AMTA 2010 is proud to present this year's Technology Showcase, featuring the following companies and software. See inside for more information on each exhibit. The AMTA 2010 Technology Showcase takes place on Tuesday, November 2 at the Westin Tabor Center (conference) hotel from noon until 5:00 pm. The exhibit is open to the public at no charge.

| | |
|---|---|
| Abraxis | Virtus Translator for Russian |
| Apptek | TranSphere<br>Talk2Me |
| Army Research Laboratory | Dictionary Work |
| Basis Technology | Arabic/Afghan Desktop Suite<br>Rosette Entity Extractor<br>Rosette Name Translator<br>Rosette Name Indexer |
| CACI | Document and Media Exploitation |
| Center for Applied Machine Translation | CYBERTRANS |
| Columbia University | Cross Linguistic Blog Alerts (COLABA) |
| Forte Communications Associates | Handheld Machine Translation |
| Integrated Wave Technologies, Inc. | Voice Response Translator |
| Microsoft | Microsoft Translate |
| MITRE | Language Technology Resource Center |
| Northrop Grumman | Foreign Media Collaboration Framework (FMCF) |
| ProMT | PROMT Translation Server 9.0<br>PROMT LSP 9.0 |
| Pan American Health Organization | PAHOMTS |
| Sakhr Software USA | Mobile Translation Technology:<br>Mobile Speech-to-Speech Translator<br>Mobile OCR and Translator |
| SDL | SDL Trados Studio 2009 |
| SDL LanguageWeaver | Automated translation software |
| Spoken Translation | Converser for Healthcare |
| Systran | SYSTRAN Enterprise Server 7 |
| Technology Development Group | Online Multimedia Parallel Corpora dB (OMPC)<br>Linguist Language Broker (LLB Light) |
| National Research Council of Canada | WebBiText |

| | |
|---|---|
| **Company or organization** | Abraxas Corporation |
| **Products or software** | Virtus Translator for Russian |
| **Version** | Early Access Version 0.8 |
| **Available for licensing Y/N** | Y |
| **Kind of product** | Machine Translation |
| **Description** | The Virtus Translator for Russian enables users to automatically translate structured foreign data containing named entities (i.e., proper nouns and proper noun phrases) and phrases describing named entities into English. |
| **Languages or language pairs supported** | Russian-to-English |
| **Unique about this system** | Performs translation on structured data |
| **New this year** | Yes |
| **Part of a larger system? Describe** | No |
| **Made up of other systems? Describe** | No |
| **Standards supported** | Transliteration Standards: IC Standard, BGN-PCGN, GOST Schedule A or ISO 9:1995, GOST Schedule B and ACA-LC |
| **Presentations at AMTA— what/when** | Presentation at Workshop on Developing, Updating, and Coordinating Terminologies, Dictionaries, and Lexicons for Terminological Consistency on October 31, 2010 |
| | |
| **Contact for more information** | |
| **Name** | Jen Doyon |
| **Telephone** | 703-821-8930 x564 |
| **Email** | jen.doyon@abraxascorp.com |

<table>
<tr><td>Company or organization</td><td>Applications Technology, Inc. (AppTek)</td></tr>
<tr><td>Products or software</td><td>TranSphere® (Hybrid Machine Translation System)</td></tr>
<tr><td>Version</td><td>2.0</td></tr>
<tr><td>Available for licensing Y/N</td><td>Yes</td></tr>
<tr><td>Kind of product</td><td>Machine Translation</td></tr>
<tr><td>Description</td><td>TranSphere® is AppTek's state-of-the-art Machine Translation (MT) application. AppTek developed TranSphere® after years of extensive linguistic research throughout the world. Its technology reaches beyond other, more simplistic approaches—some of which verge on mere word-by-word substitution—toward real-life, high-volume industrial and commercial applications. AppTek's most advanced and industry-leading MT platform is a hybridized system of two fully integrated mature engines using Statistical and Rule-based MT. AppTek's Hybrid MT is a unique and patent-pending approach for achieving the best quality in automated translation output. Our MT system is also fully integrated with PlainSpeech-Automatic Speech Recognition (ASR) and our other language technologies for complete solutions to translating foreign information whether textual or spoken.</td></tr>
<tr><td>Languages or language pairs supported</td><td>
Bi-directional Translation
<table>
<tr><td>English <-> Arabic</td><td>English <-> Polish</td></tr>
<tr><td>English <-> Farsi</td><td>English <-> Portuguese</td></tr>
<tr><td>English <-> Spanish</td><td>English <-> Russian</td></tr>
<tr><td>English <-> Dari</td><td>English <-> Italian</td></tr>
<tr><td>English <-> Chinese</td><td>English <-> Turkish</td></tr>
<tr><td>English <-> Korean</td><td>English <-> Ukrainian</td></tr>
<tr><td>English <-> Japanese</td><td>French <-> Italian</td></tr>
<tr><td>English <-> Hebrew</td><td>German <-> French</td></tr>
<tr><td>English <-> German</td><td>German <-> Italian</td></tr>
<tr><td>English <-> French</td><td>Spanish <-> French</td></tr>
<tr><td>English <-> Dutch</td><td>Spanish <-> Italian</td></tr>
</table>
One-way Translation
<table>
<tr><td>Tagalog > English</td></tr>
<tr><td>Urdu > English</td></tr>
<tr><td>Indonesian > English</td></tr>
<tr><td>Pashto > English</td></tr>
<tr><td>Persian > Arabic</td></tr>
<tr><td>Greek>English</td></tr>
</table>
</td></tr>
</table>

| | |
|---|---|
| **Unique about this system** | AppTek's TranSphere Hybrid Machine Translation is the only system to enable fully-automatic, high quality machine translation, to integrate rule-based and statistical features. AppTek's HMT solution provides a full integration of both MT methodologies, rather than simply adding rules to the statistical system or a minor statistical module to the rule-based engine. |
| **New this year** | AppTek Releases Arabic, Chinese, German, Greek, Spanish, Farsi, French, Hebrew, Italian, Japanese, Pashto, Russian, Turkish, Urdu Language Engine for its Hybrid Machine Translation System |
| **Part of a larger system? Describe** | Yes,<br>Northrop Grumman – Foreign Media Collaboration Framework<br>Northrop Grumman – Language Now<br>DIA- DOCEX<br>NASIC- DLIPS |
| **Standards supported** | TMX and XLIFF |
| **Used by** | Commercial & Government Institutions |
| **Presentations at AMTA— what/when** | "Arabic Dialect Handling in Hybrid Machine Translation"<br>**DAY ONE:** *Research Track,* 2:30 pm<br>" Improving Reordering in Statistical Machine Translation from Farsi"<br>**DAY THREE:** *Research Track,* 2:00 pm |
| | |
| **Contact for more information** | |
| **Name** | Maria Gutierrez |
| **Telephone** | 703-394-2323 |
| **Email** | maria.e.gutierrez @apptek.com |

| | |
|---|---|
| **Company or organization** | Applications Technology, Inc. (AppTek) |
| **Products or software** | Talk2Me |
| **Version** | 1.0 |
| **Available for licensing Y/N** | Yes |
| **Kind of product** | Computer Assisted Interview & Training System |
| **Description** | The AppTek Talk2Me is an Interview prototype system sponsored by TSWG and responds to the needs of users (Army) working with interpreters. The application provides automatic transcription of English and foreign language speech for interview documentation and interviewer training. Users are able to view automatically-created transcripts synchronized with video and audio interview recordings. The transcripts are divided into turns by speaker, simplifying the task of navigating through a record. |
| **Languages or language pairs supported** | Arabic will be demonstrated but other languages are available where Speech Recognition is mature |
| **Unique about this system** | Without the capabilities provided through Talk2Me, interviewers, interpreters, and interviewer trainers must rely on audio-visual recordings and handwritten notes made during interviews in order to prepare reports, create training scenarios and give meaningful feedback. With the Talk2Me system, users can search automatically-created transcripts for important words and immediately find relevant points in the synchronized recordings. They can also make notes that will be preserved at relevant points in the transcript and recording, export portions of a transcript or aligned transcript and recording, and correct transcription errors. The transcript search and processing capabilities provided by the Talk2Me system greatly enhance existing interview analysis and interviewer training mission performance. The system is under development and has not yet been tested by the users. |
| **New this year** | Novel Computer Assisted Interview & Training System |
| **Made up of other systems?** | Automatic Speech Recognition (ASR) & Machine Translation (MT) |
| **Used by** | Commercial & Government Institutions |
| **Presentations at AMTA** | "Arabic Dialect Handling in Hybrid Machine Translation" " Improving Reordering in Statistical Machine Translation from Farsi" |
| | |
| **Contact for more information** | |
| **Name** | Maria Gutierrez |
| **Telephone** | 703-394-2323 |
| **Email** | maria.e.gutierrez @apptek.com |

| | | | | | |
|---|---|---|---|---|---|
| **Company or organization** | Basis Technology | | | | |
| **Products or software** | Arabic/Afghan Desktop Suite<br>Rosette Entity Extractor<br>Rosette Language Identifier<br>Rosette Name Translator<br>Rosette Name Indexer | | | | |
| **Available for licensing** | Yes | | | | |
| **Kind of product** | Document exploitation and text analytics | | | | |
| **Description** | Basis Technology is the leading provider of desktop applications and software development tools for exploiting foreign language and English documents and digital media. Applications of interest to government agencies include text mining, foreign language information retrieval, document triage, and name resolution. For more information, visit www.basistech.com | | | | |
| **Languages or language pairs supported** | Arabic<br>Dari<br>Pushto<br>Albanian<br>Bengali<br>Bulgarian<br>Catalan<br>Chinese, Simplified<br>Chinese, Traditional<br>Croatian<br>Czech | Danish<br>Dutch<br>English<br>Estonian<br>Finnish<br>French<br>German<br>Greek<br>Gujarati<br>Hebrew<br>Hindi<br>Hungarian<br>Icelandic<br>Italian | Indonesian (Bahasa)<br>Japanese<br>Kannada<br>Korean<br>Kurdish<br>Kurdish (transliterated)<br>Latvian<br>Lithuanian<br>Malay<br>Malayalam<br>Norwegian<br>Persian | Persian (transliterated)<br>Polish<br>Portuguese<br>Romanian<br>Russian<br>Serbian<br>Serbian (transliterated)<br>Slovak<br>Slovenian<br>Somali<br>Spanish<br>Swedish | Tagalog<br>Tamil<br>Telugu<br>Thai<br>Turkish<br>Ukrainian<br>Urdu<br>Urdu (transliterated)<br>Uzbek<br>Uzbek (transliterated)<br>Vietnamese |
| **Unique about this system** | Identify resolution technology focuses on names across Middle Eastern and Far East languages in their native scripts and Romanized forms | | | | |
| **New this year** | New IME improvements, language capability | | | | |
| **Part of a larger system? Describe** | Rosette 7 is a suite of components designed to help you examine raw data, process it intelligently, and put it to work. These building blocks can be assembled into flexible solutions that fit the requirements of your application, working seamlessly within your current workflows while handling many different languages, character sets, and data sources. | | | | |
| **Standards supported** | IC, SATTS, MELTS, BGN | | | | |
| **Used by** | Numerous agencies & departments in Defense & Intelligence communities | | | | |
| **Presentations at AMTA** | "Arabic/Afghan Desktop Suite (ADS)"<br>**DAY ONE:** *Government Track*, 4:00 pm | | | | |
| | | | | | |
| **Contact for information** | | | | | |
| **Name** | Bill Ray | | | | |
| **Telephone** | 617-386-7184 | | | | |
| **Email** | bill.ray@basistech.com | | | | |

Professional language analysts leverage a myriad of tools in their quest to produce accurate translations of foreign language material. The effectiveness of these tools ultimately affects resource allocation, information dissemination and subsequent follow-on mission planning – all three of which are vital, time-critical components in the intelligence cycle. This presentation will highlight the need for interactive tools that perform jointly in an operational environment, focusing on a dynamic suite of foreign language tools packaged into a desktop application and serving in a machine translation role.

***Basis Technology's*** **Arabic/Afghan Desktop Suite (ADS)** supports DOMEX, CELLEX, and HUMINT missions while being the most powerful Arabic, Dari and Pushto text analytic and processing software available. The ADS translates large scale lists of names from foreign language to English and also pinpoints place names appearing in reports with their coordinate locations on maps. With standardization output having to be more accurate than ever, the ADS ensures conformance with USG transliteration standards for Arabic script languages, including IC, BGN/PCGN, SATTS and MELTS. The ADS enables optimization of your limited resources and allows your analysts and linguists to be tasked more efficiently throughout the workflow process.

| | |
|---|---|
| **Company or organization** | Center for Applied Machine Translation (CAMT) |
| **Products or software** | CYBERTRANS |
| **Version** | 10.6.4 |
| **Available for licensing Y/N** | For USG |
| **Kind of product** | Language and Encoding ID, MT |
| **Description** | Integrated suite of automated tools for translation, language and encoding ID, spelling and text enhancement, and encoding conversion. Available as a client-server based application or standalone. |
| **Languages or language pairs supported** | 70 languages into English – GOTS product |
| **Unique about this system** | Translation support for multilingual documents; LCTL translation not commercially available; Automatic language and encoding ID; text normalization; Romanization module for non-Latin script languages |
| **New this year** | Improved Farsi, Indonesian, Arabic, Arabic dialect handling for Levantine and Iraqi; New GOTS engines for Pashto, Farsi, Afrikaans and Portuguese |
| **Part of a larger system? Describe** | Easy integration |
| **Made up of other systems? Describe** | GOTS engines; HOTSPOT (Language and Encoding ID); TRANSCODER; BETTERMENT (spelling correction for 22 languages); TEXT NORMALIZATION MODULE, DICTIONARY MANAGER; ROMANIZER |
| **Used by** | USG |
| **Presentations at AMTA—what/when** | Demonstration only |
| | |
| **Contact for more information** | |
| **Name** | Kevin Cohea |
| **Telephone** | 301-688-5058 |
| **Email** | kmcohea@ugov.gov |

| | |
|---|---|
| **Company or organization** | Columbia University |
| **Products or software** | Products and Software |
| **Available for licensing Y/N** | Y for Research Purposes |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Enabling technologies for MT: Dictionaries, Annotation tool, Dialect identification, NER |
| **Description** | Preprocessing tools for Modern Standard Arabic and Arabic dialects |
| **Languages or language pairs supported** | Modern Standard Arabic, Dialects of Arabic |
| **Unique about this system** | Robust handling of the dialects of Arabic used in social media linguistic communication |
| **Part of a larger system? Describe** | All of it is part of a larger effort under the COLABA project |
| **Used by** | COLABA (Cross Linguistic Blog Alerts) Project |
| **Presentations at AMTA— what/when** | Cross Lingual Arabic Blog Alerting, K, Egan (TSWG)<br>**DAY ONE:** *Government Track*, 2:30 pm |
| | |
| **Contact for more information** | |
| **Name** | Mona T Diab |
| **Telephone** | 2128701290 |
| **Email** | mdiab@ccls.columbia.edu |

| | |
|---|---|
| **Company or organization** | Forte Communications Associates |
| **Products or software** | Hand held machine translation |
| **Version** | 2.8 |
| **Available for licensing Y/N** | Y |
| **Kind of product** | OCR,MT, Audio |
| **Description** | Voiscan is a hand held unit about the size of a smart phone that is design specifically for the user in the field. The unit offers an all-in-one solution for immediate document translation in the tactical environment. |
| **Languages or language pairs supported** | See attached data sheet |
| **Unique about this system** | All in one scanner and MT translation + audio |
| **New this year** | Yes, new scanning technology. |
| | |
| **Contact for more information** | |
| **Name** | Dan Erlin |
| **Telephone** | (916) 543-5322 |
| **Email** | fortcom@starstream.net |

FORTÉ COMMUNICATIONS

# VOISCAN®

## *Handheld Machine Translation for the Tactical Operator*


- **Scans documents using advanced optical character recognition (OCR)**
- **Converts scanned image to translated text**
- **Translated text displayed on large, bright integrated LCD**
- **Internal software converts text into audio voice output**
- **Recognizes 20+ printed languages**
- **Integrated camera available**

Designed specifically for the operator in the field, ***Voiscan*** offers an all-in-one solution for immediate document translation in the tactical environment.

The user simply passes the ***Voiscan*** unit over a printed document. Advanced optical character recognition (OCR) converts the scanned image to translated text data which is displayed on the unit's large LCD screen. The text is stored in a separate 1 GB module and can be transmitted to another device via USB, WiFi or Bluetooth.

Patented software converts the translated text data into an audio voice output. Audio can be heard through an integrated speaker or through plug-in ear buds if silence is required.

An SD memory stick contains the ***Voiscan*** language translation module and OCR reading technology. Memory sticks are available in 20+ written languages and easily can be changed in the field.

***Voiscan*** gives field operators a unique capability by enabling onsite translation of acquired documents for immediate assessment of their potential intelligence value. Instant access to crucial information, rather than relying upon distant centralized translation facilities, affords the field operator formerly unavailable insight to the tactical situation.

## PhotoScan Page Processing Technology

undistorted image of a scanned

document even when the scan path is not straight. ***Voiscan*** is capable of scanning documents up to 17" x 24" in size and stitching together multiple scan paths into a single, seamless text file.

Removable, language-specific SD cards enable ***Voiscan*** to instantly translate scanned document text into English for immediate onsite analysis. SD cards are currently available to provide translation into English from:

- Arabic
- Farsi
- Dutch
- German
- Japanese
- Polish
- Russian
- Swedish
- Hebrew
- Chinese
- French
- Italian
- Korean
- Portuguese
- Spanish


English text appears on a large, backlit LCD screen for easy viewing.

### Technical Specifications

***Size:*** 5.25"W x 3.25" H x 1" T
***Weight:*** 12 oz.
***Power:***

***Battery Life:***
4 hours scanning
15 hours playtime
***LCD Size:*** 4.3" diagonal
***Peak capture speed:***
16 IPS
***Max. sustained speed:***
12 IPS
***Max. capture length:***
23 inches
***Scan resolution:***
300 DPI
***Memory size:***
8 GB

***For further information, contact Dan Erlin at:***

**Forté Communications Associates**
1589 Summerhill Lane, Lincoln, CA 95648
(916) 543-5322
E-mail: fortcom@starstream.net

| | |
|---|---|
| **Company or organization** | Integrated Wave Technologies, Inc. |
| **Products or software** | Voice Response Translator |
| **Version** | 50+ languages |
| **Available for licensing Y/N** | |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Voice to voice machine translator |
| **Description** | MilTrans Voice Response Translator (VRT) is the only combat proven, eyes-free, hands-free foreign language translation system allowing for complete situational awareness and weapons readiness. The wearable, phrase-based device is miniaturized, holds 200 languages and recognition accuracy is over 98 percent. www.miltrans.com |
| **Languages or language pairs supported** | Up to 200. Current list below* |
| **Unique about this system** | Only true eyes-free, hands free system. We can mark the beginning and end of an utterance. |
| **New this year** | More languages including Punjabi, Romanian, Polis |
| **Used by** | Military, Law Enforcement, First Responders |
| **Presentations at AMTA—what/when** | Last year in Canada we had a table (thank you!) |
| | |
| **Contact for more information** | |
| **Name** | Laura Madonna |
| **Telephone** | 202 441 0791 |
| **Email** | laura@miltrans.com |

**CURRENT LANGUAGES:** Iraqi, Kurdish, Modern Standard Arabic, Pashto, Urdu, Farsi, Dari, Afar, Mandarin, Cantonese, Korean, Thai, Tagalog, French, Somali, Swahili, Bahasa, Amharic, Dutch, German, Greek, Italian, Turkish, Albanian, Serbian, Spanish, Portuguese, Cambodian, Afar, Kurmanji, Haitian Creole, Georgian, Magindanao, Tausug, Hebrew, Maranao, Croatian, Hausa, Vietnamese, Mongolian, Yoruban, Hindi, Oromo, Tetem, Tok Pisin, Ilocano, Cebuano, Hiri Moto, Egyptian, Burmese, Polish, Punjabi, Romanian.

| | |
|---|---|
| **Company or organization** | Microsoft Corporation |
| **Products or software** | Microsoft Translator |
| **Version** | |
| **Available for licensing Y/N** | Y |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Machine Translation Web Service |
| **Description** | Microsoft runs a web service performing automatic translation, with a rich API and wide language coverage, offering speech synthesis and translation memory functionality besides translation. Functionality can easily be embedded into any web site or application. |
| **Languages or language pairs supported** | 33 languages, any language to any language. |
| **Unique about this system** | High performance web service, collaborative features, speech |
| **New this year** | Collaborative features, API, speech |
| **Part of a larger system? Describe** | Microsoft Translator can be used by any developer in their own systems, or as part of a Microsoft service or application: Internet Explorer, Windows Live Messenger, Microsoft Office, Bing. |
| **Made up of other systems? Describe** | Microsoft Translator is a web service not relying on other services. |
| **Standards supported** | TMX |
| **Used by** | Microsoft services and applications, worldwidescience.org, blogs |
| **Presentations at AMTA—what/when** | User track: Better translations with user collaboration – Integrated MT at Microsoft |
| | |
| **Contact for more information** | |
| **Name** | Chris Wendt |
| **Telephone** | +1 425 706-4254 |
| **Email** | Chris.Wendt@microsoft.com |

Microsoft®
Translator

| **Company or organization** | MITRE |
|---|---|
| **Products or software** | Language Technology Resource Center (LTRC) |
| **Version** | 2 |
| **Available for licensing Y/N** | N |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Web service |
| **Description** | The provides information about machine translation and other Human Language Technology. |
| **Languages or language pairs supported** | NA |
| **Unique about this system** | U.S. |
| **New this year** | SharePoint for U.S. Government users |
| **Part of a larger system? Describe** | No |
| **Made up of other systems? Describe** | Linked to many other systems |
| **Standards supported** | ISO 639, ISO/IEC 10646 |
| **Used by** | Worldwide |
| **Presentations at AMTA—what/when** | No |
| | |
| **Contact for more information** | |
| **Name** | Jennifer DeCamp |
| **Telephone** | (703) 347-5137 |
| **Email** | jdecamp@mitre.org |

| | |
|---|---|
| **Company or organization** | PROMT |
| **Products or software** | PROMT Translation Server 9.0 and PROMT LSP 9.0 |
| **Version** | 9.0 |
| **Available for licensing Y/N** | Y |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Machine Translation |
| **Description** | PROMT has just released a new generation of PROMT translation server, that combines rule-based transfer translation with statistical engine tuning. In December 2010 PROMT will also be releasing a hybrid server, and it will be available for demo at the time of AMTA |
| **Languages or language pairs supported** | English, French, German, Spanish, Russian, Portuguese, Italian, Chinese |
| **Unique about this system** | High quality of translation, advanced tuning rules, hybrid approach |
| **New this year** | 9.0 and hybrid |
| **Made up of other systems? Describe** | Uses CCID Chinese engine |
| **Standards supported** | XLIFF, HTML, OLIF, TMX, TBX, all Office formats |
| **Used by** | Multiple enterprise clients |
| | |
| **Contact for more information** | |
| **Name** | Olga Beregovaya |
| **Telephone** | 415-425-5033 |
| **Email** | olgab@promt.com |

| | |
|---|---|
| **Company or organization** | Pan American Health Organization |
| **Products or software** | PAHOMTS® |
| **Version** | 4.6 (July 2010) |
| **Available for licensing Y/N** | Yes (international and intergovernmental organizations, NGOs, educational institutions & government agencies only) |
| **Kind of product** | Machine Translation (rule-based) |
| **Description** | PAHOMTS® is a fully automatic Machine Translation (MT) program, developed and maintained by PAHO's computational linguists, translators, and programmers. The Translation Services unit has used PAHOMTS® in a production setting since 1980. Staff and freelance translators postedit the raw output to produce high-quality translations with a 30-50% gain in productivity and a 33% cost savings. At PAHO, MT has proven to be especially effective for conference documents, scientific papers, training materials, technical abstracts, and Web pages. The output preserves the format of the original document, including tables, and is easy to polish using PAHO's posteditng macros. It is also compatible with Translation Memory software. |
| **Languages or language pairs supported** | English-Spanish, Spanish-English, English-Portuguese, Portuguese-English, Spanish-Portuguese, Portuguese-Spanish |
| **Unique about this system** | Each of the MT dictionaries contains over 150,000 words, phrases, and context-sensitive rules. Both the programs and the dictionaries are constantly being enhanced based on feedback from the users of the systems. The system includes PAHO's posteditng macros and toolbars for Microsoft® Word and Microsoft® PowerPoint®. It can be easily integrated with Translation Memory (TM) systems. It creates aligned translations that can be fed into a translation memory or a bilingual corpus. The software includes a completely trilingual graphical user interface. The extensive context-sensitive online help in English, Spanish, and Portuguese, simplifies the process of browsing and updating the PAHOMTS® dictionaries. |
| **New this year** | Upcoming integration with Multitrans |
| **Standards supported** | PAHOMTS® runs on Windows NT®, Windows 2000, Windows XP, and Windows Vista™. It can process files in Rich Text Format (RTF) version 1.9.1 (applies to Microsoft® Office 2000+ suite and programs that read/write RTF in general), MIME HTML (MHT) (applies to Microsoft® Office 2002+ suite and programs that read/write MHT in general), HTML, SGML, XML, standard ASCII and ANSI text formats, Corel® WordPerfect® for MS-DOS® (5.1, 6.0), and Corel® WordPerfect® for Windows (8.0+) formats. |

| | |
|---|---|
| **Used by** | PAHO (Headquarters and Regional Offices), WHO Headquarters, Health Agencies, International Organizations, Educational Institutions, Government Agencies, NGOs, etc. |
| **Presentations at AMTA—what/when** | – J. Aymerich and H. Camelo. 2006. Post-Editing of MT Output in a Production Setting: Experiences at the Pan American Health Organization. AMTA 2006. Cambridge, MA. August, 2006.<br>– J. Aymerich. 2004. Machine Translation in Practice at PAHO. Tutorial Notes, at AMTA 2004: Sixth biennial conference of the Association for Machine Translation in the Americas. September 2004, Georgetown University, Washington, DC, USA.<br>– M. León. 2000. A New Look for the PAHO MT system. In: Envisioning machine translation in the information future: 4th conference of the Association for Machine Translation in the Americas, AMTA 2000, Cuernavaca,Mexico, October 2000; ed. John S. White (Berlin: Springer Verlag, 2000); pp.219-222.<br>– G. A. Silva. 2000. The Use of ENGSPAN at the Pan American Health Organization: a Reviser's Perspective. Paper presented at the workshop "MT in Practice: The User Experience". AMTA 2000: Envisioning Machine Translation in the Information Future, organized by the Association for Machine Translation in the Americas (AMTA). Cuernavaca, Mexico, October 10, 2000. |
| | |
| **Contact for more information** | |
| **Name** | Julia Aymerich |
| **Telephone** | (202) 974-3565 |
| **Email** | aymericj@paho.org |

| | |
|---|---|
| **Company or organization** | Sakhr Software USA |
| **Products or software** | Mobile Translation Technology: Mobile Speech-to-Speech Translator and Mobile OCR and Translator |
| **Available for licensing Y/N** | Y |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Mobile MT, Mobile OCR |
| **Description** | **Mobile Speech-to-Speech (S2S) Translator** is available in multiple mobile platforms, the S2S translator facilitates dialogue between speakers of Arabic and English. The Mobile S2S Translator processes free-form dictation and is speaker-independent.<br><br>**Mobile OCR Document Translator** will scan, ocr, and translate documents on the fly utilizing a mobile device. It works in **multiple languages** (including Arabic, Portuguese and Russian). |
| **Languages or language pairs supported** | Arabic, English (Mobile Translator)<br>Multiple Languages (Mobile OCR) |
| **Unique about this system** | Smartphone Agnostic; also the first of its kind |
| **New this year** | **Mobile OCR and Translation Capability** |
| **Part of a larger system? Describe** | N/A |
| **Made up of other systems? Describe** | MT, OCR, TTS, ASR |
| **Standards supported** | N/A |
| **Used by** | Multiple Organizations |
| **Presentations at AMTA—what/when** | --- |
| | |
| **Contact for more information** | |
| **Name** | Kemp Gouldin |
| **Telephone** | 202-429-2981 |
| **Email** | Gouldin@sakhr.us |

| | |
|---|---|
| **Company or organization** | SDL |
| **Products or software** | SDL Trados Studio 2009 |
| **Version** | SP2 |
| **Available for licensing Y/N** | Yes |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Desktop Translation productivity software with integrated machine translation |
| **Description** | Designed for translators, SDL Trados Studio 2009 is the ultimate translation software. It provides all your translation tools in one integrated editing, reviewing, project management and terminology environment. It combines the leading translation memory technology with a host of powerful features including integrated machine translation to help you translate new content faster.<br><br>Using machine translation as part of the SDL Trados Studio 2009 SP2 environment, you can translate more content and deliver it faster than before. When translating, segments that are not leveraged from translation memory can automatically be machine translated for the translator to review, then accept and amend if necessary, or decide to manually translate instead. As a translator, you can determine how much machine translation you want to use in your process as well as select your machine translation engine of choice.<br><br>SDL Trados Studio 2009 SP2 supports 3 machine translation engines that are available over an internet connection – SDL Enterprise Translation Server, Language Weaver, and Google Translate. It is also possible to link with trained SDL and Language Weaver machine translation solutions that your customers or LSPs provide you with access to further increase your productivity. |
| **Languages or language pairs supported** | 52 languages and more than 2.704 language pairs to suit your project |
| **Unique about this system** | SDL Trados Studio 2009 SP2 supports a total of 3 machine translation engines - SDL Enterprise Translation Server, Language Weaver, and Google Translate.<br><br>The complete product offering now allows everyone in the translation supply chain to leverage the SDL Trados Studio translation memory infrastructure and distribute files and information across the whole process of translation - from translators to developers and enterprises. |
| **New this year** | Yes – SP2 is new this year |
| **Part of a larger system? Describe** | See above. |
| **Used by** | Ltconsult, Proverba Translation & Technology |

| | |
|---|---|
| **Presentations at AMTA—what/when** | Evaluating vendors for MT and post-editing at Avaya<br>DAY ONE: Commercial Track, 4:30 pm |
| | |
| **Contact for more information** | |
| **Name** | Jeremy Harpham |
| **Telephone** | +44 (0)1628 417260 |
| **Email** | jharpham@sdl.com |

| | |
|---|---|
| **Company or organization** | SDL Language Weaver |
| **Products or software** | Automated translation software |
| **Version** | SaaS, Server v5.2 for Government |
| **Available for licensing Y/N** | Yes |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Statistical Machine Translation |
| **Description** | Automated translation + TrustScore<br>Manage all forms of automated, real-time, multilingual communication in a central platform<br>No installation required, point to a URL & go<br>Trusted Translation via TrustScore™<br>Simple, flexible integration options |
| **Languages or language pairs supported** | 80+ language combinations |
| **Unique about this system** | SDL Language Weaver offers translation software with an automatic, quality ranking algorithm (TrustScore). |
| **New this year** | TrustScore |
| **Part of a larger system? Describe** | SDL Language Weaver is a standalone solution, but is also integrated into multiple applications for added value and benefit |
| **Made up of other systems? Describe** | No |
| **Standards supported** | XML,TMX, XLIFF |
| **Used by** | Fortune 500 companies, Global web properties, Defense and Intelligence organizations. |
| **Presentations at AMTA— what/when** | **Commercial Track:** Trusted Translations Deliver Compelling Results for the Travel Industry<br>**Government Track: Utilizing Automated Translation with Quality Scores to Increase Productivity** |
| | |
| **Contact for more information** | |
| **Name** | Hannah Grap |
| **Telephone** | 310-437-7300 |
| **Email** | hgrap@languageweaver.com |

| | |
|---|---|
| **Company or organization** | Spoken Translation, Inc. |
| **Products or software** | Converser for Healthcare |
| **Version** | 3.0 |
| **Available for licensing Y/N** | Y |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | MT, with speech input and output. Input via typing, handwriting, or point-and-click is also enabled. |
| **Description** | Interactive speech translation with multi-media input and customizable Translation Shortcuts™ |
| **Languages or language pairs supported** | English<>Spanish |
| **Unique about this system** | STI's proprietary technologies enable users to (1) interactively monitor and correct speech recognition and machine translation and (2) translate flexible, open-ended conversations, instead of furnishing only pre-packaged translations of fixed phrases or translations restricted to narrow domains.<br>Numerous efforts to enable automatic translation of wide-ranging dialogue have so far failed to provide widely applicable and commercially sound solutions because users have had no way to monitor or control accuracy of translation. Consequently, translation quality and user confidence have remained too low. Converser, however, can substantially raise accuracy and confidence through its patented technology for interactive feedback and correction. |
| **New this year** | Additional human-factors and ergonomic features as developed for a large healthcare organization. |
| **Part of a larger system? Describe** | N/A |
| **Made up of other systems? Describe** | Contains licensed components for MT and TTS. ASR is performed by a third-party component installed separately but controlled from Converser. |
| **Standards supported** | Windows XP or 2000. Windows Vista and Windows 7 are under development. |
| **Used by** | A pilot program is under way at a nationwide healthcare organization. |
| **Presentations at AMTA—what/when** | We have previously presented at Technology Showcases in 2009 in Ottawa and 2008 in Honolulu. |
| | |
| **Contact for more information** | |
| **Name** | Mark Seligman, Ph.D. |
| **Telephone** | 510 843-9800 |
| **Email** | mark.seligman@spokentranslation.com |

| | |
|---|---|
| **Company or organization** | SYSTRAN Software, Inc. |
| **Products or software** | SYSTRAN Enterprise Server 7 |
| **Version** | 7 |
| **Available for licensing Y/N** | Yes |
| **Kind of product** | Machine Translation |
| **Description** | Hybrid Machine Translation<br>SYSTRAN Training Server allows corporate customers and Language Services Providers to independently train SYSTRAN Enterprise Server 7 to any selected domain or business objective and to produce publishable-quality translations. Through use of the SYSTRAN Training Server, SYSTRAN Enterprise Server 7 automatically learns from existing and validated translations and updates itself as these translations are reused. |
| **Languages or language pairs supported** | **52 Commercially Available Language Pairs**:<br>English to/from: Arabic, Chinese, Dutch, French, German, Greek, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish, Swedish<br>French to/from: Dutch, German, Greek, Italian, Portuguese, Spanish<br>Italian to/from: German, Portuguese<br>Portuguese to/from: German<br>Spanish to/from: German, Italian, Portuguese<br>**25 Language Pairs, Limited Availability**:<br>Albanian to English<br>Arabic to French<br>Bengali to English<br>Bulgarian to English<br>Chinese to French<br>Czech to English<br>Chinese to/from Korean<br>Dari to English<br>English to Hungarian<br>Farsi to English<br>Hindi to English<br>Hungarian to French<br>Japanese to/from Chinese<br>Japanese to/from Korean<br>Norwegian to English<br>Pashto to English<br>Polish to French<br>Serbo-Croatian to English<br>Slovak to English<br>Tajik to English<br>Ukrainian to English<br>Urdu to English |

| | |
|---|---|
| **New this year** | SYSTRAN Training Server is part of the SYSTRAN Enterprise Server 7 solution suite which automatically learns from existing and validated translations. As a result, translation quality improves and continuously adapts to customer needs. The translation engines update themselves as new translations are added to the system.<br><br>SYSTRAN Training Server has two components. SYSTRAN Corpus Manager helps users acquire, manage and store parallel texts, existing translations and all other language assets. SYSTRAN Training Manager lets users run iterative training cycles based on predefined tasks. The training tasks include resource extraction, dictionary validation and document alignment. |
| | |
| **Contact for more information** | |
| **Name** | Jin Yang |
| **Telephone** | 858-320-2408 |
| **Email** | jyang@systransoft.com |

| | |
|---|---|
| **Company or organization** | The Technology Development Group (www.thetdgroup.com) |
| **Products or software** | OMPC and LLB Light |
| **Version** | OMPC 3.0 / LLB Light 1.0 |
| **Available for licensing Y/N** | Y |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | OMPC = Online Multimedia Parallel Corpora database<br>LLB Light = Linguist Language Broker, a Dynamic social terminology management application |
| **Description** | OMPC is an online web-application that serves as a resource for parallel corpora associated with multimedia files, available for assisting in USG NLP tool creation<br>LLB Light: is a light desktop application that allows users to seamlessly interface with the main LLB terminology management system |
| **Languages or language pairs supported** | OMPC: Arabic, Chinese, English, French, Russian, Spanish<br>LLB Light: All Unicode languages |
| **Unique about this system** | OMPC: This system allows for the import, alignment/segmentation, search and sharing of multimedia-associated parallel corpora in a variety of languages<br>LLB Light: This application provides users with the ability to quickly store, collaborate on, search and share unique foreign language terms and phrases |
| **New this year** | OMPC: This version includes additional parallel corpora across a variety of languages as well as improved search capabilities. Users can now view and download targeted corpus collections.<br>LLB Light: this brand-new version is a light application that allows users to seamlessly interface with the main LLB terminology management system. |
| **Part of a larger system? Describe** | OMPC: encompasses the entire system<br>LLB Light: interfaces with the main LLB terminology management system |
| **Made up of other systems? Describe** | OMPC: Solr 1.3, AppTek-Aligner, HotSpot,<br>LLB Light: Solr 1.3, |
| **Standards supported** | |
| **Used by** | OMPC: US-Gov users and associated researchers/academics<br>LLB Light: US-Gov users and associated researchers/academics |
| **Presentations at AMTA— what/when** | Mike O'Malley is presenting a paper on "Challenges of a Distributed Parallel Corpora" |
| | |
| **Contact for more information** | |
| **Name** | Jon Phillips |
| **Telephone** | 571-262-2699 / 571-525-7232 |
| **Email** | jphillips@thetedgroup.com |

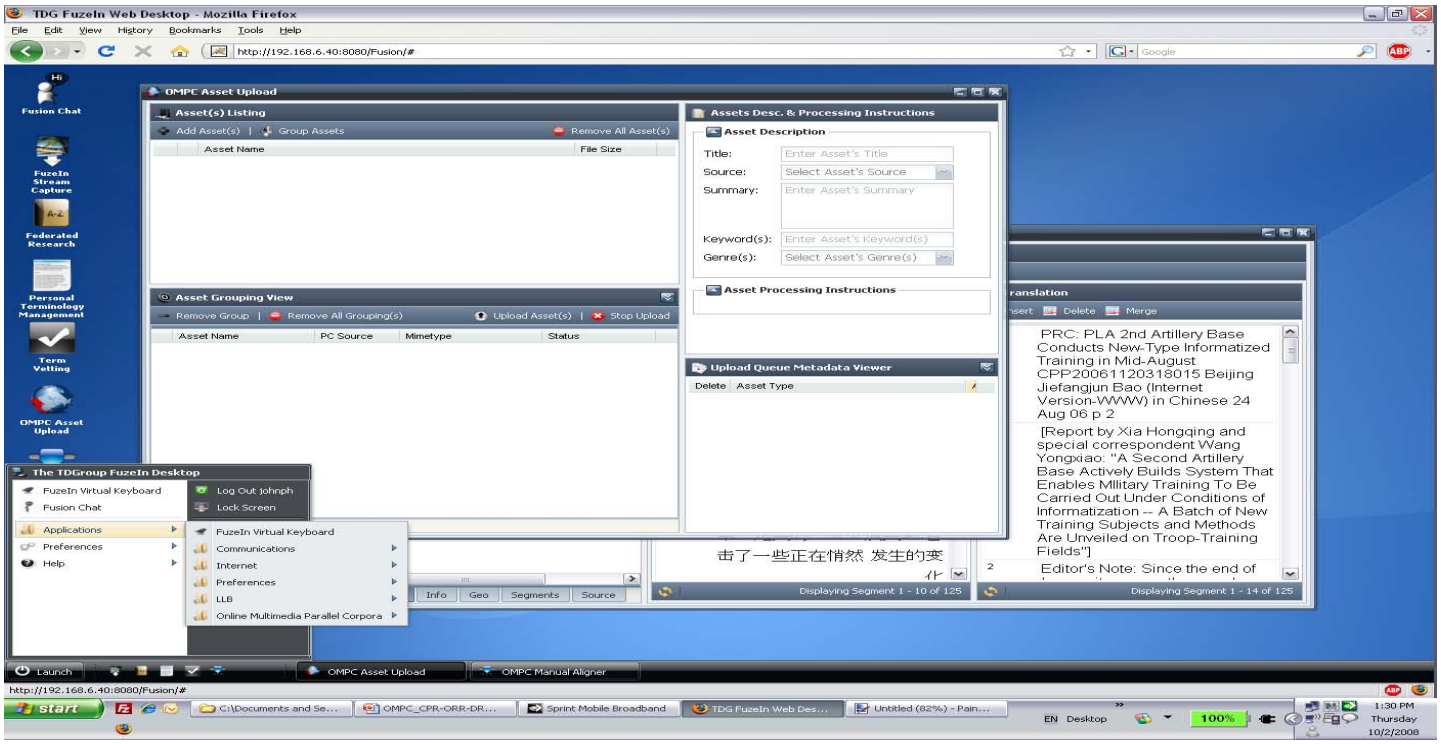
## OMPC

There is increasing interest and need to have access to a foreign language corpus of original source data and the corresponding translated data that represents what the Intelligence Community (IC) collects, processes and analyzes on a daily basis. Computer based language technologies have made great strides in the past several years, especially the ability to save data in its original form and/or convert it into Unicode based text. This capability makes accessing and archiving the data feasible and of great value to various groups in the IC: the researchers, developers, analysts, instructors and learners of foreign languages.

This type of data archiving is not unique. The Linguistic Data Consortium (LDC) is a successful example in academia. "The LDC was founded in 1992 to provide a new mechanism for large-scale development and widespread sharing of resources for research in linguistic technologies. Based at the University of Pennsylvania, the LDC is a broadly-based consortium that now includes more than 100 companies, universities, and government agencies. Since its foundation, the LDC has delivered data to 197 member institutions and 458 non-member institutions (excluding those who have received data as a non-member and later joined)." However, the data that LDC collects and archives is not representative of internal IC data.

Like the academic and research community, the IC needs access to an enormous variety of linguistic data — speech, text, lexicons, and grammar — to improve the processes by which it is able to understand and apply automated linguistic solutions to the data. In addition, the IC also has a need for specialized linguistic data based upon its specific and unique mission. Such databases are expensive to create and document. The most valuable data is that which has been validated within the IC. Further value is added by associated human translation (HT), machine translation (MT), and/or metadata. Having a centralized repository for such data is highly desirable because sharing the resources across agencies provides benefits that go beyond just having access to the data. It enriches the systematically documented repository and offers savings for all participants. In addition to the specific benefits to each entity in the Community, the overall benefits in language processing research and development will have a positive impact on the community as a whole.

The Online Multimedia Parallel Corpora (OMPC) System is designed to ingest, segment and align, store and index, and provide newly created and existing parallel corpora (PC) documents at the sentence or paragraph level. In many cases, these parallel corpora are linked to multimedia artifacts; for example they are transcriptions of video broadcasts.

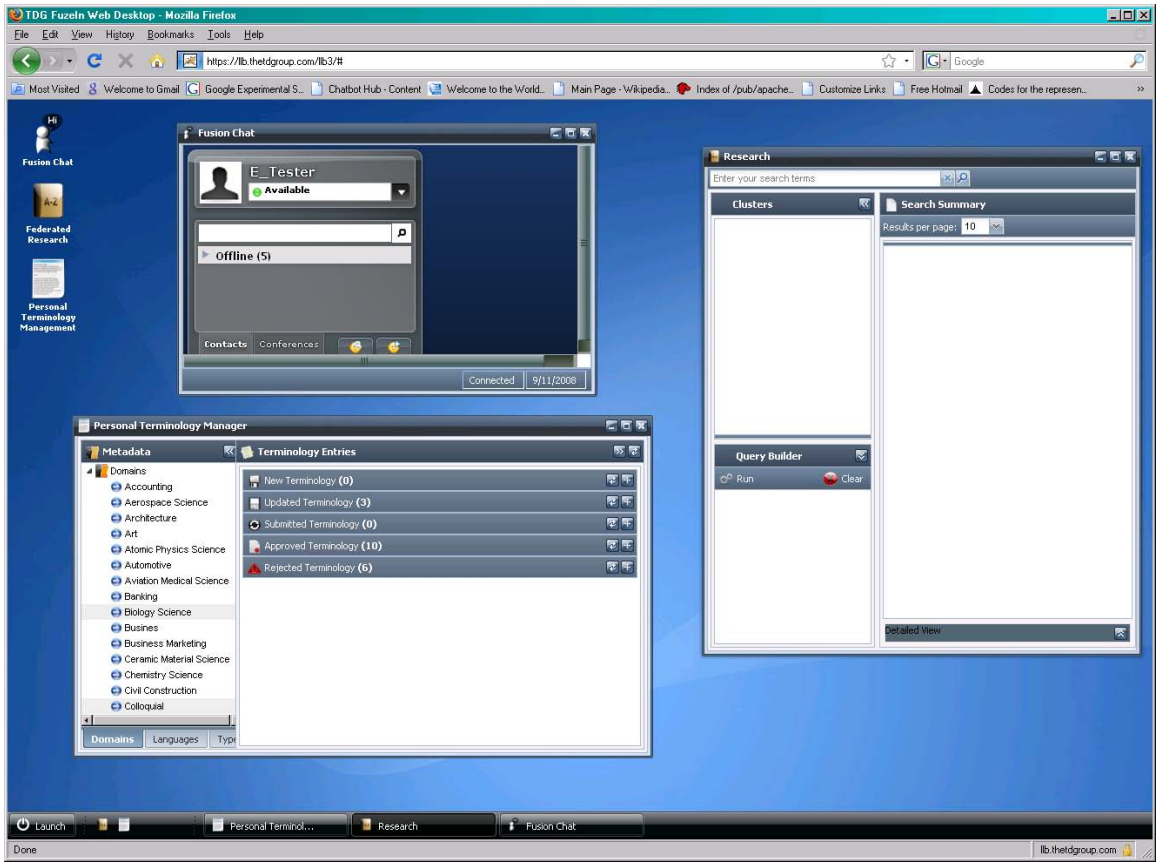


## LLB Light

The Linguist Language Broker (LLB) is a highly-customized web-based terminology management system. The LLB system includes general terminology management, and also integrates several features that are desired by today's intelligence community (IC) translators.

LLB is a "translator's toolbox," an integrated workspace where translators can collaborate on and research definitions and translations of uncommon foreign-language words or phrases. These can be added to the LLB data-store, allowing IC translators to learn and benefit from their colleagues' knowledge and prior work. The LLB also utilizes links to existing IC dictionaries and terminology systems. LLB further aids translators by filtering their search results (by usage domain, for example) to find the most relevant ones.

LLB Light is a small desktop application that allows user to easily access the robust terminology management functions of the main LLB system.



| | |
|---|---|
| **Company or organization** | National Research Council of Canada |
| **Products or software** | WeBiText |
| **Available for licensing Y/N** | Free web-based service |
| **Kind of product (e.g., MT, OCR, Dictionary, etc.)** | Free concordancer which allows searching in large multilingual web sites. |
| **Description** | WeBiText is a free, web-based concordancer, that allows translators to search in large, high quality, multilingual web sites like those of the Government of Canada, European Union organizations and other international organizations. |
| **Languages or language pairs supported** | 29 languages, mostly European, but including some less common ones like Inuktitut (language of the Inuit people of Canada) and Haitian Creole. |
| **Unique about this system** | Contrarily to conventional Translation Memories or concordancers, WeBiText comes pre-loaded with large amounts of high quality multilingual content, covering a wide range of topics. |
| **New this year** | No. It has been available since 2008, but never advertised widely until December 2009. |
| **Used by** | Several hundreds of translators. Overall, 64% of queries originate from freelance, but with 27% originating from translators working for the Canadian federal government. |
| **Presentations at AMTA— what/when** | We are submitting a full paper on this system at the workshop on parallel corpora. |
| | |
| **Contact for more information** | |
| **Name** | Alain Désilets |
| **Telephone** | 819-712-2813 |
| **Email** | Alain.desilets@nrc-cnrc.gc.ca |

WeBiText (www.webitext.ca), a free concordancer developed at the National Research Council of Canada, allows translators to search in large, public, high quality multilingual web sites like those of the Government of Canada or European Union institutions (Désilets et al, 2008).

We validated the product concept using field data collected in a Contextual Inquiry study where we observed professional translators while they carried their day to day work. In that study, we noticed that translators often used web search engines like Google to manually search in large multilingual web sites. It was also apparent that this was a time consuming process (approximately 2 minutes for a single pair of sentences) which was amenable to automation.

An overview of the system's functionality shows how it automates the time-consuming manual search process and allows translators to find several pairs of sentences in a few seconds, with a single click of the mouse. In particular, we highlight the following interesting points:

- Search in 29 languages (including some small languages like Inuktitut and Haitian Creole)

- Search in a growing list of 63 web sites from reputed organizations (ex: government of Canada, European Union institutions, international organizations) which cover a wide range of domains (ex: legal, technology, health, policy, public administration)
- View parallel sentences in context (i.e. see sentences that preceded or followed)
- View a pair of web pages side by side

The results of a log analysis have shown how WeBiText is currently used by translators and that it is currently experiencing rapid growth in the number of daily searches. An informal evaluation of the economic value of this traffic, in terms of hours of work saved to translators, is estimated to be $1 million annually as of June 2010. A look at the provenance of those queries indicates that home-based freelancers account for 64% of them, with 28% for Canadian Government users (both federal and provincial), 4% from private LSPs, 0.5% universities and 3.5% miscellaneous. WeBiText offers advantages to each of these various constituencies.

Based on a sample of queries from our logs, WeBiText produces at least one pair of correctly aligned sentences for 84% of the cases. Looking at the nature of queries being submitted by users, we see that they correspond to a somewhat even mix of general and specialised language translation problems. This is in contrast to the predominance of general language queries that was found in the logs of TransSearch, a similar tool based on the Canadian Hansard (Macklovitch et al, 2008, Simard and Macklovitch, 2005). We have also identified the common types of feedback received from users, through the Contact Us and Feedback links.

We plan to continue developing and improving the system.

References:

Désilets, A., Farley, B., Patenaude, G., Stojanovic, M. “WeBiText: Building Large Heterogeneous Translation Memories from Parallel Web Content.”, Proc. of ASLIB Translating and the Computer (30), London, UK, Nov 27-28, 2008.

Macklovitch, A., Lapalme, G., Gotti, F., (2008), “TransSearch: What are translators looking for?”, in Proc AMTA'08, Waikiki, Hawaii, October 21-25, 2008.

Simard, M., Macklovitch, E., (2005). “Studying the Human Translation Process Through the TransSearch Log-Files.”, In Proc. AAAI Symposium on Knowledge Collection from Volunteer Contributors, Stanford, USA, March 2005.