



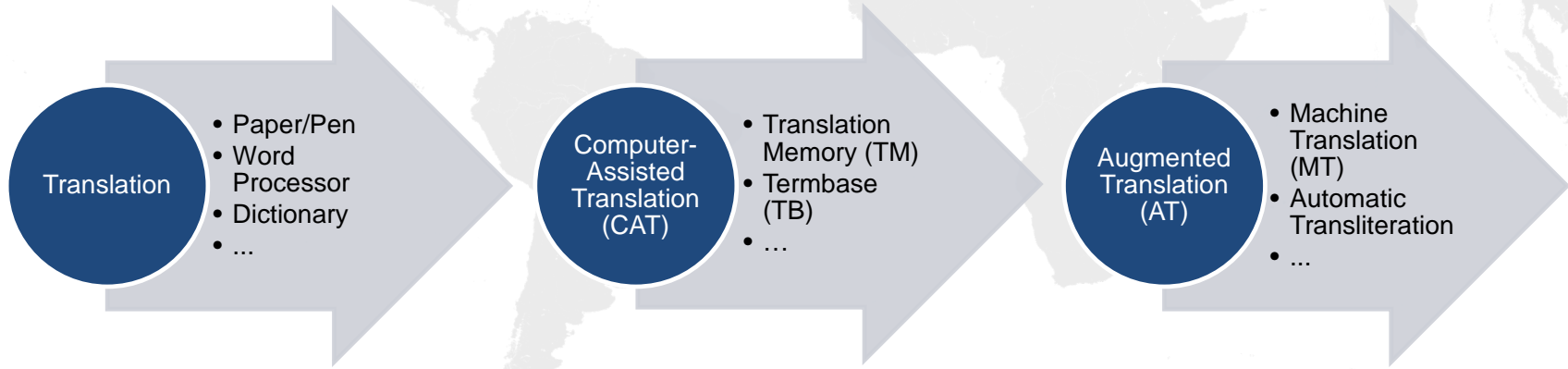
NVTC
National Virtual Translation Center

Plugging into Trados: Augmenting Translation in the Enclave

**Corey Miller, Chiara Higgins, Paige Havens, Steven Van Guilder,
Rodney Morris and Danielle Silverman**

October 8, 2020

The Evolution of CAT into AT



AT with MT on the Internet



- **SDL Automated Translation**
 - BeGlobal
 - Could be used in an enclave
 - Language Cloud
 - Google Cloud
- **Lilt**
 - On prem solution in development

AT in the Enclave



- NVTC uses Symphony
 - MT portal featuring several COTS and GOTS engines
 - Transliteration portal featuring Basis Rosette Name Translator (RNT)
- NVTC developed Trados plugin
 - Symphony MT when no sufficient TM match is found
 - Person & place name transliterations on demand

Machine Translation Interface



Translation Results - Persian Translation Memory **MT engine names listed here** Term Recognition

Project Settings

بعد پرچم را زیر سرچایش گذاشتند و به طرف میز رفتند و او زد و آن یکی زد .

1	بعد پرچم را زیر سرچایش گذاشتند و به طرف میز رفتند و او زد و آن یکی زد .	AT	After the flag below and on the table and went back down and the other one.
2	بعد پرچم را زیر سرچایش گذاشتند و به طرف میز رفتند و او زد و آن یکی زد .	AT	They put after flag under srjaish and they went toward table and he hit and that monogamist.
3	بعد پرچم را زیر سرچایش گذاشتند و به طرف میز رفتند و او زد و آن یکی زد .	AT	Then flag below tent fringe/saddle+put put and towards table they went and he stroke and that one stroke .

MT for selected segment

MT and TM auto-populated

Translation Results - Persian Translation... Fragment Matches - Persian Translation... Concordance Search Comments TQAs (0) Messages Term Recognition Termbase Search

sound_and_fury.docx.sdxliff [Translation*]

1	خشم وهیاهو	CM	The Sound and the Fury	T
2	هفتم آوریل 1928	CM	April Seventh, 1928.	H
3	از لای نرده و لا بلای گلهای بیچایبج میتوانستم زدن آنها را بنم.	100%	Through the fence, between the curling flower spaces, I could see them hitting.	P
4	داشتند بطرف جایی که پرچم قرار داشت پیش می آمدند و من از کنار نرده راه می رفتم .	AT	They headed to where the flag was ago came and i went away from the rail way.	
5	لاستر کنار درخت گل توی علف ها را می گشت .	AT	Next to the trees and grass in your lastr.	
6	آن ها پرچم را بیرون آوردند و داشتند می زدند .	AT	It brought out the flag and if they were.	
7	بعد پرچم را زیر سرچایش گذاشتند و به طرف میز رفتند و او زد و آن یکی زد .	AT	After the flag below and on the table and went back down and the other one.	
8	بعد دنبالش را گرفتند و من از کنار نرده راه رفتم .			
9	لاستر از کنار درخت گل آمد و ما به کنار نرده رفتیم و آن ها ایستادند و ما ایستادیم و من از لای نرده نگاه کردم و لاستر میان علف ها را می گشت .			
10	دوم ژوئن 1910	100%	June Second, 1910.	H
11	وقتی سایه پنجره روی برده ها افتاد ساعت بین هفت و هشت بود و من دوباره با بند زمان بودم و صدای ساعت را می شنیدم.	CM	When the shadow of the sash appeared on the curtains it was between seven and eight o'clock and then I was in time again, hearing the watch.	P
12	این ساعت پدر بزرگ بود و وقتی پدر آنرا بمن داد گفت، کونتین من بقیه همه امیدها و آرزوها را به تو می دهم.	CM	It was Grandfather's and when Father gave it to me he said I give you the mausoleum of all hope and desire;	
13	به طرز عذاب دهنده ای شایسته است که آن را برای تحصیل پوچی تجارت بشری ببری که همان قد، بد، احتیاجات، شخصیت، بخت، که بد، احتیاجات، بدت، با بد، بدت، خه، د.			

Transliteration Interface Examples



The screenshot displays a document with Persian text. A blue arrow points to the name "حسن روحانی" (Hasan Rouhani) with the text "User selected person name". A context menu is open over the text, listing various actions such as Cut, Copy, Paste, and Transliterate person name. Three dialog boxes are shown, each titled "Choose a transliteration result". The first dialog shows "Hasan Ruhani" selected. The second dialog shows "Khīābān Pāstūr" and "Khīābān-e Pāstūr" as options. The third dialog shows "BGN", "IC", and "UND_BGN" as transliteration schemes, with "BGN" selected. A "Cancel transliteration" button is present in each dialog.

Operational Pilot



- October – February 2020
- Integrated with operational work – no staged tasks
- 11 linguists, 43 documents, 9 languages
- User survey
- BLEU scores calculated against pilot documents

Post-editing MT at NVTC



- Small 2016 NVTC study showed slight productivity declines with PEMT
 - MT quality issue
 - Linguists found MT results useful but not forced post-editing
- Why pilot?
 - Ensure plugin would achieve user acceptance and enhance NVTC translation workflow
 - Learn best practices to optimize the use of MT via the plugin
 - Better understand the quality of available MT
 - Better understand how the quality of MT provided by the plugin impacts NVTC translation outcomes

User Acceptance Results



- All pilot linguists reported that the MT plugin was helpful
- Specific reported benefits
 - The display of variant translations by different MT engines saved time in either word or grammar choices and sometimes provided better alternatives that linguist would not have considered
 - When MT quality was good it was a big time-saver especially combined with TM and terminology resources already in Trados
- Unhelpful aspects of MT reported
 - Most of these were about the scenarios where MT is poor
 - Sample of specific issues individual linguists reported
 - Lack of appropriate transliteration by MT (should be solved by version 2)
 - Source document still has post-OCR errors that propagate to MT

User Post-editing Results



- Is auto-population useful?
 - All available MT results are always displayed at the top of the CAT editor window and can be pasted into the target translation cell for post-editing
 - Auto-population inserts the results of the first listed MT engine in translation field forcing a post edit or delete decision
- Results
 - 8 of 11 linguists chose having MT auto-populated at least part of the time
 - Linguists deleted the auto-populated MT segments between 20% and 100% of the time depending on the document.
 - On average, pilot linguists who chose auto-populate MT chose to post edit 36% of the time and deleted 64% of the auto-populated segments
 - 3 linguists never chose auto-population but 2 of those 3 indicated that they sometimes pasted the MT segment into the target translation field and post-edited

Machine Translation Quality



- The linguists rated each MT engine available to them on a scale of 1 to 5 with 5 being the highest
- The number of raters varied due to the availability of MT per language and how many MT results linguists chose to display
- BLEU scores were calculated for 27 of 43 pilot translations using the finished translation as the single reference translation
- BLEU scores are averages across all languages
- BLEU results roughly map to linguist judgment

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*
GOTS MT F	5	2.4	10.68

Looking Ahead



- Pilot demonstrated that MT plugin can be implemented without disruption to current NVTC workflows
- Actual PEMT will likely be around 35%, with all plugin-enabled jobs benefiting from multiple MT outputs for reference*
- Version 2 with transliteration to be operationalized in 2021
- NVTC is considering other efforts in the direction of AT including incorporation of named entity recognition (NER) into the plugin
 - Person and place names would be automatically identified and colored differently to alert the linguist that transliteration is needed
 - Would facilitate linguists' subsequent use of the transliteration plugin

*Only applies to jobs that can be done in Trados in languages for which MT engines exists.

UNCLASSIFIED



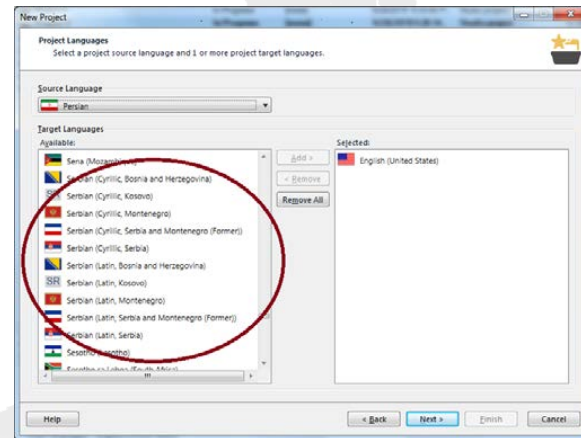
Backup Slides

UNCLASSIFIED

Transliteration Challenges



- Mapping between Trados locales (flags) and language pair script information
 - Only a subset of languages have relevant script-based locales, like Serbian Cyrillic vs. Latin



“Overloading” Locales



- We require some non-canonical scripts like Latinized Russian, Arabic and Persian that don't have such locales, so we repurpose locales like English (St. Vincent and the Grenadines)
 - We're already using arbitrary locales per language, e.g. English (United States)

