, deletion, and access to the editing page.

The tool supports Moses-based servers able to provide an enhanced CAT-MT communication. In particular, the GT API is augmented with feedback information provided to the MT engine every time a segment is post-edited as well as enriched MT output, including confidence scores, word lattices, etc. The developed MT server supports multi-threading to serve multiple translators, properly handles text segments including tags, and instantly adapts from the post-edits performed by each user.

During post-editing the tool collects timing information for each segment, which is updated every time the segment is opened and closed. Moreover, for each segment, information is collected about the generated suggestions and the one that has actually been post-edited. This information is accessible at any time through a link in the Editing Page, named Editing Log. The Editing Log page

[6] http://mymemory.translated.net

[7] http://docs.oasis-open.org/xliff/v1.2/os/xliff-core.html

(Figure 2) shows a summary of the overall editing performed so far on the project, such as the average translation speed and post-editing effort and the percentage of top suggestions coming from MT or the TM. Moreover, for each segment, sorted from the slowest to the fastest in terms of translation speed, detailed statistics about the performed edit operations are reported. This information, with even more details, can be also downloaded as a CSV file to perform a more detailed post-editing analysis. While the information shown in the Editing Log page is very useful to monitor progress of a translation project in real time, the CSV file is a fundamental source of information for detailed productivity analyses once the project is ended.

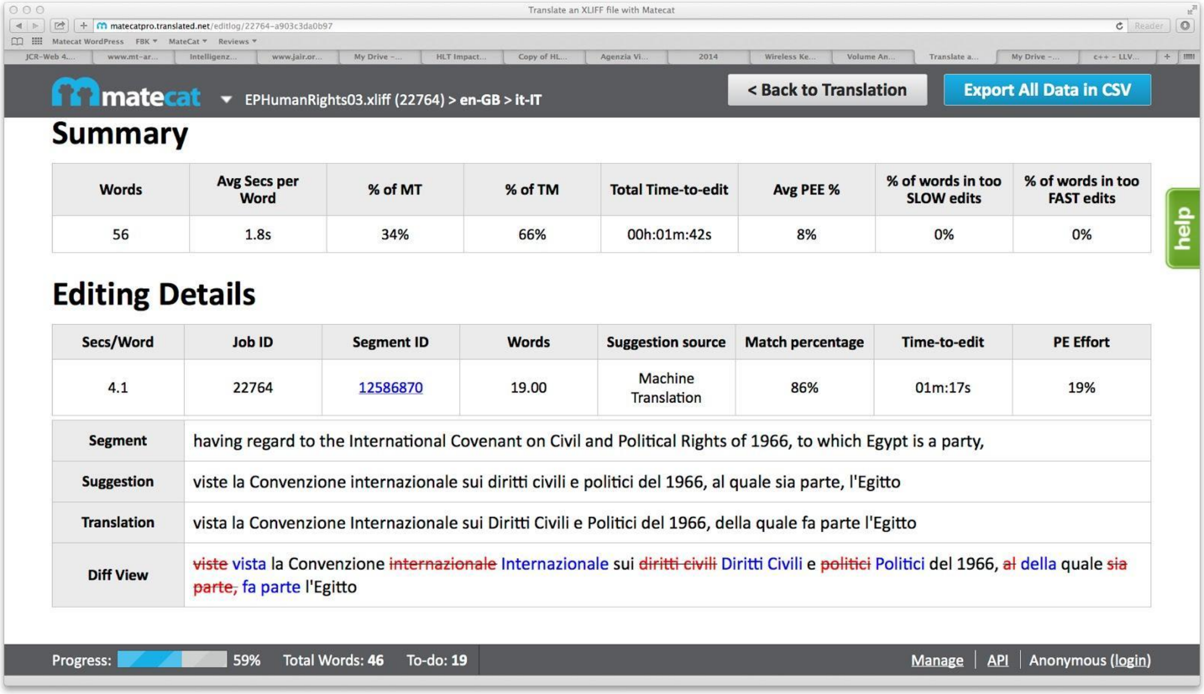

Figure 2: The MateCat Tool editing log page.

4.1.3.1.2 Applications

The MateCat tool has been exploited by the MateCat project to investigate new MT functions and to evaluate them in a real professional setting, in which translators have at disposal all the sources of information they are used to work with. Moreover, taking advantage of its flexibility and ease of use, the tool has been recently exploited for data collection and education purposes (a course on CAT technology for students in translation studies). An initial version of the tool has also been leveraged by the Casmacat project[8] to create a workbench, particularly suitable for investigating advanced interaction modalities such as interactive MT, eye tracking and handwritten input. Currently, the tool is employed by Translated for their internal translation projects and is being tested by several international companies, both language service providers and IT companies. This has made possible to collect continuous feedback from hundreds of translators, which besides helping us to improve the robustness of the tool is also influencing the way new MT functions will be integrated to supply the best help to the final user.

[8] http://www.casmacat.eu

4.1.3.2 Self-tuning MT

By *Self-tuning MT* it is meant the set of processes for adapting a general purpose MT system integrated in a CAT framework to a translation project before the translators start their job. The MT server exploits knowledge about the documents to be translated (topic, genre, etc.), performs automatic data acquisition and selection from existing TMs, parallel or comparable corpora. In the following, an overview on how self-tuning MT has been implemented in MateCat is provided.

4.1.3.2.1 Introduction

The adaptation involved in a self-tuning MT can be distinguished in two types:

- *Domain adaptation*: This type of adaptation encompasses all activity aimed at adapting a general purpose SMT system to a particular domain; as case studies, we considered legal, information technology and TED talks.
- *Project adaptation*: This adaptation is performed iteratively during the lifetime of a translation project. Typically, after a day of work by a human translator, knowledge about the newly translated text is injected into the SMT systems so that they will propose improved translations the next day. This procedure is continued until the end of the translation project.

Baseline systems were first developed in order to have references for the experimental assessment of the adaptation schemes explored during the project. The various approaches were carefully compared in joint efforts by FBK and LIUM. Such activity is summarized in the following.

4.1.3.2.2 Adaptation methods

Several techniques were developed during the MateCat project to achieve efficient domain and project adaptation. In particular, data selection has been exploited to train specialized translation and language models, which can be combined with generic models by means of fill-up and interpolation techniques, respectively. Moreover, continuous space methods applied to language modelling have been investigated to take advantage of their ability in handling efficiently very long contexts.

Data selection - It has been believed for a long time that just adding more training data always improves performance of a statistical model, e.g. a n-gram LM. However, in general this is true only if the new data is relevant enough to the task at hand, a condition that is rarely satisfied. The main idea of data selection is to use just a subset of the generic corpus that is most relevant for the task of interest.

In the framework of the MateCat project, data selection is used twice: first, to adapt a generic system to a specific domain, i.e. legal, IT or TED, before the human translator starts working; and second, to integrate it with the daily translations made by the users (project adaptation). In both cases, we applied a recently proposed cross-entropy based algorithm for the selection of the most appropriate monolingual data, and its extension to bilingual texts. Within MateCat, we implemented and made publicly available the techniques through the XenC and IRSTLM tools. The method assumes the availability of a seed corpus, that in our case is representative of the specific domain or project and is named foreground corpus, and of a large generic corpus, from where to extract task-relevant sentences, called background corpus. The first step consists in creating two LMs, one foreground and one background, used to compute a single score for each sentence of the background corpus. The score is the difference between the cross-entropy calculated with the foreground LM and the cross-entropy calculated with the background LM. The background sentences are then ordered according to their scores: the rationale behind this choice is that the larger the score, the closer the specific sentence is to the foreground corpus and - at the same time - the more distant it is from the distributions of the background corpus. The selection of useful sentences from the background corpus is finally achieved by determining the best splitting point of its sorted version. The estimation of the optimal split is performed by minimizing the perplexity of a development set on LMs estimated on growing percentages (10%,20%,...,100%) of the sorted corpus. Usually, the perplexity decreases when less, but more appropriate data is used, typically reaching a minimum at around 20% of the sorted generic data. As a side effect, the models become considerably smaller, which is also an important aspect when deploying MT systems in real applications.

As it is easy to image, the seed corpus plays a fundamental role in the process of data selection. The choice of the seed corpus depends on the adaptation task. For domain adaptation, a domain specific corpus was employed in our experiments. For project adaptation, several options were available. After one day of work on a project, to perform data selection on a bilingual seed the portion of the source text processed and the corresponding human translation – produced either by post-editing or direct translation – can be used. On the other hand, the source text of the whole translation project is usually available at the beginning of the process; then, it could be used to select project-specific data by a source-side-only but larger seed. This has turned out to be better in some of our experiments.

Fill-up - Data selection is a very effective method to adapt the translation model on the most relevant data. However, by discarding some of the available resources, we take the risk to miss some translations, which are not present in the selected data. To avoid this, we complete the translation model with the fill-up technique. In practice, the fill-up technique merges the generic background phrase table with the specific foreground phrase table by adding only phrase pairs that do not appear in the foreground table.

The entries of the filled-up model correspond to the union of the two phrase tables, while the scores are taken from the more reliable source whenever possible. To keep track of a phrase pair's provenance, a binary feature is added that fires if the phrase pair comes from the background table. The weight assigned to this feature acts as a scaling factor for the scores from the generic phrase table. The fill-up technique can be also applied to the lexicalized reordering model.

We chose the fill-up technique because it performs as good as other popular adaptation techniques but generates models that are more compact and easier to tune. Actually, we applied an even simplified version of the fill-up method, called back-off, in which the indicator feature is omitted in the filled-up phrase table as well: this means that no additional weight (that of the binary feature) has to be estimated when the combined phrase table is used, while in practice no significant performance degradation is observed.

LM combination - As concerns the LM adaptation, we employed the mixture of LMs since it is a well-established and good-performing method. The mixture model can be used to combine one or more background LMs with a foreground LM representing new features of the language we want to include. The technique consists of the convex combination of the LMs; the mixture weights are estimated on the training data of the foreground LM by applying a cross-validation scheme that simulates the occurrence of new n-grams. The method is available in the IRSTLM toolkit. Alternatively several LMs can be combined in the log-linear framework, their weights being optimized by MERT. This was used in some of our experiments to combine foreground and background models.

Continuous space methods - In the last years, the interest in using neural networks in SMT is continuously increasing. They are mainly used for improved language modelling, but there is also some recent research to use them for the translation model. Continuous space language models (CSLMs), also called neural network LMs, have shown to achieve significant improvements in many different applications.

The usefulness of continuous space methods has been explored in the framework of the MateCat project as well, by continuing the development of the CSLM toolkit. In particular, a new version with many additional features was released on March 26 2014: simplified configuration, support of deep

architectures, improved training algorithms, and faster training on GPU. The CSLM has been also integrated directly into Moses by performing n-best list rescoring of Moses' internal n-best list. This enables its use in the MateCat tool without any impact on the whole processing pipeline. This integration was successfully tested in the third MateCat field test.
The CSLM was applied to the English-German (lab tests) and English-French (lab and field test over 5 days) language pairs. In both cases, it achieved significant improvements of the translation quality. Two developments were important to achieve these results: deep neural network architectures and very long contexts (up to 30 words!).

4.1.3.3 User-adaptive MT

The goal of user-adaptive MT is to let machine translation take as soon and as much as possible advantage of user feedback, in order to learn from corrections and to hence avoid repeating the same mistakes in future sentences. The pair of input sentence and post-edit is a valuable feedback to improve the quality of next suggestions. How to exploit it for improving the SMT system was one of the main research objectives of MateCat.
In online MT adaptation specific issues have to be addressed, which distinguish it from the domain/project adaptation task. First of all, MT should adapt very quickly, because the time between two consecutive requests is usually short and correcting the same error again would negatively impacts on the translator. Other crucial matters are which information to extract from the user feedback, how to extract it, and how to exploit it to update the MT system. Finally, another problem is how to modify the system from very limited information that is the correction/translation of a single sentence. The work on user-adaptive MT focused on three aspects related to the exploitation of the user feedback, namely (i) the extraction of information from the post-edits (Word alignment section), (ii) the modification of the statistical models to embed such information (Model adaptation section), and (iii) the tuning of the system to better fit the user preferences (Tuning section).
In the following, an outline of the online adaptation framework is provided and the investigated approaches are briefly described.

4.1.3.3.1 Online adaptation framework

In the CAT workflow, source documents are split into chunks, typically corresponding to sentences and called segments that are in general translated sequentially. When the translator opens a segment, the CAT tool tries to propose possible translation suggestions, originating from the TM and/or from a MT engine. Depending on the quality of the suggestions, the translator decides whether to post-edit one of them or to translate the source segment from scratch.
Completed segments represent indeed a valuable source of knowledge, which is in fact readily stored in the TM for future use. In fact, given the compositional nature of SMT, it comes natural to believe that also text fragments observed in the post-edited or translated segments can be potentially useful to improve SMT output of future segments. From a machine learning perspective, the CAT scenario perfectly fits the online learning paradigm, which assumes that every time a prediction is made, the correct target value of the input is discovered right after and used to improve future predictions.
We conveniently transpose the online learning concept to our CAT scenario as follows. We assume an initial MT engine M_1 and a source document $x_1,...,x_D$. For t=1,...,D the following steps are performed:

1. The source segment x_t is received
2. A translation y_t is computed with M_t
3. A post-edited translation y'_t is received
4. A new M_{t+1} is created by adapting M^t with features extracted from (x_t, y'_t)

Next sections report on different aspects of the online procedure to update the MT engine: (i) the extraction of information from the user feedback, (ii) the modification of the MT models to embed such information, and (iii) the tuning of the system to best fit the user preferences. These three operations eventually happen in Step 4 of the online procedure.

4.1.3.3.2 Word alignment

Our investigation mainly focused on the methods to quickly build a highly precise word alignment from a source sentence and its translation. This is an important and challenging problem in the online scenario, in which the user interacts with the system and expects that it learns from the previous corrections and does not repeat the same errors again and again. In particular, we were interested in the cases where the given sentence pairs contain new words, for which no prior information is available. Indeed, they are one of the most important sources of errors in model updating. We considered the approaches briefly described below.

Constrained-based search - The first considered approach exploits the constrained search technique which aims at optimizing the coverage of both source and target sentences given a set of translation options. The search produces exactly one phrase segmentation and alignment, and allows gaps such that some source and target words may be uncovered. Unambiguous gaps (i.e. one on the source and one on the target side) can then be aligned. It differs in this respect from forced decoding, which produces an alignment only when the target is fully reachable with the given models.
From such phrase alignment, three types of phrase pairs can be collected: (i) new phrase pairs by aligning unambiguous gaps; (ii) known phrase pairs already present in the given model; (iii) full phrase pairs consisting of the complete source sentence and its user translation. Since preliminary experiments showed that using cached phrases consisting only of function words had a negative impact on translation quality, we restricted the phrase extraction to phrases that contain at least one content word. We found that in our experimental data cached phrases tended to be between one and three words in length, so we limited the length of new and known phrases to four words to speed up the annotation. The full phrase pair, which can have any length, is added to mimic the behaviour of a TM.

Standalone word aligners - We investigated on four word-aligners, namely Berkeley Aligner, Fast-align, Mgiza++, Bitext-tokaligner. We analysed their pros and cons in the online MT adaptation scenario, and compared their performance in aligning sentence pairs with specific attention to unknown terms. We evaluated them in terms of (i) their capability to align the whole sentence pairs (using Alignment Error Rate, AER), (ii) to align the out-of-vocabulary (OOV) words, i.e., words not present in the training data used to train the respective word aligner (using OOV-AER), and (iii) their impact on online adaptation process and on the translation quality (using BLEU and TER). We investigated three language pairs (English–Italian, English–French, English–Spanish) in three domains: JRC (i.e. the corpus of European Union Law), Canadian Hansard, and EuroParl, respectively. We found that (i) all aligners, but Bitext-tokaligner, have similar performance in aligning the whole sentence pair; (ii) Mgiza++ outperforms the other aligners in terms of OOV-AER on the English–Italian and English–Spanish data up to OOV noise levels of 32% and 16%, respectively, while fast-align performs better on the English–French data; (iii) when the OOV rate is very high, fast-align always performs better; and (iv) differences among aligners do not results in consistent and reliable difference in the performance of the corresponding online-adapted systems. For such investigation, we exploited public data sets and a new benchmark of 200 human-checked word-alignment sentence pairs in the English-Italian JRC-Acquis corpus.
Bitext-tokaligner, which relies on a raw automatic translation as a pivot alignment between source text and post-edit, was developed from scratch. Mgiza++ was enhanced with the facility of aligning new sentence pair on the fly without the need of reloading its models; this online version is hence well suited for the exploitation in the real-time CAT scenario.
We also developed an additional module aiming to refine the word alignment of OOV words. This module applies to any available word aligner. First, a symmetrized word alignment is produced using the chosen word alignment method. Then, a set of phrase pairs, possibly including OOV words, is extracted using the standard software distributed with the Moses toolkit. Finally, by exploiting the overlaps among phrase pairs a new word alignment is created.
The experimentation performed shows that the refinement module slightly but consistently improves the overall performance in terms of alignment quality, especially for high OOV rates.

Pivot-based approach - We also considered the pivot-based approach to align a source text and its corresponding post-edit. This approach relies on the availability of the translation hypothesis of the source produced by an MT engine, like Moses. First, the alignment links between the source sentence and the automatic translation are gathered from the MT system. Then, the links between the automatic translation and its post-edit are calculated by means of Translation Edit Rate (TER). Finally, the source-to-post-edit alignment links are deduced by an alignment combination based on both alignment sets computed before.

4.1.3.3.3 Model adaptation

A challenging research issue raised by the integration of MT in the human translation work is how to dynamically adapt phrase-based SMT from user post-editing. In MateCat we proposed and investigated three main techniques for such purpose, namely a cache-based adaptation approach, a discriminative re-ranking and a method for adapting the statistical models on-the-fly by exploiting a suffix array-based implementation of phrase tables. Other MateCat activities related to the online model adaptation issue regard the online tuning of SMT systems, the optimization of hyper parameters of online learning algorithms and the online learning from multiple post-editors.

Cache-based approach - The main idea behind cache-based models is to mix a large global (static) model with a small local (dynamic) model estimated from recent items observed in the history of the input stream. In our local models, both phrase-pairs and target n-grams used by the translator are cached. The local translation and language models are built to reward the phrase pairs and the n-grams found in post-edited translations according to the following policy: “the more recent, the more rewarded”. The cache-based models are implemented as additional features of the log-linear SMT model. During decoding, translation alternatives are searched both in the global static and in the local dynamic phrase tables. For each pair of source and post-edit segments, phrase pairs are extracted by means of any method presented in section on word alignments, and simultaneously added to the local translation model.

Discriminative re-ranking
The discriminative method is based on a structured perceptron to refine a feature-based re-ranking module applied to the n-best translations generated by the SMT system. More in detail, the learner is a structured perceptron and lexicalized sparse features are used, defined by two feature templates: First, all phrase pairs found by the decoder (for system translations) or by the constrained search (for the user translation) are used as features. Second, we use features defined by target-side n-grams in the user translation. Our features are not indicator functions, but use the number of source words (for the first type of features) and the number of words in target-side n-grams (for the second type of features) as values. Given a feature representation $f(x,y)$ for a source-target pair (x,y), and a corresponding weight vector w, the perceptron update on a training example (x_t,y_t) where the prediction $\hat{y}=argmax_y \langle w,f(x_t,y)\rangle$ does not match the target y_t is defined as:

$$w=w+f(x_t,y_t)-f(x_t,y')$$

The constrained search allows us to perform updates even on translations that are not reachable by the decoder. For the purpose of discriminative training, in our set-up all references are reachable since we can extract features from them and assign them model scores.

Suffix array-based phrase tables - One of the major drawbacks of conventional MT system design in any interactive scenario is the static nature of the systems: the underlying models are expensive to train and update. Phrase tables in particular are a crucial resource that we would like to be able to update interactively, so that (very) recent translations can be used as soon as possible to improve translation quality. Ideally we would like to be able to make use of approved translations/post-edits in the same post-editing session in which they are produced. However, the conventional method of phrase table construction consists of a batch processing pipeline involving (1) word-alignment of the parallel data; (2) phrase pair extraction; (3) phrase pair scoring based on global statistics over the collection of phrase pairs (e.g., smoothed conditional phrase-level translation probabilities in both translation directions); and finally (4) storing them in a file or other data structure.

An alternative to this pipeline is to collect and score phrase table entries on the fly by sampling available information sources directly rather than computing a large range of conceivable requests.

Suffix arrays are a space-efficient way of indexing large text corpora with reasonably fast retrieval times. In the context of natural text corpora, a suffix array is an array of all word positions in the corpus, sorted in lexicographic order of the word sequence starting at the respective position (hence the name suffix array). Thus, the computational cost for building a suffix array for an initial corpus is $O(n_1 \log n_1)$, where n_1 is the number of tokens in the corpus; the computational cost for finding all occurrences of a given text span is $O(\log n_1)$; and the cost of adding n_2 tokens to an indexed corpus is $O((n_2 \log n_2) + n_1)$, i.e. the cost of indexing the additional material and the linear cost of merge-sorting the two indices.

Even though suffix arrays have been used in hierarchical phrase-based translation for years, adoption into the Moses toolkit has been slow, due to concerns about translation speed and translation quality: speed, because sampling is often perceived as being slow, and translation quality, because certain global information that is used to compute phrase table scores in the conventional setting is not available when sampling, e.g. global count information used for Good-Turing or modified Kneser-Ney smoothing of phrase-level conditional translation probabilities.

This gap has now been closed thanks to MateCat: in fact, our implementation of memory-mapped suffix array-based phrase tables, which is fully integrated into the current stable master branch of the official Moses repository, successfully address all of these concerns.

Tuning and hyper parameter optimization- Moses, the SMT engine employed in MateCat Tool, relies on a weighted linear combination of several features, whose weights are tuned to fit the application domain and genre. In a CAT scenario, feature weights can be also updated online exploiting user feedback. We investigated the use of the Margin Infused Relaxed Algorithm (MIRA) for handling such a problem. On the other side, any algorithm, and therefore also the MIRA-based online learning method we proposed, itself brings some issues, related to the tuning of hyper parameters involved in the process. In fact, at first we optimized hyper parameters by means of the Simplex algorithm once, and the same values were then re-used for any possible number of iterations of the online learning. Since disregarding the dependence between the number of iterations and the hyper parameters does not seem a good assumption, successively we deeply investigated the issue in order to select (i) the optimal hyper parameters which control the rate of learning, (ii) a stopping criteria for online learning, and (iii) the optimal number of iterations of online learning being performed by the system.

Multi-user online adaptation - We also addressed the problem of adapting in a CAT framework a single SMT system to multiple post-editions, i.e. to an incoming stream of feedback from different translators. In such a situation, standard online learning methods can lead to incoherent translations by the SMT system. As a solution we proposed to adopt a multi-task learning scheme, which relies on the correlation amongst the translators computed using prior knowledge; the online learner is then constrained to take into account the relatedness amongst the translators. Whenever not enough information about the correlation amongst the translators is available, our experimental outcomes suggest to use multi-task learning with half-updates, which is a good generalization of the interaction between the translators.

We also compared the multi-task approach to that of independent systems where each translator has been allotted an online learning SMT system; evidently, multi-task also fared better against this system setup.

4.1.3.4 Informative-MT

In addition to improvements to the overall quality of MT output in general, MateCat aims at assisting translators in their task by providing them with information beyond the raw MT output. In the following you can find a summary of the research activities of the project devoted to that issue, grouped in three tasks: terminology help, sentence- and word-level confidence estimation, and enriched MT output, that is the actual integration of our findings into the MateCat tool.

4.1.3.4.1 Terminology Help

Terminology management is a crucial aspect of the work of professional translators. Here, we present the results of our work on supporting translators in this respect in two ways: first, by automatically acquiring bilingual terms from existing data, either the results of the post-editors' own (or their colleagues') work, or from monolingual data; and second, by using the acquired terms to improve the performance of an MT system at the back-end of the MateCat tool.

Bilingual Term Extraction - We designed a framework for extracting bilingual terms from a post-edited corpus and using them to enhance the performance of an MT system embedded in a collaborative CAT environment, where a large translation project is split across different translators, and where each translator post-edits a limited amount of sentences per day. Our approach takes advantage of such post-edited data to gather bilingual terms specific to the domain. The parallel data produced each day is then used to continuously improve a generic MT system by injecting the bilingual terms into the MT system, and re-optimising the system's parameters on this specific data.
Bilingual term extraction is performed in two steps. First, a keyword extractor processes the source and the target sides of the corpus in order to identify the most relevant terms in each language. Taking advantage of the parallelism of the data, each monolingual term in the source language is then paired with a term in the target language. We compare different techniques to perform this step and show that simple approaches based on word alignment and term translation are more robust and more efficient than a reference state-of-the-art method, which even relies on supervised classification. The impact of the acquired terminology on translation quality has been investigated embedding the bilingual terms into the MT system. Differently from the previous approaches that append the bilingual terms at the end of the training data or the phrase table, in the MateCat project we test, for the first time, the capability of the cache-based model to correctly manage the terms. The cache-based model makes it possible to add bilingual terms into a running MT system, without the need to stop and restart it. We experimentally compared this approach to alternatives involving at a different extent the context of acquired terms. Overall, our results suggest that:

1. An SMT model enriched with the identified bilingual terms substantially improves translation quality in terms of BLEU score over a generic MT system. Incrementally tuned systems always outperformed those whose weights were left fixed after the initial pre-deployment parameter tuning.
2. Strategies to integrate terminology also need to consider the context of a translated term. Hard translation pegging in its straightforward implementation forces a particular translation regardless of context. For proper lexical choice, the cache-based model offers a better integration of the acquired terms into the final translation.

Bilingual Term Extraction from Monolingual Documents - We also addressed the problem of automatically identifying terms in a source language document, retrieving translations from monolingual Wikipedia articles that are linked across languages, and embedding them into the MT system.

This work builds and improves on the Wiki Machine[9], a tool for identifying monolingual terminology in a text, disambiguating ambiguous terms and linking them to the corresponding page in WikiPedia. The proposed approach extends the Wiki Machine adding the following steps:

- a domain filter that removes terms from the list returned by the Wiki Machine that are not domain-specific;
- a cross-lingual linking algorithm that associates the extracted terms with their corresponding translations.

4.1.3.4.2 Sentence and Word Level Confidence

MT Quality Estimation (QE) is the task of determining the quality of an automatic translation given its source sentence and without resorting to reference translations. In MateCat, we developed QE systems able to provide predictions at either sentence or single word level, and methods for adaptive QE that address the specific challenges posed by the integration of this functionality into a CAT environment. We conducted experiments on binary (good/bad) QE with real users operating with the MateCat tool to explore the impact of human subjectivity on data annotation. We developed an effective method to assess the impact of different types of MT errors on translation quality. These activities are briefly sketched below; it is worth to noticing that tools and data sets that we developed in the course of this work were released as open-source resources.

Quality Estimation System - In the past three years, the shared tasks on QE organized within the WMT workshop series[10] have served as a major venue to measure improvements on QE. MateCat participated to some of the shared tasks in a joint collaboration with a partner of the Casmacat project (University of Valencia), obtaining excellent results. In particular, our system ranked at the top in two out of the five subtasks where we submitted runs, namely those concerning the sentence-level and word-level QE.

Adaptive Quality Estimation - The need for QE systems to adapt to the behaviour of specific users and across domain changes is an aspect of the QE problem that has been mostly ignored so far. A common trait of all current approaches is the reliance on batch learning techniques, which assume a "static" world where new, unseen instances that are encountered will be similar to the training data. In order to develop QE models for realistic scenarios where such assumptions might not hold (first of all the CAT framework), part of the MateCat QE activities focussed on tackling the task in situations where training datasets are small and/or not representative of the actual testing domain. For these situations, which are particularly challenging from the machine learning perspective, we investigated the potential of online and multitask learning in comparison to the batch learning algorithms used currently.

A. Online Learning for Adaptive QE:

In contrast to batch learning scenarios, where models are trained once and then applied to unseen data, online learning algorithms update their models with every single training instance. This allows them to respond to user feedback and to changes in the overall structure of the data that they encounter. Incorporation of online learning into the QE system requires the adaptation of its standard batch learning workflow to:

1. Perform feature extraction from a *⟨source, target⟩* pair one instance at a time instead of all at once for an entire training set;
2. Produce a quality prediction for the input instance;
3. Gather user feedback for the instance (i.e., calculate a "true label" based on the amount of user post-edits);

[9] http://bitbucket.org/fbk/thewikimachine/

[10] http://www.statmt.org/wmt14

4. Send the true label back to the model to update its predictions for future instances.

For adaptation to user and domain changes, we experimented with:

- Online learning algorithms, specifically OnlineSVR[11] and the Passive-Aggressive Perceptron[12]
- Alternative learning strategies; we compared adaptive and empty models against a system trained in batch mode using the Scikit-learn implementation of Support Vector Regression (SVR).[13] The adaptive model is built on top of an existing model created from the training data, and exploits the new test instances to refine its predictions in a stepwise manner. The empty model only learns from the test set, simulating the worst condition where training data is not available at all. The batch model is built by learning only from the training data and is evaluated on the test set without exploiting information from the test instances.
- Several datasets, language combinations and domains (either provided with the WMT shared tasks or created with the MateCat
- Different testing conditions. These range from the easiest situation, where training and test data feature homogeneous label distributions (adaptation capabilities are not required and batch methods operate in the ideal conditions), to an intermediate situation (user changes within the same domain) and the most difficult one (user and domain changes at the same time) where the differences between training and test call for adaptive solutions to the learning problem.

Our results show that the sensitivity of online QE models to different distributions of training and test instances makes them more suitable than batch methods for integration in a CAT framework. Despite slight variations across the different testing conditions (e.g., neither of the two online learning algorithms performs consistently better than the other), global mean absolute error (MAE) scores for the online algorithms in both training regimes - adaptive and empty - significantly outperform the results achieved by the batch models.

B. Multitask Learning for Adaptive QE:

A possible alternative to cope with data heterogeneity is to share information across domains. This would allow learning a QE model for a specific target domain by exploiting training instances from different domains. Then, we investigated approaches that allow not only learning from one single source domain, but also from multiple source domains simultaneously, leveraging the labels from all available data to improve results in a target domain. For this purposes, we applied the Multitask learning (MTL) framework, which uses domain-specific training signals of related tasks to improve model generalization on a target task. An important assumption in MTL is that different tasks (domains in our case) are correlated via a common structure, which allows for knowledge transfer among tasks. MTL has been demonstrated to improve model generalization over single task learning (STL) for various problems in several areas. Our experiments confirmed that even under different and hard conditions, MTL outperforms competitive QE baselines on different domains.

In addition, we investigated the effectiveness of multitask and online learning to QE. Our results confirm the effectiveness of both methods (they both produce better models than the single task learning and pooling strategies) and suggest the possibility of an integration of the two strategies as a future research direction: i.e. an online MTL method that combines the capability of MTL to transfer knowledge from different domains with the capability of online learning to train incremental models that can leverage also the test data.

[11] http://www2.imperial.ac.uk/~gmontana/onlinesvr.htm
[12] https://code.google.com/p/sofia-ml/
[13] http://scikit-learn.org/

Binary quality estimation - In MateCat, we also investigated binary QE as a variant of the classic MT quality estimation that predicts real values. First, we performed a task-oriented analysis of the usefulness of the available human-annotated datasets for binary quality estimation, and developed an automatic method to re-annotate with binary judgements the English-Spanish QE corpus that was used for the WMT 2012 shared task on QE. Then, extensions to that work and research activities were conducted that can be summarized as follows:

1. The automatic annotation procedure was experimentally evaluated for new language pairs, covering different domains and produced by different post-editors.
2. We were able to show that thresholds separating useful from useless MT suggestions can be empirically estimated, even from a relatively small amount of data and under various conditions (language pairs, domains).
3. Such thresholds are always significantly lower than the values proposed by existing QE data annotation guidelines.
4. Our empirical findings have been confirmed by the results of a verification involving several human post-editors operating with the MateCat Tool.
5. The automatic annotation method has been applied to release a freely available binary QE corpus for three language pairs.

Assessing the impact of translation errors on MT output quality - Learning from errors is a crucial aspect of improving expertise. Based on this notion, we designed a robust statistical framework for analysing the impact of different error types on MT output quality. Our approach is based on linear mixed-effects models, which allow the analysis of error-annotated MT output taking into account the variability inherent to the specific experimental setting from which the empirical observations are drawn. Our experiments were carried out on different language pairs involving Chinese, Arabic and Russian as target languages. Interesting findings were found, concerning the impact of different error types both at the level of human perception of quality and with respect to performance results measured with automatic metrics.

Delivered Tools and Resources - In terms of delivered tools and resources, the major outcomes of MateCat with respect to QE are:

- Batman. An open-source tool for aligning monolingual terminology, extracted from parallel texts, across different languages.
- BitterCorpus: an English-Italian corpus with annotated bilingual terms in the Information Technology domain
- AQET. A tool for adaptive QE, integrating all the components and the algorithms, released as open source.
- BinQE. A freely available corpus for binary QE that covers different language combinations.

4.1.3.4.3 Enriched MT Output

The purpose of QE in a post-editing scenario is to provide assistance in deciding whether the raw MT output is good enough to be post-edited, or whether it is more efficient to ignore it - if MT quality is too poor, assessing and editing the suggestion can be more time-consuming than translating the segment in question from scratch. Then, we added to the MateCat tool a "traffic light" indicator:

- A green flag indicates that it is worth to post-edit a given MT suggestion (its quality is good enough to guarantee a minimal work for the post-editor);
- A red flag indicates that it is better to rewrite from scratch the proposed suggestion (due to its poor quality, post-editing would require more effort).

Defining a criterion to assign green/red flags to each suggestion is an important step to develop such traffic light mechanism. As a possible solution to this problem we opted for limiting the HTER of each training point and using the binary data to train a QE classifier. Our experiments showed that reasonable thresholds separating “good” from “bad” translations could be set.

4.1.3.5 Field tests

MateCat evaluated utility and usability of MT-enhanced CAT technology with field tests involving professional translators. Three main field tests were organized during the project, each with the aim of measuring the progress on the new MT functionalities derived from the research activity, in particular by measuring their impact on the user productivity. Here the components evaluated in the three sets of experiments:

- First field test (year 1): self-tuning MT
- Second field test (year 2): self-tuning, user-adaptive, informative MT
- Third field test (year 3): full-fledged system

In each field test, the latest stable release of the MateCat tool developed by the industrial partner and the MT engines developed by the research partners were employed. They were carried out with professional translators, under working conditions very similar to the usual practice.
The field-tests usually involved four professional translators for each domain and translation direction (see section *Tasks* below). Each user worked with the MateCat tool equipped with a private TM (initially empty) and one or more MT engines. The work of each translator (user) was organized in two sessions. During the first session (warm-up session), the user translates a (portion of the) document by receiving suggestions from the reference baseline MT system. After the warm-up phase, in the test phase (field-test session) the user translates another (or the rest of the) document by receiving and post-editing suggestions generated by MT engines enhanced with new functionalities, unless differently specified.

4.1.3.5.1 Evaluation criteria

For quantity measure, we relied on information gathered during the various sessions. Once a translator completed the assigned task, we downloaded and processed the editing log file generated by the MateCat tool. In particular, for each collection of segments we computed the average translation speed (words/hour) and the post-edit effort (PEE) by means of the HTER score. Translation speed is computed after removing possible noisy observations. In particular, the average speed of the union of all segments and edit times is computed for each condition after removing the 5% fastest and the 5% slowest post-edits. Finally, the HTER scores for each type of suggestion are computed as a weighted mean of TER scores by taking into account the length of each segment. Statistical significance of the differences in post-edit speed and effort is assessed via approximate randomization, a statistical test well established in the NLP community.

4.1.3.5.2 Tasks

According to the MateCat plan, the field tests would had to cover two translation directions, English-Italian and English-German, and two application domains, legal and information technology (IT) documents. Actually, the outcome of a preliminary dry run performed at the very beginning of the project suggested to skip the evaluation on the German to English direction for the IT domain in the first field test. Indeed, results of the first field test pushed us to fully replace German by French in the two successive field tests due to the inability of MT systems to provide German suggestions worth to be post-edited. Moreover, in the third field test TED domain was added as suggested by the reviewers at the end of the second year of the project.

4.1.3.5.3 Results

In the following, main results of the three field tests are sketched; without claiming to be exhaustive, the goal here is to provide a general idea of the main outcomes of the three sets of experiments.

First Field Test - The aim of first field test was to measure the impact on the productivity of self-tuning MT. In this respect, during the warm-up session MT suggestions came from a reference baseline MT system (BSL), while during the field-test session a self-tuned MT (ADA, from *adapted*) was employed. For the English-Italian direction, the ADA MT yielded significant TTE and PEE improvements with respect to the BSL system. Considering the average productivity, on the IT domain we observed a 11.2% improvement in TTE and a 6.5% in PEE, while on the Legal domain we observed a 22.3% gain in TTE and a 10.8% in PEE. Quite different figures were observed for the English to German translation direction. While all translators improved their translation speed (TTE) from Day 1 to Day 2, only half of them indeed showed to take advantage of better MT suggestions. Moreover, PEE figures of all users are in fact all very high, above 41%, and even increase between warm-up and field-test sessions on the average (3.4%). The explanation for this is that users disregarded most of the time the provided MT suggestions and basically translated the segments from scratch, because for this language pair the MT quality is not yet suited for post-editing. This contributed to push us to replace German by French in the successive field tests.

Second Field Test – The aim of second field test was to measure the impact on the productivity of self-tuning, user-adaptive and informative MT. In this respect, during the warm-up session MT suggestions came from a reference baseline MT system; the post-edits generated in this stage were exploited to perform the project adaptation of the MT (self-tuning). During the field-test session, both the baseline and an MT enhanced with those three functionalities were employed, in such a way to compare performance measured on the same documents. For the legal domain, the overall results for both translation directions were quite positive. All translators but one improved their productivity in terms of post-edit effort, thanks to project adaptation and on-line learning. Observed improvements range from 6% to 13% relative, and are statistically significant at level 0.01 in three cases out of five. Concerning translation speed, five translators out of six seem to benefit from the improved MT engine. Results on the IT domain were more controversial for French translators, while for the Italian direction they look slightly better. In absolute terms, translation speed figures look higher than for the Legal domain, which means that the task was less complex for the translator (probably less domain-specific terminology and complex sentences to translate). From the post-edit effort side, the HTER figures were in the range of those observed for the legal document.

Third Field Test - On the basis of the experience of the first two field tests, for the final test we decided to adopt a still different experimental design, where each translator post-edited the same document twice, with one month interval in-between, a lapse of time we considered sufficient to forget very detailed information about the document. The goal was to tackle the problem of learning effects due to experience with the CAT tool and with post-editing. Anyway, each post-editing exercise was arranged in two sessions as usual, the warm-up session and the actual test session, during which the translator post-edits the same document, the first time with MT suggestions from the static baseline and the second time (one month later) with MT suggestions from the full-fledged system.

Experimental results showed that the full-fledged adaptive systems improved significantly both key performance indicators for all the official tasks. Improvements were in the range of 12%-18% for PEE, and 10%-37% for TTE. All gains in PEE and TTE were significant at level $p < 0.001$, but one gain in TTE, which was significant at level $p < 0.01$. Across all official tasks and languages, the overall PEE of the static condition was 28.63% while for the adaptive condition it was 24.66%. This resulted in a global gain in PEE by 13.87% ($p < 0.00115$). The corresponding figures for TTE were 1,891 w/h with the static condition and 2,298 w/h with the adaptive condition, resulting in an overall TTE gain by 21.52% ($p < 0.001$).

Finally, the additional experiment on TED, the non-official task, resulted in a little 3% loss in PEE ($p < 0.01$), and a 16% gain in TTE ($p < 0.001$). These results showed that even though for this task

adaptation provided suggestions are not better in terms of percentage of words to fix, they could be post-edited significantly faster than those provided by the static MT system.
The final field test also included additional ad-hoc experiments for assessing the contribution of two specific Informative MT functionalities: binary QE and bilingual terminology extraction.
Concerning the first issue, we were able to isolate the contribution of binary QE labels to support translators' work: measures done by means of a questionnaire indicated the achievement of the original objective of "60%" user acceptance. In addition, to make the evaluation more complete and gather quantitative evidence about the usefulness of QE, we also measured post-editing time variations with and without binary QE labels. Overall, the results of this evaluation were positive, showing that QE information brings significant post-editing time reductions when the quality of the MT suggestions is in the medium-high range ($0 \leq HTER \leq 60$).
Concerning the terminology help, our module proved to be reliable in extracting from a monolingual document *relevant* and *useful* bilingual terms. The results of an *ad-hoc* field test on this task revealed that the extracted terms are considerably above the 60% user acceptance threshold in terms of correctness of the translation, domain relevance, and usefulness to the terminologists.

4.1.3.6 Open Source Release

The ultimate goal of the dissemination and exploitation activities of the MateCat project was to promote its outcomes among the scientific, industrial, and user communities. This goal has been pursued through the publication of technical papers in the top scientific venues and journals (cf. 4.1), the presentation of the MateCat tool versions and the field test results among the industrial players, and the promotion of the MateCat tool among professional translators as well as occasional translators of Web communities (cf. 4.1.4). Finally, all the implemented software has been documented and distributed as open source, to foster their rapid exploitation.

We list below all the software developed and released along the whole duration of the project, providing for each of them a brief description and a pointer.

MateCat Tool. MateCat is an enterprise-level, web-based CAT tool designed to make post-editing and outsourcing easy and to provide a complete set of features to manage and monitor translation projects. Thanks to the integration of the largest collaborative translation memory and statistical machine translation, users will likely get significantly more matches than with other CAT tools and translate faster. Pointer: www.matecat.com

Adaptive MT server. This is a translation web service able to return a list of translation alternatives for a given input, automatically generated by an MT engine, and to adapt on-the-fly the MT engine models according to any provided feedback, consisting of a parallel sentence pair, i.e. a source text and its translation. Adaptive MT server is also able to support terminology, either pre-loaded or inserted on the-fly. Pointer: www.mt4cat.org

Online-adaptive Moses with cache-based models. This is an enhanced version of the well-known state-of-the-art SMT Moses Toolkit, which permits the dynamic (on-the-fly) adaptation of its statistical models without the need of their reloading, according to any suggestions coming from users. Online adaptation is achieved by means of cache-based language and translation models, which reward their content using either a parameterizable time-decaying scoring function or pre-defined constant values. The caches can be populated at any time from input using an xml-based annotation or pre-populated from file. Pointer: www.mt4cat.org

Online-adaptive Moses with suffix-array models. The phrase-based version of Moses has been also enhanced with a fast and reliable implementation of virtual phrase tables based on sampling word-aligned bitexts at query time. The training data is indexed for full text search via suffix arrays; phrase table entries are constructed on-the-fly by extracting translations from a sufficiently large sample of source phrase occurrences in the word-aligned parallel data.
Pointers: github.com/moses-smt/mosesdecoder (code)
www.statmt.org/moses/?n=Moses.PhraseDictionaryBitextSampling (documentation)

OnlineMGIZA++. This is an enhanced version of MGIZA++ that is well-suited for the real-time applications in which the sentence pairs are required to be word-aligned, one at a time. Nevertheless, it inherited the functionality to train alignment models from a parallel training corpus. Pointer: www.mt4cat.org.

Bitext-tokaligner. This software consists of a Perl script that implements a simple algorithm to align at word level a source sentence and its human-generated reference using a raw translation version as pivot. It has been designed to align sentence pairs in a real-time post-editing context. Both its usability and its limited time consumption make this software a suitable approach for this purpose.
Pointer: https://github.com/fredblain/bitext-tokaligner

AQET. This is an open-source package for performing Quality Estimation for Machine Translation, i.e. the task of determining the quality of an automatic translation given its source sentence and without recourse to reference translations. AQET is able to continuously learn from post-edited sentences and is reactive and robust to user and domain changes. AQET has been developed to support professional translators during their daily work and it is suitable for being embedded in a CAT tool. Pointer: www.mt4cat.org

CSLM Toolkit. During the project, the existing CSLM toolkit has been integrated in Moses, extended with project adaptation functions and extensively experimented. CSLM is an open-source software which implements the so-called continuous space language and translation model. The basic idea of this approach is to project the word indices onto a continuous space and to use a probability estimator operating on this space. A neural network can be used to simultaneously learn the projection of the words onto the continuous space and to estimate the n-gram probabilities. This approach was successfully used in large vocabulary continuous speech recognition and in phrase-based SMT systems. Pointer: www-lium.univ-lemans.fr/cslm/

XenC. This is a C++ pre-processing tool aimed at selecting textual data. It has applications to NLP and particularly Statistical Machine Translation (SMT) and Automatic Speech Recognition (ASR). It can perform language-independent monolingual as well as bilingual data selection. The goal of XenC is to allow selection of relevant data regarding a given task or subject, which then will be used to build the statistical models for a NLP system. Pointer: github.com/rousseau-lium/XenC

MT-EQuAl. MT-EQuAl is a toolkit for the human assessment of Machine Translation output. The web-based toolkit provides an annotation interface to carry out three annotation tasks, namely quality rating of translations (e.g. adequacy/fluency, relative ranking), annotation of translation errors, and word-alignment. MT-EQuAl supports the following browsers: Chrome, Safari, Firefox, and Internet Explorer. Pointer: www.mt4cat.org

BATMAN. BilinguAl TerM AligNer is an open-source tool for aligning monolingual terminology, extracted from parallel texts, across different languages. BATMAN requires in input monolingual terms from the source and target languages and the parallel documents from where the terms have been extracted. It outputs a list of aligned bilingual terminology. Pointer: www.mt4cat.org

4.1.3.7 Dissemination and user-group

Dissemination of project related news and scientific results (cf. table in Section 4.2) was done mainly via the project website (www.matecat.com) which serves as the main collector for all information regarding the project. The website was recently redesigned and now has a more commercially-oriented structure; all the information about the project, however, are key for the commercial exploitation and have been moved to a dedicated section called “Open source” from where users can access all the scientific publications, the deliverables and reports, and also download the source code to install MateCat on their servers.

As a support to the website and to create an online presence, the MateCat team has also been using accounts on the main social networks:

- Twitter - https://twitter.com/MateCat
- LinkedIn - https://www.linkedin.com/company/matecat
- Facebook - https://www.facebook.com/joinMateCat
- YouTube - https://www.youtube.com/channel/UCiFufs09x88AuI1D0kvKR6w

The social networks are used to further spread any relevant information about the project and, in the case of Twitter, to offer live coverage of the events that the MateCat team attends. Furthermore, the key information about the project are also sent to all users registered on our website (currently 1,835) via a newsletter.

A significant effort to promote the MateCat project was concentrated in a series of events hosted in Vancouver (Canada) at the end of October 2014: Workshop on Interactive and Adaptive Machine Translation (22 Oct), AMTA Main Conference (23-25 Oct), TAUS Annual Conference 2014 (27-28 Oct), MateCat User Group Event (28 Oct), Localization World Conference (30-31 Oct). In all these events, MateCat was present with invited talks, scientific papers, exhibitions and booths (see list above). In particular, the MateCat User Group Event (Milestone 6) was organized to officially launch the MateCat Tool. Approximately 60 people, including current and new users, attended the event during which Marcello Federico, Marco Trombetti and Alessandro Cattelan presented the goals of the project and gave a demo of the MateCat 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 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. 3: The MateCat team at the MateCat Final User Group Event in Vancouver.

4.1.4 Potential impact

After three years of development and of use at Translated srl, the MateCat technology is now been promoted to professional users outside the consortium as an alternative to traditional CAT tools. As of October 2014, Translated srl entirely replaced previous CAT tools it had been using with the MateCat 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.

In the last months of the project, the partners started pushing the technology and especially the hosted version of the MateCat tool to a wider audience, creating a user base of over 3,000 professional translators and a dozen of language service providers.

The technology created during the project will continue to be developed and exploited after the conclusion of the project. The partners of the consortium intend to sign a new agreement for the continuation of the development and the commercialization of the MateCat technology.

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 this creates new opportunities for integration in other platforms, such as integrated development environments (IDE), translation mobile apps, translation workflow systems, etc.

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.

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.

Exploitation and dissemination activities are supported by the website and other promotional material.

The website (www.matecat.com) has been recently redesigned to focus more on the commercial aspect of the project, especially by putting the MateCat tool as the central piece of its design. Other sections include the “Features” page, which functions as a landing page presenting the main characteristics of the software and the advanced Moses integration; the “About” section presents the goal of the project and the people who worked on it; the “Open source” section is one of the key areas of the website presenting all the scientific publications, deliverables and reports along with a link and instructions to download the open source code; the “Release notes” is intended as a list of the latest features of the MateCat tool to keep users up to date with the latest releases; finally, the “Support” section contains the instructions needed to use and install the technology.

Along with the website, some videos have been created and are used as a promotional aid. Videos have been published on the MateCat channel on YouTube. 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.

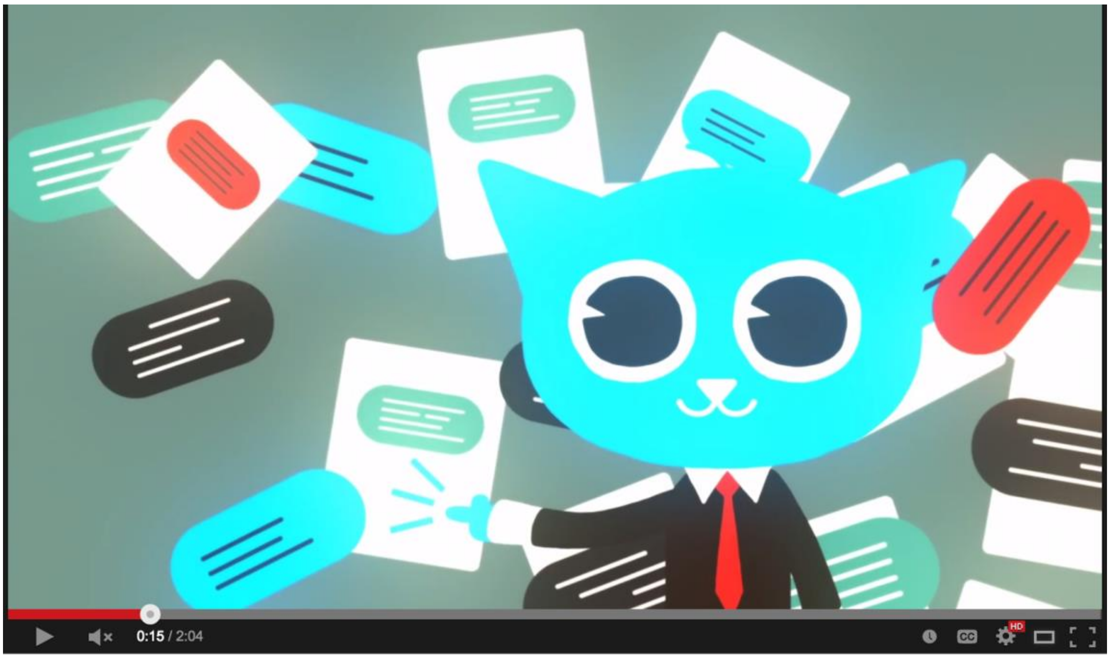

Figure 4: Longer promotional video (http://youtu.be/4MXW35duI3c).

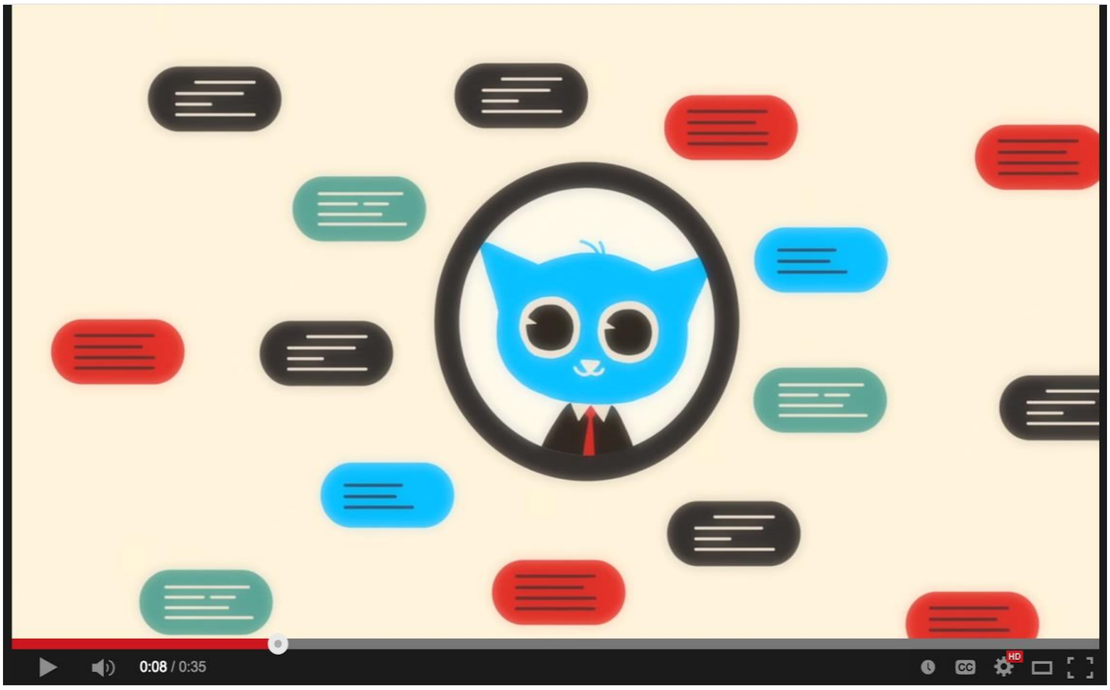

Figure 5: Shorter promotional video (http://youtu.be/a0BRq7oLA3M).

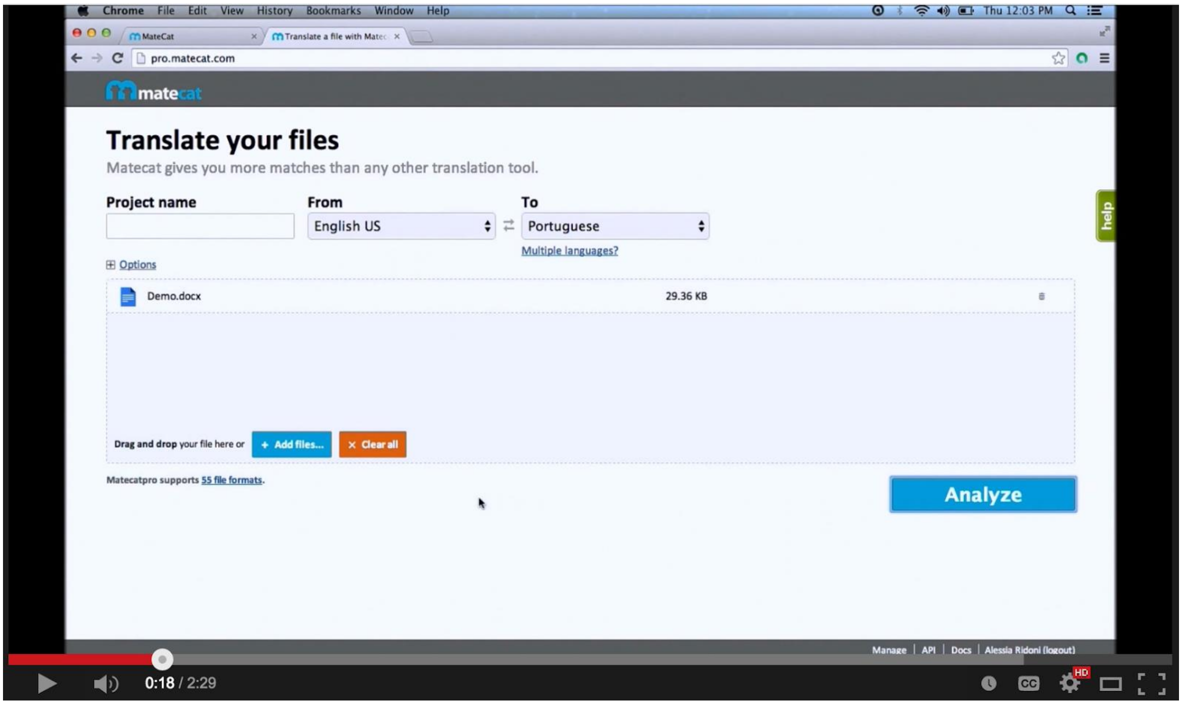

Figure 5: Video tutorial (http://youtu.be/LQSnBQWEZeU).

Finally, we created and published a video tutorial to show users how to get started translating with MateCat.
As a support for promotional activities, we also created a number of gadgets to be distributed at the events where the MateCat partner presented the results of the projects. Namely, we designed and produced MateCat polo shirts which we distributed during our events in Vancouver, bamboo pens and lollipops with the MateCat logo, leaflets which describe the advantages to using MateCat.

Figure 7: The MateCat mascot and some of the gadgets.

Moreover, the MateCat team has been active in a number of social networks where it worked to disseminate information about the product and to engage current and potential users. The team has worked mainly with LinkedIn, Twitter, YouTube and, especially, Facebook where a MateCat page and a private group for users has been created. The private group has 188 members who engage in the development of the tool as beta testers and providing feedback and, in some cases, voluntarily producing tutorial to use the software.

Apart from the interactions with users on social networks, the MateCat team has worked closely with a number of language service providers and corporations to make sure that the MateCat technology fits their needs. Among these language service providers, we count companies in the Netherlands, the United States and Brazil who are using MateCat for their production and testing out the outsourcing feature to prove the business model. Corporations interested in MateCat include eBay and PayPal who have initially been testing the software for the machine translation capabilities and now see it as an interesting technology to

4.1.5 Address of the public website

Project website: www.matecat.com

Contact address:

Marcello Federico
Fondazione Bruno Kessler
Via Sommarive, 18
38123 Povo (Trento) - Italy
Tel: +39 0461 314552
Fax: +39 0461314591
E-mail: federico@fbk.eu

4.2 Use and dissemination of foreground

Section A (public)

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[14] (if available)	Is/Will open access[15] provided?
1	*Translation Project Adaptation for MT-Enhanced Computer Assisted Translation*	*Mauro Cettolo*	*Machine Translation Journal*	*Volume 28, Issue 2 (2014)*	*Springer*	*Dordrecht*	*2014*	*pp. 127-150*	*Link*	*no*
2	*Data-driven Annotation of Binary MT Quality Estimation Corpora Based on Human Post-editions*	*Marco Turchi*	*Machine Translation Journal*	*To appear*	*Springer*	-	-	-	-	*no*
3	*Online Adaptation to Post-Edits for Phrase-Based Statistical Machine Translation*	*Nicola Bertoldi*	*Machine Translation Journal, Special issue on Post-editing*	*To appear*	*Springer*	-	-	-	-	*no*
4	*Recent Advancements in Human Language Technology in Italy*	*Bernardo Magnini*	*Intelligenza artificiale*	*Volume 7, Issue 2 (2013)*	*IOS Press*	*Amsterdam, The Netherlands*	*2013*	*pp. 91-100*	*Link*	*yes*
5	*Exploiting Qualitative Information from Automatic Word Alignment for Cross-lingual NLP Tasks*	*José G. C. de Souza*	*ACL 2013*	*4-9 August*	*ACL*	*Sofia, Bulgaria*	*2013*	*pp. 771–776*	*Link*	*yes*
6	*A Multi-Domain Translation Model Framework for Statistical Machine Translation*	*Rico Sennrich*	*ACL 2013*	*4-9 August*	*ACL*	*Sofia, Bulgaria*	*2013*	*pp. 832–840*	*Link*	*yes*
7	*Adaptive Quality Estimation for Machine Translation*	*Marco Turchi*	*ACL 2014*	*22-27 June*	*ACL*	*Baltimore, USA*	*2014*	*pp. 710-720*	*Link*	*yes*

[14] A permanent identifier should be a persistent link to the published version full text if open access or abstract if article is pay per view) or to the final manuscript accepted for publication (link to article in repository).
[15] Open Access is defined as free of charge access for anyone via Internet. Please answer "yes" if the open access to the publication is already established and also if the embargo period for open access is not yet over but you intend to establish open access afterwards.

8	*Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models*	*Marcello Federico*	*EMNLP 14*	*25-29 October*	*ACL*	*Doha, Qatar*	*2014*	*pp. 1643-1653*	Link	*yes*
9	*Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation*	*Kyunghyun Cho*	*EMNLP 14*	*25-29 October*	*ACL*	*Doha, Qatar*	*2014*	*pp. 1724-1734*	Link	*yes*
10	*Continuous Space Translation Models for Phrase-Based Statistical Machine Translation*	*Holger Schwenk*	*COLING 2012*	*10-14 December*	*ACL*	*Mumbai, India*	*2012*	*pp. 1071-1080*	Link	*yes*
11	*The MateCat Tool*	*Marcello Federico*	*COLING 2014*	*23-29 August*	*ACL*	*Dublin, Ireland*	*2014*	*pp. 129-132*	Link	*yes*
12	*MT-EQuAl: a Toolkit for Human Assessment of Machine Translation Output*	*Christian Girardi*	*COLING 2014*	*23-29 August*	*ACL*	*Dublin, Ireland*	*2014*	*pp. 120-123*	Link	*yes*
13	*Machine Translation Quality Estimation Across Domains*	*José G. C. de Souza*	*COLING 2014*	*23-29 August*	*ACL*	*Dublin, Ireland*	*2014*	*pp. 409-420*	Link	*yes*
14	*Evaluating the Learning Curve of Domain Adaptive Statistical Machine Translation Systems*	*Nicola Bertoldi*	*WMT 2012*	*7-8 June*	*ACL*	*Montreal, Canada*	*2012*	*pp. 433-441*	Link	*yes*
15	*Black box features for the WMT 2012 quality estimation shared task*	*Christian Buck*	*WMT2012*	*7-8 June*	*ACL*	*Montreal, Canada*	*2012*	*pp. 91-95*	Link	*yes*
16	*Coping with the Subjectivity of Human Judgements in MT Quality Estimation*	*Marco Turchi*	*WMT 2013*	*8-9 August*	*ACL*	*Sofia, Bulgaria*	*2013*	*pp. 240–251*	Link	*yes*
17	*Online Learning Approaches in Computer Assisted Translation*	*Prashant Mathur*	*WMT 2013*	*8-9 August*	*ACL*	*Sofia, Bulgaria*	*2013*	*pp. 301–308*	Link	*yes*
18	*FBK-UEdin Participation to the WMT13 Quality Estimation Shared Task*	*José G. C. de Souza*	*WMT 2013*	*8-9 August*	*ACL*	*Sofia, Bulgaria*	*2013*	*pp. 352–358*	Link	*yes*
19	*Findings of the 2013 Workshop on Statistical Machine Translation*	*Ondřej Bojar*	*WMT 2013*	*8-9 August*	*ACL*	*Sofia, Bulgaria*	*2013*	*pp. 1–44*	Link	*yes*
20	*FBK-UPV-UEdin participation in the WMT14 Quality Estimation shared-task*	*José G. C. de Souza*	*WMT 2014*	*26-27 June*	*ACL*	*Baltimore, USA*	*2014*	*pp. 322-328*	Link	*yes*
21	*MateCat: Machine Translation Enhanced Computer Assisted Translation*	*Project paper*	*EAMT 2012*	*28-30 May*	*EAMT*	*Trento, Italy*	*2012*	*p. 202*	Link	*yes*

22	*Measuring User Productivity in MT Enhanced Computer Assisted Translation*	*Marcello Federico*	*AMTA 2012*	*28 October – 1 November*	*AMTA*	*San Diego California*	*2012*	*-*	[Link]()	*yes*
23	*Online Multi-User Adaptive Statistical Machine Translation*	*Prashant Mathur*	*AMTA 2014*	*22-26 October*	*AMTA*	*Vancouver, Canada*	*2014*	*pp. 152-165*	[Link]()	*yes*
24	*Enhancing Statistical Machine Translation with Bilingual Terminology in a CAT Environment*	*Mihael Arcan*	*AMTA 2014*	*22-26 October*	*AMTA*	*Vancouver, Canada*	*2014*	*pp. 54-68*	[Link]()	*yes*
25	*The Repetition Rate of Text as a Predictor of the Effectiveness of Machine Translation Adaptation*	*Mauro Cettolo*	*AMTA 2014*	*22-26 October*	*AMTA*	*Vancouver, Canada*	*2014*	*pp. 166-179*	[Link]()	*yes*
26	*CSLM - A modular Open-Source Continuous Space Language Modeling Toolkit*	*Holger Schwenk*	*Interspeech, 2013*	*25-29 August*	*ISCA*	*Lyon, France*	*2013*	*pp. 1198-1202*	[Link]()	*yes*
27	*Generative and Discriminative Methods for Online Adaptation in SMT*	*Katharina Waeschle*	*MT Summit 2013*	*2-6 September*	*EAMT*	*Nice, France*	*2013*	*pp. 11–18*	[Link]()	*yes*
28	*Cache-based Online Adaptation for Machine Translation Enhanced Computer Assisted Translation*	*Nicola Bertoldi*	*MT Summit 2013*	*2-6 September*	*EAMT*	*Nice, France*	*2013*	*pp. 35–42*	[Link]()	*yes*
29	*Project Adaptation for MT-Enhanced Computer Assisted Translation*	*Mauro Cettolo*	*MT Summit 2013*	*2-6 September*	*EAMT*	*Nice, France*	*2013*	*pp. 27–34*	[Link]()	*yes*
30	*N-gram Counts and Language Models from the Common Crawl*	*Christian Buck*	*LREC 2014*	*26-31 May*	*ELRA*	*Reykjavik, Iceland*	*2014*	*pp. 3579-3584*	[Link]()	*yes*
31	*Automatic Annotation of Machine Translation Datasets with Binary Quality Judgements*	*Marco Turchi*	*LREC 2014*	*26-31 May*	*ELRA*	*Reykjavik, Iceland*	*2014*	*pp. 1788-1792*	[Link]()	*yes*
32	*Online and Multitask learning for Machine Translation Quality Estimation in Real-world scenarios*	*José G. C. de Souza*	*CLIC-it 2014*	*To appear*	*-*	*-*	*-*	*-*	*-*	*yes*
33	*Adattamento al Progetto dei Modelli di Traduzione Automatica nella Traduzione Assistita*	*Mauro Cettolo*	*CLIC-it 2014*	*To appear*	*-*	*-*	*-*	*-*	*-*	*yes*

34	*Project Adaptation over Several Days*	*Frédéric Blain*	*Translation in Transition Conference 2015*	*To appear*	-	-	-	-	-	*yes*
35	*Incremental Adaptation Using Translation Information and Post-Editing Analysis*	*Frédéric Blain*	*IWSLT 2012*	*6-7 December*	*IWSLT*	*Hong Kong, China*	*2012*	*pp. 229-236*	*Link*	*yes*
36	*Online Word Alignment for Online Adaptive Machine Translation*	*Amin Farajian*	*EACL Workshop on Humans and Computer-assisted Translation*	*26 April*	*ACL*	*Gothenburg, Sweden*	*2014*	*pp. 84–92*	*Link*	*yes*
37	*Identification of Bilingual Terms from Monolingual Documents for Statistical Machine Translation*	*Mihael Arcan*	*COLING Workshop on Computational Terminology*	*23 August*	*ACL*	*Dublin, Ireland*	*2014*	*pp. 22-31*	*Link*	*yes*
38	*Towards a Combination of Online and Multitask Learning for MT Quality Estimation: a Preliminary Study*	*José G. C. de Souza*	*AMTA Workshop on Interactive and Adaptive Machine Translation*	*22 October*	*AMTA*	*Vancouver, Canada*	*2014*	*pp. 9-19*	*Link*	*yes*
39	*Optimized MT Online Learning in Computer Assisted Translation*	*Prashant Mathur*	*AMTA Workshop on Interactive and Adaptive Machine Translation*	*22 October*	*AMTA*	*Vancouver, Canada*	*2014*	*pp. 32-41*	*Link*	*yes*
40	*Dynamic Phrase Tables for Statistical Machine Translation in an Interactive Post-editing Scenario*	*Ulrich Germann*	*AMTA Workshop on Interactive and Adaptive Machine Translation*	*22 October*	*AMTA*	*Vancouver, Canada*	*2014*	*pp. 20–31*	*Link*	*yes*
41	*Issues in Incremental Adaptation of Statistical MT from Human Post-edits*	*Mauro Cettolo*	*MT Summit XIV Workshop on Post-editing Technology and Practice*	*2 September*	*EAMT*	*Nice, France*	*2013*	*pp. 111–118*	*Link*	*yes*

TEMPLATE A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities[16]	Main leader	Title	Date/Period	Place	Type of audience[17]	Size of audience	Countries addressed
Y3	*Conference*	*LEMANS*	*GALA Annual Conference*	*23-26 March 2014*	*Istanbul (Turkey)*	*Scientific Community*	*N/A*	*International*
2	*Conference*	*FBK*	*TAUS Quality Evaluation Summit*	*4 June 2014*	*Dublin (Ireland)*	*Industry*	*~50 people*	*International*
3	*Exhibition*	*LEMANS*	*Journées d'Études sur la Parole*	*26 June 2014*	*Le Mans (France)*	*Scientific Community*	*N/A*	*France*
4	*Exhibitions*	*ALL*	*COLING*	*26 August 2014*	*Dublin (Ireland)*	*Scientific Community*	*~ 200 people*	*International*
5	*Workshop*	*FBK and Translated*	*MTM*	*9 September 2014*	*Trento (Italy)*	*Scientific Community*	*~ 100 people*	*International*
6	*Workshop*	*FBK and UEDIN*	*IAMT at AMTA*	*22 October 2014*	*Vancouver (Canada)*	*Scientific Community*	*~ 60 people*	*International*
7	*Exhibition*	*FBK and Translated*	*AMTA*	*24 October 2014*	*Vancouver (Canada)*	*Scientific Community, Industry*	*~ 200 people*	*International*
8	*Exhibition and Presentation*	*Translated*	*TAUS Annual Conference*	*27-28 October 2014*	*Vancouver (Canada)*	*Industry*	*~ 100 people*	*International*
9	*Exhibition*	*Translated*	*Localization World Conference and Exhibits*	*30-31 October 2014*	*Vancouver (Canada)*	*Industry*	*~ 200 people*	*International*
10	*Presentation*	*FBK*	*University of Sassari*	*12-13 December 2013*	*Sassari (Italy)*	*Scientific Community*	*~ 20 people*	*Italy*
11	*Presentation*	*FBK*	*University Institute ISIT*	*November 2013 - January 2014*	*Trento, (Italy)*	*Scientific Community*	*~15 people*	*Italy*

[16] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[17] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

12	*Presentation*	*Translated and FBK*	*EU translators, DGT*	*on 16 January 2014*	*Luxembourg*	*Other*	*~20 people*	*Europe*
13	*Exhibition*	*FBK*	*Festa dell'Europa*	*8 May 2014*	*Trento (Italy)*	*Civil Society*	*~100 people*	*Italy*
14	*Other*	*FBK*	*Internships: translation students*	*12-24 May 2014*	*Trento (Italy)*	*Scientific Community*	*~10 people*	*Italy*
15	*Other*	*FBK*	*Summer Internships at FBK*	*June - September 2014*	*Trento (Italy)*	*Scientific Community*	*~ 5 people*	*International*
16	*Exhibition*	*LEMANS*	*Le Mans exhibition*	*11-15 September 2014*	*Le Mans (France)*	*Civil Society*	*N/A*	*France*
17	*Presentations*	*FBK*	*eBay, San Jose*	*17-18 September 2014*	*San Jose, California (USA)*	*Industry*	*~20 people*	*USA*
18	*Presentation*	*Translated*	*User Group Event , Pan Pacific Vancouver Hotel*	*28 October 2014*	*Vancouver (Canada)*	*Industry*	*~60 people*	*International*
Y2	*Conference*	*FBK and UEDIN*	*AMTA*	*October 28 to November 1 2012*	*San Diego California (USA)*	*Scientific Community, Industry*	*~200 people*	*International*
20	*Workshop*	*LEMANS*	*IWSLT*	*6 and 7 December 2012*	*Hong Kong*	*Scientific Community*	*~100 people*	*International*
21	*Conference*	*LEMANS*	*COLING*	*8-15 December*	*Bombay, (India)*	*Scientific Community*	*~200 people*	*International*
22	*Conference*	*UEDIN*	*Deutscher Sprachtechnologietag*	*24 January 2013*	*Berlin (Germany)*	*Scientific Community*	*N/A*	*Germany*
23	*Conference*	*LEMANS and FBK*	*ACL*	*4-9 August 2013*	*Sofia (Bulgaria)*	*Scientific Community*	*~300 people*	*International*
24	*Workshop*	*FBK, UEDIN, LEMANS*	*WMT at ACL*	*8 August 2013*	*Sofia (Bulgaria)*	*Scientific Community*	*~100 people*	*International*
25	*Conference*	*LEMANS*	*Interspeech*	*25-29 August 2013*	*Lyon (France)*	*Scientific Community*	*~1,200 people*	*International*
26	*Conference*	*ALL*	*Machine Translation Summit*	*2-6 September 2013*	*Nice, (France)*	*Scientific Community, Industry*	*~200 people*	*International*
27	*Workshop*	*LEMANS and FBK*	*Post-Editing Technologies and Practice*	*4 September 2013*	*Nice (France)*	*Scientific Community, Industry*	*~50 people*	*International*
28	*Workshop*	*FBK*	*User Centric Machine Translation & Evaluation*	*3 September 2013*	*Nice (France)*	*Scientific Community, Industry*	*~50 people*	*International*
29	*Workshop*	*FBK and LEMANS*	*MTM*	*9-14 September 2013*	*Prague (Czech Republic)*	*Scientific Community*	*~100 people*	*International*

30	*Conference*	*UEDIN*	*Trikonf*	*18-20 October 2013*	*Freiburg (Germany)*	*Scientific Community*	*N/A*	*International*
31	*Presentation*	*UEDIN*	*Microsoft*	*1 November 2012*	*Redmond (USA)*	*Industry*	*~20 people*	*USA*
32	*Presentation*	*FBK*	*Trentino School of Management*	*21 December 2012*	*Trento, (Italy)*	*Scientific Community*	*~20 people*	*Italy*
33	*Presentation*	*FBK*	*University Institute ISIT*	*16 May 2013*	*Trento, (Italy)*	*Scientific Community*	*~15 people*	*Italy*
34	*Other*	*FBK*	*RHoK*	*1 June 2013*	*Trento (Italy)*	*Civil Society*	*~10 people*	*Italy*
35	*Exhibition*	*LIUM*	*Researchers' Night*	*27 September 2013*	*Le Mans (France)*	*Civil Society*	*N/A*	*Italy*
36	*Exhibition*	*FBK*	*Researchers' Night*	*27 September 2013*	*Trento (Italy*	*Civil Society*	*~10,000 people*	*Italy*
37	*Presentation*	*Translated*	*TAUS Translation Technology Showcase Webinar*	*2 October 2013*	*Web*	*Industry*	*~100 people*	*International*
38	*Presentation*	*FBK*	*Internet Festival*	*October 12th, 2013*	*Pisa (Italy)*	*Civil Society*	*~ 50 people*	*Italy*
Y1	*Conference*	*FBK*	*EAMT*	*28-30 May 2012*	*Trento (Italy)*	*Scientific Community, Industry*	*~120 people*	*International*
40	*Workshop*	*UEDIN and FBK*	*WMT*	*7-8 June 2012*	*Montreal (Canada)*	*Scientific Community*	*~100 people*	*International*
41	*Presentation*	*LEMANS*	*Meta Exhibition 2012*	*20-21 June 2012*	*Brussels (Belgium)*	*Scientific Community, Industry*	*~100 people*	*Europe*
42	*Workshop*	*Translated*	*Expertise in Translation and Post-editing: Research and Application*	*17-18 August 2012*	*Copenhagen (Denmark)*	*Scientific Community*	*N/A*	*International*
43	*Workshop*	*UEDIN*	*MTM*	*7 September 2012*	*Edinburgh, Scotland (UK)*	*Scientific Community*	*~100 people*	*International*
44	*Presentation*	*FBK*	*QTLaunch Pad*	*13-14 September 2012*	*Berlin (Germany)*	*Scientific Community, Industry*	*~30 people*	*Europe*
45	*Exhibition*	*LEMANS*	*Salon e-CNET*	*September 2012*	*Paris (France)*		*N/A*	
46	*Conference*	*FBK, Translated, UEDIN*	*AMTA*	*28 October – 1 November 2012*	*San Diego California (USA).*	*Scientific Community, Industry*	*~200 people*	*International*
47	*Presentation*	*FBK and Translated*	*Towards the integration of human and machine translation*	*31 October 2012*	*USC/ISI Marina del Rey, California*	*Scientific Community*	*~20 people*	*USA*

Section B (Confidential[18] or public: confidential information to be marked clearly)
Part B1

TEMPLATE B1: LIST OF APPLICATIONS FOR PATENTS, TRADEMARKS, REGISTERED DESIGNS, ETC.					
Type of IP Rights[19]:	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Application reference(s) (e.g. EP123456)	Subject or title of application	Applicant (s) (as on the application)
N/A	N/A	N/A	N/A	N/A	N/A

[18] Note to be confused with the "EU CONFIDENTIAL" classification for some security research projects.

[19] A drop down list allows choosing the type of IP rights: Patents, Trademarks, Registered designs, Utility models, Others.

Part B2

Type of Exploitable Foreground[20]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[21]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	*Field-test data*	*NO*	*NO*	*Language resource*	*M72.1*	*2014*	*CC BY-NC-ND*	*FBK (owner)*
General advancement of knowledge	*BinQE*	*NO*	*NO*	*Language resource*	*M72.1*	*2014*	*CC BY-NC-SA*	*FBK (owner)*
General advancement of knowledge	*BitterCorpus*	*NO*	*NO*	*Language resource*	*M72.1*	*2014*	*CC BY-NC-SA*	*FBK (owner)*
General advancement of knowledge	*Word-alignment Gold Reference*	*NO*	*NO*	*Language resource*	*M72.1*	*2014*	*CC BY-NC-ND*	*FBK (owner)*
Commercial exploitation of R&D results	*MateCat Tool*	*NO*	*NO*	*Software*	*M74.3.0*	*2014*	*GNU LGPL*	*TRANSLATED (owner)*
Commercial exploitation of R&D results	*Adaptive MT Server*	*NO*	*NO*	*Software*	*M74.3.0*	*2014*	*GNU LGPL*	*FBK (owner)*
Commercial exploitation of R&D results	*Online-adaptive Moses with cache-based models*	*NO*	*NO*	*Software*	*M74.3.0*	*2014*	*GNU LGPL*	*FBK (owner)*
Commercial exploitation of R&D results	*Online-adaptive Moses with suffix-array models*	*NO*	*NO*	*Software*	*M74.3.0*	*2014*	*GNU LGPL*	*UEDIN (owner)*

[19] A drop down list allows choosing the type of foreground: General advancement of knowledge, Commercial exploitation of R&D results, Exploitation of R&D results via standards, exploitation of results through EU policies, exploitation of results through (social) innovation.
[21] A drop down list allows choosing the type sector (NACE nomenclature) : http://ec.europa.eu/competition/mergers/cases/index/nace_all.html

Type of Exploitable Foreground[20]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[21]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
Commercial exploitation of R&D results	*OnlineMGIZA++*	*NO*	*NO*	*Software*	*M74.3.0*	*2014*	*GNU GPL v2*	*FBK (owner)*
Commercial exploitation of R&D results	*Bitext-tokaligner*	*NO*	*NO*	*Software*	*M74.3.0*	*2014*	*GNU LGPL*	*LE MANS (owner)*
General advancement of knowledge	*AQET*	*NO*	*NO*	*Software*	*M72.1*	*2014*	*GNU GPL v3*	*FBK (owner)*
Commercial exploitation of R&D results	*CSLM Toolkit*	*NO*	*NO*	*Software*	*M74.3.0*	*2014*	*GNU GPL v3*	*LE MANS (owner)*
General advancement of knowledge	*XenC*	*NO*	*NO*	*Software*	*M72.1*	*2014*	*GNU GPL v3*	*LE MANS (owner)*
General advancement of knowledge	*MT-EQuAl*	*NO*	*NO*	*Software*	*M72.1*	*2014*	*Apache 2.0*	*FBK (owner)*
General advancement of knowledge	*BATMAN*	*NO*	*NO*	*Software*	*M72.1*	*2014*	*GNU LGPL*	*FBK (owner)*

MateCat Tool

- Purpose: computer assisted translation tool
- Exploitation: by industry, free lance translators, public, from now
- IPR: freely available for download and online use
- Status: fully developed
- Impact: 100 million words translated per year (??)

MT-EQuAl

- Purpose: software toolkit for human assessment and annotation of MT output
- Exploitation: by researchers and industry, from now
- IPR: freely available

- Status: fully developed
- Impact: 100 downloads per month

CSLM Toolkit
- Purpose: software toolkit for continuous space language and translation models
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: fully developed
- Impact: 100 downloads per month

AQET
- Purpose: software tool for online adaptive MT quality estimation
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: ready to use, subject to further improvements
- Impact: 50 downloads per month

Bitext-tokaligner
- Purpose: software tool to word-align source, target and reference sentences
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: fully developed
- Impact: 50 downloads per month

XenC
- Purpose: software toolkit for data selection of texts related to a given task
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: fully developed
- Impact: 50 downloads per month

BATMAN
- Purpose: software tool for aligning terminology across languages

- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: fully developed
- Impact: 50 downloads per month

Adaptive MT Server
- Purpose: software module to connect a Moses server to a CAT tool
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: ready to use, subject to further improvements
- Impact: 50 downloads per month

Online-adaptive Moses with cache-based models
- Purpose: software module enhancing Moses
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: ready to use, subject to further improvements
- Impact: 100 downloads per month

Online-adaptive Moses with suffix-array models
- Purpose: software module enhancing Moses
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: ready to use, subject to further improvements
- Impact: 100 downloads per month

OnlineMGIZA++
- Purpose: software module enhancing MGIZA++ for online use
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: ready to use, subject to further improvements
- Impact: 100 downloads per month

Field-test data

- Purpose: benchmark data for MT technology evaluation
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: fully developed
- Impact: 50 downloads per month

BinQE

- Purpose: benchmark data for MT quality estimation tasks
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: fully developed
- Impact: 50 downloads per month

BitterCorpus

- Purpose: benchmark data for automatic terminology extraction evaluation
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: fully developed
- Impact: 50 downloads per month

Word-alignment Gold Reference

- Purpose: benchmark data for automatic word-alignment evaluation
- Exploitation: by researchers and industry, from now
- IPR: freely available
- Status: fully developed
- Impact: 50 downloads per month