PEMT in the Public Sector: Discovery, Scoping, and Delivery

Konstantine Boukhvalov and Eileen Block

ManpowerGroup Public Sector, Inc

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

Page 484

ManpowerGroup Public Sector

- **25+** years supporting commercial and government clients in over 200 languages and dialects

  ➢ Translation and l10n, transcription, interpretation, language technology support, linguist placements, multimedia analysis and reporting

- **Super user of Human Language Technology (HLT)**

  ➢ 20 years customizing language automation/HLT tools in 60 languages to achieve efficiencies, process voluminous materials, and provide cost savings

  ➢ Leverage and adapt commercial products and combined tools to optimize technology to best meet customers' needs

  ➢ Translation Management System (TMS), Machine Translation (MT), CAT/Localization Tools, Authoring, eLearning, Desktop Publishing/Graphics Design, Audio/Video Production, Lexical Data Management, Optical Character Recognition

2

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

Page 485

Presentation Objective

- Concerns about protecting data and challenges with implementation and measuring ROI have historically prevented public sector clients from using MT, CAT, and TMS

- We will show how our team successfully met client objectives while addressing data protection concerns to develop a practical, domain-specific Post-Edited Machine Translation (PEMT) solution to enable implementation by Public Sector clients.

Key Takeaways:
- How to develop a customized PEMT solution for Public Sector

- How to build and optimize TM corpora for statistical and neural MT training

- How to measure technological and procedural efficiencies for overall program success and scalability

3

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

Page 486

Historical HLT Challenges for Public Sector (PS) Clients

- **Limited HLT use due to various contract constraints**

  ➢ No co-mingling of data, no data in cloud

  ➢ No data (TM/TB) retention

  ➢ CONUS resources with citizenship, various clearance levels

    • HLT use was not widespread among PS linguist base (freelance)

- **No process automation**

  – Longer production timelines

  – Project-based translation

4

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

Page 487

Early Steps

- Secure isolated **IT infrastructure**
- Dedicated **enterprise-level CAT** setup
- **Centralized TM/TB**
- Training for resources, e.g. **CAT-trained linguists/project managers**
- **TM/TB corpora** included as a **deliverable**

5

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

Page 488

Case Study

- **Objective:** Translate multiple domain-specific content streams with more automation and increased speed

- Large-volume legacy material alignment

- Geographically dispersed workforce

  ➢MGPS

  ➢Client stakeholders

  ➢Linguists

ROCKETS AND ALGORITHMS DID NOT LAUNCH
THE SPACE AGE.

6

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

Page 489

Program Requirements

- Centralized HLT Resources
  - ➤ Projects
  - ➤ Integrated Domain-Specific Machine Translation
  - ➤ Translation Memories/TermBases
  - ➤ Tech Support/Strict IT Infrastructure Requirements
- Integrated Project Management
- Data and Personnel Security
  - ➤ Dedicated HLT resource instance
  - ➤ Controlled human access
- Continuous MT improvement cycle
- Process automation
- Seamless integration of cloud and local-install HLT solutions

7

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

Page 490

ANSWER?
Cloud-Based Post-Edited Machine Translation

- Post-Editing CAT/MT hybrid solution in an integrated TMS environment

- The Benefits of PEMT
  - ➢ Faster processing time than CAT alone
  - ➢ Greater consistency of terminology and style
  - ➢ Future leveraging and ROI
  - ➢ Workflow customization and efficiency

8

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

Page 491

Define Stakeholders and Budget

- Dedicate a representative team of production experts – include the client!
  - ➢ Get early buy-in from the future production team
  - ➢ Start building the TMS operations culture
  - ➢ Let the production-side stakeholders define a business case and the best solution
- Align Budget and HLT options
  - ➢ Define Scope and Level of Effort - manage budget and expectations
  - ➢ Calculate HLT costs (CAT/TMS/MT)
  - ➢ Determine IT setup (local install vs. SaaS)

9

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

Page 492

Challenge

There are a growing number of strong HLT solutions. How do you select the right one, and how do you implement effectively?

10

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

Page 493

Choosing/Validating the Right Solution

- Perform preliminary research
  - ➢ *What do I need?*
  - ➢ *What are my options (commercial/custom/open source)?*
  - ➢ *What are my community peers saying?*
- Choose solution candidates
- Set up orientation calls with solution developers
  - ➢ *Identify dedicated contacts for technical and contractual questions*
  - ➢ *Explore data security options for data and support*

11

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

Page 494

Choosing/Validating the Right Solution (cont.)

- Create an evaluation matrix
  - ➤ *Use the same criteria to evaluate all products*
  - ➤ *Standard criteria include:*
    - Key features
    - Benefits
    - Shortcomings
    - Technical and contract support
    - Deployment options
    - Costs

12

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

Page 495

PUT THE DATA ASIDE –
TIME FOR WHITEBOARDING!

13

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

Page 496

Whiteboard Your Workflows

- Define and document/update existing production processes
- Do *not* adjust workflows based on the solutions' limitations
  - ➤ *If it doesn't fit, it's not right for you*
- Generate a master workflow that addresses the variations
  - ➤ *Define production steps as "required" or "optional"*
- Whiteboard other business requirements/expectations
  - ➤ *Manage expectations*

14

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 – 9, 2020, Volume 2: MT User Track
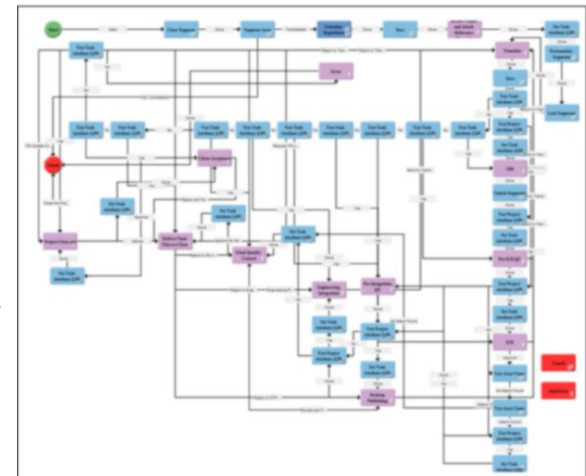
Page 497

See What "Fits" – Select and Acquire

- Combine your research with your master workflow and business requirements

- Identify the solution that provides the most value

- Generate TMS/MT Selection Report:

  ➢ Fund the acquisition and deployment

  ➢ Maintain technology knowledgebase

  ➢ Validate your decision

- Finalize the deployment plan

- Minimize the time between the acquisition and production deployment

15

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

Page 498

Initial Configuration

- Master production workflow

- Sample business rules

- Sample linguistic resources (TMs, TBs, baseline MT)

- Optional/custom components and workflow steps (forms, fields, etc.)

- Production pilot

  ➢ *Test the workflow, not just the filters*

16

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

Page 499

Document TMS/PEMT Production Procedures – Role-Specific Instructions

- Project Managers
  - ➢ *Production*
  - ➢ *Offline procedures*

- Linguist Users
  - ➢ *Production*

- Client Users
  - ➢ *Portal access/request*
  - ➢ *Production*

- Other Production Roles (as applicable)

17

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

Page 500

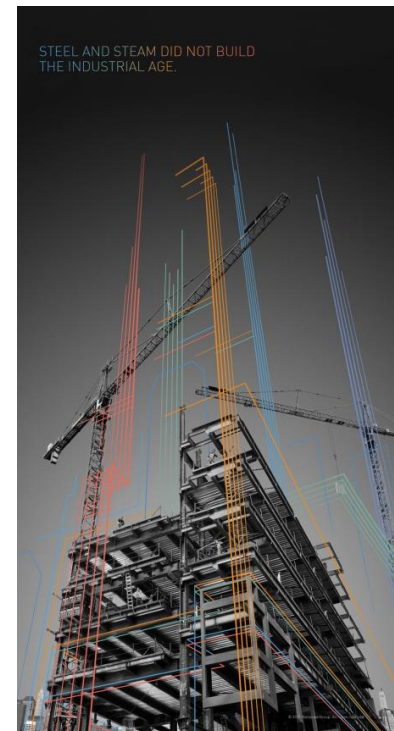# Develop and Implement Training

Client stakeholder participation and buy-in is key to project success.

- Develop a reusable curriculum
- Provide a general system overview
- Provide role-specific training

Production Deployment

- Configure and deploy client portal and other auxiliary components

- Align legacy content

- Optimize TM corpora for MT training

  ➢ Segmentation

  ➢ Markup

- Perform initial training of Domain-Specific MT engines/language pairs

- Perform first automated and human evaluation of MT – start measuring
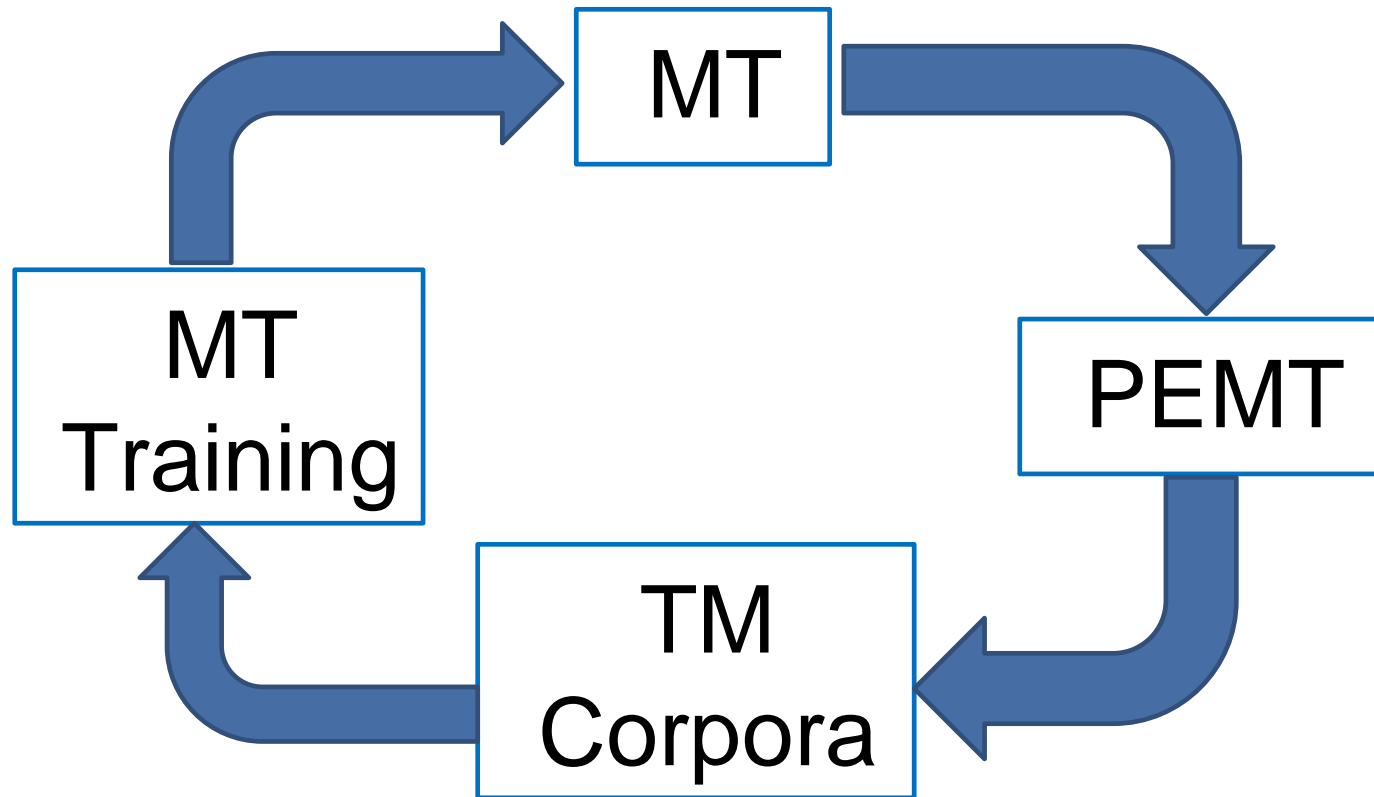
STEEL AND STEAM DID NOT BUILD
THE INDUSTRIAL AGE.

19

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

Page 502

Start PEMT!!!

- Start production for the selected Task Orders/Programs

- Adjust configuration, procedures, and documentation, as applicable
  - ➢ *Deliver the updates to the appropriate parties*

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

Page 503

Continuous MT Improvement Cycle

21

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

Page 504

Additional Automated Workflow Options

TM-MT Only

```
┌──────────────┐
│    Input     │
└──────┬───────┘
       ↓
┌──────────────┐
│ Segmentation │
└──────┬───────┘
       ↓
┌──────────────┐
│      TM      │
└──────┬───────┘
       ↓
┌──────────────┐
│      MT      │
└──────────────┘
```

TM-MT with Edit

```
┌──────────────┐
│    Input     │
└──────┬───────┘
       ↓
┌──────────────┐
│ Segmentation │
└──────┬───────┘
       ↓
┌──────────────┐
│      TM      │
└──────┬───────┘
       ↓
┌──────────────┐
│      MT      │
└──────┬───────┘
       ↓
┌──────────────┐
│     Edit     │
└──────────────┘
```

Output

22

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

Page 505

Program Launch

23

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

Page 506
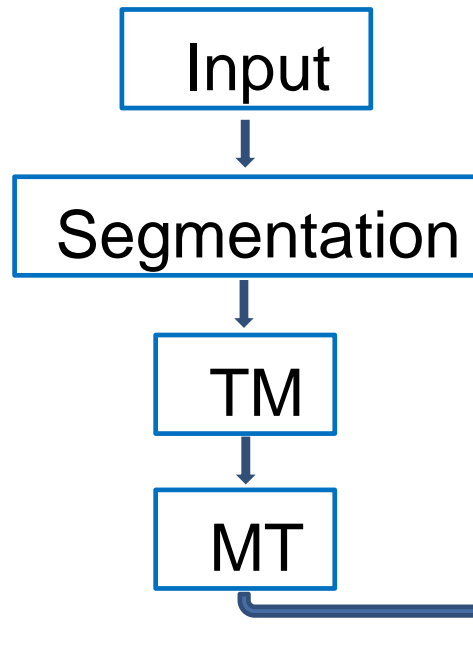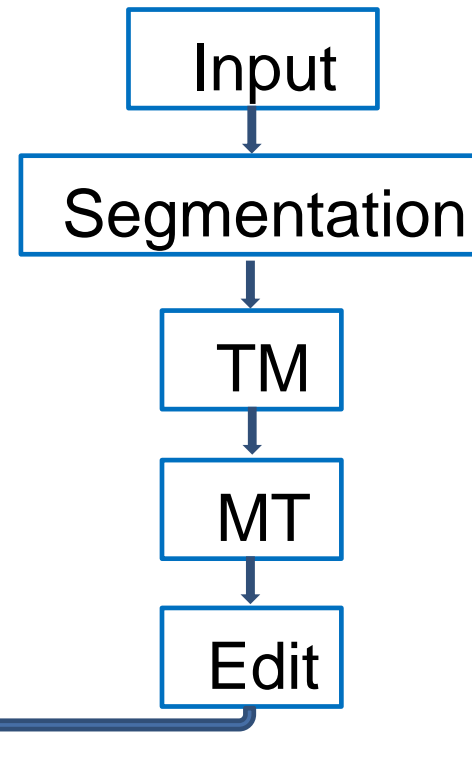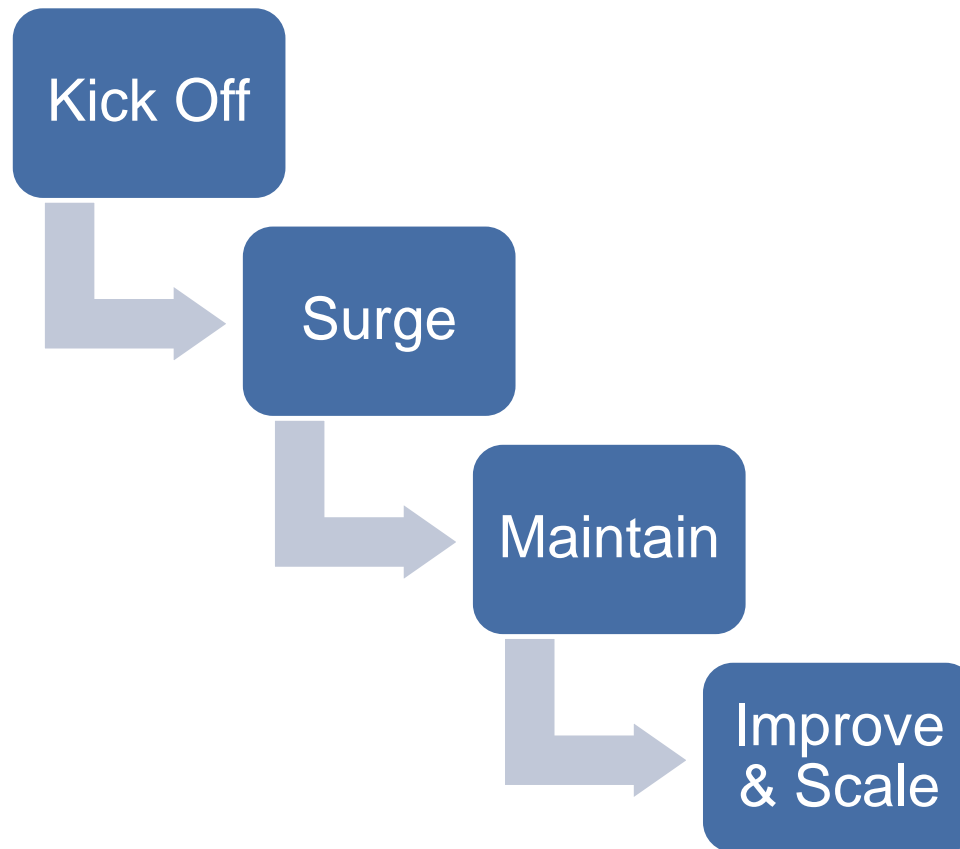
Kick Off

- **Review Statement of Work (SOW)**

  ➤ Hold kick-off meeting to set expectations, and clarify parameters and assumptions with stakeholders

- **Inform Stakeholders – before, during, after kickoff!**

  ➤ "Engineer for success" with source selection, MT training corpus

  ➤ Manage expectations for productivity, timeline

  ➤ Confirm client's priorities, preferences, and level of involvement

- **Set Goals and Key Performance Indicators (KPIs)**

  ➤ Linguist productivity

  ➤ Tool effectiveness

24

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

Page 507

Surge

- **Build a Team**
  - ➢ Linguists, Engineers, PM
- **Train**
  - ➢ PEMT objectives and workflow
  - ➢ Client-specific tools and style guides
  - ➢ HLT tools/resources including CAT, TMS, TM/TB, MT
- **Baseline**
  - ➢ Translate sample set of material (larger = better) outside of PEMT environment to measure productivity sans HLT
- **Document / Track Everything!**
  - ➢ Client communications
  - ➢ Workflow adjustments
  - ➢ Technology data
  - ➢ Performance data

25

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

Page 508

Maintain

- **Prioritize Knowledge Sharing**
  - Training materials, lessons learned, documentation
  - Meet regularly
- **Monitor, Report, Adjust**
  - Provide reports and recommendations monthly
  - Metrics
- **Evaluate Linguists' Performance**
  - Define and share performance and productivity metrics based on collected data
- **Review Client Level of Engagement**
  - Client involved too much or too little?
  - Client requests within contract scope?

26

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

Page 509

Improve & Scale

- **Monitor technology developments and provide recommendations as necessary**
  - ➢Escalate questions/issues to software developers as needed
  - ➢Test and troubleshoot
  - ➢Receive PM and linguist feedback on potential implementations

- **Evaluate MT output monthly; experiment and make adjustments as needed**
  - ➢Capture qualitative and quantitative data

- **Communicate success stories and lessons learned**
  - ➢Continually demonstrate ROI

- **Scale with additional domains and locales**
  - ➢Ensure HLT solution can accommodate growth and address locale-specific criteria

27

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

Page 510

Recap

- Review historical challenges

- Describe big picture and take incremental steps

- Receive client buy-in

- Customize the HLT solution -- one size does not fit all

- Document/track for reusability and scalability

- Develop talent through training

- Ask for client feedback and evaluate your success

- Continue to improve

28

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

Page 511

Future Enhancements

- Neural MT Non-Formal Language Support

- Post-MT Automated Editing

- Dynamic Learning NMT

- Substring tokenization

- Integrated speech-to-text supported by TMS/CAT/MT

29

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

Page 512

Thank you

Konstantine G. Boukhvalov
Operations Manager, Engineering
ManpowerGroup Public Sector
1-703-245-9372
Konstantine.Boukhvalov@Experis.com

Eileen Block
Engagement Manager
ManpowerGroup Public Sector
1-703-245-9363
Eileen.Block@manpowergroupsecure.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 513