The Language Translation Interface
and automated language translation tools
for the Australian Defence Organisation

Dominique Estival and Jennifer Biggs

Human Systems Integration Group, Command and Control Division
Defence Science and Technology Organisation
PO Box 1500, Edinburgh SA 5111
AUSTRALIA
Outline

– Goals of the task: Language translation for the ADO
  – why – what – who
  – 2 military examples
  – C2, coalition

– Automated Translation Tools for the ADO
  – LTI: The Language Translation Interface
  – Integration with AuTM (Automatic Transcriber of Meetings)

– Conclusion and issues for further research
Goals of the task (high-level)

“Investigate and report on feasibility of applying language translation and audio mining technologies to operations within the headquarters” (by June 2004)

⇒ How can the ADO deal with foreign languages?
Language Translation for the ADO

- **Why:** intelligence gathering, information sharing, joint coalition, foreign operations.

- **What:** document translation or screening, interview and meeting interpretation, command and control.

- **Who:** ADF LANGS, Language Sponsors, translators in individual units.

- **Two military examples:**
  - JWID ’02 C2IT-13
  - The PTS / Phraselator
TrIM Instant messaging server, C2IT13 JWID’02

Clients are web based; can be anywhere on CFBL. Flow can also go from Client B to Client A.

Dominique Estival & Jennifer Biggs, DSTO
The Phraselator (DARPA PTS)

- Military trial in Afghanistan, 2002
- one way translation for now
- modules of 1000+ phrases translated into multiple languages
- speech recognition or menu selection
- pre-recorded speech output
- phrase module builder available
- Phraselator doubles as a rugged pocket PC
Automated Translation Tools for the ADO

Goals and constraints

- provide ADO personnel with automated translation tools when there is a need for rapid translation and no human translator is readily available
- must use existing systems (COTS or freely available)

⇒ 1. Provide easy access to existing translation engines: Language Translation Database (LTDB)
2. Make access transparent to the users: Language Translation Interface (LTI)
Language Translation Database (LTDB)

- Product and research system information repository for members of the ADO
- Search by parameters of translation task
  - Language pairs
  - System requirements
- Currently contains 90 of an estimated 150 machine translation systems.
LTDB example: the Phraselator

One-way phrase-based voice-to-voice machine translator. Input phrase is matched with a preencoded output phrase and played through a speaker. Developed under a DARPA SBR grant.

Features:
- Translation memory
- Text-to-speech
- Speech recognition

Details:
- Machine translation
- Speech-to-speech

Categories:
- Commercial
- PC
- Macintosh

Languages:
- English
- Arabic
- Dutch
- Dari
- Pashto

System:
- Phraselator
- DARPA One-Way Phrase

Translation type and speed:
- Batch Translation
- Real-Time Translation

Translation Speed Per Page:
- 500 words per minute

Machine Translation Languages:
- English to Arabic
- English to Dutch
- English to Dari
- English to Pashto

Price:
- US $2,000.00

Notes:
- Price is for one unit. Module Builder software is available for the PC.

Release Date:
- 1/92/2002

Screenhots:
- soldier.jpg

Origin:
- Soft Lemnox, Operation Enduring Freedom, Afghanistan 2002. One of the MPs who guarded the detention. They will be using the Phraselator for prisoner processing and control.
Language Translation Interface (LTI)

- **Single interface for multiple translation engines:**
  - from the Web, from servers or local applications
  - Full MT, MAT, or Translation memories
- **Windows for pre-editing the input and post-editing the output**
- **View a number of translation results:**
  - choose the best one manually
  - default selection available
- **Word or HTML record of translation session**
- **Ability to interface with multiple applications, eg. AuTM**
La dirigeante de l'opposition birmane Aung San Suu Kyi, arrêtée vendredi dans le nord du pays, a été ramenée par la junte à Rangoun et se trouvait basé dans un complexe des renseignements militaires de la capitale, ont annoncé des sources de son parti, la Ligue nationale pour la démocratie (LND)....

1. The leader of the opposition birmane Aung San Suu Kyi, stopped Friday in the north of the country, was brought back by the junta to Rangoun and was Monday in a complex of the military information of the capital, announced sources of its party, the national League for the democracy (LND)....

2. The leading one of the opposition birmane Aung San Suu Kyi, stopped Friday in the north of the country, was brought back by the junta to Rangoun and was Monday in a complex of the military information of the capital, announced sources of its party, the national League for the democracy (LND)....

Dominique Estival & Jennifer Biggs, DSTO
LTI: Main form

The leader of the opposition birmane Aung San Suu Kyi, stopped Friday in the north of the country, was brought back by the junta to Rangoun and found itself Monday in a complex of the military pieces of information of the capital, announced sources on his part, the national League for the democracy (LND).

Dominique Estival & Jennifer Biggs, DSTO
La dirigeante de l'opposition birmane Aung San Suu Kyi, arrêtée vendredi dans le nord du pays, a été renommée par la junta à Rangoun et se trouvait lundi dans un complexe des renseignements militaires de la capitale, ont annoncé des sources de son parti, la Ligue nationale pour la démocratie (LND).

The leader of the opposition birmane Aung San Suu Kyi, stopped Friday in the north of the country, was brought back by the junta to Rangoun and found itself Monday in a complex of the military pieces of information of the capital, announced sources on his part, the national League for the democracy (LND).

The leading one of the opposition birmane Aung San Suu Kyi, stopped Friday in the north of the country, was brought back by the junta to Rangoun and was Monday in a complex of the military information of the capital, announced sources of its party, the national
MAT: translation memory

Translation memory containing translation units: previously translated matching Source / Target segments

IN: Source language segment

MAT tool compares legacy and new segments for % match

OUT: Target language segment, if % match found in translation memory

* Post-edit and insert to translation memory with source segment

Dominique Estival & Jennifer Biggs, DSTO
MT: translation processing

IN:
Source language
* Controlled language
* Pre-processing edit

OUT:
Target language
* Post-processing edit

Terminology databases

Dominique Estival & Jennifer Biggs, DSTO
LTI: translation methods

- Access to Translation Memories
- Wordfast CAT macro set
- Web browser automation
- Internet Explorer
- Translation request forms over TCP/IP

Dominique Estival & Jennifer Biggs, DSTO
LTI: Web translation engine

Internet users

WWW-server

Translation Server

Application-server

Mall-server

Dominique Estival & Jennifer Biggs, DSTO
LTI: Web translation form

Engine / Input

Web browser

Automation list
<table>
<thead>
<tr>
<th>Time</th>
<th>Translator</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:22:54</td>
<td>Transcend</td>
<td>&quot;Nous avons vu le convoi entrer dimanche dans la prison d'Insein&quot; (près de Rangoun), a indiqué une source de la LND, &quot;ensuite elle a apparemment été séparée de son entourage et amenée dans un complexe du centre de Rangoun&quot;.</td>
</tr>
<tr>
<td>10:21:49</td>
<td>Transcend</td>
<td>&quot; Saw us the convoy to enter Sunday into the insein prison&quot; (close to Rangoun), indicated a source of the LND, &quot;next she apparently was separated of his entourage and brought in a complex of the center of Rangoun&quot;.</td>
</tr>
<tr>
<td>10:17:51</td>
<td>SystranAltavista</td>
<td>Le prix Nobel de la paix avait été placé vendredi &quot;temporairement sous la protection des autorités&quot; dans le nord du pays avec 18 membres de son parti après des affrontements qui avaient fait quatre morts entre ses partisans et des éléments favorables à la junta au pouvoir. The price Nobel of peace had been placed Friday &quot;temporarily under the protection of the authorities&quot; in the north of the country with 18 members on his part after the clashes that had done four deaths between its supporters and elements in favor of the in power junta.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>La dirigeante de l’opposition birmane Aung San Suu Kyi, arrêtée vendredi dans le nord du pays, a été ramenée par la junta à Rangoun et se trouvait lundi dans un complexe des renseignements militaires de la capitale, ont annoncé des sources de son parti, la Ligue nationale pour la démocratie (LND). The leading one of the opposition birmane Aung San Suu Kyi, stopped Friday in the north of the country, was brought back by the junta to Rangoun and was Monday in a complex of the military information of the capital, announced sources of its party, the national League for democracy (LND).</td>
</tr>
</tbody>
</table>
Accepted translations

Language Translation Interface session
Translation session began at 10:17AM on 3 June 2003

Translation: French > English

Translation engines:

<table>
<thead>
<tr>
<th>Engine</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM</td>
<td><a href="C1FR_EN_political_TM">C1FR_EN_political_TM</a></td>
</tr>
</tbody>
</table>

Translations:

<table>
<thead>
<tr>
<th>Source text: French</th>
<th>Translations: English</th>
<th>Translator</th>
<th>Started</th>
<th>Finished</th>
</tr>
</thead>
<tbody>
<tr>
<td>La signature de l’opposition annula Aung San Suu Kyi, amende vendredi dans le nord du pays, a été rendue par la justice à Rangoun et a été traduite dans un compte des attaques militaires dans la capitale, parmi les sources de son parti, le NLD (le national League for democracy).</td>
<td>The leading one of the opposition dissidents Aung San Suu Kyi, arrested Friday in the north of the country, was brought back by the justice to Rangoun and was Monday in a complex of the military information of the capital, announced sources of its party, the national League for democracy (NLD).</td>
<td>Transcient</td>
<td>10:17:51</td>
<td>10:21:27</td>
</tr>
<tr>
<td>Le prix Nobel de la paix avait été placé vendredi &quot;temporairement sous la protection des autorités&quot; dans le nord du pays avec 18 membres du parti après des affrontements qui avaient fait quatre morts entre ses partisans et des membres de la police.</td>
<td>The price Nobel of peace had been placed Friday &quot;temporarily under the protection of the authorities&quot; in the north of the country with 18 members of the party after clashes that had done four deaths between its supporters and elements of the police force.</td>
<td>Transcient</td>
<td>10:21:49</td>
<td>10:22:31</td>
</tr>
<tr>
<td>&quot;Nous avons vu le convoi entrer dimanche dans la prison d’Insein&quot; près de Rangoun, a indiqué une source de la NLD.</td>
<td>* Saw us the convoy to enter Sunday into the prison of Insein&quot; (close to Rangoun), indicated a source of the NLD.</td>
<td>Transcient</td>
<td>10:22:54</td>
<td>10:27:11</td>
</tr>
<tr>
<td>Cette information n’a pas été confirmée par une source officielle. Aung San Suu Kyi et son entourage ont apparemment été recodés à Rangoun dans le but de s’arrêter à Chirchmar, et les autorités auraient demandé aux journalistes de ne pas se trouver là-bas, selon cette source.</td>
<td>This information is not confirmed source official. Aung San Suu Kyi and his entourage apparently were removed to Rangoun in the Saturday night to Sunday, and authorities had asked the populations alongside the train carrying understanding an important military escort of not to find itself autobsiders, according to this source.</td>
<td>Transcient</td>
<td>10:33:09</td>
<td>10:56:48</td>
</tr>
</tbody>
</table>

Translation session concluded at 03:12PM on 3 June 2003

Dominique Estival & Jennifer Biggs, DSTO
LTI: further work

- Input method editors
- Language detection tools
- Input sources
- Speech input/output
- LOCALISATION
  - Dictionaries
  - Thesaurus
  - Terminologies
- LTI
  - Acronym management
  - Connect to the LTDB

Dominique Estival & Jennifer Biggs, DSTO
Integrating LTI with other applications

use of the LTI with other applications:

– output translation saved in HTML, Word or TMX

⇒ LTI can be used with standard document production tools by a single translator

– also, e-mail, database interface, etc…

– example of collaborative environment, the *Automatic speech-to-text Transcriber for Meetings and Interviews* (AuTM)
AuTM: Automatic Transcriber of Meetings

collaborative meeting tool developed at DSTO
client-server architecture, up to 8 meeting participants

**aim:** record and transcribe meetings or interviews

**input:** Automatic Speech Recognition (ASR);
one microphone per participant

**output:** session transcript (text with audio files);
time-stamped; Word or HTML
LTI + AuTM

- Multilingual meetings are possible
  with or without interpreter

- Multilingual Transcript
  - view by language, or
  - view complete transcript with translations

- Multilingual Records
  - original language input
  - names of translators / translation engine
LTI + AuTM: Issues for further research

- Combining ASR and MT/MAT technologies
  ⇒ compound transcription and translation errors
  ⇒ need dialogue understanding in addition to speech recognition

- Language issues:
  - localization
  - availability of ASR for more languages
  - availability of MT systems for more language pairs
  - language pair path resolution (most MT are English-centric)

Dominique Estival & Jennifer Biggs, DSTO
Conclusions: the LTI

- simple easy-to-use translation tool when human translator not available
- single interface for multiple translation engines (full MT, MAT, or Translation Memories) from the Web, from servers, or from local applications on a secure network
- allows users to view a number of translation results and to choose the most appropriate
- input and output editing panels

=> the same interface can be used by human translators to create manual translations
Conclusions: Translation tools for C2

- automated batch translation of incoming documents (e.g. e-mail msgs) can be used for rapid intelligence gathering and information processing during high tempo operations
- useful way to rapidly access translation tools in those situations
- integrated with a speech transcriber:
  - allow meetings to be conducted in more than one language
  - useful in coalition exercises and operations
Further research

- Translation Memories: greater implementation of TMX
- Optimise MT for specific tasks:
  - some vendors do it (e.g. Systran for the EC)
  - users can do it themselves
  - in the limit, build bank of specific translation examples for TMs
    => can be classified
  - already “Acronym Manager” available for military acronyms at DSTO
- Spoken language translation:
  - Forced phrase recognition (cf. Phraselator)