

## TRANSLATING SUBTITLES WITH TRANSLATION ENVIRONMENT

Attila Imre\*

### Abstract

*The paper deals with the new possibility of translating subtitles with translation environments. Previously translated texts are preserved in translation memories offering consistency when new subtitles are translated. Samples of subtitle translations (feature films, documentaries) from English into Romanian are described using MemoQ translation environment, highlighting those aspects which may make the life of a subtitle translator easier.*

### Introduction

The modern age of translation industry seems to have forgotten pencil and paper (Pym 2002). The main reason for that is not the environmental issue, but the so-called *computer assisted translation tools* (CAT-tools). Those involved in translation business must have heard of *SDL Trados*, which is the present-day leader of this type of software. As translation seems to be a thriving business, more and more CAT-tools emerge out of the blue, and some of them are serious competitors to the leader, claiming a fair share.

The popularity of translation industry may be backtracked to three main reasons. The unifying Europe has brought about changes never imagined before in translation industry. The European Union has twenty-three official languages<sup>1</sup> and the Directorate-General for Translation (DG Translation) translates texts for the European Commission in all these possible combinations (506 in all). In 2010 their permanent staff is around 1750 linguists and 600 support staff, which enables DG Translation to state that they are among the largest translation services in the world. For instance, in 2008 alone they translated more than 1,800,000 pages.<sup>2</sup> Although they translate hundreds of thousands of pages and their standards are very high, they only cover certain areas of translations, such as legal or technical documents.

On the other hand, we can mention *globalization* as such, which resulted in finding the same product in many different parts of the world. However, as mainstream products were filling up the global market, local authorities expected the producers to have their product descriptions translated into local languages. This resulted in a translation boom, both in quantity and quality. Although we can easily find poor quality translations, there is a tendency for better quality (cf. quality assessment in translation), which seems to be possible with the help of translation environments able to provide both translation memory (TM) and term base management (TB).

Last but not least we can mention the immense potential of the internet. For example, the free Google Translator, which provides online translation for more than 50

---

\* Ph.D., Senior Lecturer, Sapientia University, Târgu-Mureș, Romania, amireital@yahoo.com

<sup>1</sup> Source: [http://ec.europa.eu/dgs/translation/whoweare/index\\_en.htm](http://ec.europa.eu/dgs/translation/whoweare/index_en.htm), 13.04.2010.

<sup>2</sup> Source: [http://www.translationconference.com/2009\\_eload/JancsiB\\_HU.pdf](http://www.translationconference.com/2009_eload/JancsiB_HU.pdf), 12.03.2010.

languages<sup>1</sup>, is in the focus of ardent debates regarding its quality, but one thing is sure: translation service is brought into our homes and we can decide for ourselves whether or not we are satisfied with its quality for private purpose. However, the drawbacks of Google Translator are already challenged by other translation services. At this point, the writer of an article about Google machine translation states:

Translate is a work in progress, so not all of the translations will be clean, crisp and accurate. But as with everything else Google does, Translate is an iterative technology that will Google will advance over time (Boulton 2010).

The "advance over time" seems very realistic, if one aspect is taken into consideration, namely terminology. An adequate terminology enables translators to unify their translations on the one hand, and on the other hand new texts are translated much faster with the help of a term base. The European Union term base interactive terminology (IATE, Inter-Active Terminology for Europe) has been active since 2007, including various fields, such as space exploration and forestry<sup>2</sup>, and translation environments are able to create new term bases, handle existing ones, as well as adding simple lists of words in two languages to support term bases in the middle of translation. This is one of the major advantages of translation environments as any translator using a translation environment will surely improve regarding both quality and quantity of translated characters per unit of time.

### Subtitling

According to Buda (1979), the primary importance of mass-media devices lies first and foremost in their capacity to broadcast live messages to long distance and to large masses of people at the same time. TV and cinema have taken the press over, as the latter can convey neither live information nor to masses of people. The political role of mass communication has been evident since the world wars, and today we can talk about different types of politics (e.g. health, environmental). We have observed that during the last two decades these types of politics may be found in Hollywood productions as well, as they are often subservient to the public, serving them with 'hot' products. Once they are subtitled in various languages, the success is guaranteed. There are very many Romanian subtitles available on the internet compared to the number of Romanian native speakers. One of the reasons may be the comparatively small number of films dubbed in Romanian, which is typical for South-American TV series. However, the equation is not as simple, as there are many other factors worth taking into account when explaining the phenomenon (e.g. small number of filmgoers, easy internet access, knowledge of English, etc.), but this is beyond our interest now.

We can easily test the relatively high number of Romanian subtitles if we start playing a film with video players which have embedded subtitle finders (e.g. SM Player, BS Player); the newer the film, the more likely to have fewer subtitles, but interestingly, the majority of new films already have Romanian subtitles. In the table below we offer 11 titles of films released in the last 15 years and their plot is connected

<sup>1</sup> As of April 2010, these languages are: Afrikaans, Albanian, Arabic, Belarusian, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Filipino, Finnish, French, Galician, German, Greek, Haitian, Creole, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Irish, Italian, Japanese, Korean, Latvian, Lithuanian, Macedonian, Malay, Maltese, Norwegian, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swahili, Swedish, Thai, Turkish, Ukrainian, Vietnamese, Welsh and Yiddish.

<sup>2</sup> Source: [http://www.translationconference.com/2009\\_upload/JancsiB\\_HU.pdf](http://www.translationconference.com/2009_upload/JancsiB_HU.pdf), 12.03.2010.

to the future of mankind, human relationships or health. The data has been collected from a major subtitle site, [www.opensubtitles.org](http://www.opensubtitles.org), and one film from 1939 is also listed for the sake of comparison. The table clearly shows the emerging popularity of subtitles, whose quality is often questionable, although their communicative function is as clear as crystal. The majority of subtitles are in English, Spanish, Russian or Brazilian Portuguese, which is not surprising due to the number of speakers. However, we could find a large number of Polish, Dutch, Greek, Hebrew and Romanian subtitles. The table includes the Internet Movie Database ranking (IMDB) of the film, its total available number of subtitles as well as the number of English, Romanian and Hungarian subtitles (collected on 13th March 2010 from [www.opensubtitles.org](http://www.opensubtitles.org)):

Film title	Released	IMDB	Subtitles	English	Romanian	Hungarian
<i>Gone with the Wind</i>	1939	8,2	162	31	13	3
<i>Waterworld</i>	1995	5,7	140	21	7	2
<i>I, Robot</i>	2004	7,0	491	44	18	2
<i>The Day After Tomorrow</i>	2004	6,3	343	25	14	4
<i>Hitch</i>	2005	6,8	594	9	15	5
<i>Thank you for smoking</i>	2005	7,8	84	16	5	3
<i>The Secret</i>	2006	5,3	65	8	7	1
<i>An Inconvenient Truth</i>	2006	7,9	91	6	4	1
<i>Slumdog Millionaire</i>	2008	8,4	441	34	31	7
<i>Avatar</i>	2009	8,5	188	16	11	7
<i>2012</i>	2009	6,0	247	48	14	5

Table 1. Films and Subtitles

The ratio of English–Romanian subtitles obviously does not indicate the number of native speaking population of the two languages, and the relatively high number of Romanian subtitles may be compared to the Hungarian subtitles as well. The explanation is partially connected to the fact that dubbing is thriving in Hungary (there are hardly any films on with original audio), whereas Romanians tend to watch the film with the original audio format (mainly English) and subtitled in Romanian. One direct consequence is that Romanians are often ‘exposed’ to English; the majority of the younger population knows English at an acceptable or even higher level, thus many people try out their knowledge by offering Romanian subtitles.

We cannot help not observing that there are more and more films connected to contemporary global problems trying to startle the simple earthbound humans. The day of doom seems inevitable, although the exact time and form of it is not clear. The films are in fact a quick reaction to our global problems leading even to mass-hysteria. We can easily shoo away these signs, but subtitles are like barometers in this respect: it often happens that the more subtitles there are regarding a film, the more ‘hot’ topic it covers. And 2012 is almost here...

#### Subtitles and translation environment

In this part we would like to present the possibility of creating subtitles with the help of a translation environment, namely *MemoQ* (version 4.0.24). This translation environment is one of the most top-notch products in the field, as it is extremely user-friendly and offers an outstanding online technical support as well as free webinars and updates once the product is purchased. It can handle various document types, such as

Microsoft Office files, HTML files, XML files or XLIFF (XML Localization Interchange File Format) files, which can be processed in other translation tools such as *SDL® Trados™* and translation editors (cf. MemoQ Quick Start Guide 2010). The main parts of this software are the translation memory, the term base, the aligner and the translation editor (word processor).

When a subtitle is going to be translated, the typically *.sub* or *.srt* type of subtitle files cannot be fed into the translation environment directly, but they can be easily turned into *.doc* or *.docx* format files. A right mouse click on the file offers in a drop-down list the possibility of opening the file with Microsoft Office (Word), then after opening we can save it in *.doc* or *.docx* format. Now we can feed the document (add) to the MemoQ translation project and create a translation memory and a term base for the glossary. As MemoQ developers explain:

The translation memory stores pairs of sentences. As you proceed in the text, the translation workspace checks if the translation memory contains a sentence similar enough to the one you are translating. If there is such a sentence, it will appear as a translation hit, and the workspace will indicate the difference between the current sentence and the stored sentence. ... If you want to build a glossary from certain expressions and their translations, you can use the term base. Terms are expressions in multiple languages, optionally complemented with additional data. The word processor in the translation workspace automatically displays the translations of the terms occurring in the text, to be inserted using a single keystroke. (MemoQ Quick Start Guide, 2010).

In order to test the efficiency of this translation environment, we have selected various fragments from feature films and documentaries. Our samples are from *An Inconvenient Truth*, *The Age of Stupid*, *Home* and *2012*. The effective use of translation software presupposes that the translator is specialized in certain fields, thus sooner or later specific word and phrases will be repeated, which should be highlighted by the translation memory and / or term base. As we are interested in the future of mankind as well as concerned about the environmental issues, we selected samples from films dealing with these two interrelated topics. The selected key words are presented in the table below, explaining our choice as well:

Film	Keywords in source text	Reason
<i>An Inconvenient Truth</i>	<i>ice cap, Earth, degree, increase, global warming, ice shelf</i>	<i>Topic specific nouns/expressions</i>
	<i>Melt</i>	<i>Topic specific verb</i>
	<i>miles, Fahrenheit</i>	<i>Culture specific words</i>
<i>The Age of Stupid</i>	<i>fossil fuels</i>	<i>Topic specific expression</i>

Table 2. Films and Keywords

These words were added to the term base (glossary), so when the word appeared in the source text again, it was highlighted and shown in a special column how it had been translated before; if there were more translations, all were listed and we had the liberty to insert any of them in the target segment or give additional translations. Additional translations are usually needed due to changes in number, person, voice (cf. change due to agreement between noun and predicate), etc. The more samples we added

to the translation project, the more hits we were able to find, consequently leading to a much more effective translation in time and quality.

### Conclusions

It is common knowledge that technical texts belonging to certain areas often contain repetitive elements, especially when newer versions of various products appear on the market (improved hardware or software upgrades), and MemoQ developers claim that effectiveness in these cases can reach up to 70%, and in cases of academic or other texts this percentage is about 10 – 30%.<sup>1</sup> However, in the case of English – Romanian translations we can further refine the possibilities of a translation environment.

The good parts of using a translation environment saves the translator from translating twice or even more things that somewhere before appeared in the text (or previous texts), although these words and expressions have to be added one by one to the term base. If a term base is handled properly, it can help in creating new term bases (it offers an extra source, no need for a dictionary or another meticulous search) and even offers the chance of easily revising and improving previous translation databases. In case of *nouns* we could observe that the term base gave impressive results: basically, the addition of the plural form or the genitive form of the word proved very effective. One major drawback we could observe is the high sensibility of the translation software, as the capitalization of the particular word was clearly different from the non-capitalized form; *Earth* and *earth* are not the same in MemoQ, as it is shown in the picture below, and once we entered *Earth* in the term base, this was not useful in the case of *earth*:

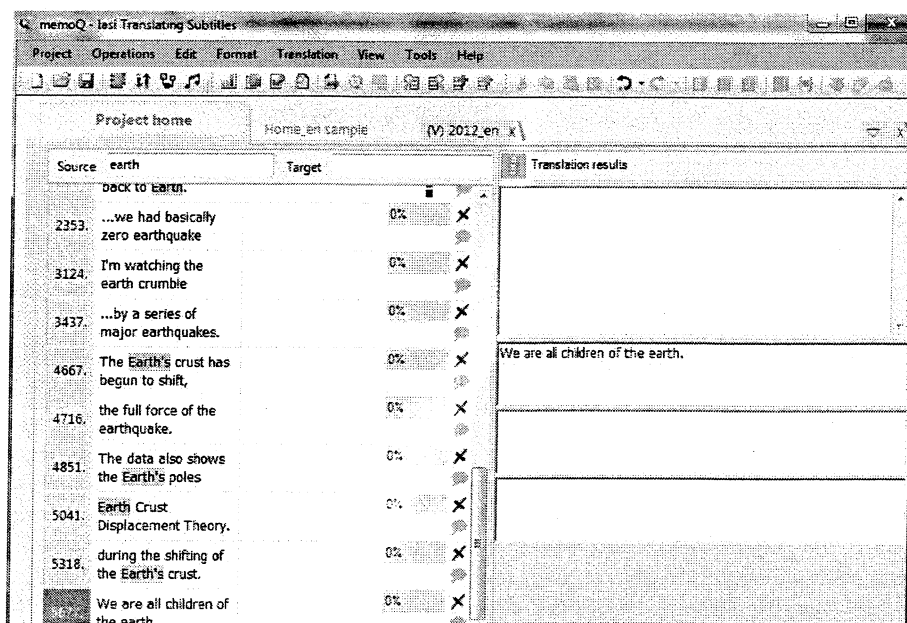


Figure 1. Earth versus earth

<sup>1</sup> Further details may be found on [www.kilgray.com](http://www.kilgray.com).

As there are two Indo-European languages involved, similarities were expected to appear, but Latin languages differ from English in verb conjugation a lot. So our suspicion that the verb *melt* is not easy to be added to the term base to be highly functional turned out to be true, and we conclude that only the root of the translated word is worth adding to the term base: *se top(i)*, *se top(e)te*, *se top(esc)*, and they are actually completed when the context is clear. If we have the English third person singular (*melts*), only the infinitive part is highlighted, as well as in the case of its continuous or participle/ gerund form (*melting-topind*, *topire*), which is not a problem, if only the root of the verb was translated originally:

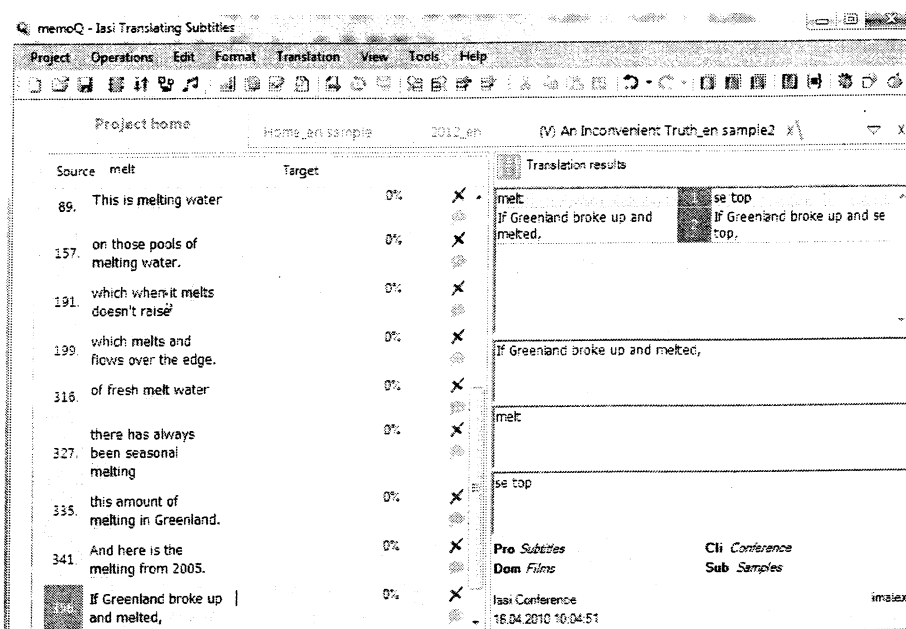


Figure 2. Verbs and MemoQ

The translation of culture-specific words was very successful; as the word *degree* goes both with *Fahrenheit* and *Celsius*, its addition to the term base seemed logical. Romanians understand 10° *Celsius* far better than 50° *Fahrenheit*, thus we would suggest this change to all translators; similarly, *miles*<sup>1</sup> should be changed to *kilometers*, unless we have a very good reason to preserve them (source of humour, geographical or cultural flavour, etc.).

Two further issues have to be mentioned: one of them is the possibility of adding modal verbs to a term base, the other is the translation of taboo words, which is highly culture specific combined with the educational background of the translator. Modal verbs are typical in English, but we tend to believe that only some of them worth adding to a term base (for instance *should* – *ar trebui să* or *must* – *trebuie să*), as other modal verbs have too many possible translations: *could* – *poate (că)*, *puteam*, *puteai*, *putea*, etc., and instead of verifying the possibilities and finding the exact match it is faster to

<sup>1</sup> An extra care should be taken to differentiate *mile* (1.609 km) from *nautical mile* (1.852 km).

type the proper one. However, the possibility of modal verbs within a translation environment is still open, as there are full sentences which are often repeated, even if translated with shades of meaning:

*You must be joking/ kidding.*

Cred că glumești./ Acum nu vorbești serious. [antonym translation]

*May all your dreams come true.*

Să-ți îndeplinești toate visele.

The problem of translating taboo words is not new. First of all, people have used taboo words since language is used to express metaphorical meanings, then as language and society changes (evolves?) some taboo words may lose their restrictedness (by having overused them or due to historical, political or geographical changes). Up to the recent past, Romanian translators new hardly more than one swear (*La naiba!*), and this phenomenon seems to be getting more and more obsolete. One reason is 'evolution', but we can also mention the emancipation of Romanian filmgoers who expect more than that as their 'repertoire' has extended as well. More and more subtitle translators hide behind untraceable nicknames, and psychologists proved that people can drastically change if they do not have to bear the consequences of being traced via the internet. More and more try their luck out on subtitling films and they often try to prove their skills by focusing on taboo/ swear words. The tendency may be regarded universal, as Hollywood productions or new Romanian films are not shy at all verbally... But this can already be the topic of another paper.

A final thought would be the future of film subtitling. *The Age of Stupid* may offer very startling evidence about the effectiveness of CAT-tools. We tend to believe that it is only machine translation (it has been recently released), as no real inflection is to be observed in the Romanian translation, and it is full of English words, probably the ones which were not added to a term base yet. Let us finish the article with a sample from the subtitle:

<i>The Age of Stupid</i> , English	<i>The Age of Stupid</i> , Romanian
Welcome to the global archive, a vast storage structure located 800km north of Norway. It contains the artwork from every national museum. There are pickled animals stack up two by two. Every film, every book, every scientific report, all stored on banks of servers. But the conditions we are experiencing now, were actually cause by our behavior in the period leading up to 2015. In other words, we could've saved ourselves We could've saved ourselves. But we didn't, it's amazing! What state of mind were we in, to face extinction, and simply shrug it off?	Bine ați venit la arhiva nivel global, o structură vastă de stocare situat 800 km la nord de Norvegia. Acesta conține opera de arta de la fiecare Muzeul National. Nu sunt animale murat stiva de până doi câte doi. Fiecare de film, fiecare carte, fiecare raport științific, toate stocate pe băncile de servere. Dar, în condițiile ne confruntăm acum, s-au datorat de fapt comportamentului nostru, în perioada de până în 2015. Cu alte cuvinte, ne-am fi putut salva. Am putea-ați salvat pe noi înșine. Dar noi nu, e uimitor! Ce stare de spirit au fost la noi, să se confrunte cu dispariție, și pur și simplu shrug it off?

## References

- Boulton, C., **Google Chrome Gets Machine Translation, New Privacy Features**, 2010. Downloaded from <http://www.eweek.com/c/a/Web-Services-Web-20-and-SOA/Google-Chrome-Gets-Machine-Translation-New-Privacy-Features-839506/> on 3<sup>rd</sup> March 2010.
- Buda, B. **A közvetlen emberi kommunikáció szabályszerűségei**, Tömegkommunikációs Kutatóközpont, 1979. Downloaded from <http://mek.niif.hu/02000/02009/02009.pdf> on 28<sup>th</sup> June 2009.
- Dearing, S., **No decision yet if WHO and big pharma 'faked' the H1N1 epidemic**, 2010. Downloaded from <http://www.digitaljournal.com/article/287123> on 12th February 2010.
- Jancsi, B. **Öt évvel a csatlakozás után: fordítás az Európai Bizottságnál**, 2009. Downloaded from [http://www.translationconference.com/2009\\_eload/JancsiB\\_HU.pdf](http://www.translationconference.com/2009_eload/JancsiB_HU.pdf) on 12th March 2010.
- Pym, A. **Translation Studies as Social Problem-Solving**. Paper for the conference *Translating in the 21st Century: Trends and Prospects*, Thessaloniki, Greece, 2002. Downloaded from [http://www.tinet.cat/~apym/on-line/research\\_methods/thessaloniki.pdf](http://www.tinet.cat/~apym/on-line/research_methods/thessaloniki.pdf) on 10th April 2010.
- XXX. **MemoQ Quick Start Guide**, 2010. Downloaded from [http://www.kilgray.com/files/user-guide/memoQ\\_QuickStartGuide\\_4\\_0\\_EN\\_0.pdf](http://www.kilgray.com/files/user-guide/memoQ_QuickStartGuide_4_0_EN_0.pdf) on 10<sup>th</sup> April 2010.
- XXX, **The European Commission's in-house translation service**, 2010. Downloaded from [http://ec.europa.eu/dgs/translation/whoweare/index\\_en.htm](http://ec.europa.eu/dgs/translation/whoweare/index_en.htm) on 13.04.2010.
- Zetzsche, J.O. **Let's Talk: Trados and the Google Translator Toolkit**, 2009. Downloaded from [https://www.atanet.org/chronicle/feature\\_article\\_october2009.php](https://www.atanet.org/chronicle/feature_article_october2009.php) on 9<sup>th</sup> November 2009.