Free and Open Source Software for translators

Corinne McKay*

Abstract: Free and open source software (FOSS) is gaining popularity in the translation industry, due to its low cost, flexibility, reliability, security and freedom from licensing hassles. With a range of free and open source translation memory suites, web browsers and office software to choose from, FOSS is an attractive option for freelancers and agencies alike. The past year has seen many new developments in the world of FOSS for translators. This article will address three main topics: what is FOSS; why is FOSS important to translators; and how can translators get started with FOSS?

Software de fuente abierta y libre para traductores

Resumen: El software de fuente abierta y libre (FOSS por sus siglas en inglés) está ganando popularidad en la industria de la traducción debido a su bajo costo, flexibilidad, fiabilidad, seguridad e independencia de los problemas de licencias. La amplia gama de memorias compatibles, navegadores y aplicaciones comerciales actual hace de los sistemas FOSS una atractiva opción para traductores autónomos y agencias de traducción. El año pasado se produjeron muchos nuevos diseños para traductores. Este artículo aborda tres temas principales: qué es el software FOSS, por qué es importante para los traductores y cómo pueden éstos comenzar a usarlo.

Key words: open source project, translation tools, Linux. ***Palabras clave:*** Projectos de fuente abierta, herramientas de traducción, Linux
Panace@ 2006; 7 (23): 95-98.

What is Free and Open Source Software?

Many people think that software that doesn't cost money is automatically Free Software. In fact, FOSS might be free or might cost money, but the real distinction is that its source code is publicly available. Free Software Foundation <www.gnu.org)>founder Richard Stallman famously paraphrased this as «free as in free speech, not as in free beer,» although this distinction was recently blurred by the production of the world's first open source beer, «Our Beer,» whose recipe is available at <www.voresoel.dk>. Free and Open Source Software is often contrasted with proprietary software, whose source code is owned and normally kept from public view by the company that sells the software. FOSS is also licensed differently from proprietary software. The most common FOSS licenses such as the GNU General Public License allow the user to:

- Install the software on as many machines as he/she wants.
- Allow any number of people to use the software at once.
- Copy the software and give it to anyone
- Modify the software, as long as certain features are kept intact (most commonly the licensing agreement).
- Freely (in the sense of «without restrictions») and for any purpose distribute or sell the software without paying royalties to the original developer.

So on the most basic level, FOSS is software whose source code is publicly available; the software might or might not cost money, but any user is free to look at the source code, change it, and release the changes to the public. So for example in October, 2000, Sun Microsystems gave away the code to an office suite that it owned called StarOffice, which became the office suite that we know today as OpenOffice.org.

FOSS is also almost always distributed under a very liberal licensing agreement; for example most FOSS licenses are good for an unlimited number of computers, so you can buy one CD of OpenOffice.org and install it on as many computers as you want. The user is also not obligated to keep records of when and where the software is installed, and the user may make modifications to the software and release those modifications to the public. So, of particular interest to translators is that anyone is free to create a language-specific version of any piece of free and open source software; you can translate it yourself and release your translation to the public, as opposed to asking a proprietary software company to do this for you.

The opposite of FOSS is proprietary software, which means software whose source code is owned by a person or company, and which is almost always sold or given away under a more restrictive End-User License Agreement (EULA).

It's important to understand that the real restriction behind proprietary, or closed-source software, is not so much price, but the fact that the owner of the software's code dictates how the software can be used. For example most proprietary software EULAs stipulate that the software may be installed on a specified number of machines, and that the user must track when and where the software is installed and uninstalled. Many proprietary software EULAs also prohibit activities that are expressly permitted by law, such as reverse engineering. Some of these provisions have been highlighted in the news in recent years due to cases such as Jon Johansen, a 16 year old Norwegian who was arrested for reverse-engineering the copy protection software

---

* French to English translator and open source software enthusiast based in Boulder, Colorado, USA. Correspondence to: corinne@translatewrite.com

on DVDs, and subsequently won both cases brought against him, and Microsoft's license compliance audits of public school systems in Philadelphia and Portland, Oregon, and the city of Virginia Beach, Virginia, which estimates that it paid about $129,000 during a Microsoft audit, to acquire missing licenses for software that it had purchased legally but had not saved the license paperwork for (http://tinyurl.com/83tvm)**.**

The most common license under which FOSS is released is called the GNU General public license, or GPL. Probably the most important provision is the fact that it prohibits anyone from taking GPL code and making a proprietary program out of it. The GPL is also sometimes known as «copyleft,» because it allows the source code's copyright holder to require that the code be distributed under the terms of the GPL. In April, 2004 the GPL was confirmed as a legally enforceable license when a court in Germany issued a final ruling upholding its validity after the German company Sitecom refused to stop distributing software produced by a company called Netfilter, in violation of the GPL's provisions.

Probably the best-known example of FOSS is the Linux operating system, sometimes called GNU/Linux, because some of its essential components come from the earlier, and also free and open source, operating system called GNU. GNU was developed as a Unix-like operating system based on free software, by Richard Stallman of the Free Software Foundation. Linux was first released in 1991 as a hobby project by Linus Torvalds, a 22 year old Finnish computer science student who had such modest goals that he didn't even name the project Linux, planning instead to name it Freax, for «Free Unix.» The name «Linux» was selected by the network administrator at the University of Helsinki who needed to name the folder where the software was available for download, and chose the name «Linux,» for «Linus Unix,» judging it to be more professional than «Freax.»

Due to some of the argued advantages of Linux such as lower cost, fewer security holes and lack of vendor lock-in, Linux has been the most popular Internet server operating system since 1999, used by 64% of Japanese corporations since 2002, selected for use by the United States National Security agency in 2003, and Linux server shipments grew 40% in 2004 (Wikipedia). Since Linux's source code is publicly available, anyone who wants to can make their own version of it, which is called a distribution. Some well-known distributions of Linux are Debian, Ubuntu, Red Hat/Fedora Core, and Knoppix. The Google search engine is probably the world's largest Linux installation, with over 100,000 Linux servers powering its searches; Google is an excellent example of one of the Internet's «killer apps» that is powered by free and open source software.

In another example of the success of FOSS, the Apache web server has been the market leader web server since statistics on this have been collected. The OpenOffice.org office suite has recorded 100 million downloads, and Mozilla Firefox web browser about 40 million downloads. Open Source Software has gained increasing momentum outside of the U.S., where countries such as Germany and Spain have migrated some or all of their national and local government desktop computers to open source software.

An additional important concept is that of open standards. Possibly the best-known example of an open standard is the PDF file format. While Adobe developed the PDF file format and Adobe's software is not open source, PDF is an open format, meaning that anyone can create a program that reads or creates PDFs without paying a licensing fee to Adobe. So, even if Adobe were to go out of business, PDF-format files would not be rendered obsolete, because other programs would still be able to read and write them. In the translation world, the most common open formats are XLIFF (eXtensible Localization Interchange File Format) and TMX (Translation Memory eXchange). Files created by CAT tools that use these formats are, at least in theory, not dependent on a single CAT tool, and can be re-used in another tool that supports these formats.

Why is FOSS important to translators?

Translators are heavy computer users, with most of us spending the majority of our work day on tasks that require the computer. In addition, most translators are self-employed and bear the cost of maintaining and upgrading their own computer systems, often at considerable cost. Translators also use, or would like to use, language-specific versions of software that are often difficult to find in proprietary distributions. Still, translators as a group have been reluctant to migrate from proprietary software toward open source alternatives.

This author can't pretend to know everything about why translators are reluctant to switch to FOSS, but here are a few theories gathered from personal experience:

- Translators as a group are somewhat technophobic and would rather concentrate on translation itself rather than on technological tools for translation.
- Translators often don't know much about FOSS, and so aren't often aware of the options available.
- Translators are afraid that using FOSS will cause them to lose clients. They are afraid that by running a different program than the client, or than other translators, they will have a hard time finding work.
- Many CD-ROM dictionaries do not work on a Linux computer.
- The market-leader translation memory software companies do not produce Linux versions of their products. FOSS TM applications are available, but they do not have a great deal of name recognition, and may require some work to make their memories compatible with the market leaders.
- FOSS applications are largely volunteer-developed and do not have the financial resources to compete with proprietary applications in terms of marketing, training and support.
- People in general, not just translators, often think that if something is free, it may be less useful or valuable than something that costs money.

As you can see, some of these reasons, such as difficulty of running CD-ROM dictionaries on a Linux machine, are entirely valid, while others, such as lack of dollars to market open source applications, are merely a matter of circumstance and not of the software's usability. Following are some reasons why translators

might consider switching to FOSS applications.

- The software is there. FOSS applications exist for every step of the translation work flow; translation memory, word processing, web browsing, accounting, website development, and more.
- Open source is cost-effective. While «it's cheap» is rarely the primary motivation cited by avid FOSS users, «value for money» is often a reason for choosing FOSS. FOSS won't free you from the need to pay for documentation, support and training, just as you do with proprietary software. But it will in many cases free you from the need to pay high prices for software, upgrades, data recovery after a virus attack, tracking licensing paperwork, etc. In addition, having free or low-cost tools available for critical but unexciting tasks like backups makes it more likely that you will actually obtain the tools.
- Open source encourages innovation. When the Mozilla Firefox web browser made headlines in late 2004, Microsoft hadn't released a major redesign of Internet Explorer since 2001, in part because they didn't have to; with a 95% market share, there simply wasn't much of an incentive to integrate tabbed browsing or localize into Albanian. However, a community-developed project like Firefox can take a «don't you wish you had…» feature and integrate it as fast as coders can work.
- Open-source licenses are less restrictive. Most end-users, including translators, never read EULAs (end-user license agreements) before clicking «I Agree» and installing new software. In addition to restrictive conditions such as allowing the software to be installed on only one computer (meaning that a user with a laptop and a desktop is legally required to buy two copies of the software at full price), some EULAs go even further and specify that the user may not publicly criticize or evaluate the software. Network Associates, the makers of McAfee antivirus software, was criticized in 2003 for prohibiting users from disclosing the results of benchmark tests on its software without prior written permission <http://www.eff.org/wp/eula.php>. Nearly every proprietary software EULA prohibits activities that are generally allowed by law, for example reverse engineering. By contrast, the most popular open source licenses such as the GNU General Public License, specify very basic requirements, such as that any released improvement of the software must also be free software, and that anyone must be allowed to run the software for any purpose.

SDL's recent acquisition of Trados has highlighted some of the reasons why proprietary licensing agreements affect translators. For example, one respondent to the GALA (Globalization and Localization Association) survey «Language Service Provider Reaction to SDL's Purchase of TRADOS» offered the following insight: «Worst case scenario: SDL changes the licensing terms for TRADOS to require information on every client/project where TRADOS is used. Thus SDL would have a list of all the competitive vendors' clients. This is unlikely, but it just gives SDL too much power in our small industry.» As this respondent points out, while this scenario seems far-fetched from a business standpoint, it would in fact be legal, pointing out that proprietary software users are very much at the mercy of the vendors who own their software.

How much less does FOSS really cost?

Various studies have compared the total cost of ownership (TCO) of proprietary software and FOSS; the real TCO depends on how you use the software- whether you want documentation on paper, whether you pay for support and training, etc. Notably, Sun Microsystems and Microsoft have both released «independent reports» stating that their products have the lowest TCO. However in terms of cost of acquisition and upgrades, it's worth noting that FOSS isn't just a little bit cheaper than proprietary alternatives, it's a lot cheaper, and the price difference is even more apparent in a large installation. Let's look at a computer setup for a typical freelance translator, comparing proprietary software (prices obtained from Amazon.com and Translationzone.com in August, 2005) and FOSS.

Proprietary Software	Open Source Software
Microsoft Windows XP Professional: $262.99	Ubuntu Linux: $0.00
Microsoft Office XP Professional: $225.00	OpenOffice.org $0.00
Microsoft Outlook: $89.99	Mozilla Mail: $0.00
Quick Books Basic: $189.99	Gnu Cash: $0.00
Trados 7 Freelance: $895.00	OmegaT: $0.00
Total: $1,662.97	**Total: $0.00**
	OR as above, plus:
	CrossOver Office Professional: $74.95
	Microsoft Office XP Professional: $225.00
	Wordfast: $220.00
	OR as above, plus:
	Heartsome XLIFF editor: $88.00

In a larger office such as a translation company, the savings could be even more remarkable. The Australian consultancy Cybersource <www.cyber.com.au> produced an in-depth report entitled «Linux vs. Windows: The Bottom Line» which compared the setup costs for a 50 user office needing an operating system, e-mail server, and database server. The study found a total cost of $69,987 for the Microsoft option, and $80 for the Linux option, resulting in a savings of $69,907. Because FOSS licenses generally allow the software to be installed on an unlimited number of computers, a 200-person translation company could set up its computer systems for the same cost, $80, while the cost of the Microsoft option would balloon to $282,974.

The cost of upgrading an open source system is typically a

fraction of what is required for a proprietary system. For example, when Trados released Trados 7 Freelance at $895.00 (price obtained from translationzone.com in August, 2005) it offered upgrades for either $395.00 or $495.00, depending on the pre-existing version. By contrast, an open source TM application such as OmegaT or Sun Open Language Tools could be upgraded for free. An open source operating system can simply be downloaded again, usually for free, when a new release comes out.

Another important savings related to FOSS is the freedom from license management costs, especially for large offices such as translation companies. In his paper «Why Open Source Software/Free Software? Look at the Numbers!» <http://www.dwheeler.com/oss_fs_why.html>, David Wheeler summarizes «Proprietary vendors make money from the sale of software licenses, and are imposing increasingly complex mechanisms on consumers to manage these licenses. Customers who cannot later prove that they paid for every installed copy of proprietary software (e.g., due to copying by an employee or losing the license paperwork) risk stiff penalties. In short: by using proprietary software, you run the risk of having the vendor sue you. To counter these risks, organizations must keep careful track of license purchases. This means that organizations must impose strict software license tracking processes, purchase costly tracking programs, and pay for people to keep track of these licenses and perform occasional audits.» For this reason alone, FOSS is an attractive option for a translation company that wants to put its resources into language work rather than software record-keeping.

What free and open source software is available for translators?

If you're new to the idea of mixing FOSS and translation, an excellent website to browse is Marc Prior's «Linux for Translators» <www.linuxfortranslators.org>, which gives an overview of «how to» and «why to» implement FOSS options. The past year has seen major advances in the availability of FOSS for translators. Many open-source CAT tools will accept files created by proprietary CAT tools; check the individual software's documentation for more information on this. Heartsome, although it is not open source, guarantees that its software is TMX-compatible with Trados Version 6.5.5, SDLX and Deja Vu X.

OpenOffice.org <www.openoffice.org>, a cost-free and open source office suite, is compatible with Microsoft Office for most tasks that do not involve Word macros. For example, a user can use OO.o to open a document created in MSWord, edit it, save it in MSWord format using OO.o, and the document will still look as if it were created and edited with MSWord. OpenOffice.org includes applications for word processing (Writer), spreadsheets (Calc), presentations (Impress) and databases (Base). It is localized into over 45 languages and is available in Windows, Mac and Linux versions.

Mozilla Firefox <www.getfirefox.com>, a free and open source web browser, has gone head to head with Internet Explorer in browser competition. Firefox won the PC World Product of the Year award, Forbes Best of the Web, and PC Magazine Editors Choice Award. It has now logged over 100 million downloads and is available for Windows, Mac and Linux in a huge variety of languages.

OmegaT <www.omegat.org>, a cost-free and open source CAT tool, released version 1.4 on August 11, 2005. OmegaT is written in Java, so will run on Windows, Linux, and Mac OS X. It supports the traditional CAT features such as fuzzy matching, match propagation, simultaneous use of multiple translation memories, and multiple file formats. In addition, unlike CAT tools such as Trados and Wordfast that work from within Microsoft Word and are therefore dependent on it, OmegaT is an independent application, although it is most useful when paired with OpenOffice.org (also free and open source). OmegaT is compatible with other CAT tools at TMX (Translation Memory eXchange) Level 1, and is localized into Catalan, Italian, Afrikaans, and German, with documentation in English, German, French, Italian and Japanese. Prior to the fall of 2005, OmegaT segmented at the paragraph level, but in current releases it segments at the sentence level.

Heartsome <www.heartsome.net>, produced in Singapore, is currently the only commercial translation tool that is Linux-compatible. Although Heartsome is not itself open source, it supports open standard documents such as XLIFF (eXtensible Localization Interchange File Format) and Open Document Format. Like OmegaT it is a «single-layer application,» meaning that it runs on its own, rather than from within another application such as Microsoft Word. Heartsome runs on Windows, Mac OS X, Linux, Unix and Solaris, and is TMX (Translation Memory eXchange) compliant at Levels 1 and 2. Heartsome also attests that «TMX files generated with Heartsome's tools have successfully been tested with Trados Version 6.5.5, SDLX and Deja Vu X.» Like the Sun Open Language Tools, Heartsome is based on an XLIFF translation editor, and it also includes a TMX-based translation memory editor. The personal edition of the XLIFF translation editor alone is $88, and the full translation suite, consisting of the XLIFF editor plus a TMX editor and dictionary editor, is $398.

Transolution <http://transolution.python-hosting.com/> provides a suite of tools similar to the Sun Open Language Tools, but written in the Python programming language. Transolution supports the XLIFF standard and runs on Windows or Linux.

Sun Microsystems released the first installment of its **Open Language Tools** (https://open-language-tools.dev.java.net/) project, an XLIFF (XML Localization Interchange File Format) Translation Editor and an XLIFF Filter. These are also written in Java so will run on Windows, Linux, or Mac OS X. At present, the tools can be used with a variety of file formats for translating documentation files in HTML, Docbook SGML, JSP, XML, OpenOffice.org and plain text, and software files in .po, Msg, Java .properties, Java ResourceBundle, and Mozilla .DTD resource file formats.

Whether you're interested in implementing FOSS at the operating system level or trying a few applications to start out with, the applications are nearly endless. And, depending on your particular situation, the benefits of using FOSS can be equally large. For more information on how to make the most of FOSS and computers in general as a translator, several free e-newsletters are out there, including The Tool Kit <www.internationalwriters.com> and Open Source Update <http://www.translatewrite.com/foss/index.php?s=foss&p=osupdate>.