Translation Technology in Action: A US Government Use Case

Vanesa Jurica

AMTA 2014

Beregovaya et al. (Eds.) Proceedings of AMTA 2014, vol. 2: MT Users Vancouver, BC © The Authors

MITRE

Overview

- **Translation technology challenges in the US Government (USG)**
- **The story of a USG Language Services Provider (LSP)**
  - How translation technology was inserted
  - Operational tools
  - Some HLT Metrics
  - Workflow
- **Lessons Learned**
- **Questions**

MITRE

Why is translation technology adoption in USG hard?

- **CAT tools built for the commercial use case in mind**
  - Control of source documents and their format
  - Control of authorship
  - Domains and terminology narrow and well-defined
  - From one language to many
- **USG use case**
  - No control of source documents or formats
  - No control of authorship
  - Varied domains that change frequently
    - Makes terminology management problematic
  - From many languages to English (mainly)
    - Languages change constantly; many are low resource

MITRE

Why is translation technology adoption in USG hard? *(cont.)*

So, how much electronic text and repetition are we talking about in USG?

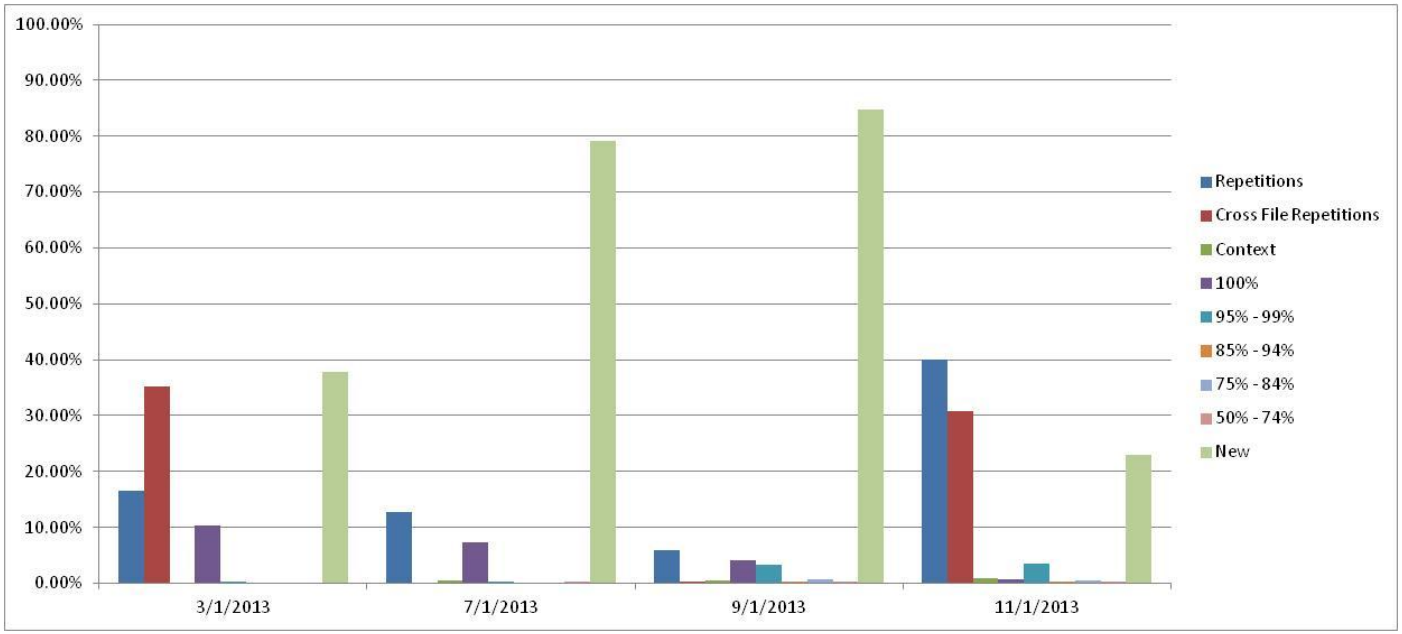
FY	PDF Percentage
2012	60%
2013	28% ← Anomaly!!
2014	68%

Percentage of source documents received in PDF

MITRE

Why is translation technology adoption in USG hard? *(cont.)*

- **What about repetition?**



Example of Variability in Repetitive Content

MITRE

Why is translation technology adoption in USG hard? *(cont.)*

- **Complicated software accreditation guidelines**
- **Translator workforce not centralized; lack of translation management staff**
- **Reluctance/concern to change existing processes**
- **Lack of awareness about translation technology in general, and translation memory in particular**

Translation Memory is NOT Machine Translation

- **Lack of knowledge about professions/skill sets needed to successfully use translation technology**

MITRE

However….

- **Even with the complications of the USG use case, translation technology *can be used successfully* if appropriately applied**
  - It will **not** be applicable to all documents, *but that's OK*!

- **Translation technology, especially translation memory, enabled this USG LSP to:**
  - Achieve significant productivity gains
  - While keeping the staff happy and engaged, and
  - Without suffering a loss in quality

MITRE

About the USG LSP

- **Small LSP for the IC (less than 40 staff total)**
- **Majority of work: full translations**
  - some summaries and audio gists
- **Varied genres: analysis reports, S&T articles, newspaper articles, diplomatic correspondence, legal materials, etc.**
- **Varied domains and languages**
  - Change depending on the geopolitical situation; can be hard to predict
- **Format: mainly PDF; some MS Office**

MITRE

Inserting Translation Technology – General Approach

- **Start small, but start in production**
- **Build incrementally on previous successes**
- **Make the smallest possible number of changes to operational workflows**

- **Embed a team of HLT experts into production to help translation managers shepherd the technology from requirements definition to selection, insertion and user acceptance**

MITRE

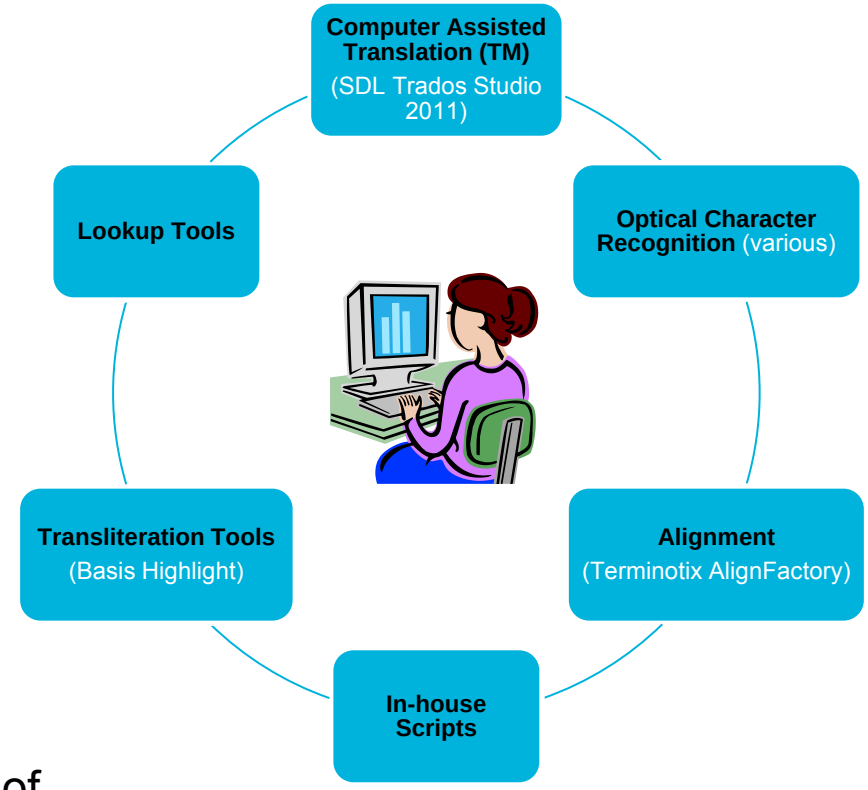
Translation Technology in Operational Use

Linguist productivity tools:

- **Computer Assisted Translation (CAT):**
  Project management, translation memory, search, autosuggest dictionaries, formatting support
- **Transliteration tools:**
  For personal names; ensure IC standard adherence
- **Lookup tools:**
  Help linguists find target equivalents faster

Supporting tools:

- **Optical Character Recognition (OCR):**
  Converts images to text
- **Alignment:**
  Creation of translation memories out of legacy data
- **In-House Scripts:**
  Facilitate use of other tools



MITRE

Computer Assisted Translation (CAT) Insertion

- **Started with a 3-month limited operational pilot (working with real data)**
  - Government selection of product (SDL Trados Studio 2011) based on extensive studies, evaluations and market surveys
  - 5 licenses (2 languages, 4 linguists, 1 task manager)
  - Pilot was successful
    - No disruptions to operations, minimal changes to workflow
    - Quality metrics somewhat higher
    - Both linguists and task manager had positive experience
    - Productivity gains obvious
- **Next, grew capabilities after evaluating pilot results**
  - More licenses and languages
  - Ongoing in-house training

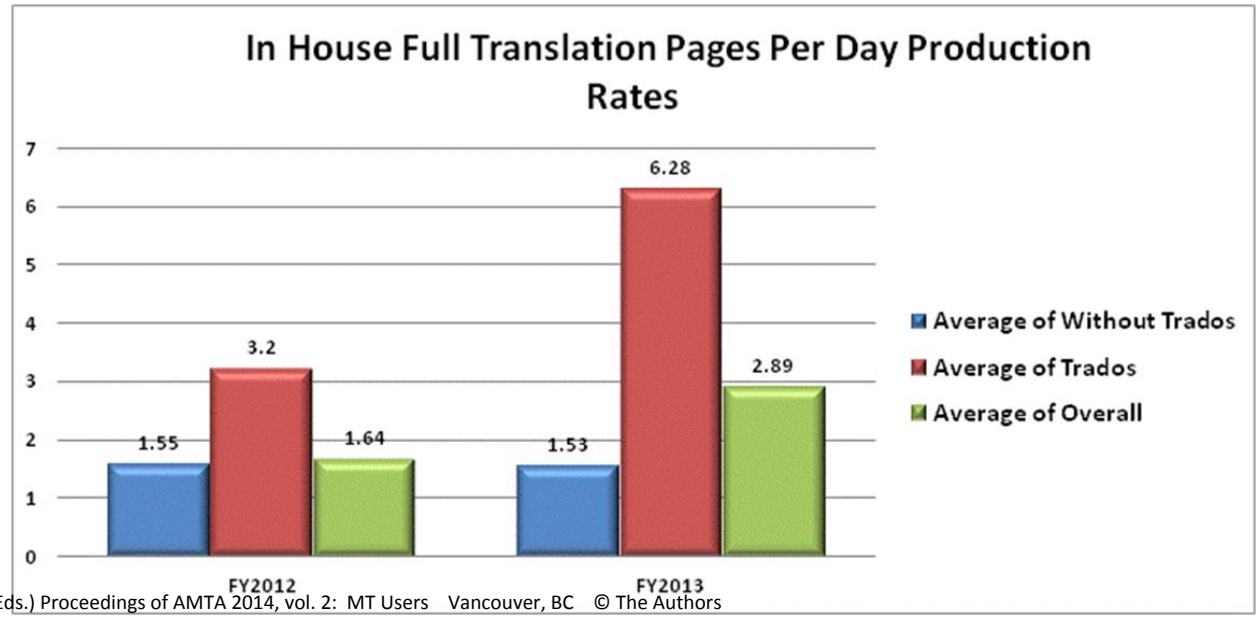
MITRE

Current State

- **Upgraded to enterprise solution - SDL WorldServer (WS)**
  - Provides centralized management of translation memories, terminologies and projects
  - Facilitates terminology management (workflow; quality control)
  - Simplifies translation project creation (no need to specify which translation memory, termbase or autosuggest dictionary to use)
- **Introduced terminology management**
  - Followed same general insertion approach
    - Defined process, metadata schema
    - Trained senior linguists (terminologists) first
    - Gradually rolling out new process and technology to all linguists
- **Integration of MT (via post-editing, with customization)**

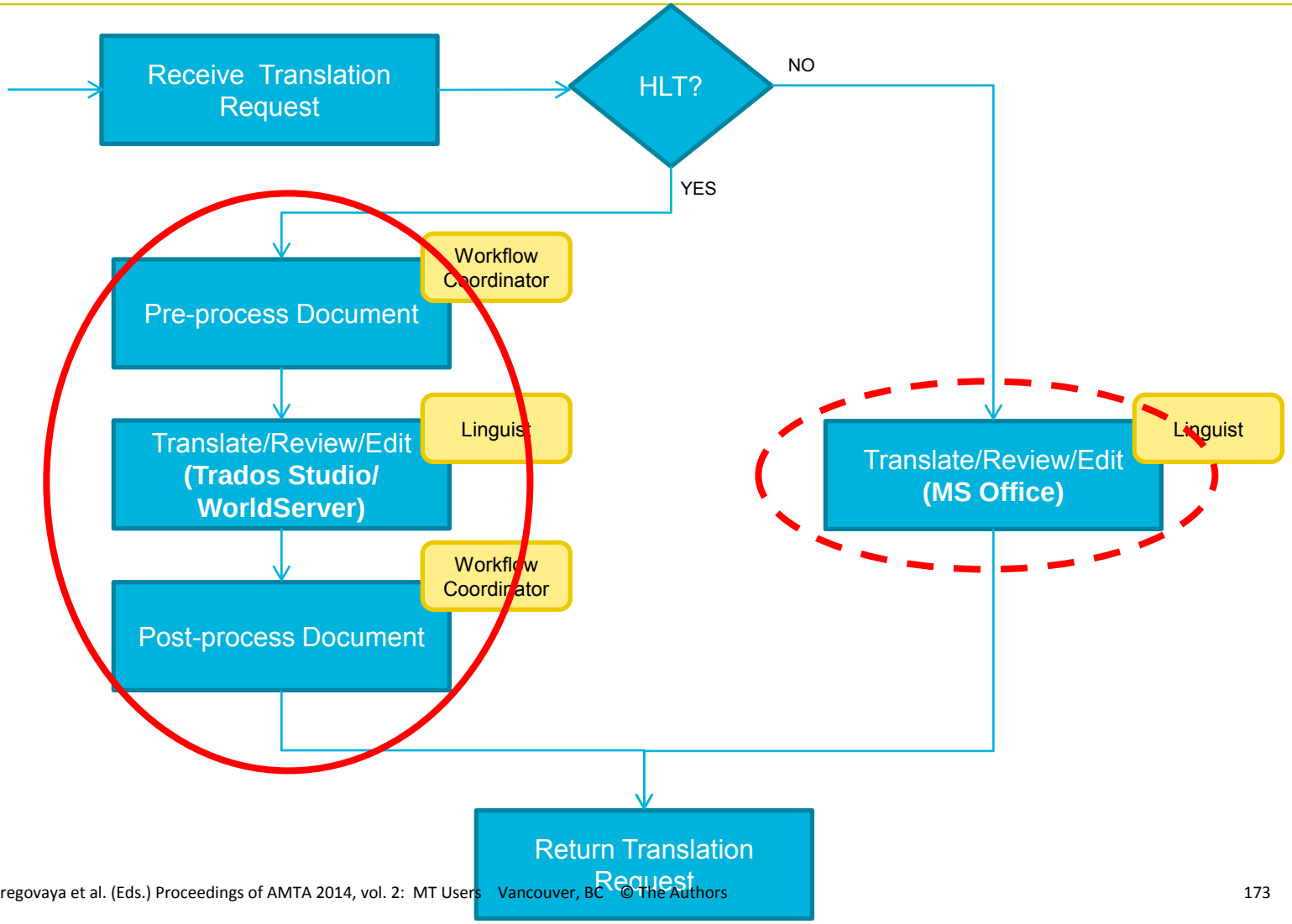
Some Metrics

- **HLT Impact Report findings (2012): For same customer, domain, document type and language direction, difference in time needed to complete translation with and without Trados was 66% (7 vs. 21 days)**
- **Overall comparison of Pages per Day (PPD) metrics for FY2012 and FY2013**



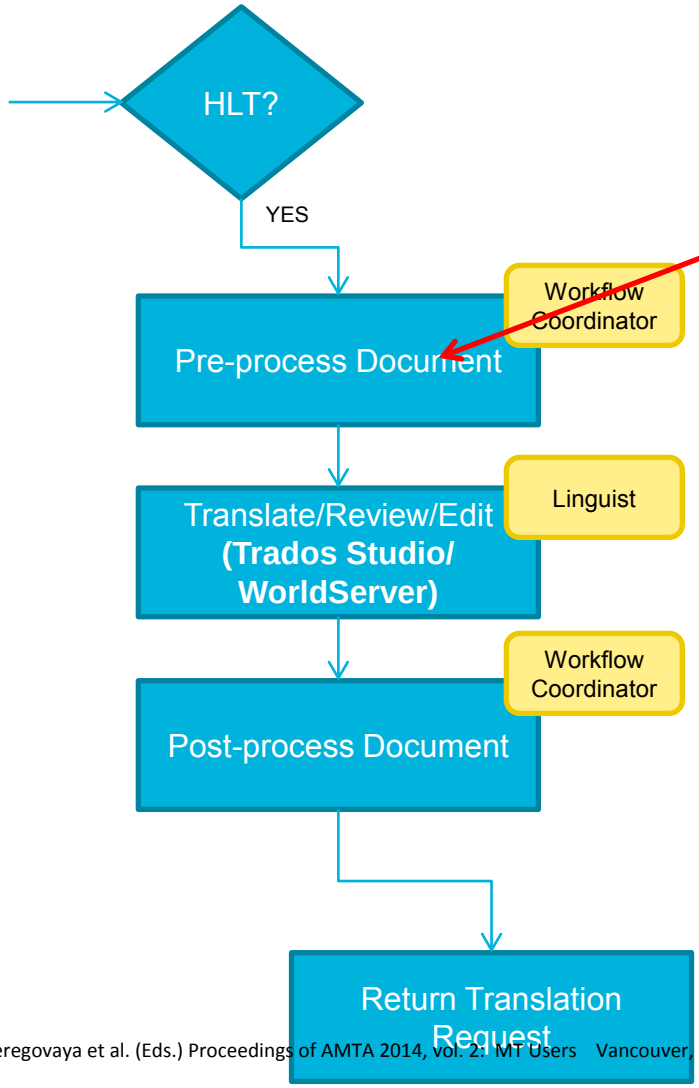
MITRE

Translation Technology Workflow



MITRE

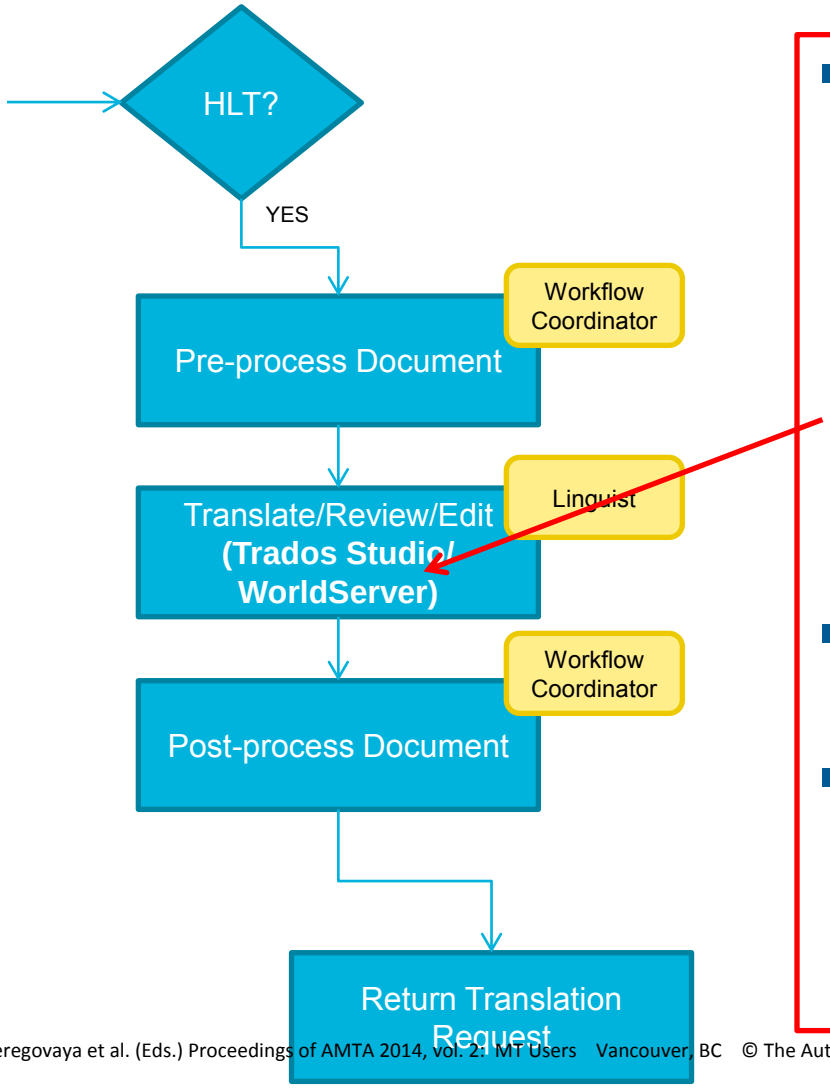
Translation Workflow with CAT *(cont.)*



- **OCR; OCR post-editing**
- **Alignment of previous translations**
- **In-House scripts**
- **Preparation for CAT**
  - Preprocessing to ensure smooth import into WorldServer/Trados Studio (e.g. kerning values, hard returns)
  - Selection of appropriate translation memories, auto-suggest dictionaries, termbanks no longer needed (Automated by WorldServer)

MITRE

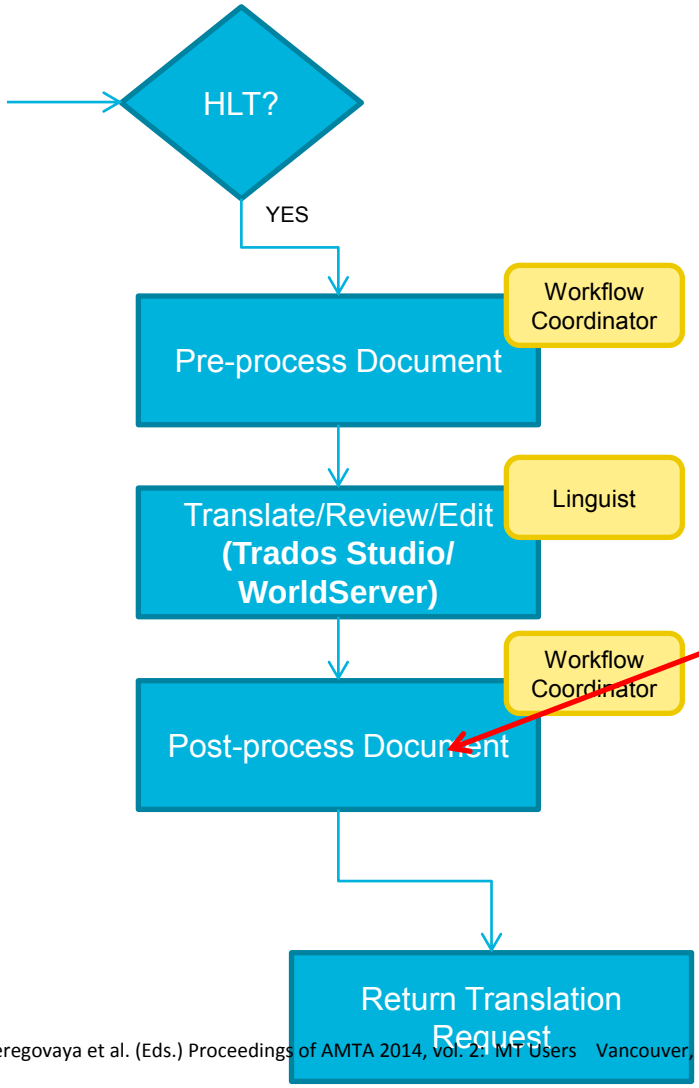
Translation Workflow with CAT *(cont.)*



- **Translation/review proceeds in Trados Studio**
  - No need to format target documents
  - Supporting language data: translation memories, autosuggest dictionaries, termbanks
- **Terminology management: MultiTerm and WorldServer**
- **TM Search in WorldServer**

MITRE

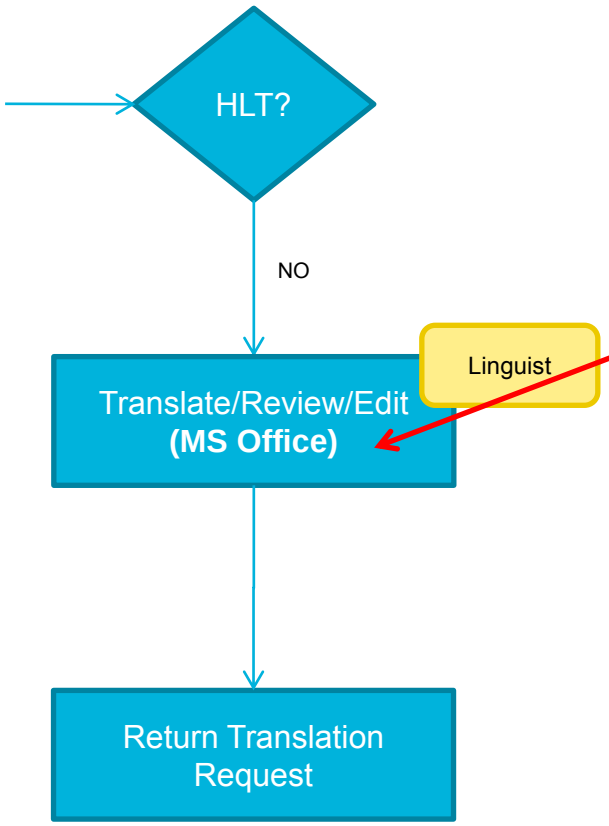
Translation Workflow with CAT *(cont.)*



- **Ensuring the look and feel is identical to the source**
  - Creating the target translation automatic in WS
- **Language data maintenance**
  - WS automatically updates master copies of translation memories
  - Re-creating auto-suggest dictionaries if needed
  - Any additional translation memory cleanup

MITRE

Translation Workflow – Limited Technology



- **Lookup tools**
- **Transliteration tool (Basis Highlight), for appropriate language**
- **WorldServer Search functionality**
- **Terminology management**

MITRE

Lessons Learned

- **Overall**
  - Start small, build incrementally
  - Stick to familiar processes wherever possible
  - Translation technology does not work for every document
  - Focus on the translation management workflow – that's the hard part!
  - Need both "strategic" and "tactical" HLT expertise
    - Strategic HLT experts use their understanding of computational linguistics and translation technology to design the overall "best practices" process
    - Tactical HLT experts help tailor the overall process to the particular organization's needs and provide day-to-day production support

MITRE

Lessons Learned (*cont.*)

- **Linguists**
  - Adjustment to technology is very fast (as soon as benefits become obvious)
  - Linguists must understand that they are not being replaced
  - Training and in-house support is crucial

- **Translation management:**
  - Needs a very specific skill set; technical skills a must
  - Workload heavier (more artifacts to maintain, pre- and post processing steps, learning new software)
  - Scripting skills an enormous advantage

MITRE

Questions?

Thank you!

Vanesa Jurica (*vjurica@mitre.org*)

MITRE