

WHITE PAPER TRANSLATION SOFTWARE



Globalization combined with the increasing use of internet and e-commerce strongly contributed to develop international exchanges. Whereas English could be considered as a “lingua franca” in most cases, still, only one third of the web pages available are in this language. Moreover, using their native language to communicate with customers, suppliers and co-workers is a strong advantage. In this international context, linguistic technologies and multilingual solutions have turned to essential tools to:

- enhance understanding and communication in foreign languages,
- simplify and support translation processes.

Several millions of users within companies already use a translation software: it has become a precious support to translations intended for publication as a tool of productivity and an effective support for communication and understanding at a time. The objective of this white paper is to understand both the technologies and the ways an instant translation software work, for an optimal use

CONTENTS

1- What is an instant translation software?

1.1- Definitions

Instant translation software

Professional solutions for corporate translation management

Electronic dictionaries

1.2- Applications :

Understanding aid

Communication support

Productivity tool

1.3 – Benefits

1.4- Components

Linguistic approach

Rules

Dictionaries

Statistical approach

Interface

1.5- Customization

2- Answers to FAQ

2.1- What is the difference between a translation software, an electronic dictionary and a translation memory ?

2.2 – Does a translation software translate literally ?

2.3 – Is it possible to connect electronic dictionaries or include one's own lexical bases?

2.4 – What sources are used to develop translation software (grammars, dictionaries)?

2.5 – Can translation software replace human translators ?

2.6 – How can you translate a text on paper ?

Can one use an OCR? (system of optical character recognition)

2.7 – How can you manage internal problem with linguistic or IT resources ?

2.8 - An opportunity for human translators?

3-Precautions of use

1. WHAT IS AN INSTANT TRANSLATION SOFTWARE?

1.1 - DEFINITIONS

➔ **An instant translation software is a computer application providing automatically and extremely rapidly a translation of any type of texts (letters, reports, articles, web sites.) from a source language to a target language. The main objective consists in providing a text conveying the information from the source document as close to the original and as fluent as possible.**

This tool can work in various environments (PC, mobile terminals, Internet, Intranet):

- either as an autonomous application,
- or as a translation function directly integrated into an application (Intranet, Internet, Word or Excel.) or in a workflow (like customer support...)

In addition, complementary tools are available offering various functionalities and bringing a number of benefits: those tools represent the ideal complement to translation software. These solutions optimize human translation management (like translation memories) and bilingual electronic dictionaries (like Collins or Lexibase)

➔ **Professional solutions for Corporate Translation Management** (CTM solutions) have been designed to provide solutions to the growing complexity of contents, the reduction of implementation delays and the increasing number of languages required. Those tools propose a global solution for control, management and overview of translation project as well as the translation process itself: definition and control of delays, availability of documents and required resources, tasks allocation to the right persons, control and filing...

These solutions are presented as a platform centralizing the set of company linguistic resources, offering total control on translation workflows and translation processes.

➔ **Electronic dictionaries** represent essential complements to professional solutions for translation: they combine the power of technology to the wealth of lexicographic basis of famous dictionaries (like Collins) for high quality results. An elaborate search mode allows instant access with a mouse click to a mine of information: large quantity of words, but also large variety of expressions and examples of contexts are available for a fast and ergonomic screen reading.

1.2 - APPLICATIONS

Three kinds of applications for a translation software can be distinguished: understanding, communication and productivity.

→ Understanding aid :

With the development of Internet, E-mails and corporate Intranet systems, users have an urgent need to understand at once information in various foreign languages. About 30% of websites are in English, the rest being shared between Spanish, German, French, and so on. Access to all information is possible only if you know a multitude of languages. Therefore you can find a great value in a translation software that will help you overcome the linguistic barriers allowing you to understand a text written in a language you do not master, but also to understand it immediately without the help of anyone: a simple click enables you to get an immediate idea of a document written in a foreign language.

A translation software gives you the autonomy that no other solution can better provide to better understand and improve reading speed.

→ Support to communication :

While facilitating the access to multilingual information, the translation software acts like a very efficient communication link: it brings down languages barriers and thus encourages foreign trade. In the setting of an internal use document, it will allow co-workers for example to capture (and to treat) the content of a message instantaneously, without appealing to a service of translators, that if he/it exists, is rarely available for this kind of tasks. The translation software offers a particularly efficient link within the company.

→ Productivity tool :

With globalization of economy, most companies have to disseminate information in several languages (internally as well as externally) and translation departments are faced with a growing number of requests. By using a translation software, human translators can accelerate translation processes and spare a lot of time on the first stage of the translation process. They can focus on the exactness and preciseness of the chosen words and become a true specialist of a specific topic.

Some examples :

- 1. To translate a contract, the translator should be a jurist and know the Law of the two countries, he must localize the translation according to the Law of both countries.*
- 2. In the United States, there are several translations for the French term "PDG": President, Chairman, or CEO. Here again it is up to the human translator to chose the correct translation.*

In both cases, the translation software will provide a first draft that will be immediately available and that can be reviewed later by the human translator if necessary.

Who can use a translation software?

Here are some types of users who might someday need a translation software. You will probably recognize yourself in one of these descriptions:

- You are German-speaking, you do not speak foreign languages and you would like to understand internet sites, newspapers, or other documents written in those languages, or you wish to send e-mails or letters in another language.
- You speak English or French, you do not speak German but you have to communicate with German-speaking people.
- You master two languages and you have to translate a lot of texts (as a translator, or a bilingual secretary) and you wish to speed up the translation process: with the use of the software, you may gain up to 40% productivity.
- You work in the English subsidiary of a German, French or Spanish company, you need to understand all documents within the company: manuals, reports, brochures, internal notes...
- You work in an international institution in an English speaking country (embassies, representations of the EEC, OECD, UNICEF...) or, the other way around, in a French delegation abroad.
- And more generally, you have to communicate once in a while in a foreign language or to do some research (students, academics, journalists, researchers...)

1.3 - BENEFITS:

The instant translation software is considered today as one of the most strategic tools of companies dealing with multilinguism. More than 5 million people make a daily use of the translation software providing:

→ Instant understanding :

The instant translation software is the only solution providing instant understanding of a text written in a language one does not master and this remains priceless. Relying on an instant translation software, the user becomes absolutely independent in its understanding of a document expressed in a language he does not or barely master.

Knowing that reading speed is between 30 to 50% higher in ones mother tongue :

Without translation software, understanding is estimated between 60 and 80%.

With translation software, understanding is estimated between 75 and 95%.

Therefore, professional translation services are less solicited so that they can focus on more strategic projects (like external use sales documentation)

→ A tool to « make oneself understood » :

With the acquisition of a translation software, we also notice new emerging usages: it is not only about “understanding”, but to make oneself ‘understood’ in a language one doesn’t necessarily master, without calling for a third party.

→ Harmonization and optimization of the corporate linguistic assets :

The implementation of a translation software enables to list, enrich, share, keep up and develop the specific terminology of the company. The translation software provides an active support for the company on its way to globalization and enables to adapt rapidly the corporate terminology, the tools (internal processes, technical glossaries) but also the statutory aspects to the next corporate languages.

→ Reduction of translation costs and time to market:

As the users of an instant translation software are more autonomous in their understanding of texts expressed in foreign languages, they do not call so often to professional translation departments. On average, we estimate a reduction of the translation budget between 10 and 30%.

→ Progress in foreign languages :

The instant translation software contributes to communicating without frontiers and encourage the users to read the documents in the original language.

1.4 COMPONENTS

Two main approaches are used today to design translation software: the linguistic approach on the one side, and the statistical approach on the other side.

The linguistic translation is usually used for common European languages while the statistical translation is kept for more unusual languages.

1.3.1- Linguistic translation:

The linguistic approach relies on a syntactic analysis, analysis, transformation and syntax rules, as well as bilingual dictionaries.

❖ **Linguistic rules :**

The translation engine is the technological core of the translation software. The stages of the automatic translation process are as follows :

- Analysis of the source text,
- Application of the linguistic rules of transformation and of desambiguation,
- Generation of the translation.

Linguistic rules applied are for instance: words order, syntactical and grammatical information, morphology, the precise analysis of the components of the sentence...

→ The finer the rules, the more precise the translation.

❖ **Dictionaries :**

Dictionaries integrated in translation software have been developed specifically for the translation software and are not a simple list of words or expressions with their translation. Each word or expression include linguistic information (morphology, semantics, syntax...) in the source language and then in the target language. This kind of information is then handled by the translation engine.

→ The richer the dictionaries, the better the translation.

Dictionaries are available in French, English, German, Spanish, Russian, Italian, Portuguese, Hebrew, Japanese, Polish, Dutch

Specialized dictionaries are intended for specific translations (computer, law, business, medicine...) to provide a precise translation as close as possible of the users terminology. Specialized dictionaries manage more than 1 million forms (words, terms, expressions and declensions) for each translation sense.

→ The performances of the linguistic translation software mainly rely on the subtlety of pre-defined rules combined with the wealth of dictionaries. Those can be customized, which means they will be adapted to the context and specific use of a company, for a significant improvement of the translation preciseness.

1.3.2- Statistical translation:

This technique of automatic translation relies on statistical models whose parameters come from the analysis of bilingual corpus of texts. It consists in comparing a same text with thousands of pages in two different languages, without calling to dictionaries, nor grammatical rules. The software is then cutting out the texts, studying the frequencies and making connections between the languages. When facing a new text, the software tries to find in its memory the elements that will best suit to provide a translation. Unlike a translation memory, statistical translation enables to translate texts that are being seen for the first

time and can be used in very short time delays to translate new languages.

Statistical tools are currently available only in Intranet environment or on very powerful PC (because they require a very important memory as well as very high calculation potential) and can be adapted to the specific environment of the user.

Conclusion :

Both approaches introduced above present a number of benefits and disadvantages:

- The linguistic model is more reliable and more predictable on the one side, but also more demanding in terms of human resources on the other side.
- The statistical model offer faster progresses but is also very demanding in IT resources and may sometimes provide very unexpected results.

The challenge for the teams dedicated to research and development in this area consist in taking the best of those two models.

1.5 TRANSLATION SOFTWARE CUSTOMIZATION :

For each translation software, either linguistic or statistic, it is strongly recommended to customize the software, which means **adapt it to the specific context of the company** to improve its performance in terms of preciseness.

1.5.1 – Linguistic based customization :

The customization of a linguistic translation software is divided into two main steps that aim at creating specialized dictionaries for specific translations (IT, medicine, military, gas, automobile). Each dictionary is the result of a long collaboration between linguists, lexicographers and specialists of the fields concerned.

- Work on glossaries: The first objective consists in evaluating the translation quality before any customization and then customize the software to perfectly match with the field thanks to the creation of a dictionary including the most commonly used expressions. Such a dictionary is made through:
 - The separate coding of the different glossaries
 - The location of multiple translations
 - The validation of the glossaries entries with the approval of the department concerned.
- Final validation of the dictionary created by the supply of corpus: this step enables to compare the translations reached with the real texts while looking for new complementary expressions and seizing the non-coded expressions. This study in corpus enables to collect a large number of additional specialized terms and expressions which means an increase of around 50% of the dictionary. The dictionary reached at the end of this customization phase is very rich and turns out to be much more comprehensive than existing dictionaries.

This work can only be completed thanks to the deep possibilities of coding provided by the translation software. The software appeared to have a large lexicographic coverage and this is confirmed by the very low percentage of vocabulary that is not recognized by the system.

Hereafter an illustration with a research project lead by the French Ministère de la Culture in collaboration with Reverso and research laboratory of TALANA University, underlining the interest of the translation software customization. The project consisted in testing out the software during a year while inserting the specific terminology dedicated to the “archivist theory”. The French Ministère de la Culture finally declared the tool was operational and brought the following benefits:

- *Productivity gain of 50%*
- *Understanding of English texts without any external support,*
- *Update of terminology..*

1.5.2 – Statistical based customization

Statistical tools can also be customized based on corpus for an improved translation quality.

2. ANSWERS TO FAQ :

2.1 – What is the difference between a translation software, a dictionary and a translation memory ?

Translation dictionaries :

Electronic dictionaries are generally CD-ROM, mobile or Internet versions of the paper dictionaries usually known. They make it possible to find a translation for a word or expression according to the context, but they do not translate complete sentences. They add up to instant translation software.

Translation memories :

Translation memories on data base allow to store sentences or already translated paragraphs and to recognize them for a new translation. To use a translation memory, you beforehand have to enter every source sentence and its translation into the base. Before using it, a long and expensive work is necessary which can usually only be achieved after a few years (according to the volume and type of the documents) Those systems are recommended for users who need to translate a large amount of documents very frequently or for users who have to work together on a same document.

Translation memories are part of more global solution for **the professional management of the corporate translation**. They are presented under the form of a platform centralizing the whole linguistic resources of a company and providing a total control on translation workflows and processes.

The software includes :

- A translation memory
- A terminological system
- Performing control tools for the project management and translations follow-up.

2.2 – Does a translation memory translate literally?

The examples below will show you that a translation software does not translate literally but analyses the context and the sentence to select the right rules of transformation to apply according to the language.

→ These sentences have been translated by an instant translation software :

French - English

1- Verbal expression ("mettre au courant de") and passive relative construction:

Il n'a pas mis ses amis au courant de l'offre spéciale qui lui avait été proposée.

He did not tell his friends about the special offer which had been proposed to him.

2- Verbal expressions ("être à son compte", "avoir affaire à") and adverbial phrase:

Je suis à mon compte depuis une semaine et j'ai souvent affaire à des fournisseurs douteux.

I have my own business for a week and I often have to deal with doubtful suppliers.

3- Expression ("bien vouloir" + infinitive verb), idiomatic expression ("sous peu" = "bientôt"):

Nous vous remercions de bien vouloir patienter, votre docteur arrivera sous peu.

We thank you for waiting, your doctor will arrive soon.

4- Construction "soit..., soit...":

Soit vous acceptez cette offre, soit vous faites une meilleure proposition.

Either you accept this offer, or you make a better proposition.

English - French

1- In the following two examples, the relative proposition is perfectly translated in French, even though there is no relative pronoun (whom, that...) in the English sentence:

The person I talked to is right here.

La personne à qui j'ai parlé est juste ici. (and not: la personne j'ai parlé à est droit ici)

I'm impressed by the software I was shown.

Je suis impressionné par le logiciel que l'on m'a montré.

2- Conditional proposition with "Should":

Should you need a car, please let me know !

Si vous avez besoin d'une voiture, s'il vous plaît faites-le moi savoir !

3- Transformation: intransitive verb -> pronominal verb:

We shook hands.

Nous nous sommes serrés la main.

German - English

1- Expressions like « entweder, oder... », « ansonsten »... are recognized by the software:

Entweder arbeite ich oder ich schlafe

Either I work or I sleep

Ansonsten lese ich

Apart from that I read

2- Whole structures are recognized:

Du verhältst dich, als ob ich blöd wäre

You behave as if I was silly

Du verhältst dich, als wäre ich blöd

You behave as if I was silly

2.3 – Can one use dictionaries or include its own lexical bases with a translation software?

A translation software already include a very rich general dictionary covering many fields to provide an answer to the most current needs. Specialized dictionary are also available to get a more precise translation in specific areas: computer, law, business, medicine

The translation can be refined then with the help of a classical dictionary (such as Collins). It is also possible to semi-automatically create its own dictionaries based on its own terminology. Those dictionaries propose a “collaborative” function allowing not only to create its own entries, but also to share them. Complementary tools (conjugator, interactive grammar, voice..) enable to sharpen its knowledge.

2.4 – What are the sources used to develop a translation software (grammars, dictionaries...)?

The dictionaries used with the translation software have a different structure from the classic bilingual dictionaries. (choice of the translation according to the context) The teams of linguists who developed translation softwares worked with the best dictionaries and grammars, but they also defined rules according to the translation engine (see chapter 1.4)

2.5 – Can a translation software replace a human translator ?

If you wish to publish a document or write an agreement, then a machine translation will not be enough

and you will have to call for a specialist either to translate the complete document or to revise the text obtained with the instant translation.

However, the machine translation enables specific usages you would otherwise never have done : understand a text while surfing on Internet for example, or translate emails...

2.6 – How can you translate a text on paper ?

Can one use an OCR (system of optical character recognition)?

You may scan your documents with a scanner and use an OCR software which transform a scanned image into a text file. Most of the scanners are now provided with this system.

You may also use an OCR system installed on your computer to use the translation software. The OCR system is however not supplied with the translation machine.

2.7 – How to manage internal problems with linguistic/ IT resources?

The publisher of the translation software provides services of linguistic experts to work on the corporate specific terminology so that the company does not bother with mobilizing internal resources to work on this project.

Furthermore, the publisher of the translation software provides integration and is committed to hosting the solution so that the company does not have to care about anything else.

2.8 – An opportunity for the translators ?

Translation software allow the translators to save a lot of time, as witnessed by the following examples:

- Fast sweeping of the text to identify words unknown or relevant terminological expressions.
- Direct recuperation of perfectly translated sentences and possibility to have the source and translated text simultaneously.
- The translated text has the same layout, so that the layout is preserved when a modification is made.
- Building-up of a terminology and guarantee of terminological coherence.
- Possibility of offering new services with high added value...

3. PRECAUTIONS OF USE

The first objective of the instant translation is to understand or make oneself understood. A pertinent criterium is consequently the reader's opinion : « does the document match with the use expected and does it translate the content of the source text ? » The answer depends on several parameters :

- Knowledge degree of the reader and efforts he is ready to make,
- context,
- nature of the source text,
- translation engine.

1/ Nature of source text.

The translation software has been created to deal as a priority with texts without spelling or punctuation mistake, expressed in a neutral style (neither literary, nor journalistic...) using the vocabulary from everyday life or business world.

To reach the best possible translation, the source text should be written clearly and avoid ambiguous structures. You should rather use short sentences than too long and complexes sentences.

2/ Correction of the source text.

You have to make sure the text to be translated is correct in the first place:

- Correct the grammar and spelling mistakes.
- Respect the typographic rules: apostrophes must be attached to the following word, the right punctuation must be respected (commas, apostrophes, dots...) as well as the accents in the languages that use them (French, Spanish...).

These few pieces of advice will allow you to avoid too many unknown words which the translation software can not analyze and which can sometimes change the meaning of the sentence.

The correction of the spelling mistake in a sentence is often enough to obtain a very good translation.

Example :

We shook hand Nous avons secoué la main.

We shook hands Nous nous sommes serrés la main.

Here: the mere omission of an "s" was enough to make the translation completely false!

3/ A translation software analyzes, but does not "understand".

A translation software can analyze the structure of a sentence, the grammatical context, and transpose them into another language, without however "understanding" the context the way a human being would. Very often, the linguistic rules can raise all the ambiguities:

Example :

The French word "comprendre" can either be translated by "understand" or by "include", but "comprendre que" will always be translated by "understand that".

But everything that is based on human intelligence (human sensibility, understanding of the world ...) cannot be analyzed by the software.

Because the software cannot understand, there are cases when the generated translation will be bad and it

will be up to the user to rectify the meaning of the text. There are other cases when the software will offer translation alternatives.

Let's have a look at a few examples where the software cannot "guess" the meaning:

1- Here are two examples in which only the reader can understand the meaning of the text:

La jeune fille a aimé cet avocat.
*The girl loved this lawyer{*avocado*}.*

the trials lasted a month
*Les procès{*essais*} ont duré un mois*

2- Here is an example of a word which can have several completely different meanings:

J'ai oublié les cartes.
I forgot the cards.

Without any context, Reverso chooses the most current translation.
If you clarify the meaning, Reverso will choose the right translation:

J'ai oublié les cartes de crédit. Peux-tu payer pour moi ?
I forgot the credit cards. Can you pay for me?

J'ai oublié les cartes routières, nous sommes perdus !
I forgot the roadmaps, we are lost!

J'ai oublié la carte des vins pour que vous puissiez commander.
I forgot the wine list so that you can order.

