

A comparison between open-source and commercial translator tools

Atiqur Rahman

Department of Computer Science and Engineering

January 2007

A comparison between open-source and commercial translator tools

Thesis report prepared by

Atiqur Rahman
ID: 02101114

Under the supervision of

Matin Saad Abdullah
Senior Lecturer
Department of Computer Science and Engineering, BRAC University

Forward

The thesis report is submitted in partial fulfillment of the academic requirements for the degree of Bachelor of Science in Computer Science and Engineering to the Computer Science and Engineering department at BRAC University, 66 Mohakhali Dhaka1212, Bangladesh.

Atiqur Rahman ID 02101114

As the supervisor of the candidates I have approved this dissertation for submission.

.....

Matin Saad Abdullah

Senior Lecturer

Department of Computer Science and Engineering

BRAC University

Declaration

The whole dissertation, unless specifically indicated to the contrary in the text, is my original work, and has not been submitted in part, or in whole for the degree or diploma to any other university.

.....

Atiqur Rahman
ID 02101114

Approval Sheet

The thesis report titled “**A comparison between open-source and commercial translator tools**” has been submitted to the following respected members of the Board of Examiners from the Faculty of Computer Science and Engineering in partial fulfillment of the academic requirements for the degree of Bachelor of Science in Computer Science and Engineering on January 07, 2007 by the following students and has been accepted as satisfactory.

Atiqur Rahman ID 02101114

.....

Matin Saad Abdullah
Senior Lecturer
Department of Computer Science
and Engineering
BRAC University

Acknowledgments

At first our heartiest gratitude goes to Almighty ALLAH, without His divine blessing it would not be possible for us to complete this project successfully. Then I am eternally grateful to my parents for their constant support throughout my academic work.

It's been a pleasure and an honor to show my respect to Matin Saad Abdullah Senior Lecturer, Department of Computer Science and Engineering, BRAC University, who has rendered continuous and encouraging guidance throughout the entire thesis period with his kind advice. His generosity and zeal for this thesis is exemplary.

I want to thank all the honorable faculty members especially Sarwer Alam, Lecturer, Department of Computer Science and Engineering, BRAC University, for their feedback and inspiration that was really instrumental in shaping this thesis.

Finally I like to offer thanks to all of our friends and well-wishers for helping us by rigorous reviews of this work and inspiring suggestion.

Table of Contents

Abstract

Chapter1 Introduction.....	9
Chapter 2 Types of Technology.....	10
2.1 Open source.....	10
2.1.1 Case studies overview.....	10
2.1.2 Some Common Open Source Applications used in the.....	11
Projects.	
2.2 Commercial translator.....	13
2.2.1 Trados.....	13
2.2.2 Language industry background.....	14
2.2.3 Why Trados Is Important.....	14
Chapter 3: The opportunity.....	15
3.1 Development and Implementation.....	15
3.2 Benefits.....	15
3.3 The Deal Diminishes Comparison.....	16
Chapter 4 SDL Could Benefit from Locking In Customers.....	18
4.1 Respondents Vent about General Concerns.....	18
4.2 Respondents Offer Advice on Technology Consolidation.....	18
Chapter 5 Conclusion.....	20
References.....	21

Abstract

When starting to analyze the tools used for translation, I have chosen Pootle as the first tool to consider is an open source tool that performs translation by means of analyzing the whole sentence first before parsing for translation. This tool is used by Ubuntu, leading distribution of Linux, and because it is open source, it can easily be modified to one's own language and hence be extended. Then for studying a commercial translation tool, the popular tool that I have found and studied is Trados. Despite being closed –source it has a good tutorial and documentation for use. It works by creating a project and keeping the file to translate then translating it. However, the creation of project is only available to the software that is bought with a certain price. Performance wise, Trados has a different way for translating, it translates each word before going for grammatical parsing. This might give rise to certain difficulties or even inconsistencies in the translation process and eventually lead to grammatically complex or incorrect sentences. However, Trados has ways of correcting for these issues but doesn't document how they do it as they are closed-source. Now, which method to use is a debate that needs to be resolved in a scientific way. Hence this project is concerned with the analysis between the two leading yet distinct tools for translation.

Chapter 1 Introduction

Usually, programs are written and documented in English, and use English at execution time for interacting with users. This is true not only from within GNU, but also in a great deal of proprietary and free software. Using a common language is quite handy for communication between developers, maintainers and users from all countries. On the other hand, most people are less comfortable with English than with their own native language, and would rather be using their mother tongue for day to day's work, as far as possible. Many would simply *love* seeing their computer screen showing a lot less of English, and far more of their own language.

Chapter 2 Types of Technology

2.1 Open Source

Over the last few years, Free and Open Source Software (FOSS) has established itself as a viable alternative to proprietary software in many areas of information and communications technology (ICT) deployment. The availability of FOSS without licence fees and its inherent characteristic of being open to modification and adaptation make it an attractive proposition to poorer communities. As a result, many projects that make use of FOSS to empower and help the people have been initiated all over the world especially in poor and developing regions.

This book contains a compilation of the case studies of several of these projects that have been considered as successful. The main focus of most of the projects chosen is not on the actual development of a particular FOSS application per se but in using FOSS applications (these may be new, enhanced or existing ones), as a means to fulfil the primary objectives of the project.

2.1.1 Case Studies Overview

The Free and Open Source Software initiatives and/or projects highlighted in this book are:

Africa

- _ Creating Educational and Business Opportunities – AVOIR, South Africa
- _ Making Legal Information Freely Available – JuriBurkina, Burkina Faso
- _ Promoting Free and Open Source Software in Africa – Meraka, South Africa
- _ Localizing Free and Open Source Software – Translate.org.za, South Africa

Asia-Pacific

- _ Connecting Farmers and Buyers – AgriBazaar, Malaysia
- _ Building an Indonesian GNU/Linux – BlankOn Linux, Indonesia
- _ Identifying and Controlling Weeds – OSCAR, India

- _ Managing Disasters – Sahana, Sri Lanka
- _ Reducing Vulnerabilities in the Pacific – Tikiwiki GeoCMS, Fiji

Europe

- _ Improving Government-Citizen Interaction – eGov Balkan, Bulgaria
- _ Inculcating ICT Usage in Educational, Social and Economic Activities

Latin America

- _ Empowering Local Communities – Colnodo, Colombia
- _ Managing the Environment – Galápagos, Ecuador
- _ Enabling Public Schools – Paraná, Brazil











As can be seen from this list, the projects come from the four corners of the globe. This is a clear indication that FOSS is being taken up worldwide, particularly in the developing countries of Africa, Asia and Latin America. The projects cover a wide range of domains ranging from specialized applications to web portals and services for e-government to full-fledged projects on the development of FOSS packages. They are initiated mainly by local organizations and, in some cases, funding is provided by international aid and development agencies.

2.1.2 Some Common Open Source Applications used in the Projects

GNU/Linux

The GNU/Linux operating system consists of the Linux kernel itself and the rest of the system software and tools/utilities which, working together, make up the operating system. Most of the system software is from the GNU Project. This operating system is commonly referred to as Linux. The GNU/Linux operating system is commonly distributed bundled together with other application software for use by end users and this is known as a Linux distribution or distro. Different Linux distros may have different applications bundled with them but they all have the same Linux kernel and core GNU utilities. Popular Linux distros include Debian, Red Hat, SuSe and Ubuntu.

Language Code: bn
Language Name: Bengali
Number of Plurals: 2
Plural Equation: (n != 1)

Name	Translated	%	Fuzzy	%	Untranslated	%	Total	Graph
 OpenOffice.org 2.1	25984	32%	0	0%	54755	67%	80739	
 Pootle 1.0	141	11%	7	0%	1132	88%	1280	
 Pootle 0.11	141	10%	0	0%	1146	89%	1287	
 Pootle 0.9	109	11%	0	0%	853	88%	962	
 Terminology	0	0%	0	0%	1073	100%	1073	

<http://pootle.wordforge.org/bn/>

GNOME and KDE

The two most popular FOSS desktop environments are GNOME and KDE. These are software that present a graphical desktop environment and interface to the user. Using the facilities offered by the desktop environment, the user can interact with the system and run applications.

OpenOffice.org

OpenOffice.org is an office application suite consisting of a word processor, spreadsheet, presentation software, graphics editor and a database program. It is compatible with the popular Microsoft Office suite, being able to read and write Microsoft Office data file formats.

The Mozilla Internet Applications

The Mozilla Internet applications include the popular Firefox web browser, Thunderbird e-mail client and the Mozilla All-in-one Internet Application suite.

Apache

The Apache web server is the most widely used web server on the Internet.

2.2 Commercial Software

2.2.1 TRADOS

In a deal with far-reaching implications, publicly-traded SDL International (SDL) announced on 20 June 2005 that it would buy privately-held TRADOS Inc. for US\$60 million.¹ When its shareholders approved the purchase in early July, SDL became by far the largest supplier of translation memory, terminology management, and translation workflow management products. SDL is also one of the world's three largest language services providers (LSP). SDL's purchase of TRADOS creates a larger company that will be more visible to a broader range of buyers dealing with global or multilingual application and content development.

This article reports on the results of a Globalization and Localization Association (GALA) survey of language service providers about the impact of SDL's acquisition of TRADOS. Its goal is not to comment on the business sense, wisdom, or moral propriety of this deal, but rather to convey the feelings of the industry about the deal.

2.2.2 Language Industry Background

This purchase by SDL continues a recently begun round of consolidation in the language services and tools industry, including Lionbridge's proposed purchase of BGS. That deal will create the world's largest LSP with annual revenue exceeding US\$400 million. Other 2005 deals of note include Irish LSP Transware's purchase of globalization management system supplier GlobalSight and Merrill acquisition of P.H. Brink, both acquiring translation workflow management tools in the transactions. The SDL and Lionbridge deals, though, captured most attention due to their size.

2.2.3 Why Trados Is Important

Nearly nine out of ten companies outsource their translation and localization work to language service providers.⁵ Buyers can choose from a wide variety of LSPs around the world, all falling into one of three categories when it comes to the technology they use to translate, localize, and manage customer projects.

- **The LSP buys most of its technology.** These companies purchase translation memory workbenches, perhaps preferring one supplier but realistically using whatever the client wants. Many layer project management and workflow applications on commercial off-the-shelf (COTS) products such as Microsoft Access or Project. Earlier this year BGS had announced that it would join this camp, standardizing core elements of its production process on TRADOS-supplied software. However, the company did not execute on this plan as TRADOS sold itself off to SDL.
- **The LSP builds some tools, but keeps them to itself.** Some firms like Connect Global, EQHO Communications, and Merrill Brink International build software infrastructure and tools that meet their specific needs better than any COTS product would. Lionbridge is the most visible example of such a company, assembling its infrastructure from commercially available and proprietary components. Earlier this year it purchased Logoport, a small software firm selling a web-based translation memory tool.⁷ A side-benefit of controlling its TM fate is that Lionbridge avoids having to buy licenses from SDL or TRADOS for the clients who don't care which TM it uses.
- **The LSP builds technology and sells it to all comers.** A few providers sell software, tools they developed (or acquired) to meet their internal needs for project management and technology needs. They hope to recoup their investment by selling them on the open market. SDL, STAR, Translations.com, and Transware exemplify this model. This approach pleases shareholders by turning a cost of doing business into an asset, but it does limit the marketability of the tools because many LSPs feel uncomfortable buying core technology from a rival. This last point is at the heart of the concerns expressed by GALA members and inspired the survey.

Chapter 3: The Opportunity

Same structure and terminology and have a high percentage of material reused from previous versions of the same, or similar, documents. The method commonly used to reduce the time spent on bulk translation was for translators to consult previous translations of relevant documents and manually identify the parts of the texts which best matched the text to be translated. This procedure proved to be time consuming and inefficient.

The TRADOS TM tools were identified as a way of improving this process by reusing the existing electronic collection of source and target texts and by automating the manual matching and retrieval procedure.

3.1 Development & Implementation

The selection of the software was made after consultations with various company representatives, extensive search of web sites, and testing of a number of other tools. Translator's Workbench was selected as a well-known and tested tool, while TRADOS offers extensive customization possibilities and direct on-site support and training. The two products are used in parallel.

TRADOS had specific text formatting and handling needs that could be met only by a significant level of customization and close cooperation with the developers. At that time, POOTLE did not offer any customization facilities, so this was only feasible in the case of TRADOS. At the end of the customization phase, which lasted 3 months and included a good deal of feedback and system tuning, the system was ready to be field-tested by users. After the end of the customization period the learning curve was relatively short and required no additional special user skills, since the POOTLE development team did provide initial on-site training.

3.2 Benefits

Although no detailed cost-effectiveness studies have been conducted, performance and quality of service was considerably enhanced through the use of translator's tools. This claim is further sustained by the fact that translator's tools have been regularly used for the past five years, and that in-house translators view them positively and have come to rely on them. More and more customers specifically ask for such tools to be used in the translation of their texts and this is often a prerequisite in calls for tender for translation projects. Customers believe that these tools can deliver faster and more consistent output.

3.3 The Deal Diminishes Comparison

Several respondents worried that this deal creates an effective monopoly in the tools area and that SDL could do as it pleases.

I worry that they will be able to price their licenses as they please, driving out their competitors. I am also worried that they will become such a big technology provider that they will actually hold a monopoly.

As SDL will become the dominant player in the translation memory technology market, the competition becomes a minor factor for them and prices are likely to rise.

SDL will have an advantage price-wise. We will have to pay whatever price they put on the product, while SDL can include the tools in the production price. If LSPs are asked to pay more for SDL's products, that will have an impact on localization prices.

Translation buyers, and vendors, basically have no choice. SDL controls the only two viable TM products and the productivity gain from using a TM tool makes it indispensable.

My first reaction has been: 'Saint Open Source, please do something for us, smaller language providers.' A month ago it was Logoport's turn, now it is TRADOS'. TRADOS has a very large install base, built on a 10-year presence, so they will make it unless something revolutionary comes along.

There are no other viable options. So whereas previously a client could turn to TRADOS for TM or workflow if they did not want to implement an SDL solution, now they have no real choice.

I think if SDL tried to control the services market through the ownership of tools, they would quickly find that customers have other options. Customers buy services mostly on price, and they are not stupid.

If SDL is able to offer a tool that is a necessity and is able to manipulate it in a way that creates an advantage for them when they offer translation services we could end up having a Microsoft of the localization world. If you don't want to use Microsoft – or SDL in the future – good luck finding an alternative. That concerns me.

This could be the tip of the iceberg. Consolidation is inevitable. It's part of the price of globalization, but my biggest concern is the degree of leverage SDL

will now have over the translation and localization industry. Is this SDL's first move in a larger industry-wide consolidation process?

Our workflows are entirely streamlined with TRADOS technology, and we'll probably have to adapt as SDL and TRADOS tool would merge into one. All tools and plug-ins we've written will become useless. If all translation service providers use the same tools and workflows, it will become more difficult to differentiate from the competition.



Chapter 4: SDL Could Benefit from Locking In Customers

The lock-in issue from the second question typifies that concern. On the negative side, respondents noted. They could be tempted to act as Microsoft did with their browsers and add something to TRADOS which could lock in users. Well, a monopoly is never a desired situation. The fewer companies there are, the bigger chance of us falling victim to a dictatorship in the world of translation tools.

My major concern is that SDL will try to slowly make clients migrate from TRADOS tools to SDL tools, without making these tools compatible with other translation tools.

But what other choice do buyers have and how much do they really care in the first place? Some viewed lock-in more positively, giving customers a one-stop solution: Clients might think it easier to buy tools and services from one vendor. This will depend on the education/viewpoint of the client. Some will raise concerns due to pricing whereas others will be happy that SDL can provide an entire package.

4.1 Respondents Vent about General Concerns

The penultimate survey question was “what are your biggest concerns about this event?” Many respondents reiterated their answers to previous questions, with the majority of the LSPs expressing concerns about buying their mission-critical tools from a service competitor; SDL using the TRADOS contacts to gain market share; the future cost of tools and/or support creating an unfair advantage for SDL; the possible lack of innovation as SDL will have a near monopoly on the tools market; and SDL generally having an unfair advantage.

Some questioned whether this is merely the tip of the iceberg of more market consolidation and whether this acquisition might concentrate too much power in one company.

4.2 Respondents Offer Advice on Technology Consolidation

The final question asked respondents “What are the top three changes you would like to see SDL make to TRADOS tools as it moves forward?” The respondents offered lots of advice to SDL as it ponders its options. Their suggestions included maintaining the status quo with separate products, making TRADOS open source, lowering the price, making a variety of specific

improvements to the product. Since this question basically asked for the respondents' Christmas lists regarding technology, the answers ranged from strategic business issues (see Table 1) to bit-twiddling technology concerns. Reflecting the diversity of the sample, some asked for changes in direction or products that contradicted other suggestions. At least one participant invited SDL to consider acquiring a project management tool that it had developed.

Chapter 5: Conclusions

Now I have almost reached to my paper. In this paper I have tried to find out different sorts of pros and cons of two types of translators. The two types of translator are open source translator and commercial translator. The discussions that I have placed above about this topic, now in conclusion I can say that commercial translator is much more efficient, powerful and more manageable in context of getting more feature and more benefit in research and other development organizations even though it needs some costs. But it relays for the long term benefits for the company. On the other hand open source translator may not be able to manage our task so friendly and also effectively.

References

Globalization and Localization Association, Andover, Massachusetts, United States of America.

<http://developer.gnome.org/doc/tutorials/gnome-i18n/translator.html>

<http://www.poedit.net/>

<http://translate.sourceforge.net/pootle-release-2004-12-16.html>

<http://en.wikipedia.org/wiki/SYSTRAN>