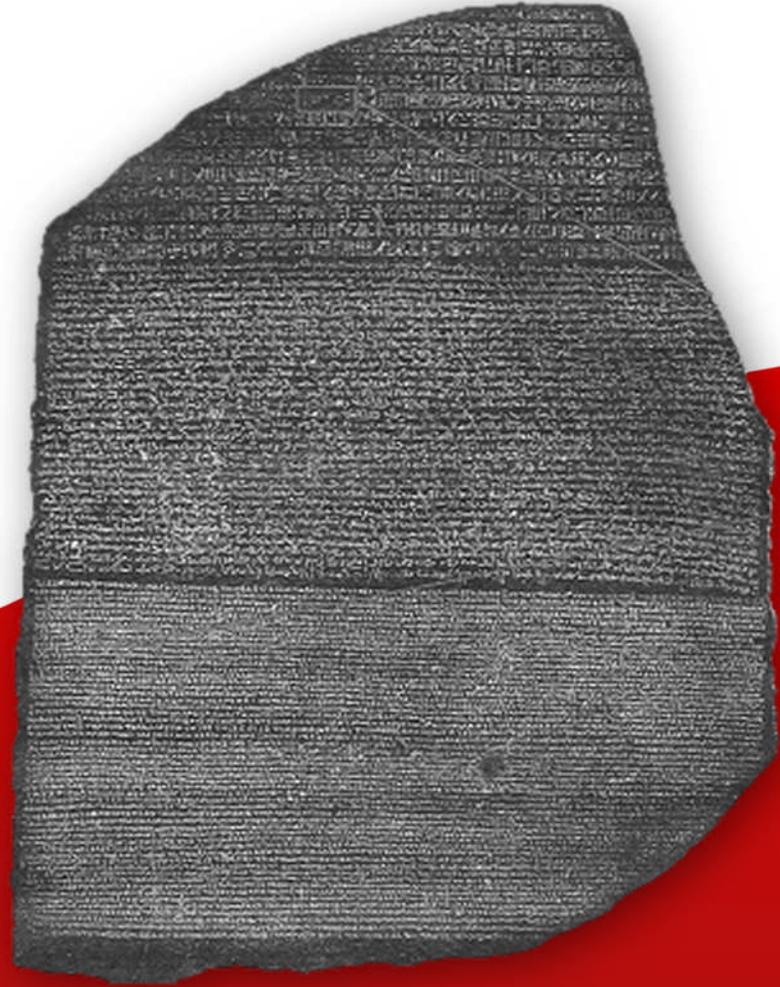


# Language Tech News

Issue 2, May 2007



[www.languagetechdivision.org](http://www.languagetechdivision.org)

**ATA • LTD**  
American Translators Association      Language Technology Division

## Table of Contents

Name this Newsletter! .....	1
Note from the Editor, by Lisa Carter .....	2
By the time..., by Michael Metzger .....	2
Freeware vs. Free Software, by Corinne McKay .....	3
The Mailing List, by Tony Roder .....	3
Finally! A Good, Affordable Terminology Management Program, by Lee Wright .....	4
A Linguistic Atlas, by Andrew Bertallot .....	5
Panelists Wanted for TCD Conference .....	9
LTD Joins TranslationSummit!, by Michael Metzger .....	9
Nominating Committee.....	10
Web-Based Translation, Pootle: Part 1 – Introduction and Installation, by Dierk Seeburg .....	10
Mac users Unite!, by Michael Metzger .....	12
Trados Tip, by Tuomas Kostiainen .....	13
Intro Seminar on SDL Trados 2006, by Emily Tell and Deborah Lockhart .....	13
LTD & TAC Contacts .....	15

### Language Tech News

Issue 2 / May 2007

#### Editor

Lisa Carter

[lisa@intralingo.com](mailto:lisa@intralingo.com)

#### Proofreaders

Barbara Guggemos, Michael Metzger,

Jill R. Sommer

#### Contributors to this Issue

Andrew Bertallot, Tuomas Kostiainen, Deborah Lockhart, Michael Metzger, Corinne McKay,  
Tony Roder, Dierk Seeburg, Emily Tell, Lee Wright

#### Layout

Jon Lee and Lisa Carter

#### Submissions

Send your submissions to the editor, including a short bio. All submissions are subject to editing.

#### Name this Newsletter!



There's a little something lacking in the title of our newsletter... i.e., an actual name! And a logo! We're thus holding a contest among members to come up with a catchy, appropriate name. So, submit your ideas for a

name, a logo or both to the editor, Lisa Carter ([lisa@intralingo.com](mailto:lisa@intralingo.com)), by June 30, 2007. We will publish submissions and hold a vote. Subsequent issues will then be published with the winning name and logo!

First and foremost, I would like to thank everyone who contributed in any way to this, our second newsletter. It takes time to contribute, but rest assured it is most appreciated by everyone who will read and benefit from what you have shared.

The next issues of the newsletter will be published in August and November of this year. Therefore, anyone wishing to submit anything for Issue 3 must do so by July 15, 2007, and by October 15, 2007 for Issue 4.

There are no specific guidelines for submissions; simply send your article to me as a DOC or RTF attachment. A submission may be as long or as short as you wish. (Well, if it is very, very long I may approach you about dividing it into parts to be published in consecutive issues.) Please also be sure to include a short bio. Ideally, this would also contain your e-mail or website contact.

Speaking of "websites," I had quite a time deciding just how to standardize this and other related words for the newsletter; i.e., is it Web site, web site or website? I opened up a discussion on our mailing list (see Tony's note later in this newsletter about how to join, if

you haven't already). Based on feedback, my decision in the end was to write compound nouns in lowercase and as one word. As for "the Web," it will remain capitalized. I'm sure not everyone will be happy with the outcome, but hopefully "majority rules"! If you're submitting and unsure, please use the American Heritage dictionary as a guide, since that's the standard I'll follow.

As a final note, I've attempted to classify each article into an appropriate "level of difficulty." That way, if you're new to language tech and pressed for time, you might want to read only certain introductory articles. On the other hand, if you've been using language tech for decades, then you can skip straight to the advanced ones.

We do hope you'll read all of the articles, though, as there's always something new to learn.

Enjoy!

Lisa Carter is currently editor of the LTD's newsletter. She translates legal, commercial and literary texts from Spanish to English. Check out her blog at [www.intralingo.com](http://www.intralingo.com)

you read this the contributors, the Editor and other "Angel Investors" of the Language Technology Division have worked hard again to make this issue possible. And what an issue it is! Where our first newsletter was a platform to broadcast division news, to call for nominations and papers, and to ask for feedback, this second issue is now on track to transition to a content-rich newsletter, filled with submissions from members like you. I am thrilled to see this happen and hope you will

enjoy this issue as well. But before you skip to the next interesting article, please allow me to thank all the contributors for sharing their time, insight, and expertise. I hope you find this inspiring and invite you to join us in shaping the future direction of both the newsletter, and the content we discuss, research and review in all division channels.

Michael Metzger is Administrator of the Language Technology Division.

Reproduced with permission from *Open Source Update*, Issue #5, March 2005 (HTML version: <http://www.translatewrite.com/osupdate5.html>)

Since *Open Source Update* debuted in December 2004, a number of readers have suggested articles on various pieces of "freeware" that are useful to translators. Most of these are add-ons to programs such as Microsoft Word, and are distributed for free, but their code is not available to the user. While no one would disagree that free is a great price for software, it's important to keep in mind the first line on the Free Software Foundation's (<http://www.fsf.org>) website, "Free Software is a matter of liberty, not price. You should think of 'Free' as in 'Free Speech'."

The Free Software Foundation defines four types of freedom that characterize free software: 0) The freedom to run the program for any purpose. 1) The freedom to study how the program works, and adapt it to your needs. Access to the source code is a precondition for this. 2) The freedom to redistribute copies so you can help your neighbor. 3) The freedom to improve the program, and release your improvements to the public, so that the whole community benefits.

So, while software that turns your cursor into a dancing peanut might be free, and cute, unless you can look at the code, change it, and give it to someone else, it's not Free Software. Recently, I interviewed the CEO of an open source audio equipment company for an article on embedded Linux in consumer products. When I asked him to name some differences between his company and others that sell embedded Linux products, he answered (this is a paraphrase), "We run Linux on our products, and we put the code out to the community so that everyone can see it, use it, and adapt it. If you're running Linux on your products but you're not releasing your code, I call that leeching off the community, not contributing to it." While this point of view might be outside the mainstream, it's a great example of a business that has truly adopted the open source paradigm.

Corinne McKay is a freelance French to English translator in Boulder, Colorado. She has presented a session on Free and Open Source Software for Translators at the past three ATA annual conferences, and edits the e-newsletter *Open Source Update*. Contact: [corinne@translatewrite.com](mailto:corinne@translatewrite.com)

The last time I looked, LTD had 979 members. Having attracted 10 percent of the Association's membership in less than one year, the Division can well be considered a success story. So why is it that our mailing list has only 103 subscribers?

On the Division's website (<http://www.ata-divisions.org/LTD/>) next to the "LTD Mailinglist" button, did we not invite you to subscribe to get the latest news, participate in discussions, ask for advice, offer support or simply eavesdrop? Did we not provide a direct e-mail link ([ataLTD-subscribe@yahoogroups.com](mailto:ataLTD-subscribe@yahoogroups.com)) for a subscription request? Did we not ask you to include your name in the e-mail so as to expedite your subscription? What else should we do to bring you into the mainstream of our profession's accelerating technological development?

It is this rapidly expanding environment that has led us to create the LTD mailing list. Because its subscribers are exclusively members of the ATA, comments about experience with translation technology products will serve as a direct source of information for the vendor-panel sessions held by the Translation and Computers Committee during the Association's conferences. The mailing list thus offers an unparalleled opportunity to help shape the tools we will be using in the future.

As your ever-loving listmaster, I eagerly look forward to a tidal wave of subscription requests before this fall's conference in San Francisco.

Tony Roder is in charge of the LTD's mailing list. Contact him at [tony@wells.com](mailto:tony@wells.com).

## A Good, Affordable Terminology Management Program

Last fall I wrote a little article for the Northeast Ohio Translators Association newsletter in which I described having purchased Jost Zetzsche's *A Translator's Tool Box for the 21st Century* and found that it contains a lot of very useful information on computers and computer software. As a bonus for that purchase, I also received a year's free subscription to Jost's biweekly newsletter, the *Tool Kit*. The 82nd edition appeared in my mailbox at the end of February, so I eagerly scanned its contents for new information. Among other noteworthy announcements, Jost gave a brief description of a new (at least, to me) terminology management system, called Logo 4.0, which he touted as being a "low cost, robust and easy-to-use dictionary creation software for translators, localizers, and businesses." Since a free 30-day trial version of the program was available for download from the publisher's website (Lexicool – [www.lexicool.com/soft\\_lingo2.asp](http://www.lexicool.com/soft_lingo2.asp)), I decided to see whether it lived up to the description. I proceeded to download the software and install it on my hard drive. This took all of about two minutes, and the program was ready to use, including an icon on both my desktop and in the Windows system tray at the bottom of the screen.

The installed program occupies just a very small amount of disk space (a little over 20 megabytes, which is miniscule compared to most software on the market these days). Nor does it appear to be "memory hungry" or cause any interference with other software running at the same time. It can be used together with any Windows program. Once installed and loaded, the program remains available until you decide to exit it altogether or until you turn off your computer, just like any other software.

Although there is no "user manual" per se, the program includes a good help file that describes and illustrates all of the basic functions. The help file contents can be printed out if desired. However, even before consulting this help file, I quickly discovered that the program's operation is both straightforward and intuitive. I easily set up a new dictionary for a test run and added a

sizeable number of term pairs to it within just a short time. Of course, all of the source-language terms are automatically alphabetized as they should be. There is only one requirement for setting up a dictionary: you must specify the language(s) to be included in it (one or more). In addition, you can provide a language code for each language, typically using the appropriate ISO two-character codes (e.g., en for English, es for Spanish, de for German, etc.), but these are not obligatory unless you plan to export the data in standard TMX (XML) format. The program includes Unicode compatibility and can handle all character sets, including non-Roman alphabets and right-to-left languages like Arabic.

Although I only entered terms in Spanish and English, the program allows for any number of language combinations, even multilingual glossaries. In addition to space for the source-language term and the target-language equivalent(s), there are also two windows for entering other data, such as definitions, notes, and subject areas for the term pair. Each of these windows will hold a reasonable amount of data – a maximum of three lines (approximately 225 characters, including spaces).

I also experimented with several of the other features, such as the ability to export a dictionary (terminology database) to any one of three different formats (XML, TXT or RTF), which can then be read into a text editor/word processor and printed. The diminutive (1" x 1¾") pop-up interface ("tool bar"), which, unless minimized, stays visible on top of whatever other application happens to be running, contains just six icons, activated either by left clicking on them or by pressing a function key. This small tool bar can also be displayed in a slightly larger size (1" x 2½") if desired. In either format, the tool bar can be easily moved around to any location on your computer screen, just like any other window.

Once a dictionary has been set up and data have been entered into it, the next essential feature is the ability to find the entry containing a given SL term and its TL equivalent(s) and then copy and paste the

equivalent into the text being prepared in the word processor. There are actually two ways to do a term lookup. One of them displays a window that lists the currently available dictionaries and a space for entering the source term to be found, while the other one is a "quick lookup" window that appears as a single line at the bottom of the tool bar ("quick search"). In this latter instance, if the SL term that you look up is found in the current dictionary, then you can display the entire entry (or entries if the term happens to appear in more than one record). In short, from my standpoint this is a very slick, timesaving feature.

Although it is unlikely that I would ever have a need to use it, the program also allows for encrypting the data contained in a dictionary and for requiring a password in order to access a data file.

In addition, Lingo 4.0 includes a means for importing data into an existing dictionary, but the data to be imported needs to be in either

TMX format or a straight text file. At this time I have not experimented with this feature, but I have several small term files created as MS Word tables that I am going to convert to straight text and see how Lingo's import function works.

Finally – and I left what I think is the best part for last – the program only costs €89 (approximately \$115 at the time of this writing)! In short, I highly recommend this excellent, affordable and well-designed software for any translator who wants to have quick and easy access to specialized terminology.

Dr. Leland D. Wright, Jr. taught Spanish at the university level for a total of ten years and worked as a full-time translator/language consultant for an international engineering and construction firm for eight years. A freelance translator since 1982, focusing primarily on Spanish legal and commercial documents, he also taught Spanish-English commercial/legal translation in the M.A. translator-training program at Kent State University from 1989 to 2004. He is currently an online instructor in Spanish-English translation for New York University.

Last summer I attended a conference on Franco-Provençal, a regional language spoken in Romandy (Switzerland), Val d'Aoste (North-Western Italy), Savoy, and other neighboring regions in France. The number of speakers of the Franco-Provençal dialects is unfortunately declining, and the linguistic departments of several local universities and some local institutions are doing notable work to document, study, and record this regional language for audio and video archives. As a translator, I was intrigued by the linguistics and, as a software engineer, by the possibilities the Web offers to access audio and video archives of speech. Of course, these possibilities exist for any language or dialect, not merely the Franco-Provençal that I regrettably failed to learn as a child. In fact, I did find examples of web applications for several different minority languages. However, since I found none that combined all of the features required by a real online atlas, even though the web technologies are currently available, I designed one myself.

## Linguistic Atlas

One important characteristic of dialects and of many minority regional languages in general is that their pronunciation and spelling (if they are written at all) are usually not standardized. Consequently, documenting the language requires studying the variations in terms across geographical areas, possibly inventing some phonetic way of spelling words, and recording actual speech in some manner in order to document usage, corresponding grammar, and syntax. A basic way to record this kind of information is to compile a linguistic atlas.

A linguistic atlas, in its simplest form, is a series of identical maps of the geographical area where a language is being investigated. One could think of a dictionary in which a map is printed for each word entry. On the map itself, the various versions of the word as used in different places are printed in the specific locations. It is particularly important to properly record in printed form the differences in pronunciation that are fundamental to this kind of research. Linguists usually invent some appropriate way of spelling the language, which may require special characters

specifically devised for the language. The characters used might be a mixture of Latin characters, IPA (International Phonetic Alphabet) or others. Linguistic atlases have been used in the study of languages and dialects, especially in Europe, for about the last two centuries. Atlases are usually limited to a specific subset of words, such as names of trees, tools, personal pronouns, greeting phrases, and the like.

## Design Issues

### Spelling

Even in this simple form, designing a linguistic atlas template for a webpage is not straightforward. The first issue that presents itself is that a standard HTML webpage cannot contain character fonts. The code in the page can only list, in order of preference, a number of fonts that the browser can use to display the page if available on the client computer. Even if IPA is used, the browser must be Unicode-compliant — version 4 or later of Netscape or Internet Explorer — and a Unicode font with IPA symbols must be installed, such as Arial Unicode MS or Lucida Sans Unicode. If a specifically designed character font is required (which is often the case), it certainly will not be available on the client computer.

There are in fact two ways to embed fonts in a webpage, but they are browser-dependent. One is Embedded OpenType for Internet Explorer, and the other is TrueDoc for Firefox, which will require installation of an ActiveX control to also work with Internet Explorer. The application that does the embedding will only work if the font designer who holds the copyright has enabled the font for embedding.

### Audio

Computers being multimedia devices, it makes sense to associate each displayed word with a corresponding audio recording. Ideally, this recording should be played by clicking on the word image. This is a straightforward idea, but what actually happens when a browser detects a clicked link to an audio file depends on the browser, the operating system, the specific way they are configured, and so on. What is most likely to happen is that a media player will invade the whole screen or part of it, hiding the atlas map and hindering the

selection of other words. Or, an appropriate media player might not be registered in the browser or available on the computer, and nothing might happen. Even if the media player behaves unobtrusively, it will take some time to download the whole file, and the audio will be played only when the download is complete. Clearly, this isn't good enough because what actually happens is uncertain and is not interactive.

### Video

As linguistics researchers often take video recordings of speakers in order to have visual and speech context for a specific word or expression, the audio recording might in fact include one or more phrases and could actually be the soundtrack of a short video clip. The issues just highlighted for playing a single word — specifically, the uncertainties regarding media player operation and the longer wait times required for loading multimedia files before they are actually played — are even more of a problem when a video clip has to be played.

### User Interface

At first glance, the technologies used for the animations that are so common in websites might appear adequate for implementing the user interface of a linguistic atlas as a web application. However, it should not be forgotten that a real atlas potentially requires access to a database of words, audio files, and video files, and this is not the case for standard animations on the Web.

Finally, the linguistic atlas has to be interactive. That is, the page should not reload each time the map for a new word is displayed. The words should change immediately and the map should stay the same.

## Implementation Requirements

All of the issues that have been described so far can be solved by means of various technologies. However, as it is unlikely that the users of a linguistic atlas will be computer geeks, the technologies used should be available on most client computers and should require little or no installation of new software and/or configuration of existing applications.

## Implementation

The first problem, displaying outlandish character fonts, can be easily solved, albeit with time-consuming and boring work: typing each word using the appropriate font and transforming it into an image file that will be displayed correctly by any browser.

To solve the remaining issues, suitable technologies should be selected to begin with. The requirement of using software that is already available on most computer rules out Java, which would otherwise be well suited for the purpose. (Java, no relation to JavaScript, is a powerful programming language executed by the browser. However, the Java virtual machine has to be downloaded and installed first.)

Playing audio and video without a potentially obtrusive media player appearing requires a browser plug-in. The most common browser plug-in available for this purpose is Flash, which can be downloaded for free from Macromedia. Using Flash also solves the problem of having to wait for the multimedia file to be loaded before actually playing it, as it is played while being progressively downloaded and the compression rate decreases the original audio/video file size. The original audio and video files have to be transformed into Flash files: again, this is easy but time-consuming on a large number of words, since each word entry requires one audio file and, if available, one video file.

Next comes the requirement of displaying and playing the words for other atlas entries interactively, without reloading the page. This requires accessing the database on the server asynchronously upon user request. It is possible to implement this with Flash, but I prefer to use the basic technologies directly, without an extra design interface layer. Ajax (Asynchronous JavaScript and XML) interactively accesses the database from the browser and a PHP script on the server reads the data from a MySQL database also on the server, and this data is sent to the browser. No proprietary server or database software is required, as PHP and MySQL are freely available on open-source Apache servers.

The database is basically an index that associates each atlas entry with the various

local versions of the current word, the names of the image files that display the word versions, the names of the audio files, and, if available, the video clips in which the local version of the word is uttered in context.

Using a database is fundamental to the design of a real linguistic atlas, because it makes the number of possible entries in the atlas almost limitless, and the speed of operating the atlas page is totally unaffected by the number of words the atlas can access as a web application.

Embedding all of the words in the webpage, using Flash or Java, is only feasible for a limited number of words, since all of the multimedia files are loaded with the page. Even with broadband, this is impractical for anything but the simplest demo atlas containing a few sample words. Most of the atlases I have been able to find on the Web are samples of this kind.

By using a database, I was able to design general purpose JavaScript code to implement the core operation of the atlas in the browser. Thus, any additions to the database can be performed by adding records into the database itself and loading the corresponding multimedia files on the server; no modifications to the atlas webpage are required.

## Atlas Demo

The atlas web application I designed is fully operational and applicable to a database containing thousands of words. However, since I am not a linguist, the demo's actual linguistic contents are fictitious, selected only to highlight the features offered by this atlas.

The linguistic area of the atlas is a map that includes six specific villages (Arvier, Courmayeur, Fénis, Gressoney, Pont-St-Martin, and Valtournenche). For the sake of illustration, it has been assumed that the local versions of a word selected from a drop-down list are Italian, English, Franco-Provençal, German, Arabic and French, respectively. This word is the entry in the linguistic atlas, and every time a new word is selected from the list, a new "page" of the atlas is presented, and the local versions of the word in each of the six villages are updated. For the sake of

interactivity, the page does not actually reload; only the words on the page change.

To illustrate how different fonts have been transformed into images, the words are spelled using the IPA for Italian, German and French, a custom font for Franco-Provençal, Merriam-Webster pronunciation for English, and standard Arabic script for Arabic. By clicking on any village, the local version of the word can be heard and in some cases a video clip is also available, including its complete script in IPA. The operation is quite fast at 128 kbps with an ISDN line, and does not necessarily require broadband.

### Multimedia Linguistic Atlas on CD/DVD

Finally, it should be noted that if a linguistic atlas were to be implemented on CD/DVD, several of the design choices that have been ruled out in this context would in fact be quite appropriate, since having direct access to multimedia files on the same computer eliminates the issues of download times. The application software could install the required character fonts on the specific computer and could play multimedia files directly in any suitable fashion, and/or use a web browser adding any required plug-in, Java, etc. Some linguistic atlases currently being developed are designed for distribution on DVD.

### References

<http://www.absoftwaredesign.com/ktatlas.php>  
My own atlas (in French). Requires a PC with a Windows OS, Internet Explorer or Firefox with Flash v. 8.

<http://www.unine.ch/dialectologie/atlas001/CadreAppelle.html>

Sample page of the Atlas Linguistique du Valais Romand by the Université de Neuchâtel (implemented on DVD). Requires QuickTime.

<http://w3.u-grenoble3.fr/ellug/livres/atlas/>  
Atlas Linguistique Parlant d'une Région Alpine by Université Stendhal of Grenoble, features some 700 words. Requires Java, QuickTime.

<http://www2.hu-berlin.de/Vivaldi/index.php?id=m4001&lang=de>  
Page for a number of Franco-Provençal phrases in Val d'Aoste by the Akustischer Sprachatlas der Dialekte und Minderheitensprachen Italiens by Humboldt-Universität Berlin. Requires Java.

<http://eastcree.org/atlas/>

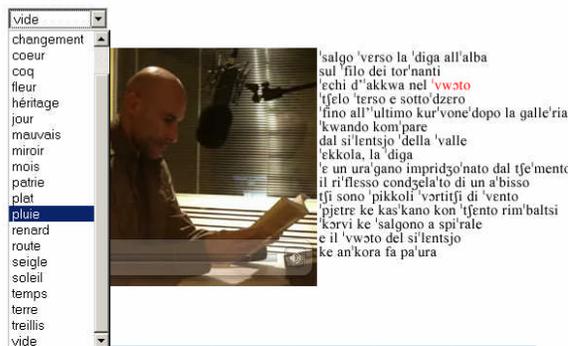
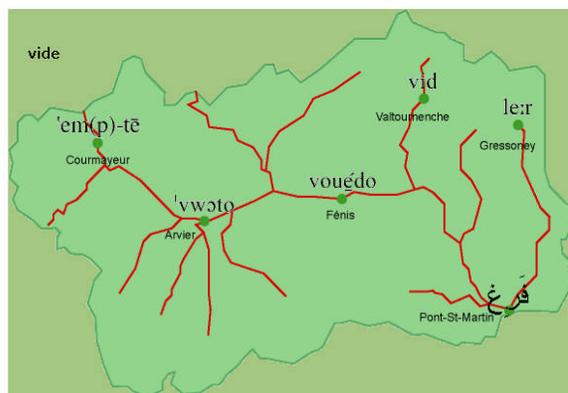
A recently updated atlas of common phrases in the Cree-Innu language by Cree Programs and Carleton University, with an impressive Google map of Canada. Implemented using Ajax.

<http://www.alepo.it>

Atlante Linguistico ed Etnografico del Piemonte Occidentale by Regione Piemonte and Università di Torino, presents a linguistic atlas of Western Piedmont in print featuring words for trees, shrubs, etc. Follow the links "Materiali", then "Ascolta gli etnotesti" for a map of Piedmont with sample multimedia clips. Implemented using Flash.

Andrew Bertalot is an ATA-certified (Italian>English, English>Italian) freelance technical translator and localization engineer.

<http://www.absoftwaredesign.com/translations.html>



Kim Vitray, administrator of the Translation Company Division (TCD), has been in touch about the exciting possibility for the LTD to hold a panel at their upcoming conference, July 26-29, 2007, in San Antonio, Texas. (See <http://www.ata-divisions.org/TCD> for more details.) The panel would consist of LTD members working in translation companies

who would like to discuss workflow and project management tools. If you would like to learn more about this, please contact Michael Metzger ([mm9@sbcglobal.net](mailto:mm9@sbcglobal.net)).

If you've always wanted to learn it all in one day, the 2nd TranslationSummit held on Monday, March 12 in Salt Lake City was surely an event not to be missed. Session after session, panel after panel, and speaker after speaker led the very interested Summit attendees through all aspects and questions one might have as a translation buyer. One could say this was "Getting it Right" on steroids – a progression from ATA's print version to a real-time event with actual experts from government, academia, and the private sector telling their peers from just those areas what to look out for, what to ask for, and what to work on. Visiting the event's website at <http://www.translationsummit.org> and looking at the posted agenda will give you an idea. Topics like Research & Training, Quality First, a T&I live demonstration, Content Authoring for Translation, Standards, and eventually Technology (the very reason LTD was present) had everybody's attention and left little time to stray from the beat in the exhibition hall.

Attendees were given solid figures, statistics, and perspectives on all things language, translation, and interpretation. And all that eventually came together in the announcement of an upcoming Language Technology Tools survey by Jost Zetzsche, currently labeled "Better Tools for Translators" and under editorial review. As members of the LTD know from our first business meeting in

New Orleans, this survey is very much at the heart of LTD's interest. We continue to play a vital role in shaping the survey by providing input, adding resources and industry know-how, helping drive it to completion. There will be an official rollout – stay tuned for more news on this in the LTD newsletter, on our website, and very likely in postings on the mailing list. As far as we can tell already, this survey is quite different from the one organized and published by Elina Lagoudaki earlier this year ("Translation Memory systems: Enlightening users' perspective"). Whereas the latter was an interesting survey on the Translation Memory segment, "Better Tools for Translators" moves in an entirely new direction by asking what tools translators actually use and how, strongly reflecting the work and ideas of Dr. Alan Melby ("Eight Types of Translation Technology"). The survey as it currently stands is very thorough and will provide detailed information, expected and unexpected, yet to be seen.

After participating in a business meeting (the second and most important reason for LTD to be at the TranslationSummit), concentrating on the overall plan for the survey, and streamlining questions and objectives, it was time to return home. What a great investment it was for the upcoming survey!

Michael Metzger is Administrator of the Language Technology Division.

The nominating committee (consisting of Bob Killingsworth and Jost Zetzsche) is looking for candidates for the following administrative positions, to be elected at the next ATA conference in San Francisco October 31 - November 3, 2007:

- Administrator (tasks include communications link between the Division, Board, and Headquarters; oversight of Division activities; communication with members; appointment of committees; oversight of elections, meetings, and social events; preparation of annual budget request; submission of requests for reimbursements; and organization of other special activities, including the maintenance of the division website and e-mail lists).

- Assistant Administrator (tasks include assisting the Administrator in all of his or her duties and substituting the Administrator in his/her absence).
- Treasurer (tasks include keeping all Division financial records and preparing financial reports).
- Secretary (tasks include preparing written records of Division meetings or other official Division business and assisting the Administrator in preparing the annual report).

If you are interested yourself or would like to propose a candidate, please e-mail Bob (robertk117@comcast.net) or Jost (jzetzsche@internationalwriters.com).

Pootle: Part 1 – Introduction and Installation

These days, the Web is all-pervasive and you can be sure someone somewhere is trying to take advantage of anything that can be done on it. It was only a matter of time before a very computer-centric activity like translation migrated to the Web as well.

### History of Web-Based Translation

Web-based translations have their origins in the Open Source gettext utilities of the Free Software Foundation's GNU project and date back as far as 1995. A group of dedicated individuals formed the Translation Project and provided a framework for internationalization and translation in order to make Native Language Support (NLS) for GNU software possible.

The GNU gettext toolset helps programmers and translators produce, update, and use translation files, mainly those files that are textual, editable files. The vehicle used by the gettext utilities and for making translation of text resources possible is the .po file. The letters PO in '.po' file stand for Portable Object. This paradigm, as well as the PO file format, was inspired by the NLS standard developed by Uniforum and first implemented by Sun in their Solaris operating system.

A step further in the evolution of these projects is the WordForge project, whose aim is to build standards-based localization tools for Free and Open Source Software (FOSS). The package enabling web-based translation is a toolset called Pootle. The acronym Pootle is derived from the term PO-based Online Translation / Localization Engine.

### What is Pootle?

According to its README file, Pootle is a web translation and translation management engine. What does that mean? Well, two things, really.

The first is web translation, or more precisely, web-based translation. In other words, translation that uses the Web to facilitate the translation activity. Using Pootle, translators can translate a document from within their favorite web browser. All they need to translate anywhere in the world is a computer with an Internet connection and a web browser – an offline component is currently in development. Like any modern Translation Environment Tool (TET), Pootle can use suggestions from a translation memory. In suggestion mode, contributions can be checked by translators with sign-off authority before being accepted into the final copy,

which is ideal for bug reporting. Pootle also performs terminology matching from glossaries in the terminology project. Pootle uses Unicode throughout, so it works with any target language, including right-to-left display for those languages that need it. Clickable characters help users who cannot type all the characters of their languages. Furthermore, Pootle allows for highlighting errors in translation and translator comments. The latter is especially useful in a translation team environment and as part of an online review process.

The second part, translation management, refers to the web-based management of translations. Unlike most other major TETs, Pootle does not rely on a backend database, but rather on Lucene, a fast text indexing engine that enables Pootle's terminology matching and translation memory recall. Updates to previously translated documents are facilitated through integration into a version control system such as CVS or its successor, Subversion. By integrating Pootle with the Translate Toolkit, translation managers can access statistics with string and word counts. The Toolkit also allows for translation of files in XLIFF format in addition to the aforementioned PO files, making it compatible with many other tools. Furthermore, Pootle can handle many different file formats by converting to PO and back again, including OpenOffice, Mozilla, CSV, Qt, plain text, HTML, XLIFF, Java properties, TMX, and TBX. Translation managers can set goals for translations teams and assign work to various translators with permissions for different functions.

### **How does Pootle work?**

On a low level, the concept works as follows: a PO file is a document that can be used directly by the translation engine, rather than by being stored in a database. In order to allow for quick access to statistics, they are cached in .stats files. This allows most checks and so on to be performed in advance and then quickly iterated.

On a high level, upon logging in, the translator is presented with a webpage containing navigational features at the top and available languages and projects, as well as account

information, in the upper right. After the source language is selected, the relevant source files are presented. Once a source file has been selected, the pertinent project data and two columns containing source and target segments appear and make up the body of the page. Segments can range from short phrases to several sentences, depending on how the PO files were generated and parsed.

When the translator is ready to start translating, he/she moves the mouse over the selected segment and clicks on the Edit button that appears next to the selected segment. This makes the target segment editable via form input and provides links for navigation, sizing of the form fields, and two ways to submit the translated segment: by way of suggestion and by submitting the translated segment. Watch for more on using Pootle in the second part of this series: [Pootle: Part 2 – Online and Offline Translation](#).

### **Requirements**

Pootle depends on several software packages to provide low-level functionality. These are all FOSS, but since there is no unified installer for the whole package, this makes installing a working copy of Pootle somewhat less than trivial.

The first requirement for Pootle is the Translate Toolkit. This software package provides a range of functions used by the Pootle software package under the hood. The Translate Toolkit is hosted by the largest FOSS repository in existence at [sourceforge.net](#). Be sure to check the README file in the Pootle software package for the required version of Translate Toolkit in order to avoid any incompatibilities.

Both Pootle and the Translate Toolkit require Python to be installed. Python is a dynamic object-oriented programming language and is the code in which Pootle is written. Again, check the README file in the Pootle and Translate Toolkit software packages for the required version of Python. Depending on your operating system, you may have to install Python from source code so the Translate Toolkit will be accessible in your Python Path.

For the web interface of the package, Pootle requires [jToolkit](#), also available from [sourceforge.net](#). [jToolkit](#) is the server engine

that enables the server- and web-based portions of Pootle. It can run applications either via the most popular web server on the planet, Apache, or standalone from the command line. For reasons I won't go into here, at the moment Pootle works best run from the command line. It's also much easier!

Downloads of PO file bundles are compressed by another software package called ZIP and this enables users to download ZIP files of directories for translation offline.

For integration with the version control systems CVS and/or Subversion, the relevant version control client must be installed and accessible from the respective directory contained in the PATH variable of the user profile.

The templating functionality of Pootle is taken care of by a software package by the name of Kid, which in turn requires the software package ElementTree for XML processing.

Installation of the software package PyLucene is optional, but helps to speed up searching. PyLucene uses the text-indexing engine Lucene to provide an effective and very fast way of indexing PO files.

Once installation of all required packages is complete, Pootle itself can be installed using the instructions from the above-mentioned Pootle README file.

## Running Pootle

Once Pootle is installed, you should edit the preferences file and specify your languages and

projects. The server is started by entering the PootleServer command in a terminal client window.

At this point you are ready to begin translating. Just point your web browser to the local address <http://localhost:8080/> and you should see the projects and languages page where you can register. After registering, you may log in and start translating.

If you would like to have a look at a Pootle server in action, please visit the good folks at the WordForge project and point your Web browser to <http://pootle.wordforge.org/>. There you can participate in various localization projects in (at the time of this writing) 94 languages and 12 projects, including the popular web browser Firefox and the office suite OpenOffice.

Thanks for your time and stay tuned for part 2 in this series, Web-based Translation, Pootle: Part 2 – Online and Offline Translation.

## References

- 1) WordForge  
<http://translate.sourceforge.net/>
- 2) Translate Toolkit  
<http://translate.sourceforge.net/wiki/toolkit/index>
- 3) Pootle  
<http://pootle.wordforge.org/>

Dierk Seeburg is a Programmer/Analyst with Choice Hotels International in Phoenix, Arizona. A native of Germany, he has been an editor and freelance translator and interpreter of English and German for 17 years.  
Contact: [info@icotext.com](mailto:info@icotext.com)

Mac Users Unite!

❖ FYI

by Michael Metzger

TransMUG (Translator Macintosh User Group) and IMUG (International Macintosh User Group) are in the early stages of determining whether there is interest in a presentation (topic to be determined) to be co-sponsored by both groups and coincide with the ATA conference in San Francisco, October 31 – November 3, 2007. At this point, the organizers only need to know how many conference attendees would be interested in such an event, to be held at a nearby location or the conference site itself. If you would like to

participate outside of the regular conference program, preferably Thursday or Friday evening, please let them know. Send a short e-mail with the subject "Mac Users Unite" to Michael Metzger ([mm9@sbcglobal.net](mailto:mm9@sbcglobal.net)) by June 30, 2007, indicating which evening would work best for you. The results of this informal survey will be forwarded on to the hopeful organizers.

Michael Metzger is Administrator of the Language Technology Division.

Anyone who has used Trados on a laptop is painfully aware of the problem with unavailable keyboard shortcuts. The most frequently used keyboard shortcut [Alt] + [Num +] (for Set/Close Next Open/Get) requires the use of the plus (+) key on the numeric keypad. However, since accessing the numeric keypad is quite cumbersome on most laptops, most people end up using the toolbar button instead. This can be really annoying if

one prefers using keyboard shortcuts when translating with Trados.

Luckily, there is a simple solution to this – but for some reason it is not mentioned anywhere in Trados manuals, menus or Help files. For both shortcuts that normally require the use of the numeric keypad, there is an alternative keyboard shortcut that uses letters instead, as follows:

Standard shortcut	Alternative shortcut	Menu Command
[Alt] + [Num +]	[Alt] + [Ctrl] + [N]	Set/Close Next Open/Get
[Alt] + [Num *]	[Alt] + [Ctrl] + [Z]	Translate to Fuzzy

If this is not enough for you, or you happen to hate some other Trados keyboard shortcuts as well, you can customize them almost any way you want. For more information on this, see the following support article: <http://tinyurl.com/ydlgtg>.

Tuomas Kostiainen (tuomas@jps.net) is an English to Finnish translator and Trados trainer, and has given several Trados workshops and presentations. For more Trados help information, see [www.finntranslations.com/tradoshelp](http://www.finntranslations.com/tradoshelp)

Language professionals were energized and excited at a Translation Tools Seminar in New York City, held on January 20, 2007, by Tellmealgo Translations and The Language Shop. Despite near-freezing temperatures and ice-covered roads that threatened to delay the seminar, everyone arrived in great spirits. The trainer, Emily Tell, confidently presented a topic that many translators fear or, at best, regard as existentially important to their practice. In a large room filled with fourteen translators that Saturday in New York City, some of the more theoretical and practical aspects of translation tools and, in particular, SDL Trados 2006, were covered.

The visually appealing PowerPoint presentation focused on defining key terms. These often confuse translators due to their inherently technical connotations and the fact that they are constantly evolving, like languages themselves. The seminar stemmed from a feeling of empathy for the difficulty that translators face when they are expected to

master two working languages and highly complex software. On the other hand, it was emphasized that, now more than ever, language professionals must take advantage of all available free resources both online and in print in order to learn how translation tools can benefit their careers. Also described was the need to embrace the union of the translation and technology fields.

A further discussion ensued concerning Translation Technology (“TT”) as an umbrella term that includes two branches. The first comprises the famous Computer Assisted Translation (“CAT”) tools such as SDLX and Trados. They are so called because they aid in the translation process performed by a human being. The second branch is Machine Translation (“MT”), which is a sub-field of computational linguistics. It is also computer software that translates text from one language to another. The main difference between CAT tools and MT is that the latter actually completes the translation process and

requires post-editing. For more information on Machine Translation and a lively discussion on the emerging profession of post-editing, refer to the article “Machine Translation: Translating Automation into New Opportunities.”<sup>1</sup>

The discussion of TT ended after CAT tools were further divided into three groups. The first is Translation Memory (“TM”), which is defined as a database, typically used in a CAT tool to aid human translators. A TM consists of text segments in a source language and their translations into one or more target languages. The segments can be blocks, paragraphs, sentences or phrases<sup>2</sup>. An example of a program that creates TMs is Trados Translator’s Workbench (“TW”). The second group is Terminology Management tools, another component of many CAT tools. Terminology Management programs catalog words and phrases, together with pertinent information, in a database. Terminology Management software can be used to create glossaries. Each word can be placed in a glossary and assigned a subject matter, translation, definition, and other characteristics. An example of a Terminology Management program is Trados Multiterm. The third group of tools is related to the software localization industry. An example of this kind of software is the whole range of ENLASO tools.

A brief, interactive discussion of typical file formats for translators was followed by the hands-on portion of the seminar. The first function taught was how to create a Translation Memory using the SDL Trados 2006 demo version of Translator’s Workbench.

After creating a TM, we discussed analyzing the sample file to be translated. Basically, this is similar to performing a word count. The difference between analyzing a file in Trados and performing a word count in Microsoft Word can be seen in Figures 3 and 4 below.

Figure 3 – Trados Analyse Function

Match types:	Segments	Words	Percent
Context TM:	0	0	0%
Repetitions:	0	0	0%
100%:	7	160	77%
95% - 99%:	0	0	0%
85% - 94%:	1	34	16%
75% - 84%:	0	0	0%
50% - 74%:	0	0	0%
No match:	1	15	7%
Total:	9	209	100%

Figure 4 – MS Word Count

Statistics:	
Pages	4
Words	706
Characters (no spaces)	3,674
Characters (with spaces)	4,395
Paragraphs	13
Lines	100

Basically, Trados compares a document to be translated with a database of previously translated segments and words. It not only provides the total number of segments and words in the document, but it also tells you (1) what percentage of those segments and words has been translated before; and (2) whether it is an exact translation or a fuzzy one. MS Word just tells you the number of pages, words, characters, paragraphs, and lines in a document.

Trados and MS Word are not the only programs that perform word counts. Practicount ([www.practiline.com](http://www.practiline.com)) counts the words in all of a job’s files, even if they are in different formats. However, it does not count scanned PDF files that are images.<sup>3</sup>

Following the morning session of the seminar, there was a lunchtime discussion concerning translation tools and other topics relevant to the linguistics business. In a cozy room known as “the Library” because of its wooden panels and bookshelves, twenty-one language professionals of diverse backgrounds took shelter from the brutal cold and broke down barriers often faced in big metropolises such as New York City. They shared experiences in differing fields such as ESL education, the performing and visual arts, and the military. As an added bonus, one lucky translator won the raffle prize donated by Jost Zetzsche: the 5th edition of his e-book, *The Translator’s Tool Box: A Computer Primer for Translators*. Many friendships and business relationships were made, renewed and strengthened. The Networking Lunch served as a reminder that these encounters are rare and advantageous. The exchange of resumes and business cards added icing to the networking cake.

Afterward, back in the classroom, the afternoon session continued with additional

functions of SDL Trados 2006. Topics included translating a file using the Trados toolbar, translating an updated file, and using the concordance search.

The seminar concluded on a positive note as participants received certificates of attendance. ATA-certified members received continuing education points for having attended. They not only shared a hands-on experience but, just as importantly, they also became more aware of the myths surrounding translation tools. Finally, they realized the importance of embracing the shift to a more technologically advanced translation environment in order to enhance their practice.

Emily Tell is the director of Tellmealgo Translations, a New York-based agency that specializes in Spanish and

Portuguese language services. She has also taught in the translation program at the Universidad de Santiago, Chile. She is currently developing a series of EFL instructional videos. Contact: [translations@tellmealgo.com](mailto:translations@tellmealgo.com)

Deborah Lockhart is the treasurer of the New York Circle of Translators and the owner of The Language Shop, a New York based Translation Company specializing in translation, interpreting, transcription and various other linguistic services. She has worked as a freelance translator and educator since 1985. Contact: [info@thelanguageshop.org](mailto:info@thelanguageshop.org)

<sup>1</sup> Ursula Schwalbach and Franco Zearo, "Machine Translation: Translating Automation into New Opportunities," *The ATA Chronicle* (May 2006): 23-26.

<sup>2</sup> [http://en.wikipedia.org/wiki/Translation\\_memory](http://en.wikipedia.org/wiki/Translation_memory)

<sup>3</sup> Donna Sandin, "Translation Tools: Getting Your Hands Dirty," *The ATA Chronicle* (July 2006): 18-19.

## LTD & TAC Contacts



### LTD Division Administrator

Michael Metzger  
[mm9@sbcglobal.net](mailto:mm9@sbcglobal.net)

### Newsletter Editor

Lisa Carter  
[lisa@intralingo.com](mailto:lisa@intralingo.com)

### Nominating Committee

Robert Killingsworth  
[robertk117@comcast.net](mailto:robertk117@comcast.net)

Jost Zetzsche  
[jzetzsche@internationalwriters.com](mailto:jzetzsche@internationalwriters.com)

### TAC Chair

Alan Melby  
[melbyak@yahoo.com](mailto:melbyak@yahoo.com)

### Mailing List

Tony Roder  
[tony@wells.com](mailto:tony@wells.com)

### Website

Michael Wahlster  
[Michael@GW-Language.com](mailto:Michael@GW-Language.com)