Demonstration of a Neural Machine Translation System with Online Learning for Translators

Miguel Domingo[1], Mercedes García-Martínez[2], Amando Estela[2], Laurent Bié[2], Alexandre Helle[2], Álvaro Peris[1], Francisco Casacuberta[1], and Manuel Herranz[2]

[1]PRHLT Research Center - Universitat Politècnica de València
{midobal, lvapeab, fcn}@prhlt.upv.es
[2]Pangeanic / B.I Europa - PangeaMT Technologies Division
{m.garcia, a.helle, a.estela, l.bie, m.herranz}@pangeanic.com

Abstract

We introduce a demonstration of our system, which implements online learning for neural machine translation in a production environment. These techniques allow the system to continuously learn from the corrections provided by the translators. We implemented an end-to-end platform integrating our machine translation servers to one of the most common user interfaces for professional translators: SDL Trados Studio. Our objective was to save post-editing effort as the machine is continuously learning from human choices and adapting the models to a specific domain or user style.

1 Introduction

Productivity is crucial in the translation industry. Nowadays, translation companies must be more competitive than ever and meet fast commercial demands. Thus, they need to produce high quality translations in shorter periods of time. Machine translation (MT) can help them to achieve this goal: instead of a linguist thinking out or "creating" translations from scratch, "humanizing" automatic translations has become a common process in the industry. This is known as post-editing (PE) and it has been shown to be effective in many cases (Arenas, 2008; Hu and Cadwell, 2016). As MT systems are continuously improving their capabilities (e.g. Hassan et al., 2018; Wu et al., 2016), this workflow has become of major relevance in the translation industry. Nonetheless, MT technology is still far from perfect (Dale, 2016; Koehn and Knowles, 2017), and there is still room for improvement.

Inherently to the PE process, new bilingual data is continuously generated (the post-edited samples). This data is typically used for the creation of domain-specific corpora, useful for adapting systems from a broader domain into a specific domain, client or style. The online learning (OL) paradigm aims to perform this adaptation during the PE process: each time the user validates a post-edited translation, the system is updated as this data is taken into account. Therefore, when the next translation is produced, the system will consider the previous post-editions. It is assumed that better translations (or translations more suited to the human post-editor preferences) will be produced.

The OL paradigm has quickly attracted the attention of researchers and industry. The CasMaCat (Alabau et al., 2013) and MateCat (Federico et al., 2014) projects—where phrase-based statistical MT systems were adapted incrementally from the user post-edits—achieved many advances in this direction. More recently, OL techniques were also applied to neural machine translation (NMT) systems (Peris et al., 2017; Turchi et al., 2017; Kothur et al., 2018; Wuebker et al., 2018; Peris and Casacuberta, 2018).

In this paper, we introduce a demo system of our in-house OL framework, in which we integrated our translation servers with the translators user-friendly interface SDL Trados Studio.

The rest of this document is structured as follows: Section 2 introduces the online learning paradigm. Next, in Section 3, we describe in detail our in-house architecture in which this paradigm is implemented. Finally, Section 4 summarizes the demo system.

2 Adapting a NMT system via online learning

We are interested in benefiting from the post-edits generated by the user during the PE process. To that end, we update the system on-the-fly, i.e, as soon as a sentence has been validated by the post-editor. Right after the user confirms a post-edit, we update the models of our NMT system, using the source sentence and the post-edit as a training pair. This adaptation can be done following gradient descent, the regular training method for neural networks.

3 Architecture

Our in-house architecture of the OL framework is composed of three main modules: the MT engine, the user interface and the translation server which links both.

Moreover, we added a logging option to keep user tracking information as keystrokes, time and mouse movements. Fig. 1 illustrates this architecture.

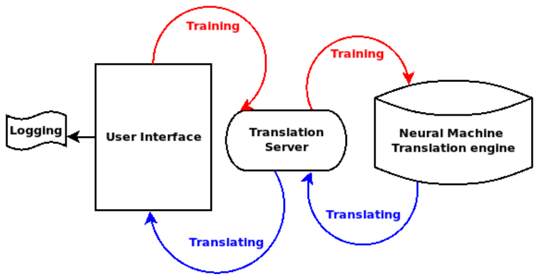


Figure 1: Architecture of our in-house OL framework. Red arrows represent how the correction made by the user arrives to the Machine Translation engine to be retrained. Blue arrows represent how machine translation is delivered to the user.

The translation process consists of delivering machine translations to the user interface and the training process retrains the MT engine with the feedback provided by the user. Both processes are performed through client-server communication. Next, we describe each module in detail.

3.1 Machine Translation Engine

The core of the MT engine is composed by the models generating translations, which can be retrained when required. Each translation project has its own model, whose architecture is set according to the project's need. All models are neural-based and are trained using `OpenNMT-py` (Klein et al., 2017).

Each MT model has its own configuration file, which contains personalized translation and OL options, such as tokenization, subword segmentation, learning rate, etc.

3.2 Translation Server

A translation server communicates with the MT models in order to generate translations and adapt the systems based on the user's post editions. This server is based on `OpenNMT-py`'s REST server and uses the HTTP protocol to define the messages in order to serve user's requests. The code of our translation server is open and available[1]. We created a branch in `OpenNMT-py` that features this server and is compatible with all its different models.

The communication between the user interface and the MT engine is performed by means of GET and POST requests. The server waits for translation requests. When received, these requests are sent to the machine translation engine in a JSON format. When a machine translation segment is corrected by the user, the correction is sent to the translation engine.

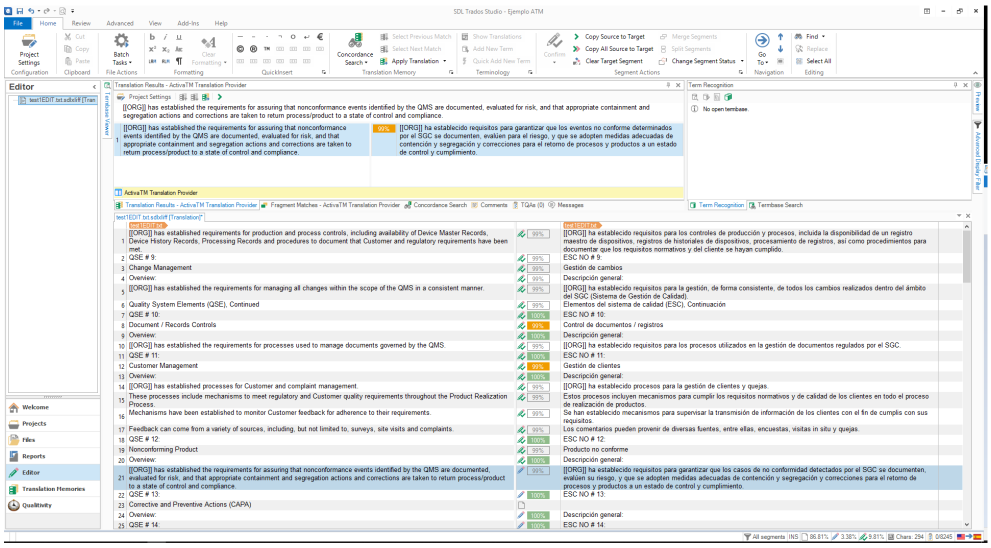
3.3 User Interface

In the translation industry, the most common user interface for translators is SDL Trados Studio[2]. Fig. 2 shows the user interface. The user gets the machine translation outputs automatically when the target part of the segment in the interface is clicked. Then, the user post-edits the segment and, when the translation is corrected, confirms it.

SDL allows the development of plugins for Trados Studio to enhance and extend the tool. Moreover, it has a large developer community[3] helping the software with add-ons and apps. We incorporated our adaptive framework as a Trados Studio plugin, that connected the user interface with Trados Studio with our translation server. As the user confirms a post-edit, the reviewed segment is sent back to the MT engine to be retrained with this new information.

In order to set up this plugin, the user fills the credentials with a username, password and URL pointing to the server (see Fig. 3 top left box). Also, the user fills the required languages and clicks in the "use machine translation" option (see Fig. 3 top right box). Finally, all the options

[1]`https://github.com/midobal/OpenNMT-py/tree/OnlineLearning`
[2]`https://www.sdl.com/es/software-and-services/translation-software/sdl-trados-studio/`
[3]`https://community.sdl.com/developers-more/developers/language-developers`



ActivaTM Translation Provider

[[ORG]] has established the requirements for assuring that nonconformance events identified by the QMS are documented, evaluated for risk, and that appropriate containment and segregation actions and corrections are taken to return process/product to a state of control and compliance.

99% [[ORG]] ha establecido requisitos para garantizar que los eventos no conforme determinados por el SGC se documenten, evalúen para el riesgo, y que se adopten medidas adecuadas de contención y segregación y correcciones para el retorno de procesos y productos a un estado de control y cumplimiento.

No open termbase.

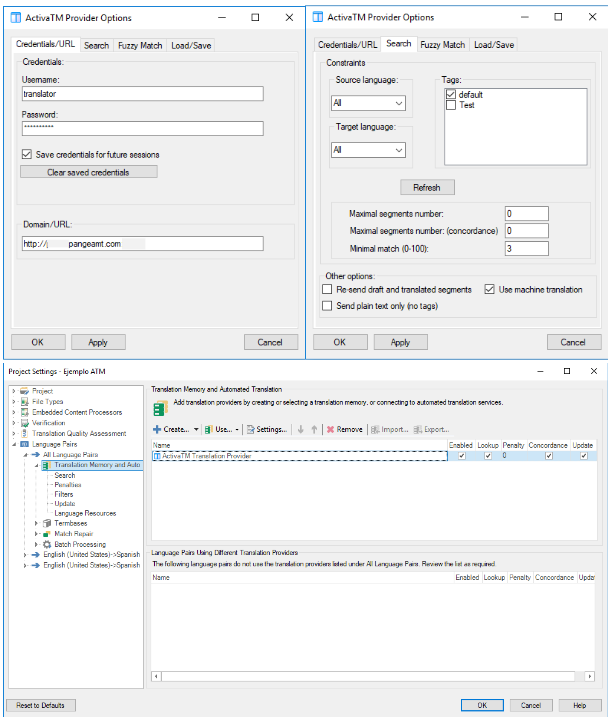
test1EDIT.txt.sdlxliff [Translation]*

	test1EDIT.txt		test1EDIT.txt
1	[[ORG]] has established requirements for production and process controls, including availability of Device Master Records, Device History Records, Processing Records and procedures to document that Customer and regulatory requirements have been met.	99%	[[ORG]] ha establecido requisitos para los controles de producción y procesos, incluida la disponibilidad de un registro maestro de dispositivos, registros de historiales de dispositivos, procesamiento de registros, así como procedimientos para documentar que los requisitos normativos y del cliente se hayan cumplido.
2	QSE # 9:	99%	ESC NO # 9:
3	Change Management	99%	Gestión de cambios
4	Overview:	99%	Descripción general:
5	[[ORG]] has established the requirements for managing all changes within the scope of the QMS in a consistent manner.	99%	[[ORG]] ha establecido requisitos para la gestión, de forma consistente, de todos los cambios realizados dentro del ámbito del SGC (Sistema de Gestión de Calidad).
6	Quality System Elements (QSE), Continued	99%	Elementos del sistema de calidad (ESC), Continuación
7	QSE # 10:	100%	ESC NO # 10:
8	Document / Records Controls	99%	Control de documentos / registros
9	Overview:	100%	Descripción general:
10	[[ORG]] has established the requirements for processes used to manage documents governed by the QMS.	99%	[[ORG]] ha establecido requisitos para los procesos utilizados en la gestión de documentos regulados por el SGC.
11	QSE # 11:	100%	ESC NO # 11:
12	Customer Management	99%	Gestión de clientes
13	Overview:	100%	Descripción general:
14	[[ORG]] has established processes for Customer and complaint management.	99%	[[ORG]] ha establecido procesos para la gestión de clientes y quejas.
15	These processes include mechanisms to meet regulatory and Customer quality requirements throughout the Product Realization Process.	99%	Estos procesos incluyen mecanismos para cumplir los requisitos normativos y de calidad de los clientes en todo el proceso de realización de productos.
16	Mechanisms have been established to monitor Customer feedback for adherence to their requirements.	99%	Se han establecido mecanismos para supervisar la transmisión de información de los clientes con el fin de cumplis con sus requisitos.
17	Feedback can come from a variety of sources, including, but not limited to, surveys, site visits and complaints.	99%	Los comentarios pueden provenir de diversas fuentes, entre ellas, encuestas, visitas in situ y quejas.
18	QSE # 12:	100%	ESC NO # 12:
19	Nonconforming Product	99%	Producto no conforme
20	Overview:	100%	Descripción general:
21	[[ORG]] has established the requirements for assuring that nonconformance events identified by the QMS are documented, evaluated for risk, and that appropriate containment and segregation actions and corrections are taken to return process/product to a state of control and compliance.	99%	[[ORG]] ha establecido requisitos para garantizar que los casos de no conformidad detectados por el SGC se documenten, evalúen su riesgo, y que se adopten medidas adecuadas de contención y segregación y correcciones para el retorno de procesos y productos a un estado de control y cumplimiento.
22	QSE # 13:	100%	ESC NO # 13:
23	Corrective and Preventive Actions (CAPA)		
24	Overview:	100%	Descripción general:
25	QSE # 14:	100%	ESC NO # 14:

All segments INS 86.81% 3.38% 9.81% Chars: 294 0/8245

Figure 2: User Interface from Trados Studio SDL.

have to be enabled in the translation provider plugin (see Fig. 3 bottom box) in the Trados Studio project settings.



ActivaTM Provider Options — Credentials/URL

Credentials:

Username: translator

Password: *********

☑ Save credentials for future sessions

Domain/URL: http:// pangeamt.com

ActivaTM Provider Options — Search

Constraints

Source language: All

Target language: All

Tags: ☑ default ☐ Test

Maximal segments number: 0

Maximal segments number: (concordance) 0

Minimal match (0-100): 3

Other options:

☐ Re-send draft and translated segments ☑ Use machine translation

☐ Send plain text only (no tags)

Project Settings - Ejemplo ATM

Translation Memory and Automated Translation

Add translation providers by creating or selecting a translation memory, or connecting to automated translation services.

Name	Enabled	Lookup	Penalty	Concordance	Update
ActivaTM Translation Provider	✓	✓	0	✓	✓

Language Pairs Using Different Translation Providers

The following language pairs do not use the translation providers listed under All Language Pairs. Review the list as required.

Figure 3: Machine translation plugin configuration.

3.4 Logging

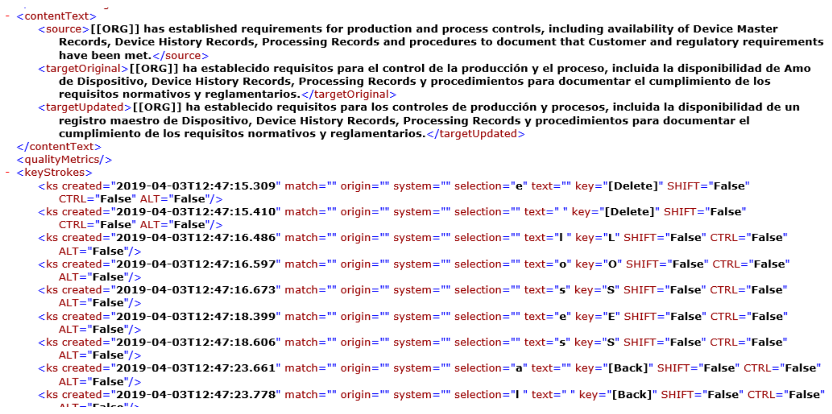
In order to measure the productivity and effectiveness of OL during the PE process, we integrated tools to log the time, keystrokes and mouse movements involved in post-editing a given file. To achieve this, we incorporated the Qualitivity[4] plugin for Trados. This plugin generates an XML logging file, which contains all the keystrokes time information per segment. An example of this logging is shown in Figure 4.

With all this log information, we can measure the effort required to post-edit a file using MT with OL. Preliminary experiments in simulated and real environments with professional translators (Domingo et al., 2019), reported significant improvements of the quality of the translations generated by the MT systems (up to 5.3 points according to hTER, and 7.8 points according to hBLEU), and a significant reduction of the PE time (up to 7.5 second per sentence).

4 Summary

We have introduced a demonstration of Pangeanic's translation framework, which incorporates on-the-fly system adaptation via online learning. This paradigm allows human translators /post-editors to produce more human-quality text, that is, be more productive—a fundamental issue in the translation industry—since the system is continuously learning from the user post-edits, avoiding repetition of the same errors. We have integrated our MT servers into the SDL Trados Studio user interface which is one of the most used in the translation industry. This system aims to improve human translators' work by saving

[4]https://community.sdl.com/product-groups/translationproductivity/w/customer-experience/2251.qualitivity



```
- <contentText>
    <source>[[ORG]] has established requirements for production and process controls, including availability of Device Master
      Records, Device History Records, Processing Records and procedures to document that Customer and regulatory requirements
      have been met.</source>
    <targetOriginal>[[ORG]] ha establecido requisitos para el control de la producción y el proceso, incluida la disponibilidad de Amo
      de Dispositivo, Device History Records, Processing Records y procedimientos para documentar el cumplimiento de los
      requisitos normativos y reglamentarios.</targetOriginal>
    <targetUpdated>[[ORG]] ha establecido requisitos para los controles de producción y procesos, incluida la disponibilidad de un
      registro maestro de Dispositivo, Device History Records, Processing Records y procedimientos para documentar el
      cumplimiento de los requisitos normativos y reglamentarios.</targetUpdated>
  </contentText>
  <qualityMetrics/>
- <keyStrokes>
    <ks created="2019-04-03T12:47:15.309" match="" origin="" system="" selection="e" text="" key="[Delete]" SHIFT="False"
      CTRL="False" ALT="False"/>
    <ks created="2019-04-03T12:47:15.410" match="" origin="" system="" selection="" text=" " key="[Delete]" SHIFT="False"
      CTRL="False" ALT="False"/>
    <ks created="2019-04-03T12:47:16.486" match="" origin="" system="" selection="" text="l " key="L" SHIFT="False" CTRL="False"
      ALT="False"/>
    <ks created="2019-04-03T12:47:16.597" match="" origin="" system="" selection="" text="o" key="O" SHIFT="False" CTRL="False"
      ALT="False"/>
    <ks created="2019-04-03T12:47:16.673" match="" origin="" system="" selection="" text="s" key="S" SHIFT="False" CTRL="False"
      ALT="False"/>
    <ks created="2019-04-03T12:47:18.399" match="" origin="" system="" selection="" text="e" key="E" SHIFT="False" CTRL="False"
      ALT="False"/>
    <ks created="2019-04-03T12:47:18.606" match="" origin="" system="" selection="" text="s" key="S" SHIFT="False" CTRL="False"
      ALT="False"/>
    <ks created="2019-04-03T12:47:23.661" match="" origin="" system="" selection="a" text="" key="[Back]" SHIFT="False" CTRL="False"
      ALT="False"/>
    <ks created="2019-04-03T12:47:23.778" match="" origin="" system="" selection="l " text=" " key="[Back]" SHIFT="False" CTRL="False"
```

Figure 4: Example of Qualitivity's logging file.

time and effort.

Acknowledgments

The research leading to these results has received funding from the Spanish Centre for Technological and Industrial Development (Centro para el Desarrollo Tecnológico Industrial) (CDTI) and the European Union through Programa Operativo de Crecimiento Inteligente (Project IDI-20170964). We gratefully acknowledge the support of NVIDIA Corporation with the donation of a GPU used for part of this research.

References

Vicent Alabau, Ragnar Bonk, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes García-Martínez, Jesús González-Rubio, Philipp Koehn, Luis A. Leiva, Bartolomé Mesa-Lao, Daniel Ortiz-Martínez, Hervé Saint-Amand, Germán Sanchis-Trilles, and Chara Tsoukala. 2013. CASMACAT: An open source workbench for advanced computer aided translation. *The Prague Bulletin of Mathematical Linguistics*, 100:101–112.

Ana Guerberof Arenas. 2008. Productivity and quality in the post-editing of outputs from translation memories and machine translation. *Localisation Focus*, 7(1):11–21.

Robert Dale. 2016. How to make money in the translation business. *Natural Language Engineering*, 22(2):321–325.

Miguel Domingo, Mercedes García-Martínez, Álvaro Peris, Alexandre Helle, Amando Estela, Laurent Bié, Francisco Casacuberta, and Manuel Herranz. 2019. Incremental adaptation of NMT for professional post-editors: A user study. In *Proceedings of the Machine Translation Summit*. Under publication.

Marcello Federico, Nicola Bertoldi, Mauro Cettolo, Matteo Negri, Marco Turchi, Marco Trombetti, Alessandro Cattelan, Antonio Farina, Domenico Lupinetti, Andrea Martines, Alberto Massidda, Holger Schwenk, Loïc Barrault, Frederic Blain, Philipp Koehn, Christian Buck, and Ulrich Germann. 2014. The matecat tool. In *Proceedings of the 25th International Conference on Computational Linguistics: System Demonstrations*, pages 129–132.

Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, et al. 2018. Achieving human parity on automatic chinese to english news translation.

Ke Hu and Patrick Cadwell. 2016. A comparative study of post-editing guidelines. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, pages 34206–353.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. 2017. Open-NMT: Open-source toolkit for neural machine translation. In *Proceedings of the Association for the Computational Linguistics*, pages 67–72.

Philipp Koehn and Rebecca Knowles. 2017. Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39.

Sachith Sri Ram Kothur, Rebecca Knowles, and Philipp Koehn. 2018. Document-level adaptation for neural machine translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 64–73.

Álvaro Peris and Francisco Casacuberta. 2018. Online learning for effort reduction in interactive neural machine translation. *Accepted in Computer Speech & Language*.

Álvaro Peris, Luis Cebrián, and Francisco Casacuberta. 2017. Online learning for neural machine translation post-editing. *arXiv:1706.03196*.

Marco Turchi, Matteo Negri, M Amin Farajian, and Marcello Federico. 2017. Continuous learning from human post-edits for neural machine translation. *The Prague Bulletin of Mathematical Linguistics*, 108(1):233–244.

Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Ł. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv:1609.08144*.

Joern Wuebker, Patrick Simianer, and John DeNero. 2018. Compact personalized models for neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 881–886.