Plugging into Trados: Augmenting Translation in the Enclave

Corey Miller   corey.a.miller@nvtc.gov
Chiara Higgins   chiara@nvtc.gov
Paige Havens   paige.havens@nvtc.gov
Steven Van Guilder   steven.vanguilder@nvtc.gov
Rodney Morris   rodney.d.morris@nvtc.gov
Danielle Silverman   daniellecsilverman@gmail.com
National Virtual Translation Center, Washington, DC, USA

Abstract

The National Virtual Translation Center (NVTC) built an SDL Trados Desktop plugin allowing access to a variety of state-of-the-art machine translation (MT) and transliteration services within its enclave. While Trados users in the open environment have had access to several internet-supplied MT services for some time, this has eluded enclave users. NVTC has had access to an MT portal within its enclave featuring several government and commercial engines in a wide variety of languages, and the plugin now enables users to get results from these engines within Trados, when the translation memory threshold is not met for a translation unit (TU). Translators can customize the order of presentation for the various engines for their language, including whether they want the engine at the top of the list to automatically populate target TU halves. In addition, NVTC has added an automated transliteration feature by which a user can highlight a proper name in the source text and ask for it to be transliterated either as a person or place name; the automated transliteration service will then populate the target text with a transliteration according to the appropriate scheme. Together, the MT and transliteration functionality constitute an important advance in the transition from baseline computer-assisted translation (CAT) to augmented translation (AT), where a fuller range of human-language technology (HLT) services are placed at translators' disposal, ideally within the CAT tool itself. In both the translation and transliteration cases, translators can post-edit the automatically produced information directly within Trados. This paper documents the user pilot of the plugin that assessed user acceptance variables such as preferences for specific MT engines and interface configuration options. The pilot results also provide data points that shed light on productivity and quality impacts of post-editing in NVTC's environment.

1. Introduction

The mission of the National Virtual Translation Center (NVTC) is to "provide timely and accurate translation services to support national intelligence priorities and protect our nation and its interests."[1] In support of that mission, NVTC has constantly sought to employ and improve the state of the art in computer-assisted translation (CAT) technology. We will describe the baseline CAT environment at NVTC. Following that, we will introduce the notion of Augmented Translation (AT), which is an evolution of CAT that introduces an array of human

---

[1] https://www.fbi.gov/about/leadership-and-structure/intelligence-branch/national-virtual-translation-center

UNCLASSIFIED

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 469

language technologies (HLT) into the CAT workflow. In particular, we will describe two CAT enhancements: one for machine translation (MT), and the other for automated transliteration. We will then describe a user pilot employing the machine translation enhancement and its effectiveness and acceptance within our translator community. Finally, we will touch upon next steps in the further evolution of AT.

2. Computer-Assisted Translation

NVTC's main CAT tool is SDL Trados 2019 (Trados) coupled with SDL Trados GroupShare 2020 (GroupShare). Trados facilitates verbatim translation of a wide range of document types by leveraging translation memories (TMs) and termbases (TBs). The TMs and TBs are stored on a GroupShare server so that a geographically dispersed community of linguists (NVTC's term for human translators) can use and modify them simultaneously. This CAT environment is situated on an isolated government network which does not have connectivity to the internet. While this enclave provides security, it does pose challenges in terms of access to software and language resources.

3. Augmented Translation

Augmented Translation emerged as a notion in 2017 in publications by CSA Research[2] and Deloitte (Eggers et al., 2017). According to CSA, "Just as 'augmented reality' uses AI [artificial intelligence] to enrich individuals' access to relevant information about their surroundings, this transformation provides linguists with more context and guidance for their projects." The Deloitte group situates AT within a program of improving government processes by incorporating AI into them. They establish a hierarchy of levels of evolution of the incorporation of AI into workflows: relieve, split up, replace, augment. In the most evolved "augment" approach to translation, "translators use automated translation tools to ease some of their tasks, such as suggesting several options for a phrase, but remain free to make choices. This increases productivity and quality while leaving the translator in control of the creative process and responsible for aesthetic judgments."

In practical terms, AT is an evolution from CAT where additional HLT and project management (PM) services are seamlessly available to participants in the translation workflow. CSA mentions neural adaptive machine translation available in the CAT process as a key feature of AT. They mention that this is available both in Lilt and SDL's BeGlobal, which are part of a set of plugins offered for Trados under the rubric of Automated Translation,[3] which include access to machine translation from both SDL and Google.

Another class of HLTs CSA envisages as part of AT is Automated Content Enrichment (ACE). ACE appears to be relevant encyclopedic information that can be made available within CAT. Although not explicitly mentioned by CSA, automated transliteration is a function important to NVTC that we propose to incorporate within our local instantiation of AT.

3.1. Machine Translation Plugin

SDL's Automated Translation plugin works on the open internet where there is access to Google Translate and SDL's Language Cloud. Since part of NVTC's CAT workflow takes

---

[2] https://csa-research.com/Insights/ArticleID/140/Augmented-Translation-Powers-up-Language-Services

[3] https://docs.sdl.com/783545/577209/sdl-trados-studio/automated-translation

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 470

place in an enclave, those MT sources are not accessible. However, NVTC has access to Symphony, which is an MT portal featuring several state-of-the-art commercial-off-the-shelf (COTS) and government-off-the-shelf (GOTS) MT engines. The portal has both a graphical user interface (GUI) as well as application programming interface (API) access. GUI users can paste in foreign language text and get English translations using each featured engine (or a subset). Both the GUI and API features offer optional text language identification as a preliminary step to translation. NVTC sought to provide access to Symphony in Trados via a plugin.

SDL has a free software development kit (SDK)[4] that developers can use to build their own plugins, and an active developer and user Community[5] which serves as a forum for discussion. NVTC engaged The MITRE Corporation, which had developed Symphony, to develop the plugin prototype. Similar to the other Automated Translation plugins, the Trados Symphony Plugin allows users to add MT engines from Symphony to their projects alongside other translation resources. Figure 1 shows the Trados editor window with plugin. When no TM match for a given source phrase is found above a user-specified threshold, the MT engines are displayed for the selected segment at the top of the editor window. Users are able to select which available engines are displayed and in what order. The engine in first position can also optionally autopopulate the target side of the editor or the user can choose to start with empty segments when no TM match is found. In this case, users are able to use mouse or keyboard to paste any of the MT results to the target segment for post-editing as they see fit.
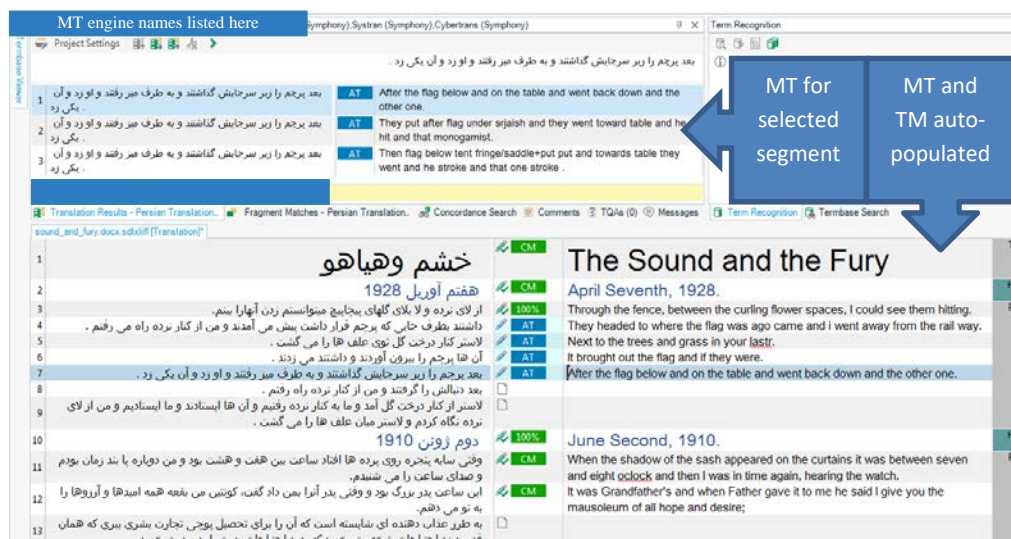

Figure 1. Trados editor with MT plugin

3.2. Automated Transliteration Plugin

Since NVTC's workflow involves the translation of multiple languages into English, the notion of transliteration is important. Transliteration refers to rendering foreign names and places, among other things, from foreign alphabets into the English alphabet. NVTC's Style Guide requires that person names be transliterated according to the Intelligence Community (IC)

---

[4] https://appstore.sdl.com/language/developers/sdk.html

[5] https://community.sdl.com/

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 471

standard for each language, and that place names be transliterated according to the Board of Geographic Names (BGN) transliteration standard for languages where those standards exist.[6]

In order to assist linguists with generating transliterations according to the standards, NVTC has employed the Rosette Name Translator (RNT, transliterator in our parlance) from Basis Technology.[7] Although RNT is available in Symphony, it was thought that incorporating its functionality directly into Trados would facilitate linguists' use of the tool. Accordingly, a second version of the machine translation plugin was created to include access to RNT. RNT supports several different transliteration standards including IC and BGN, and the plugin can be configured to recognize a preferred standard for each language.

Figure 2 shows examples of the transliteration selection windows inside the Trados editor. First the user right-clicks on a selected person or place name in the source text which pulls up a Trados menu where the user chooses whether to transliterate the name as a person or place. In the case of multiple possible transliterations, the linguist selects the most correct one and it is placed at the next position in the target text, which the linguist can then post-edit. A user-configurable option causes the preferred transliteration for a language to automatically populate the target side of the Trados editor after the linguist selects whether it is a person or a place name.

3.3. Global Plugin Configuration

The MT and transliteration functionalities are now incorporated in a single plugin whose behavior is governed by a configuration file in a JavaScript Object Notation (JSON)[8] format. The
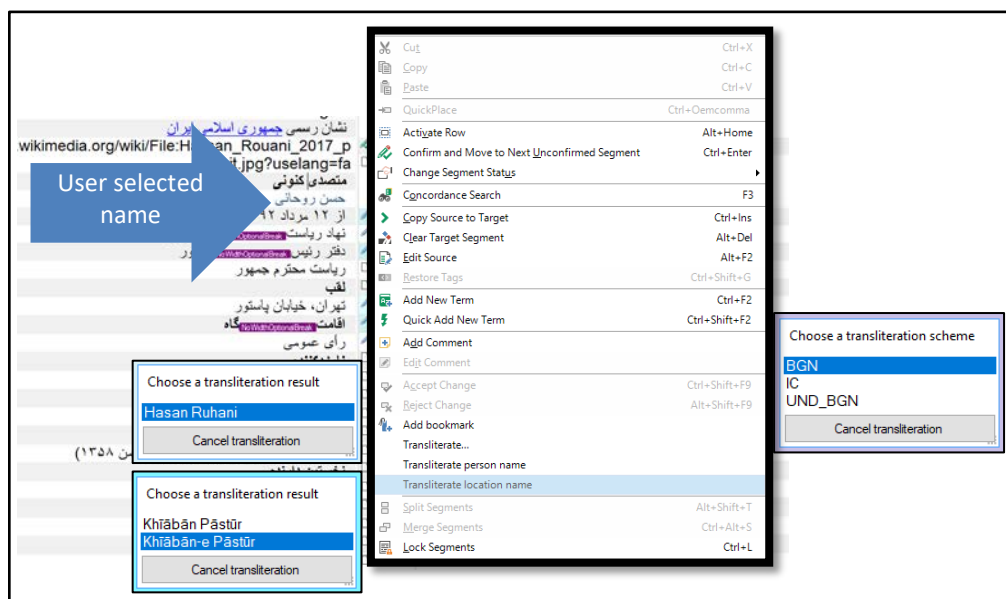

Figure 2. Sample transliteration plugin interface windows

configuration file has sections for both MT and transliteration and has been specified once for

---

[6] https://geonames.nga.mil/gns/html/romanization.html

[7] https://www.basistech.com/text-analytics/rosette/name-translator/

[8] https://www.json.org/json-en.html

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 472

NVTC, so linguists do not need to modify it. One challenge that was resolved by this configuration approach was interoperability issues caused by the wide variation in language naming conventions. The configuration file maps between language name variants used by the various MT vendors, RNT, Trados and NVTC. The language name mapping ensures that each of the plugin technologies delivers the correct language-specific functionality to the user. For transliteration, the configuration file specifies what the NVTC-approved person and place name schemas are for each language.

The project and user specific configuration options for MT and transliteration are described in sections 3.1 and 3.2.

4. Operational Pilot

From October 2019 to February 2020, NVTC conducted an operational pilot using the Symphony Trados (machine translation) plugin. The pilot participants used the MT-only version one plugin since alpha testing on version two with transliteration features was not complete until near the end of the pilot period. The pilot was fully integrated into NVTC's day-to-day operations with the linguists using the plugin on regularly assigned operational work. The pilot linguists were not required to complete extra tasks or reporting for the pilot aside from attending an initial training and then completing one or two fifteen-minute surveys. This approach provided ample opportunity to assess how full operational use of the plugin would impact linguists and their work, but somewhat limited the measurements that could be collected. For example, we did not stage translation of similar documents for comparison with and without the plugin or have linguists retranslate prior work completed with terminology and translation memory support only. These sorts of staged tests can allow detailed productivity comparisons and help determine whether linguists were over- or under-editing MT. Instead, this pilot focused on: 1) gauging whether the plugin could be integrated into NVTC's workflow effectively, 2) the readiness of the MT and plugin interface to enable post-editing of MT, and 3) identifying which configuration and implementation approaches are most effective.

The pilot involved 11 linguists who translated 43 documents in nine languages: Arabic, Chinese, English, Farsi, French, Korean, Russian, Spanish and Turkish. The most prevalent domains across pilot documents were defense, engineering, cyber and economics. Many of the source documents were MS Word or other text-based documents, but spreadsheets, PowerPoint slides and OCR output documents were also included in the pilot. The pilot included documents that were suspected to be non-ideal for MT but represented typical NVTC operational work.

The primary data collection for pilot results was a user survey (see Appendix A for the complete user survey). Fourteen surveys were collected. One linguist dropped out of the pilot before completing a survey because of an operational work change, and linguists participating in the pilot for longer periods of time completed a second identical survey later in the pilot. To the extent possible, finished translations, source documents and full MT output for all available engines were collected to calculate bilingual evaluation understudy (BLEU) scores in order to gain insights into any correlation between BLEU scores and post-editing machine translation (PEMT) outcomes.

4.1. Introducing Post-editing to Operational Linguists

Motivation for the pilot arose from previous shortcomings with PEMT on NVTC operational data, due primarily to available MT engines. For example, a 2016 NVTC PEMT study indicated that available MT would perform poorly on NVTC data and study participants experienced productivity declines when post-editing MT (Richerson, 2016). By 2018, when the idea of a custom MT Trados plugin was conceived, neural machine translation was on the rise and MT

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 473

with significantly improved performance was beginning to be available to government. Furthermore, the plugin user interface would be designed to provide several variant MT results as a resource to linguists without requiring PEMT. In cases where MT quality is poor, the user would be able to start with empty target translation cells when no TM match is available. In this case, there would be no time lost to deleting inadequate MT results or lengthy post-editing, but MT results would now be visible and potentially save time by allowing the linguist to get the gist of a segment or discover alternative wording that would not have been available without MT. The MT autopopulation feature, on the other hand, would be useful when MT performs well on a document. In these cases, PEMT of a target segment that would be otherwise empty has the potential to result in significant productivity gains. This final conclusion is based on studies of productivity gains from PEMT of high-quality MT versus starting from scratch in commercial settings (Escartin, 2015).

Before rolling out the plugin operationally for this pilot, it was important to show that the plugin would be accepted by users and improve translation outcomes. Additionally, we hoped to gather information about how to most effectively use the plugin by answering questions about the best approach to configuring the linguist MT environment. The linguists who participated in the pilot all had prior SDL Trados experience but were not screened for prior PEMT experience. Linguists were trained using custom training that included PEMT best practices from the TAUS MT Post-editing Guidelines[9].

One key objective of the pilot user survey was gaining insight into optimizing plugin configurations for the users. One set of survey questions was aimed at determining whether it was better to autopopulate MT into the target translation window so that linguists must either post-edit the segment or delete it—or instead to simply leave the linguists to view the MT options, using the mouse or keyboard command to bring the displayed MT into the translation window as needed. Other questions related to the linguists' preferred display order of MT engines and choice of MT engine(s).

4.2.   PEMT Acceptance and Preferences

All pilot linguists' surveys reported that the plugin was helpful, and many described specific benefits such as reduced time to complete translations. All linguists found it helpful to have the alternative MT engine results display as a reference and most indicated that seeing the different variants saved them time in either word or grammar choices. For certain documents, however, the survey indicated that the plugin was less helpful due to poor MT outputs.

Eight of the 11 linguists chose to try having the MT autopopulated at least part of the time. Linguists reported completely deleting anywhere from 20 – 100% of autopopulated MT segments depending on the document and language. On average, the linguists chose to post-edit 36% of the autopopulated sentences during the pilot and for the other 64%, they used the MT as a reference and wrote the sentence themselves from scratch. Of the linguists who chose to never autopopulate the MT segments, all but one indicated that they sometimes pasted an MT segment and post-edited it depending on the quality of the MT output.

4.3.   MT Quality

The Symphony MT portal provides MT from three different commercial providers in addition to one GOTS MT system. For two of the commercial providers, Symphony initially hosted both

---

[9] https://www.taus.net/think-tank/reports/postedit-reports/taus-post-editing-guidelines

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 474

a new neural MT system and an older non-neural system. One of the older solutions was removed part way through the pilot meaning that up to five MT engines were available via the plugin throughout the pilot depending on the language. The default configuration for the plugin displayed MT results in the order provided in Table 1. Since the plugin allows autopopulating of MT results from the first displayed engine only, linguists could request they be re-ordered. The surveys asked the linguists to rate each MT engine on a five-point scale with five reflecting the highest quality. Note that the default ordering of MT options within the CAT display could have influenced linguist judgment since it is generally easiest to use the first option.

Solution Name	No. of Raters	Average User Rating	Average BLEU
COTS MT A (neural)	10	4.3	31.6
COTS MT B (neural)	8	3.38	26.7
COTS MT C (neural)	7	4.14	28.6
COTS MT D	7	2.86	19.6
COTS MT E	4	3.5	N/A[10]
GOTS MT F	5	2.4	10.68

Table 1. Linguist judgment of available MT versus BLEU

In addition to survey data reflecting linguist judgment of the MT, BLEU scores were calculated on each of the available MT outputs for 27 of the 43 documents translated during the pilot. This involved saving full document MT from each available MT engine and then using the finished translation as the reference document for the BLEU calculation. We used the finished translation as the sole reference translation in the BLEU calculations, and consequently scores are likely lower than would be achieved using multiple reference translations.

BLEU results are shown in Table 1 alongside the ratings linguists provided via survey. A correlation between BLEU and linguist ratings is apparent by inspection of the table and noting that both measures rank the MT solutions in the same order. This in turn confirms that BLEU roughly maps to the NVTC pilot linguists' judgment of which MT results were useful.

The ratings in Table 1 are averages across all languages but there is significant variation in scores across and within languages. For example, COTS MT B (neural) achieved the highest BLEU scores for Chinese documents, but the scores varied from 13.2 to 39.9 between documents. This means that the tradeoff decision of whether to post-edit or use MT only as a reference will vary depending on the document even within one language.

In the 2016 NVTC study on post-editing, the BLEU scores of the MT used ranged from 9.05 to 14.98. One of the conclusions of that study was that improved MT was needed for effective post-editing, and now that has become a reality with BLEU scores peaking over 60 for some of the language/document combinations in this pilot. Furthermore, that study only covered strict post-editing where the MT was only available through prepopulating in the target translation blocks. The option, provided by the new plugin, of displaying several MT options without prepopulating, provides a viable solution for cases where the MT is performing poorly on a particular group of documents.

---

[10] COTS MT E was decommissioned prior to the end of pilot, which prevented collection of full MT outpluts for BLEU calculations.

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 475

5. Conclusions and Future Work

Overall, the pilot results of the Trados Symphony plugin were favorable. The CAT integrated PEMT interface, the display of multiple MT results to provide alternatives, and the increasing quality of available MT led to overall user acceptance of the plugin during the pilot. The pilot demonstrated that the plugin can be implemented and PEMT adopted without disruption to the existing NVTC workflow and provided information to support successful adoption of the plugin at NVTC. To the extent that the pilot results turn out to be predictive, it is likely that around 35% of plugin enabled translation will be PEMT, potentially resulting in productivity gains due to linguists choosing PEMT when the MT only needs light post-editing. The other 65% of plugin enabled translations will benefit from the addition of MT output as a reference. The results data provided a wealth of information that can be further analyzed for more granular NVTC-specific insights beyond the scope of this paper.

One important conclusion of the pilot is a reaffirmation of the need to provide training to linguists on post-editing best practices to enable them to make good decisions such as when it is worth performing PEMT versus translating from scratch with MT used for reference purposes. Despite recent dramatic improvement in MT quality, the pilot showed that MT still performs poorly in many cases. Follow-on work to better identify those cases where MT fails can enable a two-pronged approach:1) identifying in advance whether or not a document is a good candidate for PEMT could facilitate plugin configuration decisions and increase efficiency, and 2) identifying and prioritizing languages and domains for MT improvement could increase the benefit of MT over time.

The plugin transliteration features, while not operationally assessed during the pilot due to development timing, are integrated into version two of the plugin to be operationally tested in 2021. NVTC is considering other efforts in the direction of AT, including the possible incorporation of named entity recognition (NER) into Trados. The form this might take is that person and place names would be identified on the source side of the editor and colored differently, thus facilitating linguists' subsequent identification of the names and use of the transliteration features of the plugin.

References

Eggers, W., Schatzky, D. and Viechnicki, P. (2017). AI-augmented government: Using cognitive technologies to redesign public sector work. Deloitte University Press.

Escartin, C. and Arcedillo, M. (2015). Machine translation evaluation made fuzzier: A study on post-editing productivity and evaluation metrics in commercial settings. In *Proceedings of MT Summit XV, vol.1: MT Researcher's Track,* pages 131 -144, Miami, FL. U.S.A.

Massardo, I., van der Meer, J., O'Brien, S., Hollowood, F., Aranberri, N. and Drescher, K. (2016). *MT Post Editing Guidlelines.* Taus Signature Editions.

Richerson, E. (2016). On Inserting Post-Editing Machine Translation at National Virtual Translation Center: An Operational Pilot.

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 476

Appendix A: Symphony Trados Plugin Pilot User Survey

Name:
Date:
Project ID:
What language are you translating from?
What language are you translating into?
Domain/Genre of Source Document:

If MT was prepopulated
1) Did you delete any of the MT segments and translate from scratch?
2) If you answered yes to question one, approximately what percentage of MT segments did you delete?
3) Please describe the most helpful and/or unhelpful aspects of post editing MT?

If MT was not prepopulated
4) Did you insert any of the MT results into the target segments instead of translating from scratch?
5) If you answered yes to question four, which translation engine results did you insert most often?
6) If you answered yes to question four, please describe the most helpful and unhelpful aspect of post editing MT.

All pilot users
7) Did you find it helpful to have MT results displayed as a reference?
8) Please rate the quality of each MT engine where a rating of 5 is fantastic and 1 is awful.
9) Did you (or a PM or LTT rep) change the setting for MT display at any point while you were using it?
10) If you answered yes to question nine, please indicate what change was made and why.
11) Overall, did you find the MT output helpful, and if yes, how?

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track

Page 477