2012
AMTA
20Years

The Tenth Biennial Conference of the
Association for Machine Translation in the Americas

TECHNOLOGY SHOWCASE GUIDE

SAN DIEGO, CA
OCTOBER 28- NOVEMBER 1, 2012

Technology Showcase

Army Research Laboratory	When to Select, Replace, or Augment MT Engines
Asia Online	Language Studio Pro Version 3.0
Basis Technology	Rosette Linguistic Platform and Highlight
Center for Next Generation Localization	
Center for Next Generation Localization	TMT Prime
CETRA	Systran Translation Project Manager 7.3.1.3
Columbia University	Columbia Arabic Dialect Modeling
Copenhagen Business School	CASMACAT Workbench Prototype 1
Deutsches Forschungszentrum fuer Kuenstliche Intelligenz	GmbH
IPTranslator	IPTranslator
Johns Hopkins University	Joshua Decoder Toolkit 4.1
MateCat	MateCat
Microsoft Corporation	Microsoft Translate
MITRE Corporation	Clipper
MultiCorpora	MultiTrans Prism Version 5.5
National Tsing Hua University	crossGRASP Prototype 1
National Tsing Hua University	TransArts Prototype 1
Northrop Grumman	DCGS-A
Pan American Health Organization	PAHOMTS 4.7
Raytheon BBN	BBN Web Monitoring System
Safaba Translation Solutions	Safaba MT Version 2.1
SAIC	OmniFluent and Talk2Me
Sakhr	
SDL	
Spoken Translation	Converser for Healthcare 3.1, SpeechTrans
Systran	
TranslatorsFactory.net	
United States Department of Defense	Center for Applied Machine Translation (CAMT)
Virtus	Virtus
Yandex	Yandex Translate

Company Name: Asia Online Pte Ltd

Products Being Shown (Name and Version): Language Studio Pro V 3.0

Company URL: http://asiaonline.net/

POC Name(s): Kirti Vashee

POC Email: kirti.vashee@asiaonline.net; k.vashee@gmail.com

Unique Features: The Language Studio™ Suite is a comprehensive translation platform that contains a set of technologies ranging in scope from tools for linguistic analysis, data mining, real-time multilingual chat and language processing tools to fully integrated, automated, high speed translation (millions of words per day).

Years of development and research have gone into developing the core Language Studio™ platform that translates more than 500 language pairs. A large number of European, Middle Eastern and Asian languages are supported now, and more than 200 are currently under development.

Many of the technologies that have made Asia Online the quality leader in automated machine translation are available under license. Complex tasks such as word and sentence segmentation, Optical Character Recognition (OCR) and sentence alignment are made simple with the Language Studio™ platform.
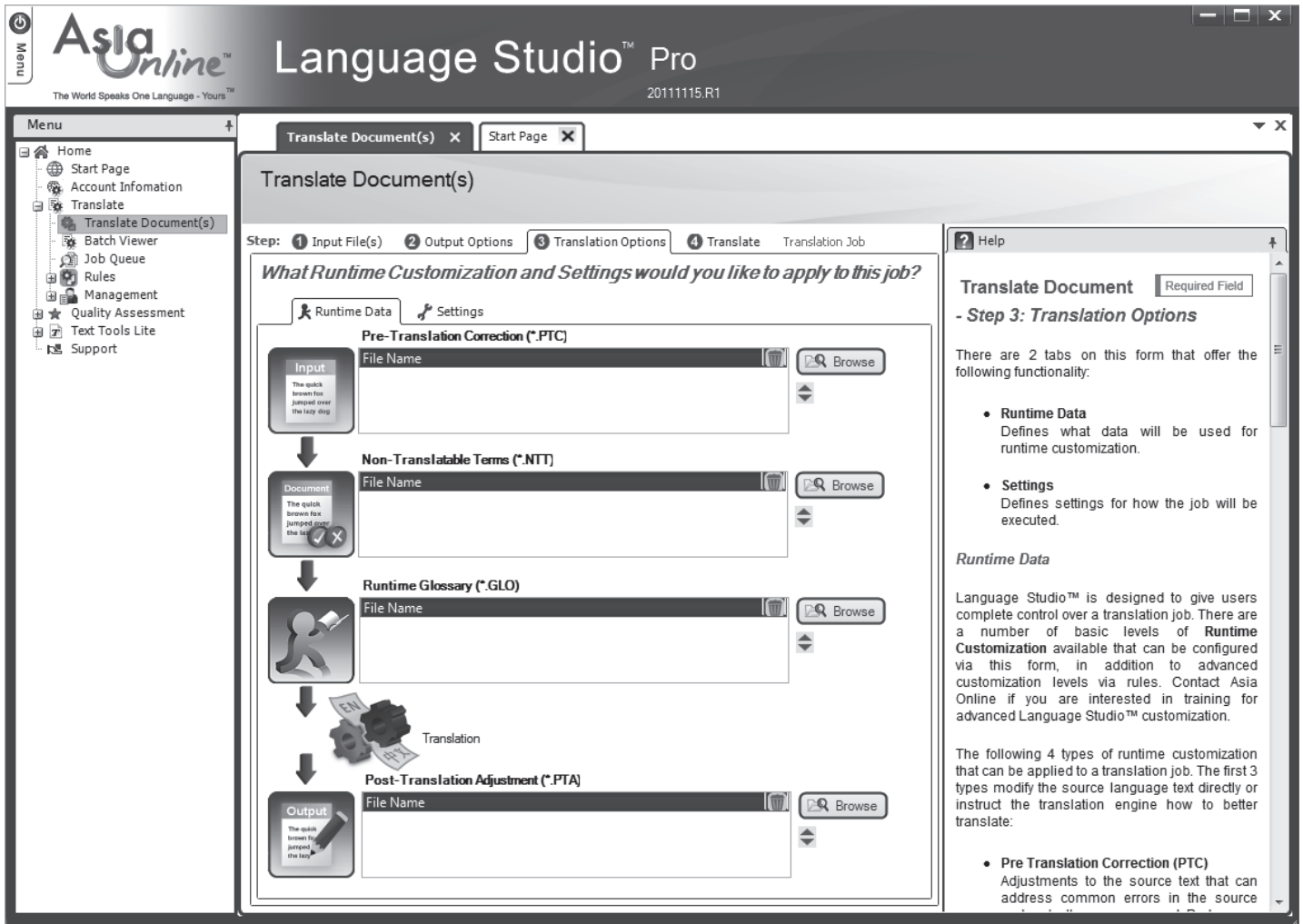
The products and technologies listed below are a subset of Asia Online's comprehensive linguistic technology portfolio. These include:

- Alignment tools to ensure that parallel data is accurately paired and aligned
- Translation memory quality assessment and cleanup tools
- Translation memory normalization and scrubbing tools
- Data format conversion tools
- Data cleaning tools
- Dictionary and glossary management tools
- Optical Character Recognition – Asia Online is the first to deliver very high accuracy OCR services for a number of complex Asian language scripts.
- Word and phrase segmentation - Asia Online software automatically segments Asian language words and phrases, which speeds automated spell-checking and indexed text searches

Advances Since Last Year: Language Studio™ Pro enables the direct submission of documents for translation right from the desktop in document formats such as TMX, XLIFF, Microsoft Word, Microsoft PowerPoint, Plain Text, XML, HTML and more. Access all of the more than 520 language pairs available in Language Studio™ right from your desktop!

Some of Language Studio™ Pro 3.0 new features include:

- A new wizard interface for submitting translations.
- Easy setting of Pre Translation Corrections, Runtime Glossaries, Non-Translatable Terms and Post Translation Adjustments.
- Default and job specific translation rules allow the same translation engine to be used for different customers.
- Editing and testing tools for rules.
- Batch translations allow for multiple documents to be submitted and monitored.
- Automated downloading of finished documents.
- In addition to TMX, every job now also produces an XLIFF file for importing back into TMS systems or editing tools.
- Translation Project manager tool.
- Automated update tool, making future updates easier.
- Step by step help and guides on every screen making it easy for even the most novice translation user.

Language Studio™ Pro
20111115.R1

Menu

Translate Document(s) ✕   Start Page ✕

Translate Document(s)

Menu
- Home
  - Start Page
  - Account Information
  - Translate
    - Translate Document(s)
    - Batch Viewer
    - Job Queue
    - Rules
    - Management
  - Quality Assessment
  - Text Tools Lite
  - Support

Step: ❶ Input File(s)   ❷ Output Options   ❸ Translation Options   ❹ Translate   Translation Job

What Runtime Customization and Settings would you like to apply to this job?

👤 Runtime Data   🔧 Settings

Pre-Translation Correction (*.PTC)

Input
The quick brown fox jumped over the lazy dog

File Name   🗑   🔍 Browse

Non-Translatable Terms (*.NTT)

Document
The quick brown fox jumped over the lazy...

File Name   🗑   🔍 Browse

Runtime Glossary (*.GLO)

File Name   🗑   🔍 Browse

Translation

Post-Translation Adjustment (*.PTA)

Output
The quick brown fox jumped this lazy

File Name   🗑   🔍 Browse

❓ Help

Translate Document   [Required Field]

- Step 3: Translation Options

There are 2 tabs on this form that offer the following functionality:

- **Runtime Data**
  Defines what data will be used for runtime customization.

- **Settings**
  Defines settings for how the job will be executed.

Runtime Data

Language Studio™ is designed to give users complete control over a translation job. There are a number of basic levels of **Runtime Customization** available that can be configured via this form, in addition to advanced customization levels via rules. Contact Asia Online if you are interested in training for advanced Language Studio™ customization.

The following 4 types of runtime customization that can be applied to a translation job. The first 3 types modify the source language text directly or instruct the translation engine how to better translate:

- **Pre Translation Correction (PTC)**
  Adjustments to the source text that can address common errors in the source

Company Name: Basis Technology

Products Being Shown (Name and Version): Rosette linguistics platform 7.6 and Highlight language analysis suite 5.0

Company URL: www.basistech.com

POC Name(s): Joel Ross

POC Email: joelr@basistech.com

Unique Features:

Rosette's Unique Features:

The Rosette® Linguistics Platform is the world's most widely used component library for multilingual text retrieval and analysis. Rosette provides automatic language identification, linguistic analysis, entity extraction, and entity translation from unstructured text, all in a single, unified framework.

- Automatically identify the language or languages present in a document for correct analysis, filtering, and retrieval of text
- Increase the precision and recall of full-text search in multiple languages
- Extract important concepts from unstructured text, such as names, locations, dates, and identifiers
- Translate names from many languages into English
- Match names against lists or databases, regardless of language or other variations
- Rosette Chat Translator: Analyze and translate casual Arabic in Latin script, also known as Arabizi, into Modern Standard Arabic.

Highlight's Unique Features:

- Automatic and interactive translation of names and places
- Standardization of foreign names and places in multiple languages to the IC standard and standards
- In-document dictionary look-up
- Access to the CIA's Chiefs of State
- Quickly look up foreign features on a map

Advances Since Last Year:

Rosette's Advances Since Last Year:

- Improvements in Language Identification for multilingual documents
- Added support for Turkish linguistic analysis, improvements to many other languages
- Added support for Portuguese entity extraction, identification of "Product" as a new entity type
- Improvements to Arabic, Farsi, Dari, and Japanese name translation
- Added support for Spanish, Russian name matching. Improvements to English, Arabic, Farsi, Dari, Japanese.
- Web Services API

Highlight's Advances Since Last Year:

- Added the Farsi and Pashto IC standards
- Improvements to Arabic, Farsi, Dari, and Pashto name translation
- Improved user interface to be more intuitive, requiring less training
- Web services API
- Single installer for all four components

Company Name: Centre for Next Generation Localisation

Product Being Shown (Name and Version): TMT^{PRIME} v1.1

Company URL: http://www.cngl.ie

Company Information:

The Centre for Next Generation Localisation (CNGL) is a dynamic Academia-Industry partnership with over 100 researchers developing novel technologies addressing the key localisation challenges of volume, access and personalisation. CNGL has been researching language technologies for a number of years and now houses one of the worlds largest academic research groups focused on machine translation. Over the past year CNGL researchers have developed a patent pending back-end system to help translators and post-editors better utilize the outputs of statistical machine translation systems.

Representative Name(s): Declan Groves, Aswarth Dara, Sandipan Dandapat

Representative Email: {dgroves,adara,sdandapat}@computing.dcu.ie

Product Description:

TMT^{PRIME} is a recommender system that facilitates the effective use of both translation memory (TM) and machine translation (MT) technology within industrial language service providers' (LSPs) localization workflows. LSPs have long used Translation Memory (TM) technology to assist the translation process. Additionally, recent research has successfully shown that MT systems can be used along with TMs to assist the translation process for certain sentences in a Computer Aided Translation (CAT) system. However, there are no commercially available solutions for integrating MT systems with existing TM systems, in terms of selecting the best quality output. TMT^{PRIME} provides a commercially-viable platform that allows for the seamless integration of MT with legacy TM systems to provide the most effective (least effort/cost) translation options to human translators, based on the TMT^{PRIME} confidence score.

Language Supported: Any

Unique Features:

- Feature-based translation recommendation to reduce post-editing effort/cost
- Multiple MT and TM systems within an unified scoring mechanism
- Estimates document-level translation costs using different MT/TM engines

Advances Since Last Year: First prototype version

TMT^{PRIME} is a self-supporting platform that provides translation recommendations based on translations produced by different available MT and TM systems. The recommendation can be either at sentence or document level. In this demonstration, we present an experimental version of the TMT^{PRIME} that integrates:

1. Three different translation systems: a TM system, the OpenMatrex (http://www.openmatrex.org/) and the Bing Translator (http://www.bing.com/translator/)

2. System-independent features and a classifier to find the confidence (*recommendation score*) of each translation

3. Translation recommendation for each sentence in a document and estimates the cost for document level translation

TMT^{PRIME} is the first commercial attempt for effectively integrating MT into localization workflows that traditionally use TM technology, in order to improve post-editor productivity and subsequently reduce the overall cost of translation. TMT^{PRIME} will also facilitate the estimation of translation costs when using different MT/TM systems both in isolation and within the recommended usage of those MT/TM engines.

AMTA Technology Showcase 2012

Company Name: Centre for Next Generation Localisation

Product Being Shown (Name and Version): MTwatch v1.0

Company URL: http://www.cngl.ie

Company Information:

The Centre for Next Generation Localisation (CNGL) is a dynamic Academia-Industry partnership with over 100 researchers developing novel technologies addressing the key localisation challenges of volume, access and personalisation. CNGL has been researching language technologies for a number of years and now houses one of the worlds largest academic research groups focused on machine translation. Recently, CNGL has started work on a number of commercial development projects with the aim of building upon our existing research in order to build commercially-relevant applications.

Representative Name(s): Declan Groves, Paraic Sheridan, Steve Gotz

Representative Email: {dgroves,psheridan,sgotz}@computing.dcu.ie

Product Description:

At the CNGL we continue our work on state-of-the-art research in Machine Translation and also on commercially-driven applications for our work. One such project is the development of MTwatch, a suite of data analytics, evaluation and visualization technologies for cloud-based MT platforms.

Cloud-based MT helps to lower the entry barrier for translators, providing them with the ability to train their own MT systems, using their own data, and to run these systems in the cloud, thus circumventing the need for previously required computing resources. However, in order to train high-quality systems users of such platforms, particularly non-expert users, require additional support that can give them an indication of the health of such MT systems and the expected translation quality.

MTwatch is built with the goal of empowering the user of cloud-based MT solutions by providing them with the ability to monitor and visualize the quality and performance of their trained systems. By providing the user with feedback as to the quality and suitability of their data for training MT systems we can both manage their expectations and provide them with guidance as to how to improve the quality of the translations produced by their MT engines.

For this first version we will focus on demonstrating the evaluation component of MTwatch as deployed within the KantanMT cloud-based MT platform.

Language Supported: Any

Unique Features:

MTwatch v1.0 incorporates a suite of integrated analytic and evaluation tools that:

- Provide an overall system score (averaged across multiple evaluation metrics)
- Provides metric gauges which display individual evaluation scores
- Displays comparisons to previous scores
- Provides access to an historical view of metrics measured against word count
- Incorporates a number of widely-used evaluation metrics, including BLEU, TER, FMeasure and METEOR
- Provides preliminary training data analytics

Advances Since Last Year: First prototype version
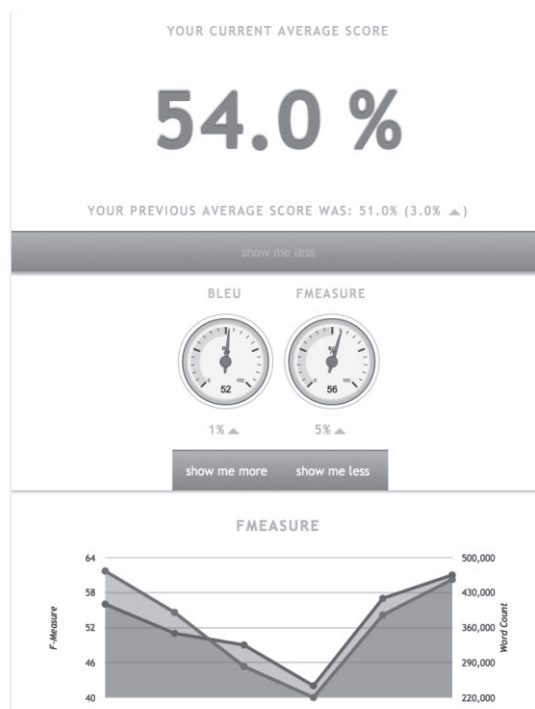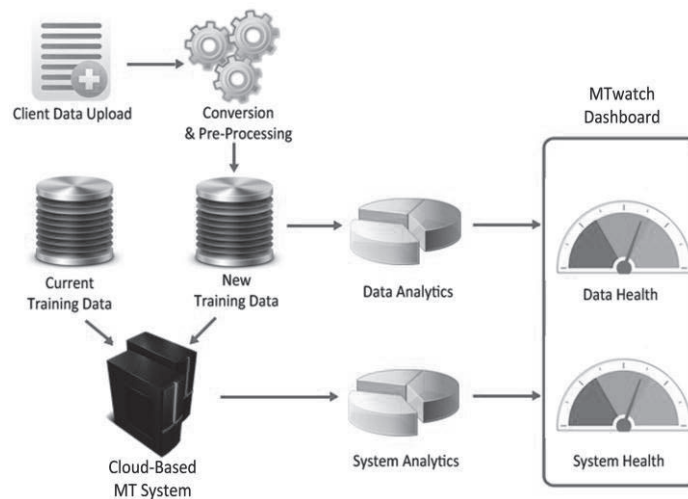
Company Name: CETRA, Inc.

Products Being Shown (Name and Version):

- Systran Translation Project Manager 7.3.1.3

Company URL: www.cetra.com

Company Information: Language Services Provider

Representative Name(s):

- Jiri Stejskal, President and CEO
- Tommy Tomolonis, Project Manager

Representative Email:

- Jiri.Stejskal@cetra.com
- Tommy.Tomolonis@cetra.com

Product Description: Use of Systran Translation Project Manager in the translation of market research surveys in xml format – Capabilities and Issues

Languages Supported: FIGS

Unique Features: N/A

Advances Since Last Year: N/A

Other Information: N/A

Company Name: CADiM: Columbia Arabic Dialect Modeling

Products Being Shown (Name and Version):
Dialectal Information Retrieval Assistant (DIRA) v2.0

Company URL: http://www1.ccls.columbia.edu/~cadim/

Company Information:
The CADiM group was established in 2005 at the Center for Computational Learning Systems (CCLS) at Columbia University. CCLS is an organization focusing on sponsored projects in multiple research disciplines such as Machine Learning and Natural Language Processing (NLP). CADiM consists of three PIs (alphabetically: Mona Diab, Nizar Habash and Owen Rambow), 12 doctoral students and three research programmers. CADiM is a world leader in research on Arabic and Arabic Dialect NLP.

Representative Name: Dr. Nizar Habash
Representative Email: habash@ccls.columbia.edu

Product Description:
DIRA is a tool that provides query expansion capabilities for Information Retrieval (IR) systems. It takes one or more lemmas (in English or Modern Standard Arabic) as input, and for each lemma it generates MSA and/or dialectal inflected forms. The forms are generated based on a set of expansion options such as user-specified preferences for variant (MSA and/or dialect) and for inflectional features (number, aspect, gender etc.). Apart from the expanded list of inflected forms, the tool also provides the ranking score for each inflected form.

Languages Supported: Egyptian Arabic, Modern Standard Arabic

Unique Features:
Performing query expansion and providing a list of expansions along with the dialect and ranking score for each of the expansions.

Advances Since Last Year:
We converted the research-based tool that was written in Perl into a much more robust, commercial grade tool developed in Java. This has well defined integration points and an architecture that enables it to be integrated with any commercial information retrieval systems.

Other Information:
Research and development of DIRA was sponsored by the Department of Defense.

DIRA v2.0

Dialectal Information Retrieval Assistant

Input	Example
English lemma	see
or Modern Standard Arabic lemma	رأى
Output	*Example*
Modern Standard Arabic inflected forms	رأيت، رأينا، يرى، ترى، نر، يراها، وسنراهم،...
Egyptian Arabic inflected forms	شاف، شفت، يشوفهم، ماشفش، ومشفتلهوش، ...

Company Name: Copenhagen Business School

Products Being Shown (Name and Version):
CASMACAT Workbench Prototype 1

Company URL: www.casmacat.eu

Company Information: CASMACAT is an EU 7th framework project

Representative Name(s): Ragnar Bonk, Michael Carl, Jakob Elming

Representative Email: {rb,mc,je}.ibc@cbs.dk

Product Description:

The CASMACAT workbench, prototype 1, is a web-based tool for machine translation post-editing and for analyzing human translation processes. The workbench is designed to examine the effects of integrating translation technology into the translation process, and how this affects the behavior of the human translator. Key features of the workbench are: web-based technology, extensive logging of user behavior including keyboard logging and eye tracking, and exact replay of the translation session. The tool has been released as open source.

Languages Supported: Any

Unique Features: web-based workbench for post-editing of MT output with extended logging of user behavior (keyboard and gaze data) as well as replay functionality

Advances Since Last Year: First prototype released

Other Information: None

The CASMACAT Workbench builds on experience from the Translog (http://www.translog.dk/). a tool designed for studying human reading and writing processes e.g. as in translation . The CASMACAT workbench extends Translog's key-logging and eye-tracking abilities with a browser-based front-end and an MT server in the backend. The main advances are:

1. it is using web-based technology which allows for easier portability across different machine platforms and versions

2. it reproduces a much more realistic translation environment by visually and functionally resembling commercial translation tools

The CASMACAT Workbench provides different layout options such as two columns with source segments on the left aligned to translation segments on the right (see Figure 1), or one column with already translated segments above the current segment and future source segments below. Shortcut keys are used for functions such as navigating between segments. The translation field can be pre-filled by machine translation through a server connection and also automatically updated online from an interactive machine translation server.
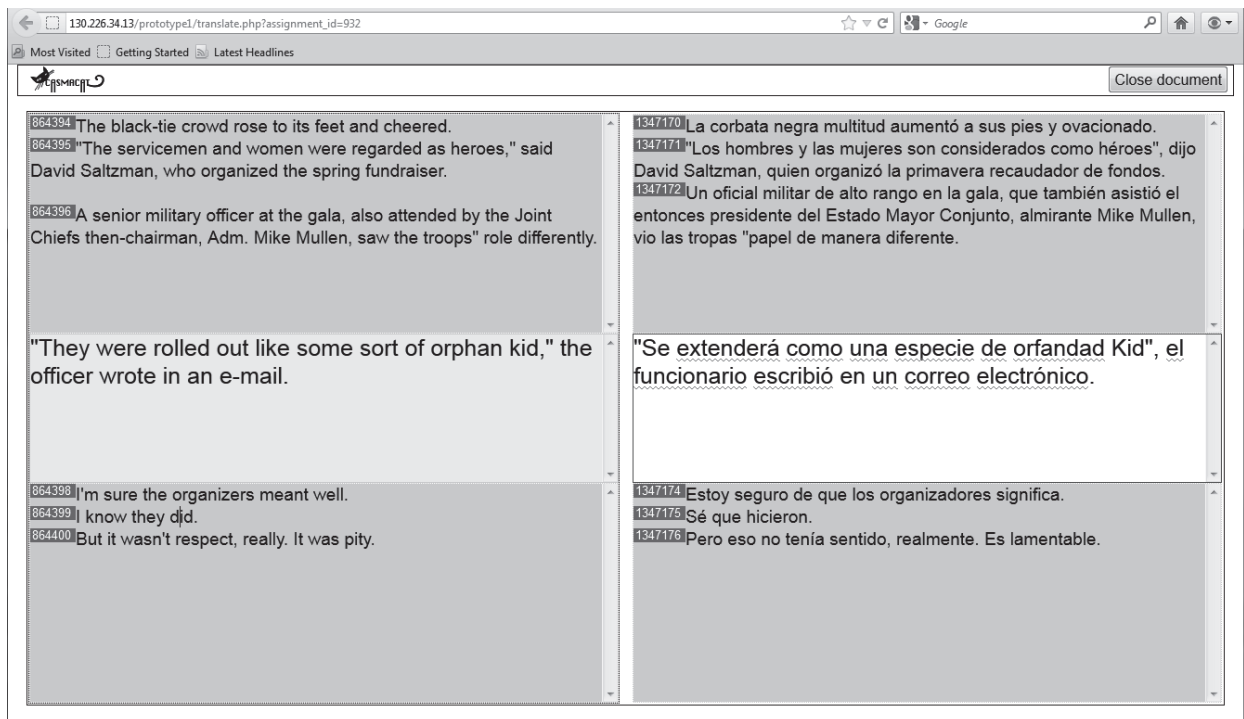
Figure 1: Screenshot of CASMACAT prototype-I: The source text appears in segments on the left and the target text on the right. The light box in the middle of the screen contains the current segment which is editable. Translators can navigate with ALT+UP/DOWN through the segments.

The main innovation of the CASMACAT Workbench is its exhaustive logging function. This opens for completely new possibilities of analyzing the translator's behavior both, in a qualitative and quantitative manner. The extensive log file contains all kinds of events, keystrokes, mouse, cursor navigation, as well as gaze behaviour (if an eye-tracker is connected) that have occurred during the translation session. The log data can be used to analyze and model the translation process in a more precise manner. Extensive logging allows also to replay translation sessions which allows us to visually gain insight into the choices made by the translator during translation and post-editing.

Company Name: DFKI GmbH, Language Technology Lab

Products Being Shown (Name and Version): Appraise

Company URL: http://www.dfki.de/lt/

Company Information: Our mission is to ease the use of human languages by machines and to improve solutions that benefit from language use. We conduct advanced research in language technology and provide novel computational techniques for processing text, speech and knowledge. We strive for a deeper understanding of human language and thought, studying the true needs of the end user and the demands of the market.

Representative Name(s): Christian Federmann

Representative Email: cfedermann@dfki.de

Product Description: Appraise is an open-source tool for manual evaluation of Machine Translation output. Appraise allows to collect human judgments on translation output, implementing annotation tasks such as

1) translation quality checking;
2) ranking of translations;
3) error classification;
4) manual post-editing.

It features an extensible XML import/output format and can easily be adapted to new annotation tasks. Appraise also includes automatic computation of inter-annotator agreement scores allowing quick access to evaluation results.

Appraise is available under an open, BSD-style license.

Languages Supported: Any language your annotators can work with.

Unique Features: Wide range of different annotation tasks available; easily extensible to new annotation scenarios; language independent; under active development sine 2008.

Advances Since Last Year: Automatic computation of inter-annotator agreement scores. Improved the different annotation tasks. Updated documentation and release package in time for the Seventh MT Marathon 2012 which took place September 3-8, 2012 in Edinburgh, Scotland.

Other Information: Software package and full source code available from GitHub at https://github.com/cfedermann/Appraise

Company Name: IPTranslator

Products Being Shown (Name and Version): IPTranslator

Company URL: http://www.iptranslator.com

Representative Name: Dr. John Tinsley

Representative Email: john@iptranslator.com

Product Description

IPTranslator is a machine translation (MT) web service aimed at intellectual property professionals which addresses a number of key issues in their daily workflow, namely: the lack of a single resource for patent MT; inconsistency of translation quality; inconsistency of user experience; inability to influence the translation process; and the inability to act on the translation output (edit, extract meaning, share, etc.).

At the core of the service is a set of MT systems for strategic languages which have been developed and optimised specifically to translate patent text. Comprehensive evaluations, including human judgements, have shown translation quality to be better than or equal to the leading machine translation systems approximately 75% of the time.

The translation offering is supplemented with a number of value-adding features, some of which offer a level of personalisation for the user not available in competing services. Most notably, users have the option of maintaining their own terminology database which can guide the translation process in a particular manner, e.g. to always translate a term in a particular way.

Languages Supported
English to/from Chinese, Japanese, Korean, French, German, Spanish and Portuguese

Unique Features
- **Domain tuned:** State-of-the-art patent specific machine translation
- **Fully integrated:** On-the-fly translation via browser plugin
- **Personalised:** Improved quality over time via translation editing and terminology management
- **Flexible:** Translate web content, PDF, and other file types
- **Added Value:** Segment highlighting, bilingual keyword extraction, download, share, and more…

Patent Search and Translation Trends

Most patent searchers make use of machine translation technology

Google Translate	79%
Online Dictionaries	39%
Babelfish	21%
Patent Office Tools	78%
Babylon	2.5%
Microsoft Bing	1.2%
Other	27%
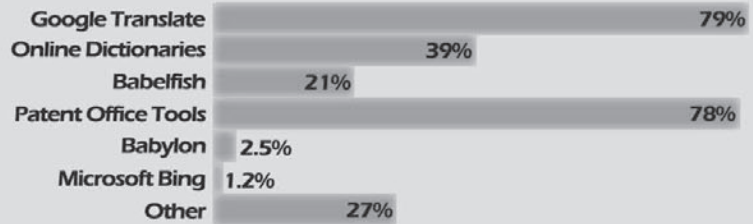
84% of patent searchers would like better machine translation

Demand for translation is highest for Asian languages but exists for others as well

Korean
Spanish
Portuguese
Chinese
Russian
Japanese
German
French

Patent searchers use multiple search tools and resources

Espacenet	70%
Patentscope	22%
Thomson	39%
PatBase	54%
Total Patent	5%
Questel	43%
STN	50%
Google Patent	17%
Other	7%
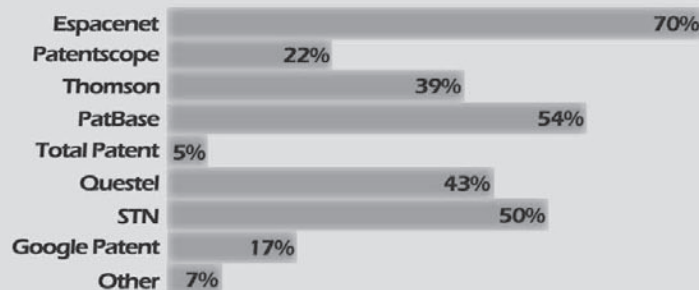
Patent searchers work across a variety of web browsers

Safari
Other
IE
Firefox
Chrome
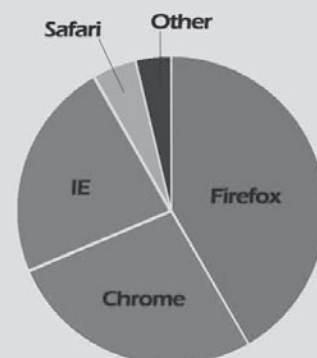
IPTranslator provides personalised on-demand patent-specific machine translation for all major languages

The translation service can be accessed on-the-fly from any web browser no matter where the patent search is being carried out

IPTranslator.com

Company Name

MateCat

Products Being Shown (Name and Version)

MateCat v. 1 (Beta)

Company URL

http://www.matecat.com/

Company Information

MateCat is a research project financed by the European Union grant 287688, under FP7-ICT-2011-7, with an investment of 2.7 million euros.

The MateCat consortium consists of:
- Fondazione Bruno Kessler
- Translated srl
- Université du Maine, Le Mans
- University of Edinburgh

Representative Name(s)

Marcello Federico, Marco Trombetti, Alessandro Cattelan

Representative Email

federico@fbk.eu, marco@translated.net, alessandro@translated.net

Product Description

MateCat sets out to fill the gap between machine translation technology and the translation industry's needs and processes. It integrates statistical machine translation and collaborative translation memories, within the human translation workflow. MateCat's goal is to define new operating conditions for statistical MT in order to better match the typical workflow and functional requirements of those using CAT tools. It aims at developing MT technology that requires minimal post-editing effort by the human translator. MateCat will advance the state-of-the-art by making MT technology aware of how it is used, attuning itself to the domain, adapting to the corrections and implicit feedback from the user, and providing useful information to the user.

The CAT tool component is an enterprise-level computer assisted translation tool designed to allow for the assessment of MT technology. With MateCat, MT researchers can collect data to measure the quality of machine translation engines by testing them in a professional translation environment. In particular, it allows to collect data on:

1. Post-editing effort, that is the average percentage of word changes applied by the translators on the suggestions provided by the CAT tool.
2. Time to edit, that is the average translation drafting speed by the translators.

Languages Supported

All western languages

Unique Features

State-of-the-art statistical MT employed in more or less interactive settings generally lacks dynamic adaptation capabilities that allow it to learn from the user's feedback. On the other hand, a very natural desire of a human translator using MT in a CAT tool would be to see a consistent use of terminology and style that is similar to his/her own throughout the text, and that once he/she corrects an error this should not occur again in the following text segments. In addition, such adaptations should happen in real time.

On-line learning

MateCat will provide methods for the automatic self-correction of MT making use of the implicit feedback of the user. The segments of text that have already been post-edited by the user will be analysed and compared with the corresponding automatic translations by the MT in order to spot the errors together with their corrections and the portions accepted by the translator. The MT models will be modified accordingly by penalizing the former and reinforcing the latter, or, more drastically, by removing the source of errors. Although ad-hoc transformations could be similar to those for the project adaptation, the goal here is to make them very precise and consistent with the actual translator. Through on-line adaptation, which is performed in real-time and sentence by sentence, MT should automatically translate the following segments more and more consistently with respect to the previous ones from the point of view of the translator's lexical and stylistic preferences.

Context-aware translation

MateCat will also focus on providing suggestions by MT which are consistent with respect not only to the already edited segments but also to the whole document. This context information will be embedded in the statistical models and will enable better disambiguation, for instance, between lexical alternatives. The context-based models will combine information about recurring terms and expressions extracted during the document analysis with the corresponding chosen and confirmed translations as soon as they become available. In particular, translation constraints related to inter-sentence and intra-sentence anaphoric expressions, to syntactic concordances, and to lexical coherence will be taken into account by means of specific statistical models.

Real-time processing

The core components of traditional MT systems, that is, the translation and the language models, are generally static: they never change after an initial training phase. This means that they are unsuitable for a dynamic environment like the one that MateCAT is designing for

translators. In order to model the dynamic changes depicted in the two previous tasks, MateCat will develop innovative data-structures that can be rapidly and effectively updated as soon as a new translation is supplied by the user, and innovative, efficient algorithms for performing this adaptation in such a way that the whole process takes place in real time and is transparent to the translator. Moreover, efficiency will be improved by taking advantage of single CPU multithreading, as well as distributed computing facilities running on private clusters or computer clouds.

Advances Since Last Year

First beta release

Other Information
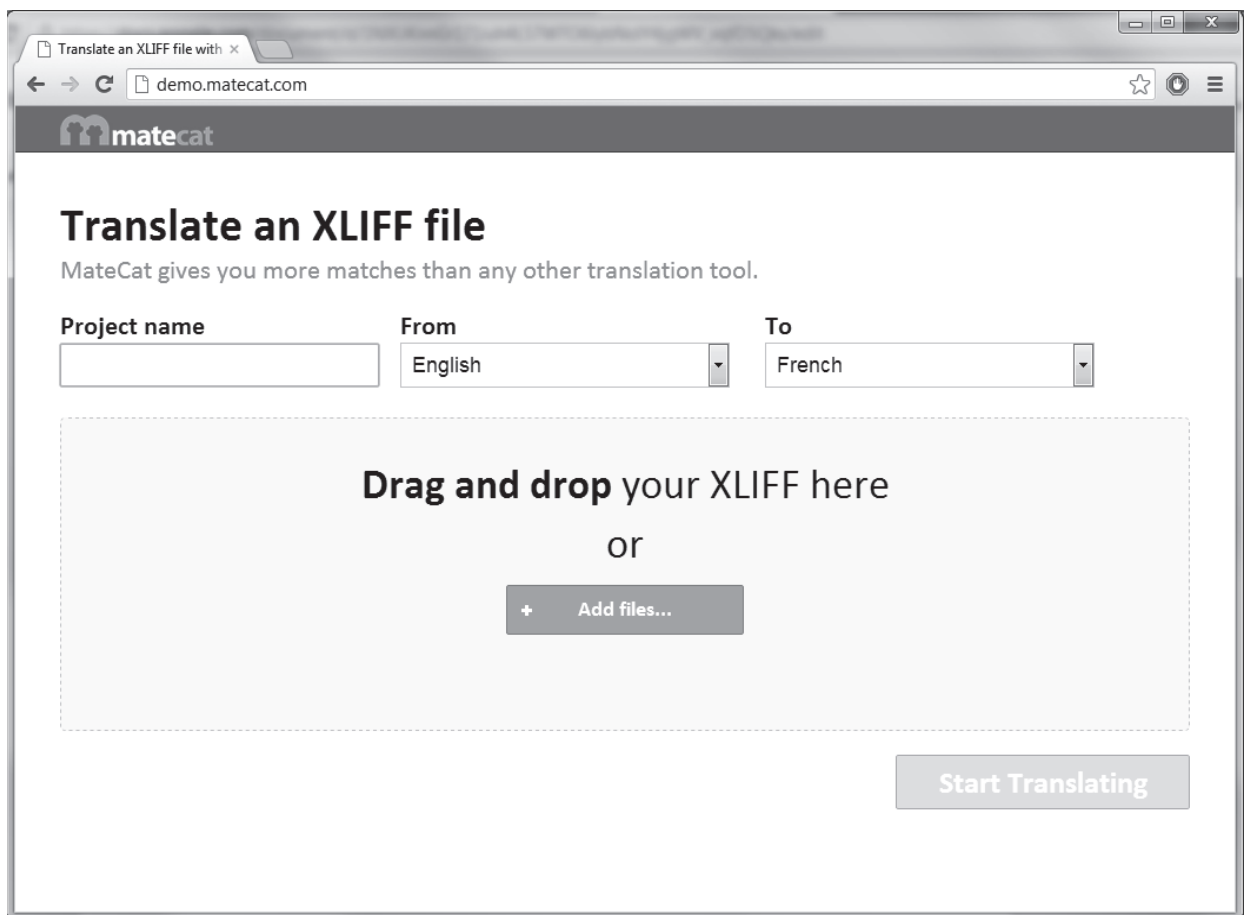
MateCat CAT tool web UI - File importer.

MateCat CAT tool web UI - Translation editor.

Company Name: Microsoft

Products Being Shown (Name and Version): Microsoft Translator

Company URL: www.microsoft.com/translator

POC Name(s): Chris Wendt, Vikram Dendi

POC Email: translator@microsoft.com

Unique Features:

- Machine Translation
- Self-service MT training
- Collaborative Translation Framework
- Wide language and domain coverage

Advances Since Last Year:

This year we are introducing the Microsoft Translator Hub, a web based tool for self-service customization of the Microsoft Translator engine. The unique quality of the Hub is that the custom data and the very large statistical models used by Bing Translator work in unison, to provide the best automatic translation, even with small amounts of custom training data.

The Collaborative Translation Framework, which is the combination of a cloud-based translation memory (TM), a set of methods to read and write to that memory, and a web page widget for interactive editing of that TM, is fully integrated into the Microsoft Translator API and the Microsoft Translator Hub, enabling you to seamlessly use your community feedback and corrections for custom MT system training, all without having to install anything locally.

All of this provides for significantly better raw MT than an untrained, generic engine. A number of translation tools integrate this functionality, making post-editing of customized MT a snap.

Company Name: MultiCorpora

Products Being Shown (Name and Version): MultiTrans Prism Version 5.5

Company
    Name: MultiCorpora
    URL: multicorpora.com
Information: MultiCorpora is one of the few language technology experts in the world with the experience and ability to understand the needs of organizations that must control multilingual communication from authoring to publication.

MultiCorpora was created in 1999 and has grown its revenue and profitability at an average of 22% since 2006. With several offices in North America and Europe, the company has now over 275 enterprise clients.

Its flagship product is MultiTrans Prism, a translation management software suite designed to provide organizations with the technology they need to empower translation business owners with controls over their translation assets management, project management and their translation process without adding extensive system administration. MultiTrans Prism 5.5 was recently reviewed by Common Sense Advisory in their TMS live tool and achieved high scores in key areas such as interoperability, business reporting, translation and web functionalities. With this newest release, its total composite score increased from 638 to 693 .

The ideal candidate for its solutions is a global enterprise, a language service provider, an international organization or a government body that needs to translate a certain volume of content in other languages. To name a few, RR Donnelley, Nomura, numerous Canadian and Swiss federal departments, and many UN organizations rely on MultiTrans Prism to manage their mission critical translation operations.

Representative
    Name: Daniel Gervais
    Email: dgervais@multicorpora.com
    Telephone: 1 877.725.7070

Product Description: MultiTrans Prism is an innovative client-server software solution available as a web based application integrating project and business management, translation memory, and terminology management to enable multilingual communication and workflow automation from authoring to publication. The solution is powered by a best-in-class terminology management system, an Advanced Leveraging Translation Memory technology (ALTM), and a project management system with business intelligence reporting. These elements represent the heart of the MultiTrans Prism solution and can be provided as a whole or as modules to support your responsibilities within the multilingual communication chain.

The Machine Translation Interface is designed to integrate third-party machine translation systems such as Systran, ProMT, Google Translate, and Microsoft Translator into the MultiTrans Prism translation process. Once connected to a machine translation system, this functionality becomes ingrained into the translation process. This is an optional process you can turn on for projects with document types that are well suited to machine translation processing.

The results from the machine translation engine become part of the standard matches returned to the translator. Machine translation matches can be categorized as less reliable than previously approved matches from the translation memory, so the translation memory matches will be displayed first, or in the case of pre-translation, replacements will first be selected from the translation memory (based on match criteria) and replacements from the machine translation engine will only be used for those without TM matches.

MultiTrans Prism is a preconfigured solution designed to give you all the benefits of a true Translation Management System (TMS) quickly and easily. It can be deployed either on your own servers (and even web enabled), or hosted and managed on MultiCorpora's servers if you do not have the required IT infrastructure. MultiTrans Prism has different licensing options: perpetual licenses configured to your exact needs; licenses to facilitate outsourced translation services; a package of licenses ideal for the business needs of a small or medium size LSP; and a convenient pay-as you-play Software as a Service (SaaS) model for those who may not have a capital budget to invest in a full system, but who need access to a state-of-the-art TMS. MultiTrans Prism also offers compatibility with conventional TM tools and supports all current standards including the most recent developments with XLIFF and TBX.

Languages Supported: All UNICODE languages

Unique Features:
- ALTM – advanced leveraging allows the recovery of all repeated segments from full paragraph to sub-segment matches and can quickly feed parallel corpora to SBMT
- Revolutionary WordAlign technology allows not only automatic concordance searches, but an exact distribution of how the concordance was previously translated. All translations can be seen in full context.
- A fully integrated terminology solution offers comprehensive terminology workflow capabilities, TBX compatibility and more. Advanced access management allows controlled dissemination while keeping all terminology together for ease of management.
- A unique graphical interface for setting up simple or complex workflows which can be specific to various document types.
- Online post-editing interface
- Automatic updating of TextBase translation memories directly following specific points in the workflow. For example if a document delivery is finalized, the translation memory would be automatically updated.
- Automatic analysis and pre-translation of documents entered into the workflow system, including the automatic creation of an industry-standard XLIFF file for translation.
- A suite of quality assurance checks, including filtering for specific issues, a validation report and live validation checks ensure top-quality results.
- Terminology and concordance lookups directly linking Microsoft Word to the terminology and TextBase TM web applications, without any necessary pre-treatment of the Word documents.
- Translation workbenches available directly in Microsoft Office (Word and PowerPoint) allowing users to work directly in familiar editing environments.

Advances Since Last Year:

MultiTrans Prism (version 5.5) was released to the market in May 2012 at the Localization World Conference in Paris. MultiTrans Prism version 5.5 carries on with our customer-oriented development path. There were three primary focus areas in the new release: usability, security, and deployment ability. We have introduced major enhancements in all three areas. As well, MultiTrans Prism version 5.5 adds two optional new functions in the areas of project management and business management: the capability of using a real-time web editing environment for translators and the capacity to automatically generate reports.

We have also recently included further enhancements to the XLIFF Editor, revamped HTML document processing, added faster and easier license deployment, and TM export in MS Word.

Company Name:

NLP lab, National Tsing Hua University

Productions Being Shown (Name and Version):
crossGRASP Prototype 1

Company URL:
nlp.cs.nthu.edu

Company information:
Natural Language Processing team at National Tsing Hua University

Representative Names:
[*]Chung-chi Huang, Mei-hua Chen, Hung-ting Hsieh, and Jason S. Chang
[*]Institute of Information Science, Academia Sinica, Taipei, Taiwan

Representative Email:
{u901571,chen.meihua,vincent732,jason.jschang}@gmail.com

Product Description:
crossGRASP is a web-based cross-lingual grammar pattern finder for language learning. It learns to translate first-language search queries and provide grammatically-motivated usage summaries for the translations, expected to help foreign language learning. It has been applied to computer-assisted language translation and the results are encouragingly promising.

Language Supported:
Chinese to English

Unique Features:
a web-based language learning tool for translating phrases and collocations and providing usage (grammatical and lexical) summary for the translations to assist English as Foreign or Second Language learners.

Advances Since Last Year:
First prototype released

A Cross-Lingual Grammar Pattern Finder for Language Learning

Many linguistic search queries (e.g., "play role") are submitted to language learning tools on the Web every day, and an increasing number of Web services specifically target linguistic search queries. For example, *Sketch Engine* provides summaries of the search word's collocational and grammatical behavior while *JustTheWord* further clusters the word's collocates. However, language learners may not know the query words in the language they are learning in the first place and they may need a usage summary of the search words for (word usage) navigation and (sentence) generation.

Consider the linguistic search query "扮演角色" (play role). The best response for this query is probably not the single top-ranked translation "the role" returned by *Google Translate*. A Good response for language learning might contain top-n possible translations with grammatical patterns characterizing their phraseological tendencies. Take the translation "play role" for instance. Its predominant pattern grammar includes "*play* DT ADJ *role*", "*play ~ role* PREP(*in*) VBG" and so on. DT, ADJ, PREP and VBG stand for determiners, adjectives, prepositions and gerunds, respectively.

We present a new system, *crossGRASP*, that automatically learns to translate first-language search queries and provide grammatically-motivated usage summaries for the translations, expected to help foreign language learning. Take "扮演角色" for example. *crossGRASP* generates "*play* DT ADJ *role*", "*play ~ role* PREP(*in*) VBG" for one of its translation "*role played*". We now describe the system in more detail.

Translation. Hong Kong Parallel Text is utilized to obtain translation equivalents. Specifically, *crossGRASP* uses IBM models for word alignment, grow-diag-final for alignment combination, and Koehn et al (2003)'s heuristics for phrase pair extraction. To accommodate flexible collocations (i.e., collocations or phrases that are not fixed) like "play role", skipped phrase pairs e.g. ("play ~ role", "扮演 ~ 角色") are also considered. Otherwise, the search of "扮演角色" for translation would normally unsuccessful.

Pattern Grammar. To generate grammatical patterns for a translation at run-time, we preprocess the target-language reference corpus (e.g., British National Corpus). The preprocessing involves lemmatizing and part-of-speech tagging the sentences, and constructing inverted files for the words in the corpus. Lemmatization reduces the effect of morphology on statistics while part-of-speech tags are used for generalization such as "oneself" of "enjoy *oneself*" and "one's" of "make up *one's* mind" used in learner dictionary. Inverted files, on the other hand, contain information on the sentences and the positions where a lemma appears and its surface forms for pattern generation.

At run-time, we first identify the sentences containing the multi-word translation. Then we replace the words in the sentences with the lemmas in the translation and the rest of the words with their corresponding part-of-speech tags, and extract fixed-window segments surrounding the translation from the transformed sentences (inspired by (Gamon and Leacock, 2010)). Finally, we will acquire grammar patterns such as "*play ~ role* PREP(*in*) VBG" and "*play* DT ADJ *role*". In implementation, *crossGRASP* further displays the lexical instantiations (e.g., "play ~ role in determining" and "play an important role") of the patterns and their example sentences for language understanding and learning. Note that the translations of the lexical phrases may also be provided.

In summary, we have introduced a method for learning to translate a linguistic search query and summarize the word usages of its translations for language learning. The method involves extracting contiguous or skipped translation equivalents and generating words' phraseology in terms of grammar patterns. We have implemented and evaluated the method applied to computer-assisted language learning. In our sentence completion task, English learners in Taiwan benefited more from the *crossGRASP*'s pattern grammar than traditional dictionary (the gain was almost tripled). The promising result prompts us to evaluate *crossGRASP* more fully in the near future.

Company Name:

NLP lab, National Tsing Hua University

Productions Being Shown (Name and Version):
TransArts Prototype 1

Company URL:
nlp.cs.nthu.edu

Company information:
Natural Language Processing team at National Tsing Hua University

Representative Names:

Hung-ting Hsieh, *Chung-chi Huang (corresponding author), Mei-hua Chen, *Ming-hong Bai, Jason S. Chang, and *Keh-jiann Chen
*Institute of Information Science, Academia Sinica, Taipei, Taiwan

Representative Email:

{vincent732,u901571,chen.meihua,mhbai.tw,jason.jschang}@gmail.com;
kchen@iis.sinica.edu.tw

Product Description:
TransArts is a web-based in-page computer-assisted translation assistant for specific domain of arts and exhibitions. It aims to help translators understand and translate source documents and provides functionalities such as term sense disambiguation and translation based on online Web data or off-line crawled corpora and Chinese segmentation while translating. The in-page translation and sense disambiguation make the tool more interactive and promising.

Language Supported:
Chinese to English

Unique Features:
A web-based computer-assisted translation tool incorporates modules of term disambiguation and translation via online or offline data sets, segmentation while translation, document translation, each of which constitutes a research topic and software application.

An In-Page Computer-Assisted Translation Tool for Specific Domain

Many translators submit texts to machine translation (MT) systems on the Web in the hope of accelerating their translation process. However, most online MT systems such as *Google Translate* are not for specific domain such as arts and so on.

Consider the text of arts "北宋徽宗政和六年的仿商宋鼎" (*The tripods in the Northern Song period mimicking those in the Shang and Song Dynasty*). The best translation for this sentence is probably not "North Song governance and six years of imitation, Song Ding" returned by *Google Translate* trained on general domain. A good response, on the other hand, might contain sense disambiguation for words like "鼎" and translation or transliteration for terms like "北宋", and "商".

We present a Chinese-to-English computer-assisted translation system, *TransArts*, that learns to translate terms or texts in the specific domains of arts and exhibitions (i.e., term or text translation for art displays of National Palace Museum in Taiwan). *TransArts* provides document translation as well as in-page term disambiguation and translation and learns translation from the off-line parallel sentences or online mixed-code sentences. We will now take a closer look at *TransArts*' functionalities, each of which is a research issue and is currently based on well-known solution. The working prototype is at http://nlp-service1.cs.nthu.edu.tw:8080/~nlplab/vincent/assistedTranslation/.

Term Disambiguation. *TransArts* disambiguates the senses of the user highlighted terms using Lesk algorithm which compares the word overlapping ratios between sense definitions and the input article. For instance, if a user highlights "鼎" in the above text, *TransArts* disambiguates it as "ancient word use for metal container" for assistance in monolingual understanding.

Online Term Translation. *TransArts* finds translations for the highlighted terms on the Web, if indicated. In this module, we submit the source terms, which may be accompanied by translations of partial source terms or relevancy feedback as query expansion, to search engines, collect the returned snippets, and extract possible in-page translations via patterns. Take a source term "資料

探勘" and its translation "data mining" for example. A pattern "(" lies between them and there are more. In implementation, our patterns are learned by submitting bilingual phrase pairs and analyzing their snippets. Such online term translation is designed to accommodate new words/senses.

Seg and Trans. *TransArts* translates at character level to bypass the word boundary issue in Chinese-to-English MT. Specifically, *TransArts* generates top-n in-page term translation while segmenting. We segment and translate the highlighted Chinese terms or sentences using greedy and dynamic programming algorithm based on bilingual translation probabilities.

Document Translation. *TransArts* integrates Moses as MT server for sentence translation. In the user interface, we allow for editing and reordering the translation to speed up the translation process and alleviate translators' burden.

To translate texts in the specific domains of arts and exhibitions, we crawl and sentence align the bilingual articles on National Palace Museum websites. Apart from the parallel corpora such as Hong Kong Parallel Corpus, FBIS and Wikipedia's language links, *TransArts* further trains on domain-related snippets containing term translation pairs in Bilingual WordNet and multi-lingual Art & Architecture Thesaurus. That is, we submit phrase pairs to search engines to collect snippets in the domain of interest for better translation variety and coverage. The functionality of **Translation Memory** in *TransArts* returns translations by referencing aforementioned data collections.

In summary, we have introduced a method for learning to assist translators in understanding and translating source documents in a specific domain. The method incorporates functionalities, each of which constitutes a research topic, such as term disambiguation, online and off-line term translation, translation while segmenting, and document translation. We have implemented the method as applied to computer-assisted translation and we look forward to its deployment among translator community and its extensive evaluation.

Company Name: Pan American Health Organization

Products Being Shown (Name and Version): PAHOMTS® 4.7

Company
    Name: Pan American Health Organization
    URL: www.paho.org
    Information: The Pan American Health Organization (PAHO) is an international public health agency with almost 110 years of experience in working to improve health and living standards of the countries of the Americas. It serves as the specialized organization for health of the Inter-American System. It also serves as the Regional Office for the Americas of the World Health Organization and enjoys international recognition as part of the United Nations system.

Representative
    Name: Mr. Hermes Camelo
    Email: camelohe@paho.org
    Telephone: (202) 974-3783

Product Description:

PAHOMTS® is a fully automatic Machine Translation (MT) program, developed and maintained by PAHO's computational linguists, translators, and programmers. The translation unit has used PAHOMTS® to process over 5 million words annually since 1980. Staff and freelance translators postedit the raw output to produce high-quality translations with a 30-50% gain in productivity.

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. 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.

Other features include:

- Fully graphic translation interface that may be invoked from the desktop, from a Microsoft® Word toolbar or from a Microsoft® PowerPoint® toolbar;
- Fully graphic Dictionary Update utility, with context-sensitive help and other user-friendly features.
- Batch Translation utility allowing the translation of several input files in one single step;
- Dictionary Browse utility, with copy-paste capabilities;
- Dictionary Merge utility to merge the user's customized dictionaries into PAHO's master dictionaries;
- Import capability to facilitate the incorporation of user glossaries;
- Export utility to create lists of terms added or modified by the user;
- ActiveX® macros to translate selected text within Microsoft® Word or PowerPoint®.

At PAHO, MT has proven to be especially effective for conference documents, scientific papers, training materials, technical abstracts, and Web pages. At sites that have access to large volumes of text in electronic format, rough translations can be obtained quickly and economically using MT. The output preserves the format of the original document, including tables, and is easy to polish using PAHO's postediting macros.

Languages Supported:

English, Spanish, and Portuguese

Unique Features:

Specialized in public health and medicine; glossaries and context-sensitive rules working together to select alternate translations; postediting macros

Advances Since Last Year:

About 10,000 new dictionary entries, minor bug fixes, and enhancements in all translation engines and format handlers.

Other Information:

PAHOMTS$^{®}$ is copyrighted by the Pan American Health Organization. It can be licensed by non-profit institutions through a signed license agreement. The one-time license fee includes one to six language directions, dictionaries, and on-line manuals. Training in dictionary updating, technical support, and upgrades are available.

More information available at http://www.paho.org/english/am/gsp/tr/machine_trans.htm

Company Name: Raytheon BBN Technologies

Products Being Shown (Name and Version): BBN Web Monitoring System (WMS) integrated with dialect query expansion and identification tools from Columbia University (prototype)

BBN Mobile MultiMedia Monitoring System (M4S) v1

Company URL: http://bbn.com

POC Name(s): Martha Lillie

POC Email: mlillie@bbn.com

Unique Features:

WMS: dialect based query expansion enables search of Arabic web media using dialect terms; highlighting of Egyptian dialect words in web text to support language understanding/learning

M4S: Laptop based system transcribes both streaming and file-based foreign language media in real time

Advances Since Last Year: Both products are new this year.

Company Name: Safaba Translation Solutions

Product Being Shown (Name and Version): Safaba MT v2.1

Company URL: www.safaba.com

Company Information:
Safaba is a provider of automated language translation and localization solutions to the world's leading global enterprises. With its innovative approach to Machine Translation (MT), Safaba helps its clients enhance their global customers' experience and boost local and global brand value. This is accomplished by transforming the growing volumes of corporate content at the velocity dictated by global markets while maintaining message integrity and value.

Safaba delivers both technology and services that maximize the return on investment for automating translation. Safaba's approach to machine translation introduces a multi-step automated translation process, proven to yield highly tuned, enterprise-specific results. Going well beyond the boundaries of standard statistical MT, Safaba's solutions enable translation automation to become a key element of an effective corporate globalization strategy.

Representative Name: Dr. Alon Lavie, CEO

Representative Email: alon.lavie@safaba.com

Product Description:
Safaba MT v2.1 is an innovative approach for developing highly-optimized enterprise-specific Machine Translation (MT) solutions. It introduces a modular multi-phase machine translation process, proven to yield highly tuned, enterprise-specific results. It is structured as a series of technologies organized in concentric circles/layers. This approach is designed to address some of the key shortcomings of standard statistical MT such as unsupervised learning and the need for massive data manipulation required to achieve adequate translation results. It also allows for faster implementation and rapid incremental learning that reduces known errors and provides cost-effective migration paths, thus enabling faster commercial adoption of translation automation.

Languages Supported: English-to-All, All-to-English

Unique Features:
- Enterprise specific target language optimization modules
- Source and Target language transformation modules
- Provided as a full-service solution (SaaS)

Company Name: SAIC

Products Being Shown (Name and Version):
Omnifluent™ Translate
Omnifluent™ Media
Talk2Me®
Company URL: http://www.saic.com/linguistics/
  Company Information:

  Science Application International Corporation (SAIC) is an $11.1B company with approximately 43,000 employees, almost 450 offices nationwide, and global reach into almost 50 different countries. SAIC provides scientific, engineering, broad linguistic and IT services and solutions to federal, state, and local governments, commercial companies, utilities, and international customers.

  Our linguistic services span hundreds of professionals in translation, transcription, interpretation, publication, web content and language analysis. In 2010 SAIC acquired the assets of Apptek, a leading provider of Machine Translation (MT) and Automatic Speech Recognition (ASR) technologies. Since the acquisition, SAIC has continued to develop world-class MT and ASR technologies and has introduced the Omnifluent family of products.

Representative Name(s):
Representative Email:
Product Description:

SAIC's Omnifluent^{TM} Human Language Technology
The Omnifluent suite of human language technology products enables automated translation and transcription of multilingual audio and text, quickly rendering meaning across languages and cultures. By combining SAIC's patent-pending hybrid machine translation (HMT) with automatic speech recognition (ASR) technologies on a single unified platform, our language products provide the context, meaning, and understanding necessary for effective communication. Omnifluent products are available through software as a service, local installation, and subscription delivery models.

Omnifluent™ Translate
Omnifluent Translate is a comprehensive multilingual translation platform that automatically translates multiple formats of both text and audio content. This configurable translation solution can reduce the costs and time associated with your translation needs. The software provides users with several optional tools to enhance your translation experience: Tailoring, Audio File Translation, Name Entity Identification, Summarization, Web Site Translations, and Microsoft Office Capabilities to include Outlook emails.

Omnifluent™ Media
Omnifluent Media is a comprehensive multimedia platform that provides automated alignment, closed captioning, subtitling, and rich metadata development for media monitoring and analytics. By automating the labor-intensive task of aligned closed captioning and subtitling, content developers and broadcast media companies can broaden their reach into new audiences without the significant cost or effort of manual transcription or translation. Through our automated rich

metadata development, media companies can quickly and efficiently improve search indexing, analytics, and search engine optimization (SEO) strategies. This scalable environment can process broadcast media, videos, webcasts, social media, telephone conversations and other forms of multimedia content.

Talk2Me: A Practical Application of Omnifluent Technology

Talk2Me® is a multilingual application with capabilities to record an interview session and perform analysis on a specialized platform. Developed specifically for government use, Talk2Me's underlying technology uses speech to text transcription and search capability to speed up the location of relevant segments of the saved interviews. Search and analysis can be done within a session and across multiple sessions. Talk2Me can be configured for a variety of uses in a variety of professional settings, including law enforcement, diplomacy, health care, and legal. The application offers both flexibility and mobility by operating on any standalone laptops/tablets and through web services.

Languages Supported:

Hybrid Machine Translation (HMT)
- Arabic
- Arabic (Egyptian)
- Arabic (Iraqi)
- Arabic (Levantine)
- Chinese (Simplified)
- Chinese (Traditional)
- Dari
- Dutch
- Farsi/Persian
- French (Canadian)
- French (Continental)
- German
- Greek
- Hebrew
- Indonesian
- Italian
- Japanese
- Korean
- Pashto
- Portuguese
- Russian
- Spanish
- Tagalog
- Thai
- Turkish
- Ukrainian
- Urdu

Automatic Speech Recognition (ASR)
- Arabic
- Dari
- English
- Farsi
- French
- German
- Italian
- Pashto
- Persian
- Urdu

Unique Features:

Machine Interpretation
SAIC's core technology is comprised of hybrid machine translation (HMT) and automatic speech recognition (ASR) on a single platform, bringing together the power of these technologies to produce accurate translations to meet text and multimedia translation needs.

Single Decision Engine
SAIC's HMT combines features of both rules-based machine translation (RBMT) and statistical machine translation (SMT) for improved consistency, fluency, and accuracy, preserving the fidelity of meaning in multilingual translation communication.

Advances Since Last Year:

- Automated closed captions and subtitling workflow solution including optional post-editing
- Web portal for easy customer access
- Integration into TMS Collaborative Translation Platform
- Release of speech-to-speech mobile applications for health and travel domains
- Self-tailoring tool for easy customizations
- Confidence scoring for translation
- Extension of OmniFluent platform through APIs

Other Information:

Company Name: SDL

Products Being Shown (Name and Version): SDL BeGlobal

Company
> **Name**: SDL
> **URL**: http://www.sdl.com/automated-translation/
> **Information**: SDL's products and solutions enable companies to communicate with their customers across all channels of communication, anytime, in the language of the customer. SDL develops best of breed technology offerings for automated translation, workflow management, translator productivity, terminology management and software localization.

Representative
> **Name:** John Ferreira
> **Email**: jferreira@sdl.com
> **Telephone**: 703.724.1817

Product Description:

SDL BeGlobal – the industry's first cloud platform for real-time automated translation

SDL BeGlobal enables business users and content owners to manage automated translation in real-time through one central interface for multiple types of content, communication and social media. The application can also be used by translation teams to increase productivity for content that requires human translation or review, or to reduce post-editing costs through crowd-sourcing.

SDL BeGlobal's translation engines automatically translate multiple types of digital content in your enterprise - from support documentation & website content, to chat conversations & email. Our translation engines are trained in multiple domains and come with SDL TrustScore® - our patent pending algorithm that delivers a quality score with each translation.

Languages Supported: 80+ language pairs

Unique Features:

Trusted branded communication across all communication channels
- TouchPoints™ enable business users to define & control the requirements for each translation project – including the language, content type, quality, and terminology lists to be used
- SDL TrustScore™ delivers a quality score for each translation Brand & Terminology Management enables users to define how brands & key terminology should be translated
- Automatic quality improvement mechanisms continuously improve future translations
- Crowdsourcing widgets enable your communities to give feedback and suggest better translations

Easy to Use
- One central place to manage all automated translation
- Intuitive, consumer grade user interface that any business user can use

· Built- in reporting and analytics

Flexible access options
· Web application - Nothing to install, just point to a URL and go
· REST API to easily integrate with your current infrastructure
· Pre-integrated with other systems via the SDL GlobalConnect plugin
· Simple desktop and in-browser applications for secure translation by all employees in a business

Reliable & secure
· Multi-tenanted SaaS, combined with a role-based privilege model, can meet different security requirements
· Premium security deployment option available if required

Advances since Last Year:

Introduced in April 2012, SDL BeGlobal Translator is the latest innovation on this platform. This new capability is a simple translation option for anyone in an enterprise needing fast, ad hoc translation of a document, text or website. As a secure, web-based application, SDL BeGlobal Translator enables all employees to have access to translation, while staying in compliance with data security requirements and ensuring that sensitive company data is not being submitted to unsecure, free translation sites for translation. Organizations can upload their own brand and terminology lists to enhance translations and customize the URL and application to brand it for their particular company.

Other Information:

Company Name: SDL/ Language Weaver

Products Being Shown (Name and Version): SDL Language Weaver Enterprise Translation Server 5.3

Company
> **Name**: SDL/ Language Weaver
> **URL**: http://www.sdl.com/products/sdl-enterprise-language-server/
> **Information**: SDL's products and solutions enable our customers to communicate with their customers across all channels of communication, anytime, in the language of the customer. SDL Language Weaver's government solutions enable government organizations to cost effectively translate all content in near real-time. All information, regardless of the original language, can then be analyzed and escalated to maximize translation resources.

Representative
> **Name:** John Ferreira, Senior Director of Federal Sales
> **Email**: jferreira@sdl.com
> **Telephone**: 703.724.1817

Product Description:

SDL LW Enterprise Translation Server – on-site, statistical machine translation software for the government

SDL LW Enterprise Translation Server enables automated translation of foreign language information, resulting in timely access to mission critical information for decision making.

Designed for government entities that need secure access to translation solutions, SDL LW Enterprise Translation Server offers scalable, on-premise deployments with straight-forward integration into existing applications and workflows. This solution is engineered to enable organizations to meet the volume, speed and accuracy requirements for on-going translation demands. Engineered from the ground up, it offers easy integration with virtually any application, such as broadcast monitoring, web monitoring, document exploitation, live chat, and more.

Languages Supported: 80+ language pairs

Unique Features:
• Integration into SDL workflow and translator productivity software
• Client-Server architecture for on-premise, large scale, distributed deployment
• Flexible input & output formats
• User dictionary
• Web Services (SOAP & Rest) API
• Side-by-side display for web page translations
• Unicode internal
• Continuous improvement in translation speed and quality
• Rapid development of new languages
• Easy language pair training to specific subject domains using existing translated content

Advances Since Last Year:

SDL Language Weaver Enterprise Translation Server 5.3 includes significant quality improvements for existing language pairs and 3 new languages:

Slovenian <> English
Estonian <> English
Lithuanian <> English

Other Information:

Company Name: Spoken Translation, Inc.

Products Being Shown (Name and Version): Converser for Healthcare 3.1; SpeechTrans Mobile App; SpeechTrans Instant Messenger; SpeechTrans InterprePhone

Company URL: http://www.spokentranslation.com
Company Information: text
Representative Name(s): Mark Seligman, Ph.D.
Representative Email: mark.seligman@spokentranslation.com

Product Description:

Spoken Translation, Inc. will display its *Converser for Healthcare* system for speech-enabled automatic translation, featuring patented reliability and customization tools: users can verify and correct translation (using Reliable Retranslation™ and Meaning Cues™); enter text via speech, typing, or handwriting; use customized Translation Shortcuts™ for frequent phrases; and save bilingual transcripts of conversations. Complementary technologies from partner SpeechTrans, Inc. will also be shown: speech translation on mobile phones and iPad-class tablets; translated instant messaging with video; land-line phone translation; and phone connection to instant messaging, suitable for the hearing-impaired.

Languages Supported: Converser for Healthcare: English<>Spanish. SpeechTrans products: dozens of languages in many combinations

Unique Features:

Converser for Healthcare offers patented reliability and customization tools: users can verify and correct translation (using Reliable Retranslation™ and Meaning Cues™); enter text via speech, typing, or handwriting; use customized Translation Shortcuts™ for frequent phrases; and save bilingual transcripts of conversations. SpeechTrans products offer multi-lingual speech translation on multiple platforms, for both in-person and remote use (including video). Photo translation is also enabled.

Advances Since Last Year: Converser for Healthcare: Numerous interface improvements in response to a successful pilot project at a major US healthcare organization. SpeechTrans products: introduction to worldwide markets.

Other Information:

Company Name: SYSTRAN Software, Inc.

Products Being Shown (Name and Version):

SYSTRAN Enterprise Server 7, SYSTRAN Training Server 7

Company URL: www.systransoft.com

Company Information:

SYSTRAN's market-leading machine translation solutions empower businesses and customers to understand multilingual information in real-time and communicate more effectively in 60+ language combinations. Organizations of all types and sizes integrate SYSTRAN solutions into any enterprise business segment, such as customer support, collaboration, content management, business intelligence, product localization, knowledge management, eCommerce, eDiscovery, and others to deliver quick relevant information in multiple languages, enhance the enterprise ecosystem and improve customer satisfaction and retention. Use of SYSTRAN's solutions maximize productivity, ensure time constraints are met and reduce translation costs.

Representative Name(s): Denis Gachot

Representative Email: gachot@systransoft.com

Product Description:

SYSTRAN Hybrid Machine Translation Solutions for the Enterprise

SYSTRAN Training Server allows corporate customers and Language Services Providers to independently train SYSTRAN Enterprise Server 7 to any selected domain or business objective to produce publishable-quality translations. Through use of Training Server, Enterprise Server 7 automatically learns from customers' existing and validated translations. Additionally, the translation engines are continuously updated as new translations are added to the system and quality improves and adapts to corporate needs. The training tasks include resource extraction, dictionary validation and document alignment.

Languages Supported:

52 hybrid language pairs	English ⇔ Arabic English ⇔ Chinese English ⇔ Dutch English ⇔ French English ⇔ German English ⇔ Greek English ⇔ Italian English ⇔ Japanese English ⇔ Korean	English ⇔ Polish English ⇔ Portuguese English ⇔ Russian English ⇔ Spanish English ⇔ Swedish French ⇔ Dutch French ⇔ German French ⇔ Greek French ⇔ Italian	French ⇔ Portuguese French ⇔ Spanish German ⇔ Italian German ⇔ Portuguese German ⇔ Spanish Italian ⇔ Portuguese Spanish ⇔ Italian Spanish ⇔ Portuguese
38 additional language pairs	Albanian → English Arabic → French Bengali → English Bulgarian → English Chinese → French Czech → English Chinese ⇔ Korean Dari → English Danish → English English → Czech English → Hungarian	Estonian → English Farsi → English Finnish → English Hindi → English Hungarian → English Hungarian → French Japanese ⇔ Chinese Japanese ⇔ Korean Latvian → English Lithuanian → Czech Norwegian → English	Pashto → English Polish → French Romanian → English Serbo-Croatian → English Slovak → English Slovenian → English Somali → English Tajik → English Turkish ⇔ English Ukrainian → English Urdu → English

Unique Features:

SYSTRAN's Hybrid Machine Translation Solutions for the Enterprise combine the predictability and language consistency of rule-based MT with the fluency and flexibility of statistical MT. This means that enterprise data may be deeply customized and adapted to any subject in context, unlike other machine translation solutions which are generic all-purpose systems and lack the linguistic rules and millions of dictionary entries which have been and continue to be integrated into SYSTRAN's engines. Additionally, SYSTRAN's hybrid engine sharply reduces the amount of data required to train the software and also reduces the size of the statistical models while maintaining high performance.

SYSTRAN solutions have been deployed in both the US Government and global enterprise marketplaces for 40+ years. Our original MT engines were developed by us and continually evolve with innovative advancements in natural language processing and related language technologies.

Company Name: Ntrepid Corporation

Products Being Shown (Name and Version): Virtus

Company
   Name: Virtus Translator
   URL: www.virtustranslator.com
   Information: 877.681.1913

Representative
   Name: Jen Doyon or Wes Babcock
   Email: jen.doyon@ntrepidcorp.com,
   wesley.babcock@ntrepidcorp.com
   Telephone: 877.681.1913
   E-mail: info@virtustranslator.com

Product Description: Virtus solves the challenge of translating foreign language data, particularly named entities, in databases and spreadsheets. Column-by-column, Virtus recognizes the foreign language and data category (such as person names or street addresses) and executes an appropriate translation and/or transliteration strategy. Virtus operates in stand-alone mode or integrated with other NLP technologies.

Languages Supported: Arabic, Chinese, French, German, Persian, Russian, Spanish.

Other Information: Watch the Virtus video and sign up to receive a free trial here: www.virtustranslator.com.

Yandex Translate

Yandex Translate (translate.yandex.com) is a new online translation service targeted at all user groups. It is based on statistical machine translation technology, and its models contain words and phrases mined from billions of web-pages. Yandex Translate provides translation for words, texts and web-pages from English, German, French, Spanish, Italian, Polish, Turkish and Ukrainian languages into Russian and backwards. It runs on the Yandex web site and promptly translates millions of web-pages and text fragments every day. Service API enables developers to integrate the translation functionality into any site or software product.

The service was launched in 2011, and over the recent year our team has deployed a number of technical and interface solutions, which we would like to showcase at the AMTA-2012 Conference.

1. Typing assistance

TypingAssistant tries to guess the word you are entering by the first few letters and gives you the most probable suggestion. The assistant not only improves the speed of typing, but also helps you avoid making mistakes, which very often lead to translation quality degradation.

2. Synchronous translation

To improve the efficiency of translation process text is automatically translated as words are being entered in the input field. In contrast to systems, executing letter-by-letter translation, our system identifies probable word endings and produces more fluent and adequate results.

3. Statistical machine dictionary

People use translation service not only to translate sentences, but also to lookup translation of some word or phrase. Unlike translation of a coherent text, where we have to find one but the best translation for each lexical unit, separate word or phrase translation must provide all possible variants. To fulfill this task we've developed a statistical machine dictionary based on the same parallel corpora as our SMT system uses. In order to improve user experience and to make machine dictionary entries more common for end users we group translations by their meanings, provide explanations and show usage examples.

4. Multiple layouts for webpage translation

Reading translated web-pages often requires referring to the original text. To help people compare translated web-page to its original we provide two layouts: side-by-side (horizontal or vertical) and overlapped. Both pages (original and translation) are loaded into browser memory at once, so switching between view modes is instantaneous.

<u>Notes</u>

Notes

Notes