

Migrating Terms to Greener Pastures

**Elias Ferguson
Office of Language Services
U.S. Department of State**

Office of Language Services (LS)

- ❑ Office of Language Services, U.S. Department of State
 - ❑ Language support for the conduct of foreign policy that dates back to 1781
 - ❑ Mission: “Facilitate communication with non-English speaking governments and people by providing high-level interpreting (spoken word) and translating (written word) support to the Executive Office of the President, the Department of State and other federal agencies.”



LANGUAGE SERVICES
Language Support for Government

Anti-Terrorism Assistance Program (ATAP)

- Mission: foster cooperation between U.S. law enforcement officers and those of partner nations.
 - Protect national borders, critical infrastructure, national leadership
 - 19 languages, 9 million words in 2010.
 - Major languages
 - Languages of “Limited Diffusion”



LANGUAGE SERVICES
Language Support for Government

Brief Project History

- ❑ Began using CAT tools in 2007
 - ❑ LS used Trados 2007 and Multiterm
 - ❑ Translators used Trados and Wordfast
 - ❑ One project manager for all projects and languages
 - ❑ Good project termbase (TB), but no central TB
- ❑ In 2010, LS began migration of its translation memories (TM) and termbases (TB) to Trados Studio 2009
 - ❑ Focused on TM cleanup and migration first
 - ❑ All ATAP TMs have been cleaned and uploaded to the server
 - ❑ Trained around 80 translators on how to use Trados Studio 2009

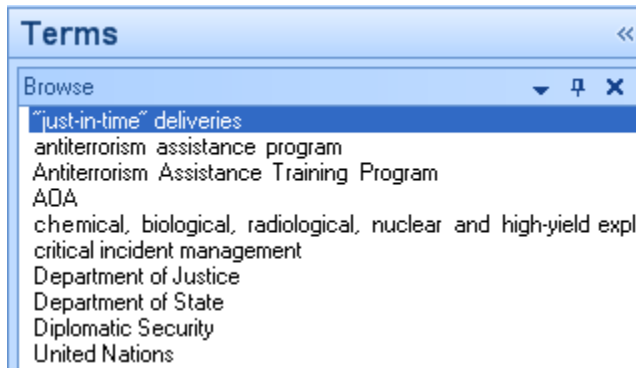


TB Migration – MultiTerm Experiments

- ▣ 17 experiments in total
 - ▣ First nine focused on import/merge settings and definitions
 - ▣ Objectives:
 - ▣ (1) Understand exactly how MultiTerm (MT) handled merging of entries,
 - ▣ (2) Determine best filter and import settings to manage terminology merging and maintenance
 - ▣ Duplication of data!?



Example of Duplication of Entries



Statistics		
Total number of entries:	10	
Total number of multimedia objects:	0	
Indexes		
Name	Terms	Entries
English:	10	10
Hindi:	10	10

"just-in-time" deliveries	???
antiterrorism assistance program	የጸረ ሽብር ድጋፍ ፕሮግራም
Antiterrorism Assistance Training Program	የጸረ ሽብር ድጋፍ ስልጠና ፕሮግራም
AOA	???
critical incident management	???
Department of Justice	???
Department of State	???
Diplomatic Security	???
United Nations	???
chemical, biological, radiological, nuclear and high-yield explosive	???

Before Import



LANGUAGE SERVICES
Language Support for Government

Example of Duplication of Entries

Terms <<

Browse ▾ ↕ ✕

- "just-in-time" deliveries
- antiterrorism assistance program
- Antiterrorism Assistance Training Program
- AOA
- chemical, biological, radiological, nuclear and high-yield expl
- critical incident management
- Department of Justice
- Department of State
- Diplomatic Security
- United Nations

Statistics		
Total number of entries:	10	
Total number of multimedia objects:	0	
Indexes		
Name	Terms	Entries
English:	20	10
Hindi:	20	10

	በግዜው የሚደርስ
	በግዜው የሚደርስ
antiterrorism assistance program	የፀረ ሽብርተኝነት እርዳታ ፕሮግራም
antiterrorism assistance program	የፀረ ሽብርተኝነት እርዳታ ፕሮግራም
Antiterrorism Assistance Training Program	የፀረ ሽብርተኝነት እርዳታ ስልጠና ፕሮግራም
Antiterrorism Assistance Training Program	የፀረ ሽብርተኝነት እርዳታ ስልጠና ፕሮግራም
AOA	ኤር ኦፕሬሽን ቦታ
AOA	ኤር ኦፕሬሽን ቦታ
chemical, biological, radiological, nuclear and high-yield explosive	የኬሚካል፣ ባዮሎጂካል፣ ሬዲዮሎጂካልና ኒውክሌር ፈንጂ መሣሪያዎች
chemical, biological, radiological, nuclear and high-yield explosive	የኬሚካል፣ ባዮሎጂካል፣ ሬዲዮሎጂካልና ኒውክሌር ፈንጂ መሣሪያዎች
critical incident management	የአስፈጻሚ ድርጊት ሂደት
critical incident management	የአስፈጻሚ ድርጊት ሂደት
Department of Justice	የዋና ድርጅት ሚኒስቴር
Department of Justice	የዋና ድርጅት ሚኒስቴር
Department of State	የውጭ ጉዳይ መስሪያ ቤት
Department of State	የውጭ ጉዳይ መስሪያ ቤት
Diplomatic Security	የዲፕሎማሲ ደህንነት
Diplomatic Security	የዲፕሎማሲ ደህንነት
United Nations	የተባበሩት መንግሥታት
United Nations	የተባበሩት መንግሥታት

After Import



TB Migration – MultiTerm Experiments

- ❑ LS has a ticket in with SDL Support
 - ❑ Good communication and troubleshooting to date – the issue is currently with SDL’s development division. Latest update: SDL has made progress. Resolved?
- ❑ The last eight experiments focused on:
 - ❑ Trying to fine-tune the import settings to avoid duplications
 - ❑ Looking for a workaround
 - ❑ Finding a new workflow that would avoid the need to merge TBs all together.



TB Migration – Continued

- ❑ Workaround found.
 - ❑ Simple export of duplicated terms and re-import into an empty TB with the proper import settings will eliminate them.
 - ❑ Future corruption?
- ❑ New Workflow developed.
 - ❑ Central, read-only TB.
 - ❑ Editable, project TB for new entries.
 - ❑ No more merging – every term is a new term.
 - ❑ No more duplicates!
 - ❑ Changes to the Central TB are done by project managers.



TB Migration – Manual Approach

- ❑ After completing the experiments and while waiting for SDL Support to troubleshoot the duplication issue..
 - ❑ Manual Merge using Excel
 - ❑ Most efficient/flexible
 - ❑ Macros/Formulas allowed for quickly making multiple changes
 - ❑ Very simple TB structure!
- ❑ First language centralized was Spanish



TB Migration - Challenges

- ❑ Lack of resources (people, time, money) and different TB structures over the years...
 - ❑ Duplication of terms
 - ❑ Multiple translations of terms
 - ❑ Incomplete entries
 - ❑ Misspellings
 - ❑ Random punctuation



Incorrectly Entered Terms

❑ Multiple terms per entry

Entry number: 1
Client: ATAP
Course: CRT (Crisis Response Team)



English

alternate route



Spanish

otra ruta, otro camino

Incorrect

Entry number: 1
Client: ATAP
Course: CRT (Crisis Response Team)



English

alternate route



Spanish

otra ruta

otro camino


Correct



Incorrectly Entered Terms

❑ Acronyms

Entry number: 11
Client: ATAP
Course: VIPP-TST (VIP Protection - Tactical Support Team)

	English probability of detection - PD
	Spanish probabilidad de detección - PD

Incorrect

Entry number: 11
Client: ATAP
Course: VIPP-TST (VIP Protection - Tactical Support Team)

	English probability of detection PD
	Spanish probabilidad de detección PD

Correct



Incorrectly Entered Terms

▣ Capitalization

Entry number: 1
Client: ATAP
Course: CRT (Crisis Response Team)

 **English**
Alternate Routes

 **Spanish**
Otras rutas
Otros caminos



Incorrect

Entry number: 1
Client: ATAP
Course: CRT (Crisis Response Team)

 **English**
alternate route

 **Spanish**
otra ruta
otro camino



Correct



Incorrectly Entered Terms



- ▣ Definitions and notes written in the main term entry

Entry number: 4
Client: ATAP
Course: EIC (Explosive Incidents Countermeasures)

	English Fins (grenades, mortars)
	Spanish Aletas

Incorrect

Entry number: 4
Client: ATAP
Course: EIC (Explosive Incidents Countermeasures)
Note: grenades, mortars

	English fins
	Spanish aletas

Correct



Cleanup

1. Get everything in Excel.
2. Remove extra spaces at the end of an entry – Excel formula (TRIM).
3. Get rid of duplicates – Identified by using Conditional Formatting to highlight duplicates.
4. Get rid of incomplete entries.
5. Clean incorrectly entered terms.
6. Correct misspellings, capitalization issues, plural vs. singular, delete unnecessary punctuation, etc.



TB Migration – Outcome

- ❑ Convert Excel files into MultiTerm termbases and upload to the server.
- ❑ TB Cleanup and Centralization Guidance
- ❑ Project manager and Translator TB Management Guidance
- ❑ New Workflow
 - ❑ Central = read only
 - ❑ Project = editable – for new, project-specific terms
 - ❑ No more duplicates
 - ❑ No more merging
 - ❑ Greater control over changes



Conclusion

□ Summary

- Experimentation
- Manual Cleanup
- Several TBs cleaned and centralized
- Goal to have all TBs centralized by the end of the summer

□ Lessons Learned

- If there is money available, dedicate someone full-time to set up the workflow in the beginning - Less pain!
- Guidance and training for project managers and translators on how to manage terminology



Thank you!

***Questions, comments,
suggestions?***



LANGUAGE SERVICES
Language Support for Government