LetsMT!

Platform for Online Sharing of Training Data and Building User Tailored MT

www.letsmt.eu

Project no. 250456

Document information

Deliverable number:	D5.5
Deliverable title:	Usability evaluation of business and financial news SMT
Due date of deliverable according to DoW:	2012-02-29
Actual submission date of deliverable:	2012-02-29 (Revised: 2012-04-10)
Main author(s):	Stefanna Suijdendorp (SEM), Raivis Skadiņš (TILDE)
Participants:	SEM, TILDE, UCPH, FFZG
Reviewer	TILDE
Work package:	WP5
Work package title:	MT usage in news translation: facilities and evaluation
Work package leader:	SEM
Dissemination level:	PU
Version:	1.2
Keywords:	web page translation, translation widget, browser plug-ins, machine translation web-service

History of versions

Version	Date	Status	Name of the author (partner)	Contributions	Description/Approval level
Draft	28th February	Draft	Semlab *Stefanna Suijdendorp*	SEM, UCPH, FFZG, TILDE	Draft of deliverable, ready for final review.
Final (v1.0)	29th February	Final	Semlab *Stefanna Suijdendorp*	SEM, UCPH, FFZG, TILDE	Final version of deliverable
Revised (v1.1)	10th April	Draft	Semlab *Stefanna Suijdendorp*	SEM, UCPH, FFZG, TILDE	Draft of revised version of deliverable
Revised (v1.2)	10th April	Final	Tilde *Raivis Skadiņš*		Revised version of deliverable

EXECUTIVE SUMMARY

This deliverable (D5.5) aims to evaluate the machine translation systems for business and finance news.

The LetsMT! widget has an option for translating the news headings from English into various smaller European languages. The news headings are automatically translated by means of an application in the widget and for this deliverable over 1000 translated sentences were tested manually to evaluate its translation quality. The overall average score of the current system is very low and leaves a lot of room for improvement. The inter-rater agreement was good as the score reflects the individual scores of 3 different and independent evaluation panels, based in 3 different countries. These panels used Absolute Evaluation tool from Linguatec to value the sentences by the adequacy and fluency.

The LetsMT! plugin is tested from the customers point of view in both Mozilla Firefox and Internet Explorer. There is a list of major issues which have to be solved before making the LetsMT! plug-in available for public usage.

After the manual translation evaluations, SemLab used ViewerPro to evaluate the Dutch to English machine translation from a functional point of view. The outcome of these translations is promising – a representing 80% of these sentences is translated with sufficient accuracy to enable event detection in ViewerPro.

Table of Contents

Introduction

The main aim of WP5 is to evaluate the SMT systems for business and financial news translation. Two MT facilities (widget and the browser plug-in) and their usability as well as translation quality in general are evaluated. The evaluation results will be used to improve translation quality incrementally in general as well as MT usability in applications.

This research report consists of 3 major parts: first of all the description and evaluation of the tested LetsMT! components – the widget and the browser plug-ins. Secondly the manual evaluation and thirdly the functional evaluation, added by Semlab.

The functional evaluation and the evaluation of the widget is being led by SEM; the evaluation of the browser plug-in has been done by FFZG, the methodology for the manual evaluation has been prepared by Tilde

The manual evaluation of the translation quality was performed by the project partners who were asking native speakers of the various target languages for support. Human evaluation metrics (fluency, adequacy, tests of understanding and overall quality of translation) were applied to evaluate translation quality.

The usage of SMT systems and facilities for business and financial news translation is evaluated and elaborated iteratively in this WP.

Detailed test results of the translation widget and browser plug-ins are written in Chapter 1, the methodology and results of the manual quality evaluation as well as a brief description of the evaluation tool can be found in Chapter 2, the methodology and results of the functional evaluation can be found in Chapter 3. Initial outcomes from the manual evaluation are clarified in Chapter 4. Conclusions and recommendations are stated in Chapter 5. The appendices are included separately and contain lists of known issues in tested LetsMT! components and all the test results of the translated sentences in 3 languages; named Dutch, Danish and Croatian.

1 LetsMT! components evaluated

1.1 Widget by SEM

The web page translation widget is an application that has been introduced in several online MT systems, e.g., Google Translate and Microsoft Translator. These widgets bring real-time, in-place translations to a web site. Users can see pages in their own language, without having to go to a separate translation web site. The web page translation widget will help users to translate the web page only if developers of the web page have added this widget to their web page.

This widget has been developed by project partner FFZG and is tested by integration into SemLab's news sentiment website www.newssentiment.eu. In order to adequately test the widget functionality and have a standard of comparison, SemLab also integrated the LetsMT webservice though the API directly (without using the Widget). The system is tested on SemLab's website to ensure positive results in dissemination and exploitation activities through other (financial news) websites. *(source: Deliverable 5.2)*
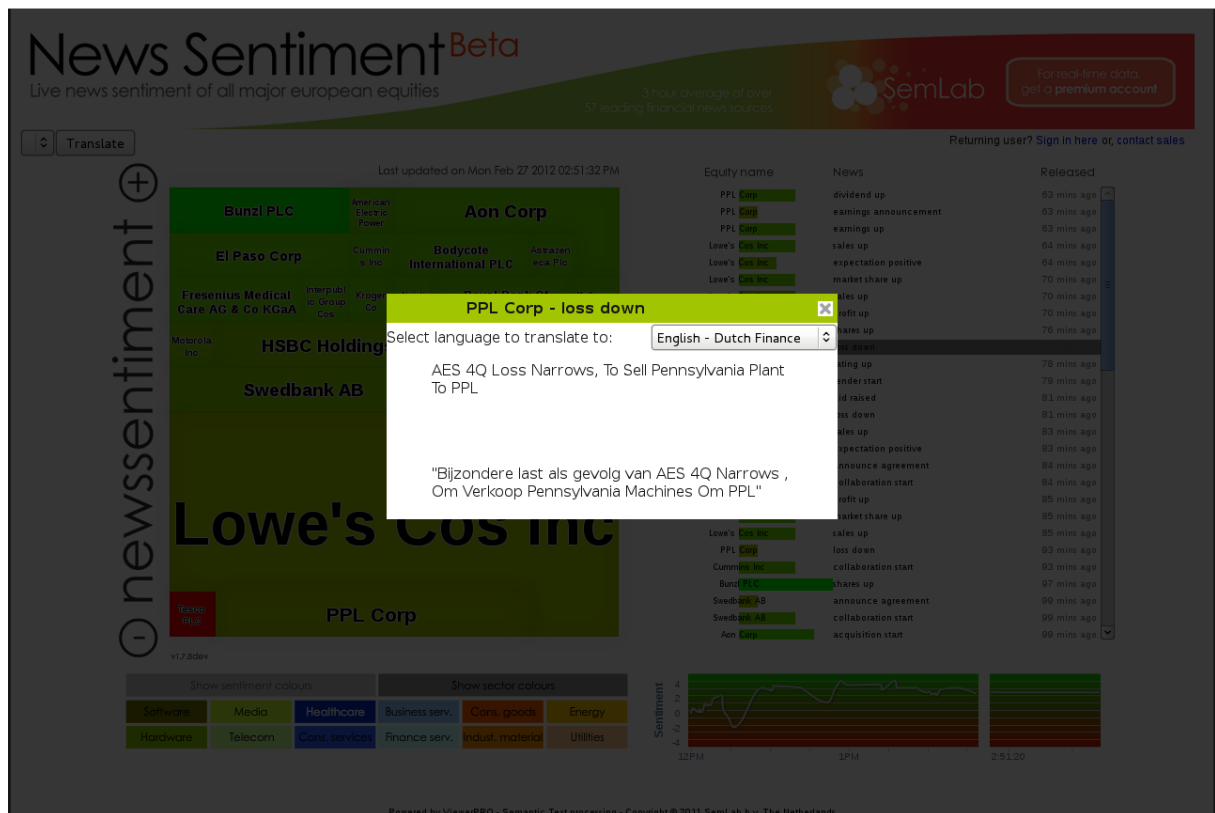

Figure 1 Screenshot of LetsMT! widget

In the above screenshot it is visible how the widget gives the translation of a selected news heading. The translation outcomes of multiple headings are all manually tested. Appendix 1 contains a list of issues found in testing.

1.2 Browser plug-in by FFZG

1.2.1 Description

The browser plug-ins are envisioned as an illustration of interfacing with the publicly available LetsMT! translation web service. However, while the web page translation widget requires integration within client websites by their developers, the browser plug-ins are planned as tools for quick and easy access to the basic translation services of the LetsMT! platform by using the user interfaces of client web browsers. Therefore, they enable the translation of websites regardless of their functionality, as long as their content is written in one of the languages supported by the LetsMT! platform.

As with the web page translation widget specification, both the conceptual and the logical design is discussed.

The functional specification also states certain non-functional requirements for the browser plug-ins:

1. The translation should be provided at rate of one sentence per second. The performance of the browser plug-in is dependent on that of the LetsMT! web-service, but it can also depend on the quality of the internet connection

2. The browser plug-ins are to be developed according to specific requirements of selected browsers. The Browser Helper Object and other necessary facilities of the Microsoft .NET platform will be developed for integration with Internet Explorer. For Mozilla Firefox, Javascript and XML will be used in concordance with the manual for the development of Firefox plug-ins. All communication will be developed over standard protocols by using the LetsMT! web service as a back-end and the facilities of selected programming languages concerning this functionality.

3. The users should authenticate implicitly, using LetsMT! web service authentication mechanism before using the SMT systems of LetsMT! platform. Authentication is necessary to identify which custom SMT systems are available for the user. If the user does not provide authentication, the default one should be used. TLS/SSL data encryption should be used to protect the user content sent over the network.

4. The browser plug-in is required to be made available to the general public from the LetsMT! public website.

As with the web page translation widget, the functionality of the browser plug-ins relies entirely on the availability and accessibility of the LetsMT! translation web service.

(source: Deliverable 5.1)

1.2.2 Functioning

In this section, usability evaluation for the LetsMT! browser plugins is provided. The user requirements for the LetsMT! browser plugins are specified in the LetsMT! functional specification (D1.2) and their implementation is presented and provided by the respective LetsMT! Deliverable (D5.1).

According to the LetsMT! functional specification (D1.2), the LetsMT! browser plugins are "envisioned as a tool for quick and easy access to the basic translation services of the

LetsMT! platform by using the user interfaces of client web browsers." The functional specification therefore states that the LetsMT! browser plugins should interface between the end user and his web browser on one side and the LetsMT! translation platform on the other (as the plugins are "implemented as software interfaces to the LetsMT! web service", "they mediate between the platform API and the client browser interface and web presentation layer"). Moreover, the browser plugins should provide access to the LetsMT! translation services in a way that is "quick", "easy" and "basic". Additionally, the functional specification states that the browser plugins "serve the purpose of demonstrating the basic capabilities of the LetsMT! platform to a wide spectrum of potential users", thus implying their basic and illustrational role within the overall functionality of the LetsMT! platform, which in turn implies the importance of their simplicity and usability rather than utilization of advanced features of the LetsMT! translation services.

The LetsMT! functional specification (D1.2) also states that "it is planned to support only Microsoft Internet Explorer and Mozilla Firefox web browsers" and thus only the browser plugins for these two web browsers are implemented (D5.1). In the following sections, a usability-oriented review of the LetsMT! Mozilla Firefox plugin for web page translation and the LetsMT! Microsoft Internet Explorer plugin for web page translation is provided separately. Each of the plugins is firstly described with respect to the LetsMT! functional specification (D1.2) and their documentation (D5.1). Usability of the plugins is then assessed for each of the plugins. The usability assessment is provided with respect to standards (ISO/TR 9241-100:2010) and (ISO/TR 16982:2002) dealing with software usability.

Included are the reviews of the installation process, configuration procedures, the usage of the plugins in actual web page translation, the process of uninstalling the tool and possibly other unclassified remarks regarding tool usage and functionality with respect to the LetsMT! functional specification (D1.2). Each of the review sections for both of the LetsMT! translation plugins –obtaining, installing, configuration, usage in translation, uninstalling and other remarks – is given with respect to the basic principles of human-centered software design and its overlap with the LetsMT! functional specification (D1.2). Human-centeredness here relates to a basic subjective notion of usage efficiency ("ease of use") and the efficiency of the learning and tool adjustment process ("ease of learning").

Following is such an assessment of the LetsMT! Microsoft Internet Explorer and Mozilla Firefox plugin for web page translation, in that particular order. Our test environment is a Lenovo ThinkPad T400s with Microsoft Windows 7 Professional Edition (64 bit) installed along with Microsoft Internet Explorer 9.0 and Mozilla Firefox 10.0.2 as browsers of choice for the testing.

1.2.3 Microsoft Internet Explorer plugin

In the accompanying documentation of the LetsMT! deliverable containing the LetsMT! browser plugins (D5.1), it is stated that "the release time of the plugin was postponed due to the announcement and appearance of the beta version of IE 9.0" and that " the consortium decided that the LetsMT! translation plugin for IE will be issued as soon as the stabile version of IE appears" due to the "utmost importance to have the LetsMT! translation plugin developed for the latest version of IE and not just the previous ones" being that the Microsoft Internet Explorer browser holds a " significant percentage" with respect to "total number of browsers used worldwide ".

Respective to the information stated in (D5.1), the LetsMT! Microsoft Internet Explorer plugin for web page translation is insufficiently documented in the current version. However, the plugin deliverable itself is now available within (D5.1) as a binary and source package and it can therefore be assessed with respect to its usability.

1.2.3.1 Obtaining and installation

At the time of conducting this usability evaluation, the LetsMT! Microsoft Internet Explorer plugin for web page translation (LetsMT! MSIE plugin further in the text) is made available to the end users through the LetsMT! public web site only by means of individual contact stating the user interest to the LetsMT! consortium through the contact page (see URL https://www.letsmt.eu/Integration.aspx). However, this is here considered to be but a minor issue, being that the overall LetsMT! platform is currently in demo version. At the time of the official release of the entire LetsMT! platform to the public, the LetsMT! MSIE plugin should be made available for download as well.

For testing purposes, the current version of the LetsMT! MSIE plugin is obtained through (https://extranet.tilde.lv/letsmt/Deliverables/Forms/AllItems.aspx) LetsMT! consortium extranet, which is not a publically available web site. The LetsMT! MSIE plugin is contained within a ZIP archive named "letsmt_d5.1_IE_plugin.zip". The archive contains the plugin source code, installation guide and the installation package itself. The source code is a Microsoft Visual Studio solution and contains the source code of the plugin written in C++ and utilizes the Browser Helper Object, as stated in (D1.2) and rephrased in (D5.1). The source code is not further discussed here, being that it is not included by the definition of usability evaluation implying usability for the end users of the browser plugins and not for prospective developers.

The documentation of the package is titled as "user documentation" and contains the sections "installation" and usage". The installation section states: "Start the installer and select the destination directory. It is recommended to leave the default installation directory. After the installation is done (it should take few seconds), you can run Internet Explorer and use the plugin".

The installation process itself is conducted precisely as described in the included documentation – quick, easy and straightforward. Figure 2 LetsMT! MSIE plugin installation provides a screenshot from the installation process. The installation results in a folder on the hard drive, in the selected folder or more specifically, a single DLL library – named "ieplugin.dll" – which is contained within that folder. The LetsMT! MSIE plugin also appears as "LetsMT" in the "Programs and Features" menu of the Microsoft Windows 7 Control Panel. It should be noted here that this short name within Control Panel should be altered in order to enable users with faster search if they would desire to uninstall the MSIE plugin. The suggested altered name could be "LetsMT! translation plugin for Internet Explorer", which is short enough to maintain its visibility in the menu, gaining informatively.
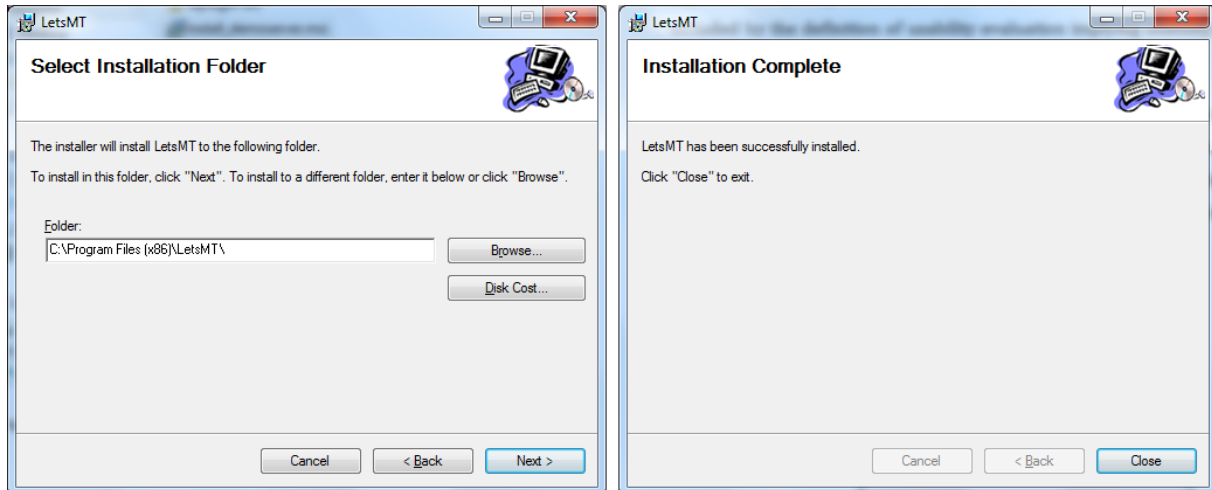
Figure 2 LetsMT! MSIE plugin installation

Once the LetsMT! MSIE browser plugin is installed, it is further accessed via the interface of Microsoft Internet Explorer.

1.2.3.2 Configuration

The LetsMT! MSIE browser plugin is configured on usage. Accordingly, options cannot be accessed via the Microsoft Internet Explorer Add-on menu. Being that the Add-on menu itself is not envisioned as to provide the users with add-on options, but simply to list additional information and options for enabling and disabling the add-ons, it is not here considered to be a LetsMT! MSIE plugin issue. Moreover, the plugin configuration itself is very simple, being that it only requires of the user to provide valid credentials for interfacing with the LetsMT! platform web service and these credentials may also be defaulted or left out entirely, depending on the implementation choices for the overall platform.

1.2.3.3 Utilization

The included documentation for the LetsMT! MSI plugin, in the usage section, states the following. (It is included here in whole, being that it is not contained in (D5.1), as previously elaborated.)

"To start using the plugin select *LetsMT translate* from *Tools* menu in Internet Explorer. Options dialog will show with list of available dictionaries and *Set credentials* button. *Set credentials* button shows another dialog with username and password fields. Valid username and password must be provided to use the service. After setting the credentials, dictionary list will reload and show available options for the given account. *Translate* button will start the page translation with selected dictionary. If there is an active selection on the page, only this selection will be translated and the result will be displayed in a separate window. Otherwise, the whole page will be translated in place (i.e. text will be replaced on the page itself). If the network connection or translation service is currently unavailable, dictionary list will show *Unavailable* and *Translate* button will be grayed out."

The MSIE plugin usage basically works as described in the documentation. A web page is loaded and the translation services accessed through the Internet Explorer Tools menu. Depending on the user choice for translation – whole page or only a specific part of the page

content, selected by mouse, similar to copy-paste procedure – the end result is a dialog box containing the translation as plain text. The process is illustrated by Figure 3 and Figure 4.
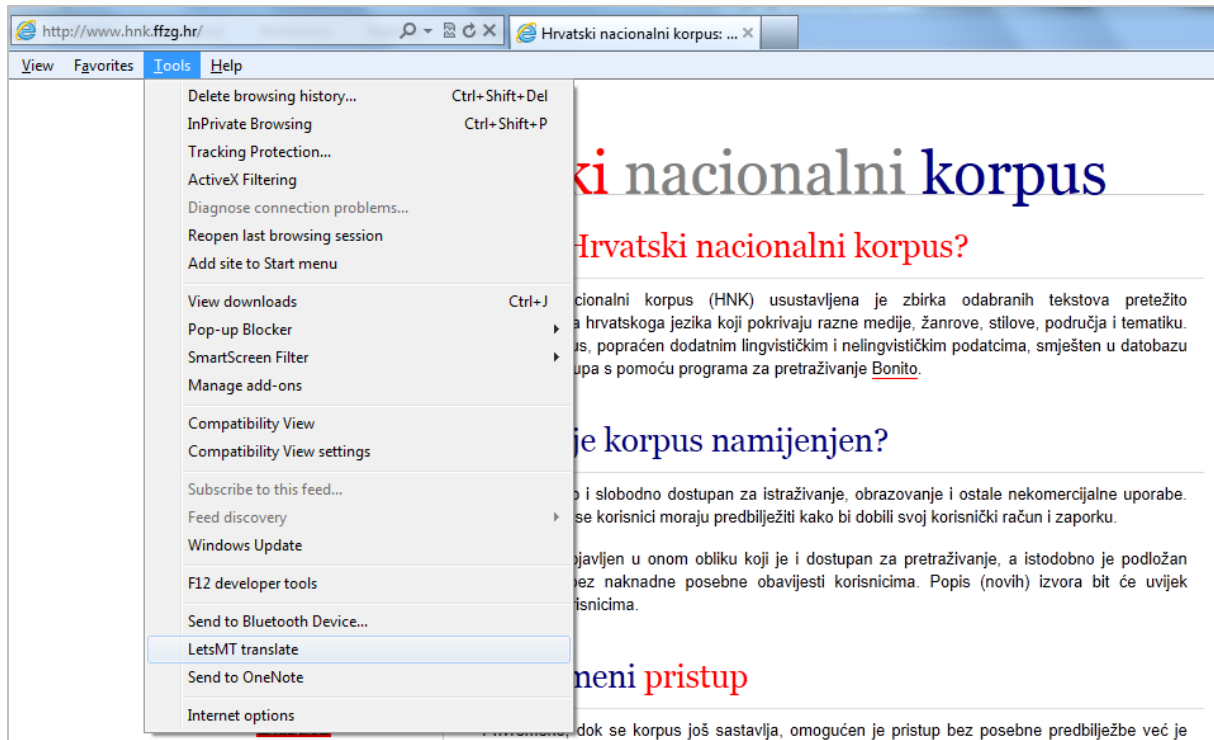

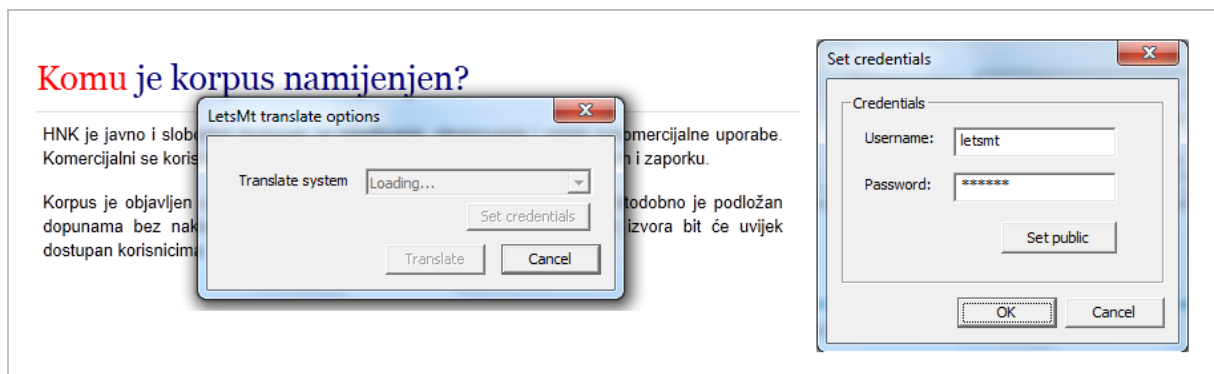**Figure 3 The MSIE plugin LetsMT translate option**

Figure 4 The translation system list and credentials option

With respect to system usability, there is a certain number of issues with this type of interaction with LetsMT! translation services, i.e. this specific approach to using the MSIE translation plugin through the provided graphical user interface from the perspective of the end user.

Firstly, accessing the plugin translation functionality through the Tools menu of the browser might be considered slower and less intuitive than, for example, right-clicking the webpage, especially for translating sections of the page content previously most probably selected precisely by using the mouse. It is considered here that, along with reaching the plugin functionality through the Tools menu, the same functionality should be offered to the end user via the right-click dropdown context menu of the browser, which could also be named "LetsMT translate".

Secondly, obtaining the translation inside a dialog box is here not considered as the best of options for providing the user with translation results. Focusing the dialog box over the browser window hides the initial content from the user and also strips the existing HTML encoding and CSS styles, thus making it harder for the user to visualize the source content from the target (i.e. translated) content. An approach suggested here, along the guidelines of the LetsMT! functional specification (D1.2) is the one with inline embedding of the translated content within the HTML source of the web page. In order to implement such functionality, the LetsMT! translation web service (D3.4) should support explicit handling of the source encoding, at least up to a point from which more specific handlers might be implemented in the plugins. Similar to other existing translation plugins and widgets, referenced in the functional specification (D1.2), the LetsMT! MSIE browser plugin should operate within the boundaries set by the source web page encoding and provide results to the users within those boundaries. Resolving this issue is considered here to be of the highest importance and priority with respect to the overall usability of the plugin.

Thirdly, feedback should be provided for the end user on the status of the translation process. Accessing the LetsMT! translation web service and obtaining a translation for the chosen source and target language is a potentially slow process and the user should be informed in its status. Currently, the browser screen freezes when requesting translation through the "Translate" button and is once again enabled only when the translation is fed to the user through the translation dialog.

Finally, a technical issue with interacting the LetsMT! MSIE browser plugin and the LetsMT! translation service (SOAP API) has also emerged during this test. This issue is mentioned in the plugin documentation, in the "Troubleshooting" section, where it is stated that "since LetsMT! server certificate is not properly signed users will currently experience difficulty in translating using the plugin. Microsoft provides instructions to how certificate can be imported into Internet Explorer". This issue is no longer relevant, since the LetsMT! platform web site (and respective web services) are now digitally signed and certified by a trusted authority. However, with regards to the interaction between the plugin and the translation web service, it should be noted that the "Set credentials" menu of the plugin, as illustrated by figure 3, does not contain an option for explicitly referencing the LetsMT! translation service URL. Being that a change within the LetsMT! platform might eventually appear, changing the URL of the service, it is considered here that the user should be provided with an option for changing that URL, since it is currently hard-coded within the LetsMT! MSIE plugin source (in source folder, file "LetsMTServerConnection.cpp", line 22).

1.2.3.4 Uninstalling

As previously mentioned, the LetsMT! MSIE browser plugin installation data can be reached through the Windows 7 Control Panel "Programs and Features" section, where it is listed as "LetsMT", as illustrated in Figure 6 and also through the Microsoft Internet Explorer 9.0 "Manage Add-ons" menu option, as shown in Figure 5. The second option, specific for Microsoft Internet Explorer, provides the user only with the information on specific plugins, and not with uninstall options. Therefore, through this menu, the user may only enable and disable the LetsMT! MSIE browser plugin and view its basic information. The uninstall process may only be initiated through the "Programs and Features" list of the Windows 7 Control Panel. The uninstall process itself is very simple and implemented through the basic

uninstall interface provided by the Windows 7 operating system. The same general remark regarding the plugin naming also applies here. The names for the plugin should be consistent across the various interfaces through which it might be accessed. For example, in Figure 5 (the Internet Explorer add-ons menu), the plugin is called "LetsMT translate" and in Figure 6 (Control Panel Programs and Features), its name is "LetsMT". Once again, it is suggested here that these should be changed to "LetsMT! translation plugin for Internet Explorer".
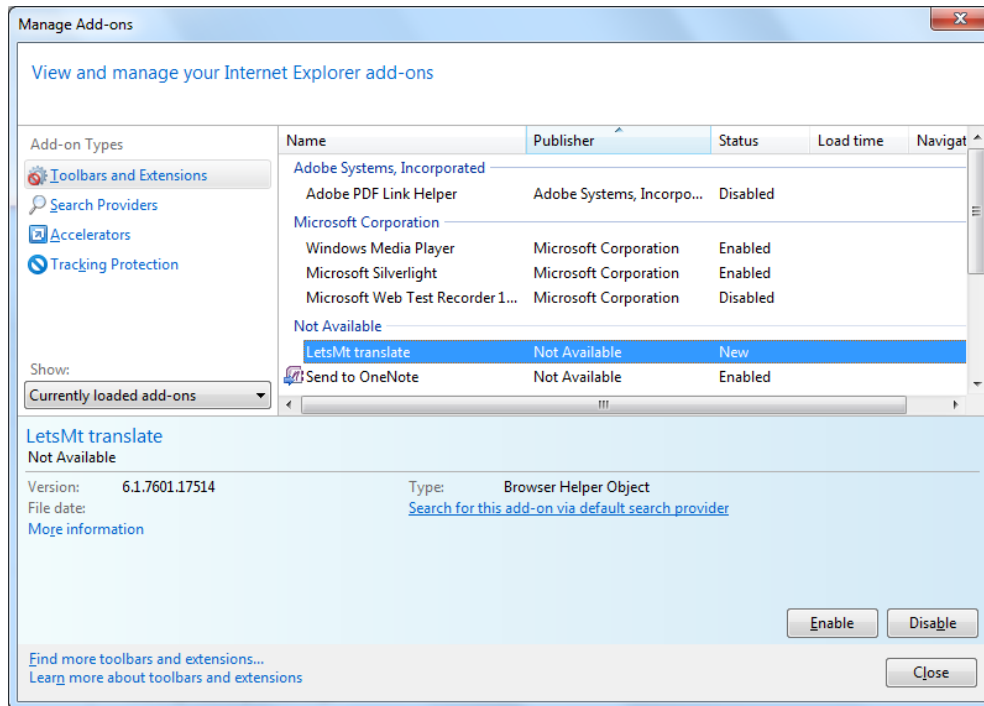

Figure 5 The MSIE plugin in Manage Add-ons menu

Figure 6 The MSIE plugin in Control Panel Programs and Features

1.2.3.5 List of issues

List of open issues encountered while testing the LetsMT! MSIE browser plugin with respect to the usability for the end user is summarized in Appendix 1 of this deliverable - List of issues, Internet Explorer plug-in. The list is given in the order of encountering the issues and describing them in the previous section. A unique code for each issue is given, along with a short description and the priority of resolution from the standpoint of usability.

1.2.4 Mozilla Firefox plugin

The LetsMT! deliverable regarding browser plugins (D5.1) differentiates the plugin terminology as used throughout the LetsMT! documentation from the one used by Mozilla developers. It states that "Mozilla Firefox uses a slightly different terminology; the browser

can be extended by installing add-ons. Add-on can be either a theme, an extension or a plugin. Plugins are usually add-ons that enable browsers to process new types of content (e.g. QuickTime, Java, Flash), so LetsMT! functionality is provided by Firefox extension." Thus, the LetsMT! Mozilla Firefox browser plugin should officially be named as the LetsMT! Mozilla Firefox browser extension, not to introduce confusion in the end user when obtaining the module.

The module is well-documented within the deliverable (D5.1), its documentation including descriptions for its installation, usage and setting the preferences. Each of these is discussed in the following sections.

1.2.4.1 Obtaining and installation

The remarks with respect to obtaining the package for the LetsMT! Mozilla Firefox extension (LetsMT! FF extension further in the text) are virtually identical to the ones for the LetsMT! MSIE plugin: it is currently unavailable through the LetsMT! public website and was obtained through the restricted consortium website for purposes of usability testing. Additional to the remarks made for the MSIE plugin, the FF extension should also be made available through the official channel for Mozilla Firefox add-ons (URL https://addons.mozilla.org/en-US/firefox/) in order to improve its future overall visibility and implicitly the visibility of the entire LetsMT! platform.

On the restricted LetsMT! web site repository, the FF extension is contained within a single file "firefox_plugin.xpi", which is in fact a ZIP archive, packed and named as instructed by Mozilla Firefox extension manuals. The development process and package content is also documented in the respective deliverable (D5.1).

In the documentation, the installation process is described as follows. "The LetsMT! Firefox extension is a single file with .xpi extension. User installs it by double-clicking the file and later clicking on the "Install Now" button in Firefox." The installation may also be initiated by dragging and dropping the extension file to the Add-ons menu within Mozilla Firefox. The browser is restarted and the extension is made available. However, the package is hardcoded for compatibility with certain versions of Mozilla Firefox, as indicated in the "install.rdf" file of the source package. In order to make it compatible with current versions of Firefox, the user must alter the "install.rdf" file within the package and change the respective fields dealing with Firefox version. In the documentation, it is stated that "the extension is tested for Firefox 3-3.6 and the latest 4.0 beta. Because in the future Mozilla Firefox might break backward compatibility, we have conservatively specified that it is compatible with versions 3.0-4.0". Current version of Mozilla Firefox is at this time 10.*, due to the upgrades and altered naming convention. It is thus recommended here that the version should be hard-coded to a higher number in order to ensure compatibility with the latest versions of the browser and also backward compatibility. The end user should not be explicitly instructed to alter the source of the extension manually.

The current version of the LetsMT! Mozilla Firefox translation extension seems to be incompatible with the current version of Mozilla Firefox (10.*). It is explicitly stated in the documentation (D5.1) that "the extension is tested for Firefox 3-3.6 and the latest 4.0 beta. Because in the future Mozilla Firefox might break backward compatibility, we have conservatively specified that it is compatible with versions 3.0-4.0". The LetsMT! FF extension should by all means be altered to support the latest versions of the browser,

being that versions 3.0-4.0 might already be considered as obsolete by the majority of prospective end users of the FF extension. Due to this issue, the testing platform was altered to run Mozilla Firefox browser, version 3.6, in order to successfully implement the usability testing procedure.

1.2.4.2 Configuration

Once the LetsMT! FF extension is installed, it is reachable through the Add-ons menu of Mozilla Firefox. Similar to the LetsMT! MSIE browser plugin, the end user may alter only the credentials. Thus, all the remarks stated for the MSIE plugin with regards to configurability apply also for the LetsMT! FF extension. The configuration interface is illustrated in Figure 7.
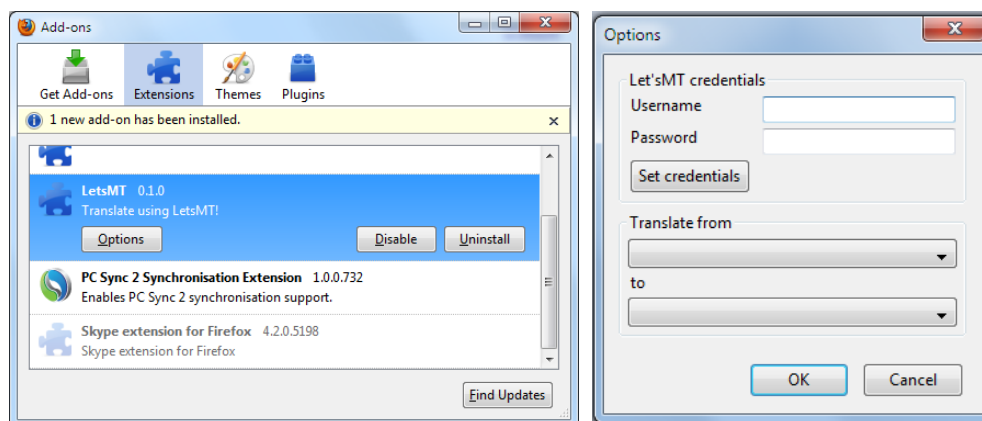

Figure 7 The LetsMT! FF extension list and preferences

Additionally, the LetsMT! FF extension naming convention suffers from the same issues as the LetsMT! MSIE plugin and the same remarks apply once again – the naming of the FF extension should be uniformly applied by using the name "LetsMT! translation extension for Mozilla Firefox".

Differing from the MSIE plugin, the FF extension has an additional translation option in the preferences menu. It is closely related to the usage of the extension in actual translation and it is thus commented in the following section. However, it is noted here that the translation option of the preferences menu does not induce any issues with respect to extension usability.

1.2.4.3 Utilization

The documentation (D5.1) states the following with respect to the actual usage of the extension in web page translation: "Most of the functionality is available via a context menu that can be accessed by a right-click (or control-click on a Mac) on the web page. If a part of the web page (e.g. some text) is selected, LetsMT! extension adds a new menu entry LetsMT! translate selection that has a submenu listing all the available systems (pairs of source and target languages). The list of language pairs is obtained by contacting LetsMT! service. If there is no selection, LetsMT! extension adds a new menu entry LetsMT! translate page that also offers the same list of language pairs."

This type of usage is here considered to be much more appropriate with respect to the overall ease of use for the translation extension, especially when compared with the LetsMT! MSIE browser plugin.

The FF extension can be used to translate the entire page and only a subsection of the page, selected by the user. Both translation functions are accessed through the same uniform and intuitive right click interface. However, similar to the MSIE plugin, translation of selected sections is provided to the end user in the form of a text box showing the plain text version of the selected content. The same request with respect to the usability therefore applies here – the LetsMT! FF extension should respect the content style boundaries when providing results to the end users. Illustration of LetsMT! extension usage is given by Figure 8. Feedback on the status of translation is also missing in the LetsMT! FF extension, as for the MSIE plugin.
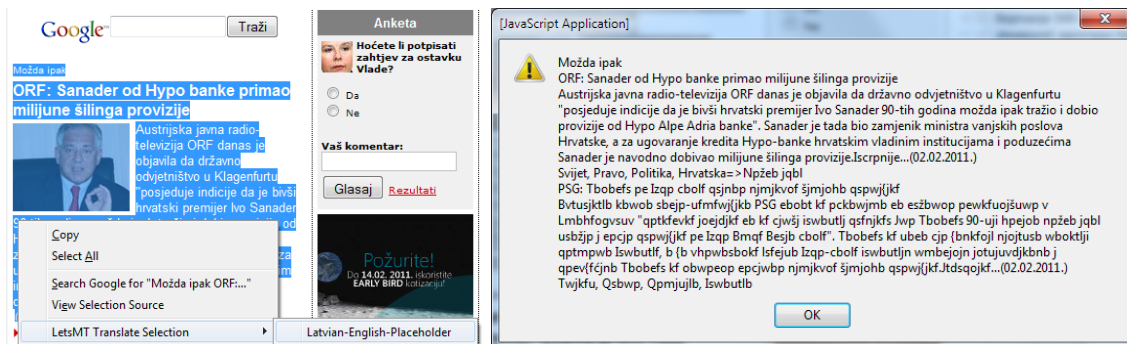

Figure 8 Using the LetsMT! FF extension for web page translation

1.2.4.4 Uninstalling

The uninstall procedure for the LetsMT! FF extension is run through the extension interface of the browser and it is simple and straightforward. This interface enables the user to install, alter, disable and uninstall the extension. The naming convention remark also applies for this interface.

1.2.4.5 List of issues

List of open issues encountered while testing the LetsMT! FF browser extension with respect to the usability for the end user is summarized in Appendix 1 of this deliverable - List of issues, Firefox plug-in. The list is given in the order of encountering the issues and describing them in the previous section. Similar to the list of LetsMT! MSIE plugin issues, a unique code for each of the issues is given, along with its short description and the priority of resolution from the standpoint of module usability.

1.2.5 Conclusions

Previous sections described a basic usability evaluation of the LetsMT! end user browser plugins (namely, Microsoft Internet Explorer plugin and Mozilla Firefox extension) for web page translation by interacting with the LetsMT! platform translation web service. Issues are described in the respective sections and listed by the respective plug-ins in Appendix 1 (Table 6 and Table 7). Common issues include providing user with adequate feedback during the translation process, making the modules publically available and informing the naming conventions, enriching the configuration options and especially respecting the style and markup boundaries of the translated content.

2 Manual evaluation of Translation Quality

2.1 Description of methodology

Evaluating the output of machine translation systems has been subject of constant debate. Preferably, evaluation should be carried out manually by human judges who rate the output according to adequacy and fluency (Koehn, 2010, sect. 8.1). Variations in human judgment can be balanced out by supplementing the subjective evaluation with a task-based evaluation. Performance could be measured on e.g. a post-editing task or an understanding task (Koehn, 2010, sect.8.4.1 and 8.4.2). However, since human evaluation is expensive and time-consuming, it is not suitable for monitoring everyday progress during system development.

It has been shown that human judgments were more consistent on a ranking task chosen over a subjective rating of adequacy and fluency (Callison-Burch et al., 2007). But in this experiment we chose rating of adequacy and fluency, because we have only one system per language pair to evaluate.

The most widely used methodology when manually evaluating MT is to assign values from two five point scales representing fluency and adequacy (Callison-Burch et al., 2007, section 3.1.). These scales were developed for the annual NIST Machine Translation Evaluation Workshop by the Linguistics Data Consortium (LDC, 2005).

The following measurements are applied to evaluate the translations in this evaluation:

2.1.1 Adequacy

Sentences have to satisfy a requirement or have to meet a need satisfactorily. The determination as to whether the scope and concept of a planned operation are sufficient to accomplish the task assigned. The adequacy is mainly tested by looking at the accuracy of the sentence: the quality of having no error. The five point scale for adequacy indicates how much of the meaning expressed in the reference translation is also expressed in a hypothesis translation:

> 0 = All
>
> 1 = Most
>
> 2 = Much
>
> 3 = None

2.1.2 Fluency

The second five point scale indicates how fluent the translation is. The translated sentences will be tested on fluency: how fluent and smooth a sentence is translated. The viscosity can be measured when reading the sentence, if you have to read it over and over to understand what it says, it is not correctly translated. Word order is a useful measure to value the fluency of the sentence.

When translating into target language the values correspond to:

> 0 = Flawless target language

1 = Good target language

2 = Non-native target language

3 = Disfluent target language

Separate scales for fluency and adequacy were developed under the assumption that a translation might be disfluent but contain all the information from the source.

2.2 SMT Systems

Evaluation was carried out for SMT engines with following characteristics:

Table 1 Evaluated SMT engines

No.	Language	BLEU	NIST	Parallel training data (segments)
1.	English - Dutch	0.529	8.817	1,125,637
2.	English - Danish	0.520	8.664	420,897
3.	English - Croatian	0.219	5.850	58,814

For more detailed information about trained SMT engines for finance domain see deliverable – "D5.4 Automatic evaluation report of business and financial news SMT".

2.3 Used evaluation tool

When translating 1000 sentences, it is helpful to have an application which supports the evaluation process. We used Absolute Evaluation tool by Linguatec GmbH[1]. The evaluation tool was originally developed by Linguatec GmbH to use for MT evaluation in EASTIN-CL project[2].

This application, makes use of 4 multiple choice options to value the adequacy and 4 multiple choice options to value the fluency of the sentence. These outcomes are saved and can be transformed into useful data for Excel statistics.

---

[1] http://www.linguatec.net/; Absolute Evaluation tool by Linguatec GmbH. Used with permission.
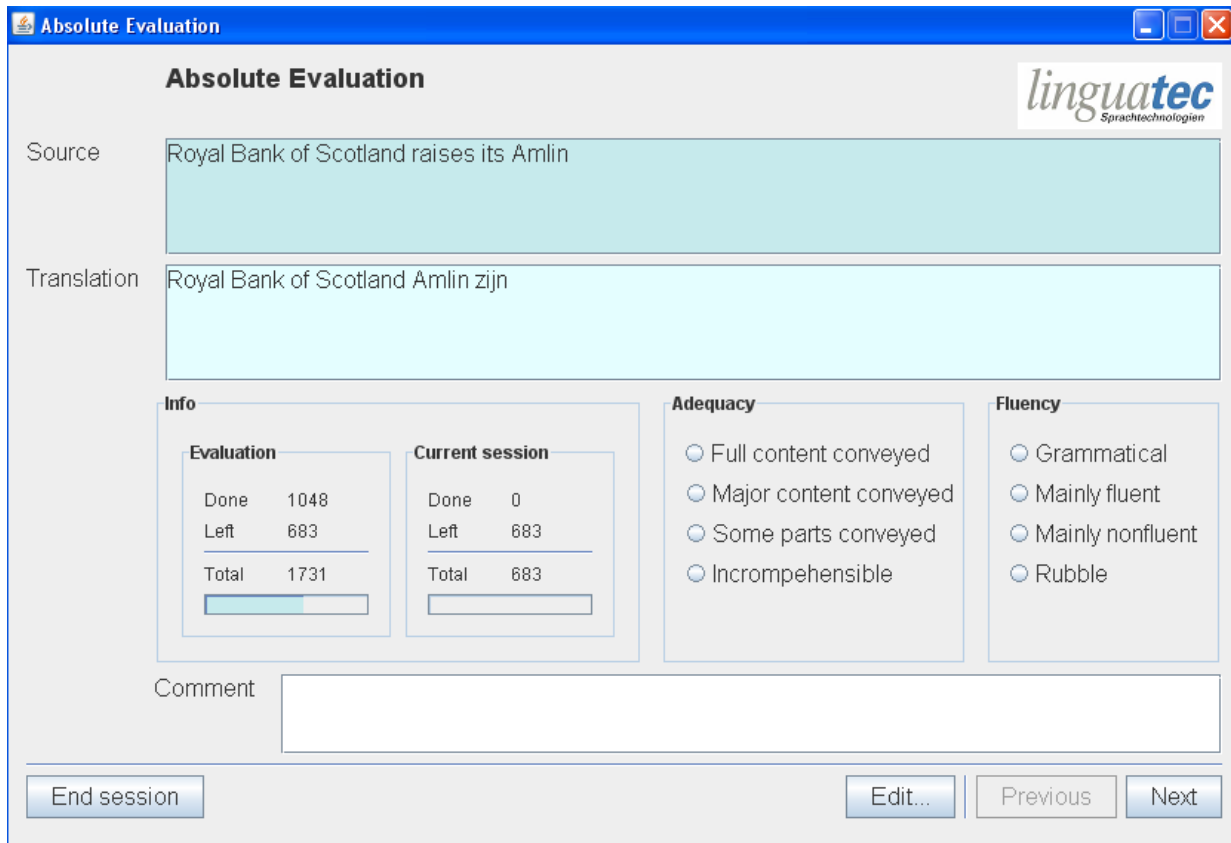
[2] http://www.eastin-cl.eu/home

Figure 9. Screenshot of used tool

This 'absolute evaluation' tool has been used by all translators and has been of great use.

The only problem was that the comments in the 'comment box' (at the bottom) are not saved and therefore do not appear in the final Excel sheets. This is unfortunate because adaptations were to be made quicker with the help of this tool.

2.4 Initial evaluation results

For the evaluation, over 1000 automatically translated sentences had to be checked manually on translation accuracy and fluency. Below one will find the outcome, recommendations and final feedback.

Table 2 Accuracy and Fluency values

Accuracy	Fluency	Value
Full content conveyed	Grammatical	0
Mayor content conveyed	Mainly fluent	1
Some parts conveyed	Mainly non-fluent	2
Incomprehensible	Rubble	3

The following tables contain results on the manual quality evaluation.

Table 3 EN-NL

Value	Accuracy		Fluency	
	Number of sentences	**%**	**Number of sentences**	**%**
0	51	3%	63	4%
1	270	17%	243	16%
2	405	26%	377	24%
3	826	53%	869	56%
Sum	1552		1552	
Average	2.3		2.3	

Table 4 EN-DA

Value	Accuracy		Fluency	
	Number of sentences	**%**	**Number of sentences**	**%**
0	170	11%	176	11%
1	195	12%	138	9%
2	366	23%	235	15%
3	887	55%	1069	66%
Sum	1618		1618	
Average	2.2		2.4	

Table 5 EN-HR

Value	Accuracy		Fluency	
	Number of sentences	**%**	**Number of sentences**	**%**
0	99	6%	65	4%
1	194	12%	78	5%
2	466	28%	249	15%
3	881	54%	1248	76%
Sum	1640		1640	
Average	2.3		2.6	

In the tables shown above, the average score is between 2 and 3 which stands for incomprehensible adequacy and rubble fluency, or in other words; rubbish translation. There are 1552 Dutch, 1618 Danish and 1640 Croatian evaluated sentences.

When looking at the translated sentences and their scores, it is clear that the same English sentences cause trouble for the SMT system. This seems to be caused by abbreviations, capital letters, specific terminology and jargon which the machine doesn't notice, nor translates. Apart from this, it appears that many of the sentences are news headings which are often summarised quotes or eye-catchers, but they are not – in the first place – meant to explain and clarify the situation. They are meant to catch the attention of the reader to read the rest of the article where the situation is further explained.

Abbreviations are a problem because the SMT translation system doesn't translate these. For the reader it doesn't cause a problem, because with help of the context and some background knowledge, everyone understands what is meant.

The overall score of the 3 independent evaluations are: for adequacy a 2.27 and for fluency 2.44, which stands for very bad translation quality.

Quite a few sentences, see several examples below, are of very bad quality in the source language.

- Johnson & Johnson, Merck Settle Arbitration Proceeding >JNJ
- Merck Still Sees 2011 Rev Up Low- To Mid-Single Digits >MRK
- Merck & Co Backs 2011 View Of Adj EPS $3.64-Adj EPS $3.76 >MRK

These sentences are a clear example of a bad input. These do not make sense when read in the source language and therefore will neither make sense in a translated version.

3 Functional evaluation of Translation Quality

The traditional evaluation of a translation system is to compare it to a human experts translation and note the discrepancies. This is precisely what is done in the previous part of this deliverable. As can be seen from the results, the current state of the systems still leave sufficient room for improvement with respect to imitating a human translator. Specifically the construction of grammatically correct sentences appears to be problematic, and this poses a significant obstacle for completely understanding the sentence.

However, there may be applications that can benefit from a less optimal translation and still obtain useful results. This hypothesis can be tested using SemLab's ViewerPro semantic analysis platform. This platform is designed to match the words of the sentence with a set of domain specific events of interest. Since this system is targeted at financial analysis, it is designed to do this as quickly as possible (typically within 1 ms per sentence) and therefore uses little grammatical information, but mainly focuses on key words.

The following sections describe the evaluation method and discuss its results. First, to support the understanding of the results, a more detailed description of the ViewerPro event detection methods is provided.

3.1 ViewerPro event detection

ViewerPro is used to transform natural language news messages to computer readable news events. This procedure is based on a domain specific repository of knowledge – The Knowledge base, which is maintained as ontologies. The ontologies in the Knowledge base contain the concepts and their lexical representations relevant to a specific domain. It contains, for instance, a Company ontology describing all companies of interest to the user in terms of their synonyms, ISIN codes, ticker symbols etc., but also a terms ontology is present, containing specific concepts relevant to, for instance, the domain of equity based finance.

3.1.1 First step – Concept matching

As a news message is received by ViewerPro, its content is parsed for the presence of references to concepts in the various ontologies of the Knowledge base (see figure below).
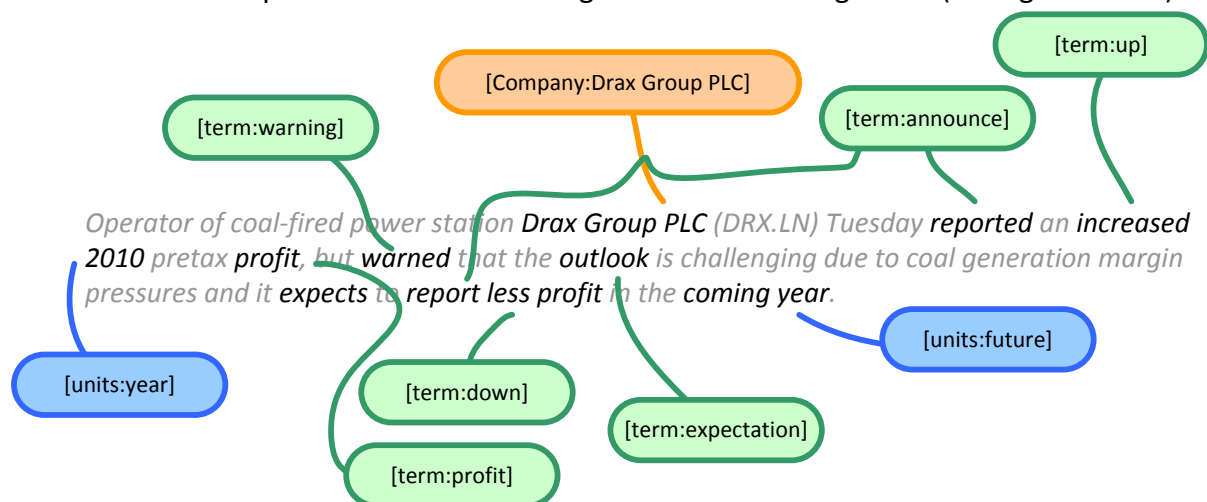

Figure 10. Knowledge base

The result of this operation is a list of all relevant concepts in a text. In this example, the list would be:

[Company:Drax Group PLC] [term:announce] [term:up] [units:year] [term:profit] [term:warning] [term:expectation] [term:expectation] [term:announce] [term:down] [term:profit] [units:future]

3.1.2 Second step – Concept chunking

Using a proprietary heuristic including semantical, morphological, syntactical and typographical inputs, the list of concepts is segmented into chunks of related concepts. In this example, the following three chunks would result:

1. [Company:Drax Group PLC] [term:announce] [term:up] [units:year] [term:profit]
2. [Company:Drax Group PLC] [term:warning] [term:expectation]
3. [Company:Drax Group PLC] [term:expectation] [term:announce] [term:down] [term:profit] [units:future]

3.1.3 Third step – Pattern matching

The output of ViewerPro primarily consists of Events. Events are pre-defined semantic concepts describing important content of news messages in the selected domain – in this case equity based finance. Examples of Events in this domain are: "Company announces dividend payment" or "Company credit rating is upgraded" etc.

All Events are defined by a set of patterns. The patterns defining the Event "Company announces an increase in profitability" or in short "Company profit up" is defined by the following patterns:

1. <ICB Industry> <term:announce> <term:up> <term:profit>
2. <ICB Industry> <term:up> <term:profit> <units:periode>
3. <ICB Industry> <term:up> <term:profit> <units:valuta>
4. <ICB Industry> <term:up> <term:profit> <units:percentage>

As now both the Events and the news messages are described in terms of the ontology, the occurrence of Events in the message can easily be determined by matching the patterns on the chunks defined in step 2.

It will not be surprising that the first chunk matches the first event pattern, resulting in the creation of an instance of the "Company profit up" Event (as graphically depicted in the figure below).
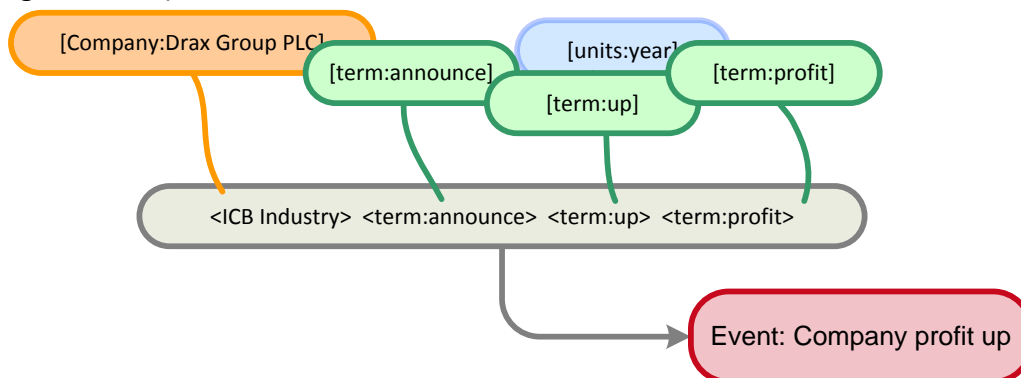

Figure 11. Event: Company profit up

3.2   Evaluation method

Critical to the evaluation procedure is the fact that the ViewerPro system regarding the domain of corporate finance is only available in the English language. The possibility to translate other language documents automatically into a "functional English" which is of acceptable accuracy regarding the key concepts in a sentence would significantly expand the reach of the ViewerPro product.

To evaluate this proposition, the following procedure was used:

- A translator was constructed using the LetsMT website. This itself is a useful evaluation as it is a clear example where a new translator is build and used in a third party software platform for a practical purpose. The following steps were taken:
    - o Step 1: Entered a name for the system, selected the source and destination language and set the permission to public as all available systems are also public. Didn't use the "add field" button.
    - o Step 2 (Parallel corpora) Selected "SemLab Business News 2 as the corpus to use.
    - o Step 3 (Monolingual corpora) accepted the default selection (SemLab Business News 2), none of the other corpora seemed relevant
    - o Step 4 (Advanced options (optional)) Skipped this step
    - o Step 5 (Training) started the training of the system with a click of a button.
    - o The total translator training took about 10 hours to complete.
- A dataset of Dutch financial news sources was obtained using available web resources. The following RSS feeds were used:
    - o http://beurs.startpagina.nl/rss.xml
    - o http://www.nrc.nl/nieuws/categorie/economie/rss.php
    - o http://www.nu.nl/feeds/rss/beurs.rss
    - o http://www.telegraaf.nl/rss/dft.xml
    - o http://www.trouw.nl/nieuws/economie/rss.xml
    - o http://fd.nl/?widget=rssfeed&view=feed&contentId=33810
    - o http://www.verzekeringen-online.nl/rss/rss.xml
    - o http://www.volkskrant.nl/economie/rss.xml
    - o http://www.ad.nl/nieuws/economie/rss.xml

  The time period used was December 1 2011 until March 22 2012. The news was passed through a filter, only allowing news items mentioning indexed companies to pass. In this period and for these 9 RSS feeds 14,445 news items passed the filter. The items were downloaded, translated using the http access of the LetsMT! server and parsed using our ViewerPro platform, resulting in 1401 events.
- The resulting dataset was analyzed by SemLab's default corporate finance configuration. This contains about 250 of the most frequently occurring events described in financial news, relevant for company equity evaluation. The company ontology consists of 1000 listed organizations and the terms ontology consists of about 800 entries, including synonyms.
- The resulting events were compared to the original Dutch language news item and their appropriateness was listed on a binary scale. i.e. an event was either classed as correct or incorrectly reflecting the content of the news message.
- To facilitate a direct comparison of the event extraction to the translation quality, also the English machine translation was evaluated. For this also a binary scale was used. The

translation was classified as either adequate or inadequate, roughly comparing to class 1 "Mayor content conveyed" or below as used in the previous section. A total of 500 events were evaluated by hand by a native Dutch speaker and a domain expert in corporate finance.

3.3 Results and discussion

The evaluation yielded the following results (see also **Figure 13**).

- 15% of the sentences were both inadequately translated and did not yield a correct event
- 4% of the sentences were adequately translated but did not yield a correct event
- 24% of the sentences were adequately translated and yielded a correct event
- 57% of the sentences were inadequately translated, but still yielded a correct event
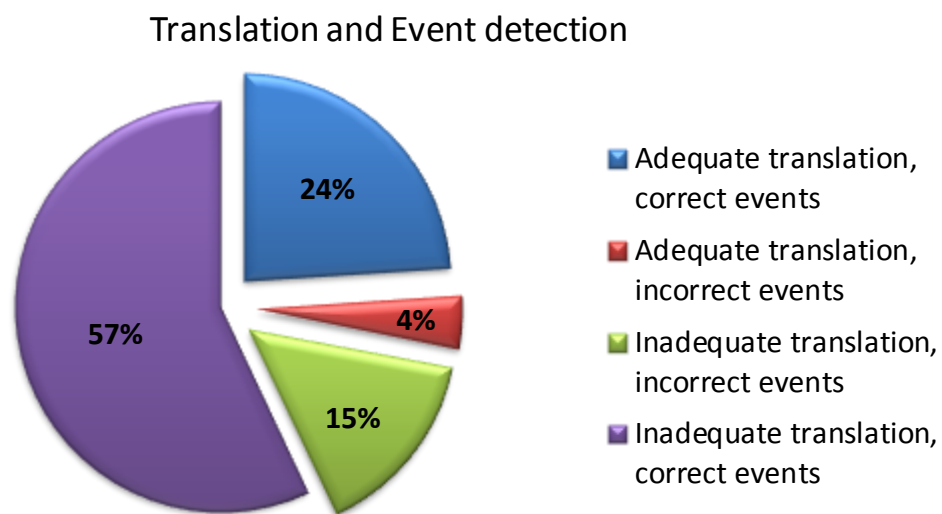
Translation and Event detection

Figure 12: Distribution of results regarding adequacy of the translation and correctness of the detected events.

Rearranging these results according to the respective quality of the machine translations and the event detection yield the following graphs:
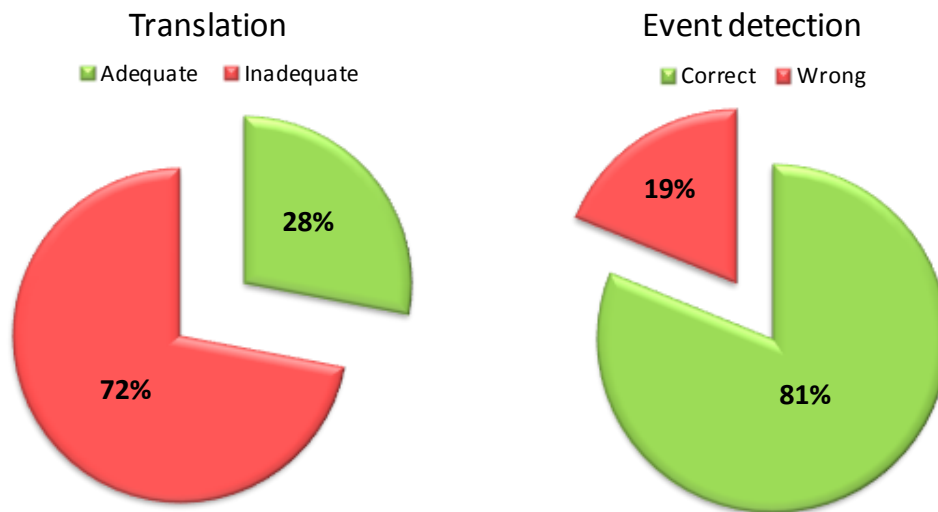
Translation

Event detection

Figure 13: Percentage of adequate translations (left) vs percentage of correct ViewerPro events (right) over the dataset of 500 random chosen messages.

These findings indicate that despite the rather poor translation quality of less than 30% adequately translated sentences, still in 80% of these sentences the correct ViewerPro events were detected. Accepting a rule of thumb that any natural language processing system needs to be about 75% accurate to begin providing useful results, these findings clearly illustrate the claim that a sub optimal translation can still be of use for further processing by ViewerPro.

To further investigate these findings, the following figure illustrates the typical results of the translation and event detection.

> *Original:* De omzet van Wolters Kluwer steeg vorig jaar met 1 procent tot 3,35 …
>
> *MT:* Wolters Kluwer ' s revenues increased during the year by 1 per cent and 3.35...
>
> *Correct:* Last year Wolters Kluwer revenue increased by 1 percent to 3,35 …
>
> *ViewerPro event:* Revenue up
>
> *Original:* TNT Express heeft over 2011 een nettoverlies gedraaid van € 270 mln.
>
> *MT:* TNT Express is a particularly good performance within Ordina posted a loss for 2010 of €270 million.
>
> *Correct:* TNT Express has posted a net loss of € 270 million over 2011.
>
> *ViewerPro event:* Loss announcement
>
> *Original:* ING maakt een einde aan het gratis storten van contant geld.
>
> *MT:* ING, an end to reorganize the value of money.
>
> *Correct:* ING ends the free deposit of cash money.
>
> *ViewerPro event:* none

Figure 14: Typical examples of the translation and event detection evaluation procedure, listing (top to bottom) the original text, the machine translation by the Lets MT system, a human generated translation and the extracted events by ViewerPro.

The first two examples illustrate the finding that the translation system is capable of translating the key concepts in the text, but fails at the finer details and the construction of a grammatically correct result. Fortunately, the ViewerPro system mainly uses the occurrence of the key concepts for its event detection. For this reason, the event detection is far better than might be expected from the translation quality findings.

To further analyses the quality of the event detection over the translated text, the following figure depicts the fraction of correct events per event sector. For this graph only the event categories of which 10 events or more have occurred in the dataset.

As not all events occur with the same frequency in any dataset, the results might be skewed by a high accuracy over the single most occurring events. The figure however indicates that the event detection on the translated text is not affected by specific events. Less than 20% (3 out of 16) of the event categories show an accuracy below the threshold value of 75%.
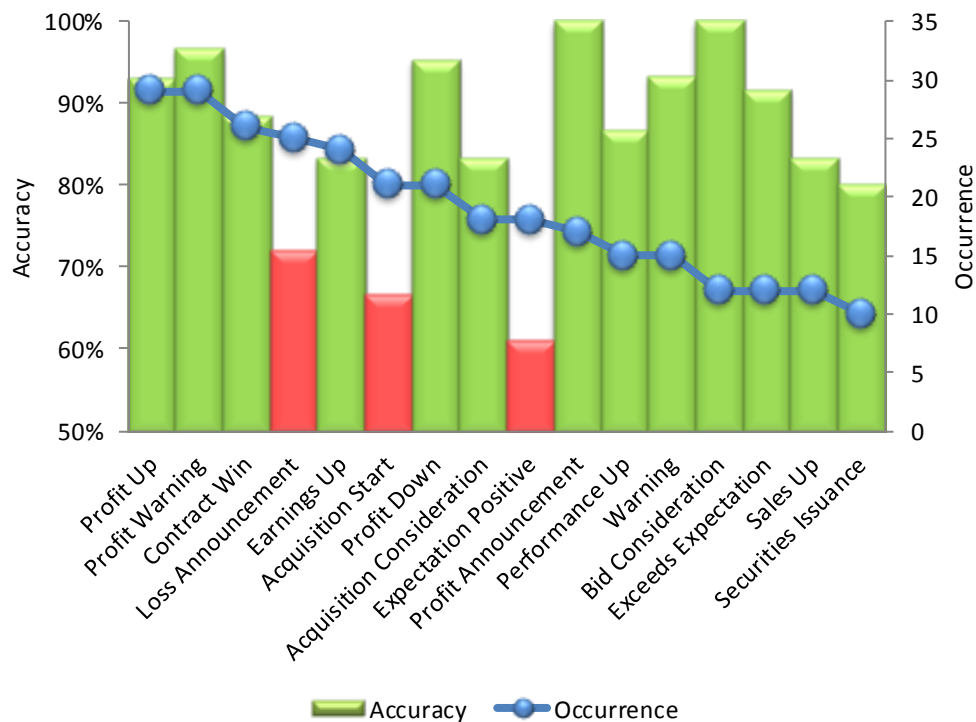

Figure 15: Distribution of correct event percentages over the various event categories compared to their occurrence in the dataset. Only categories for which at least 10 events were detected are included.

3.4 Conclusion functional evaluation

The current state of the LetsMT! platform allows a non-expert to construct a translation engine, and embed it into an external platform (in this case ViewerPro platform for semantic processing) without difficulty or extensive support. The resulting translations, however poor judging by conventional metrics, allow the ViewerPro platform to correctly identify the events in over 80% of the cases. This makes the LetsMT! system immediately useful for the extension of ViewerPro into processing non-English data sources.

4 Final feedback

During the evaluation, it was noticed that there were many duplicates in the list. This was very inconvenient for the evaluation. Though the idea behind the news sentiment was that the news is taken from 57 news sources with the risk of having the same newsheadings multiple times. This is intentionally done with the idea of 'valuing' newsitems: the more often it is mentioned in the news, the more important the newsitem becomes in the news sentiment. Though for the evaluation of the translated sentences, this caused extra work.

Another main problem with many headings was the specific professional language, or jargon used. The input (often unstructured information) was of poor quality, so translating was made extremely difficult and the outcome of the translation therefore was even worse. One of the problems are abbreviations, numbers, monetary values and abbreviations which is not being interpreted correctly by the computer and so not at all being translated.

According to Lene Offersgaard, the headings are in 'upper case'. "This seems to work badly for the machine translation system – giving bad evaluation scores."

Another problem faced by machine translation is the incorrect translation of verbs. In a sentence where there are for example 4 words, 3 of them are nouns, but the 4th is a verb, the verb (the 4th word) is often a key word. When all 3 nouns are correctly translated, but the verb isn't, it still very hard to understand the main meaning of the sentence.

In different occasions there is a standard expression in English used which will be translated literally into another language by the machine. This however, though it might be a correct translation, means the intentional message might be modified or changed because of the literal translation.

5 Conclusions

The overall translation quality of the first set of translated sentences is very poor. These evaluations come from 3 independent, different testing panels but generate the same results. The majority of the sentences was ranked with the worst score possible in combination with a 'worse than moderate' score. Only between 5 and 11% of the sentences were positively evaluated.

However, the functional evaluation of the translated text show a much more favourable result. The current state of the LetsMT! platform allows a non-expert to construct a translation engine, and embed it into an external platform without difficulty or extensive support. The resulting translations allow the ViewerPro platform to correctly identify the events in over 80% of the cases. This makes the LetsMT! system immediately useful for the extension of ViewerPro into processing non-English data sources.

The evaluation of the LetsMT! plug-in is tested by FFZG in both Mozilla Firefox and Internet Explorer. The installation process itself is conducted precisely as described in the included documentation – quick, easy and straightforward. The plugin configuration appears to be simple, though there is a certain number of issues by naming the FF extension. For both evaluations, done with Mozilla Firefox and Internet Explorer, FFZG made a list with issues which have to be resolved before making the LetsMT! plug-in available for public use.

Recommendations

Apart from the above mentioned problems, improvements could be easily made by implementing the following tips:

1. Delete all the duplicate sentences from the data lists (even though they were not deleted on purpose this time), it makes evaluating so much easier

2. Try to sift incomprehensible words, standardised quotes or specific mentionings. They are not understood by the SMT machines

References

Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., and Schroeder, J. (2007). (Meta-) evaluation of machine translation. In Proceedings of the Second Workshop on Statistical Machine Translation, pages 136–158. Association for Computational Linguistics: http://www.cs.jhu.edu/~ccb/publications/meta-evaluation-of-machine-tranlsion.pdf

Koehn, P. (2010). Statistical Machine Translation. Cambridge University Press.

Vilar, D., Xu, J., D'Haro, L. F., & Ney, H. (2006). Error Analysis of Statistical Machine Translation Output. In LREC-2006: Fifth International Conference on Language Resources and Evaluation. Proceedings (pp. 697-702). Genoa, Italy

(ISO/TR 9241-100:2010) Ergonomics of Human-System Interaction – Part 100: Introduction to standards related to software ergonomics. See URL: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=52712

(ISO/TR 16982:2002) Ergonomics of Human-System Interaction – Usability Methods Supporting Human-Centred Design http://www.iso.org/iso/catalogue_detail?csnumber=31176

Association for Computational Linguistics:

http://www.cs.jhu.edu/~ccb/publications/meta-evaluation-of-machine-tranlsion.pdf

LetsMT! project deliverables:

- D1.2 - Functional Specification
- D3.4 - SMT Web Service Ready for Integration
- D5.1 - Widget for Web Page Translation / Browser Plugins for Web Page Translation,
- D5.2 - Business and financial news website with translation widget,
- D5.4 - Automatic evaluation report of business and financial news SMT,
- D5.6 - News translation scenario: status report

Appendix 1

List of issues, Firefox plug-in

Table 6 List of issues for the LetsMT! FF extension

Identifier	Description	Priority
MOFF-01	Alter the name of the LetsMT! Mozilla Firefox browser plugin for web page translation to "LetsMT! Mozilla Firefox browser extension for web page translation" to improve visibility and avoid confusion	medium
MOFF-02	The extension should be made available through the LetsMT! platform website and the official channels for Mozilla Firefox extensions	low
MOFF-03	Hiding versioning from the end user by altering the source code ("install.rdf") of the extension in order to ensure forward and backward compatibility of the package with as much versions of the browser as possible	medium
MOFF-04	The extension configuration interface should provide an option for altering the LetsMT! translation web service location	medium
MOFF-05	Browser extension naming should be consistent throughout all the interfaces in which it appears	high
MOFF-06	Support latest versions of Mozilla Firefox	high
MOFF-07	Users should be given feedback from the translation service and/or browser extension while the translation is in process	high
MOFF-08	The extension interface for providing the results of the translation to the end user should be implemented to respect the style boundaries of the translated content	high

List of issues, Internet Explorer plug-in

Table 7 List of issues for the LetsMT! MSIE plugin for web page translation

Identifier	Description	Priority
MSIE-01	**Insufficient documentation for the LetsMT! MSIE browser plugin in LetsMT! deliverable D5.1**	**medium**
MSIE-02	**Include the documentation given in the MSIE browser plugin package to LetsMT! deliverable D5.1**	**medium**
MSIE-03	**The plugin should be made available through the LetsMT! platform website**	**low**
MSIE-04	**Browser plugin naming should be consistent throughout all the interfaces in which it appears**	**high**
MSIE-05	**The browser plugin interface navigation should be altered to be more user friendly**	**high**
MSIE-06	**The browser plugin interface for providing the results of the translation to the end user should be implemented to respect the style boundaries of the translated content**	**high**
MSIE-07	**Users should be given feedback from the translation service and/or browser plugin while the translation is in process**	**high**
MSIE-08	**The plugin configuration interface should provide an option for altering the LetsMT! translation web service location**	**medium**

Appendix 2

In a <u>separate</u> document one can find the following:

1. Translations English – Dutch
2. Translations English – Danish
3. Translations English – Croatian

All test results are included in a separate appendix document to make sure everyone has access to this data and there is no deception, cheating, deceit or something similar with the data. Of course we expect the project partners to use this evaluation data with discretion according to the privacy policy set up in the project contract.