The Translation of Documentaries: Can Domain-Specific, Bilingual Glossaries Reduce the Translators' Workload?

An Experiment Involving Professional Translators

Sabien Hanoulle,
Department of Applied Linguistics, Translators and Interpreters, UAntwerp, BELGIUM
Véronique Hoste,
Department of Translation, Interpreting and Communication, Ghent University, BELGIUM and
Aline Remael
Department of Applied Linguistics, Translators and Interpreters, UAntwerp, BELGIUM

ABSTRACT

This article investigates by means of an experiment, whether the integration of a domain-specific, bilingual glossary helps professional translators reduce translation process time and terminological errors in the translation of documentaries. It also examines the way these translators use the glossary. Special attention is devoted to the methodology for the experimental study design, the keystroke logging tool used for understanding the writing process, as well as the statistical analyses. The experiment was conducted in two sessions and involved two groups of professional translators. Both groups worked once without, and once with, a bilingual glossary. One group used a manually labelled glossary, the other an automatically extracted glossary. The results demonstrate that the glossaries have a positive bearing on the total process time, the pause time before terms and on terminological errors.

KEYWORDS*:* audiovisual translation, documentaries, keystroke logging, off-screen dubbing, terminology extraction

1. Introduction

"Documentaries constitute a fascinating field which has given rise to an academic domain in its own right, Documentary Studies, within film studies" (Espasa 2004:183). Many authors have attempted to define the genre labelling it as a 'nonfiction film' and analysing the fuzzy boundaries between fiction and nonfiction (Renov 1993:2-3; Ward 2005:23; Plantinga

1997:16-18). Others, like Nichols, believe that it is the multiple agents that produce documentaries who define the genre: "documentaries are what the organizations and institutions that produce them make. […] The context provides the cue" (2001:22). Most approaches insist on its versatility, and this view is also shared in audiovisual translation (AVT) research focusing on the translation of documentaries. This sub-field has only been attracting the attention of AVT studies in the last two decades, although the translation process of documentaries is a specifically audiovisual practice. Translators of written texts are usually specialists in a limited number of domains (e.g., legal translation), whereas documentary film translators are generally not specialized in a specific domain, but in a specific mode, i.e. 'audiovisual'. As such, they need "minimum knowledge of a maximum of topics" (Mir in Espasa 2004:190). Moreover, the mode of discourse they translate is an oral rendering of a written text with potentially diverse registers (more formal for the narrator and more spontaneous for interviewees). This great diversity has led to equally diverse translation guidelines (Gregory and Carroll 1978:61-63, Espasa 2004:191). All this makes the translation of documentaries a specific audiovisual practice.

In addition, documentaries can use various translation techniques, such as voice-over, lip-synch dubbing, off-screen dubbing[1] and subtitling, each of which requires specific translation skills. Depending on their working country, documentary translators are specialized in one or more of these translation techniques. In subtitle countries like Scandinavia, the Netherlands and Belgium, subtitling in combination with off-screen dubbing is common, whereas in dubbing countries (France, Italy and Germany, to name but a few) documentaries are mainly translated in voice-over, lip-synch dubbing and off-screen dubbing (Franco et al. 2010:85). Another characteristic of documentary translation relates to the textual functions of documentaries. According to the classification by Rosa Agost for the translation of audiovisual genres (1999:30-40), documentaries are considered to be an informative genre with narrative, descriptive, persuasive and expository functions, each of which can be more or less dominant. Espasa (2004:191) concludes that translating documentaries, requires an all-round knowledge of text types and functions, language registers, domain-specific terminology and translation techniques.

[1] Term proposed by Franco, Matamala & Orero (2010) which indicates the translation of the commentary voice heard off-screen.

In addition, documentary translators translate for a heterogeneous target audience in terms of age, cultural background and expertise, which poses challenges for terminology translation (Franco 2000:236; Espasa 2004:193). Matamala (2010:259) points out that terminological problems occupy a prominent position in the translation process and include "identifying terms, understanding terms, finding the right equivalent, dealing with the absence of, or the inability to find, an adequate equivalent, dealing with denominative variation, choosing between *in vivo* and *in vitro* terminology and avoiding wrong transcriptions".

For all these reasons, we believe it is relevant to support audiovisual translators, providing them with a domain-specific, bilingual glossary, in our case an ad hoc English-Dutch glossary, one for each episode, for the translation of documentaries. Two possibilities were envisaged in our research: a manually created glossary and a glossary drawn up by a terminology-extraction system, even though automatic systems are mainly used for texts with a high degree of repetition and a large amount of terminology, such as technical, scientific, financial and legal texts (Christensen and Schjoldager 2010:8; Lagoudaki 2010:12), while documentaries contain a mixture of general utterances and domain-specific terminology. Nevertheless, some documentaries do contain recurring terminology, i.e. some genres appear to contain more domain-specific terms than others (for instance, episodes about natural sciences), as was demonstrated by a pilot study with master students in translation (Hanoulle et al. 2015:9). In addition, this pilot study also demonstrated that the process and pause time before terms was reduced significantly when the participants worked with a bilingual glossary and that the terminological errors decreased, albeit not significantly. These promising results needed, of course, to be tested with professional translators.

The present article deals with this essential issue, explaining the set-up of the experiment for professionals and the study design. The results are subsequently discussed and research challenges for the future are explored. The over-arching research question concerns the impact of bilingual glossaries on translator's workload and workflow and has been operationalized by subdividing it into three, concrete research questions:

1) Does the integration of a domain-specific, bilingual glossary into the translation process reduce the process and pause time before terms of professional translators?

2) Does the number of terminological errors decrease when translators work with a bilingual glossary?
3) To what degree do professional translators use the glossary? Do they consult it more as they become more used to it?

2. Related Research

2.1 The system: terminology extraction

Over the past few decades, researchers have increasingly turned their attention to the investigation of terminology-extraction systems. Both the technical aspects and the applications have been the subject of careful study by researchers in different fields: computational linguistics, software engineering and translation science. Automatic terminology-extraction systems basically rely on two methodologically different approaches. The linguistic approach is based on the characteristics of term formation patterns, which are expressed as part-of-speech code sequences (e.g., N N, N prep N, Adj N) and as such, is always language dependent. The statistical approach on the other hand is language independent and is based on quantifiable characteristics of term usage, i.e. terms tend to occur more frequently in specialized texts than in general domain texts (Macken 2010:105).

As purely statistically-based methods tend to over generate terms and purely linguistically-based methods produce some noise, most state-of-the-art systems use hybrid approaches, as proposed for the first time by Daille (1996). Whereas the hybrid approach of Daille (1996) was initially targeted towards monolingual terminology extraction, a similar methodology was also ported to bilingual terminology extraction by herself and other researchers in subsequent years (Frantzi and Ananiadou 1996; Daille 2000; Vintar 2001; Wilson et al. 2009, to name but a few). The standard methodology in these hybrid approaches consists in first identifying term candidates monolingually, after which source and target terms are aligned. This approach of generating bilingual dictionaries from monolingual terminology lists was also adopted by the well-known commercial system SDL Multiterm Extract Trados®[2]. An alternative approach to bilingual terminology extraction was proposed by Macken et al.

[2] http://www.sdl.com/cxc/language/terminology-management/multiterm/extract.html (accessed 21 January 2015).

(2013), who took a multilingual perspective from the start in their TExSIS[3] system. TExSIS first generates candidate terms from linguistically motivated aligned chunks, which are based on a shallow morphosyntactic automatic pre-processing of the texts by means of part-of-speech tagging and lemmatization. By aggregating words into syntactic chunks, not only single-word terms, but also multi-word terms, which have been shown to be frequent in technical texts, are captured. Both word alignment information and syntactic chunk information are taken into account for the creation of the bilingual candidate term list. The candidate terms are filtered by means of several statistical filters for single-word and multi-word terms (Macken et al. 2013:17-18). Using a corpus of the automotive industry, TExSIS was tested against SDL Multiterm Extract Trados®, the LUIZ system from Vintar (2001) and Similis®[4], a commercial, bilingual terminology-extraction system. The researchers found that TExSIS performed favourably compared to the other systems for the creation of bilingual term lists of technical texts thanks to its integrated multilingual approach. Given its performance, we integrated the TExSIS system in our experiments.

The application of terminology-extraction systems was investigated in commercial settings by Warburton (2013). She proposed a method for the integration of terminology extraction in the translation pipeline, which is particularly beneficial to companies that translate large, terminologically-rich corpora. Positive results in terms of quality and speed were also found by Coombs (2014:53), who proved the importance of terminology management (starting from automatic terminology extraction) in a case study on biotechnological patent translation.

2.2 The mode: off-screen dubbing

The above mentioned studies apply terminology extraction to technical translation, not to audiovisual translation, the core of the present research. Yet, several projects on audiovisual translation have been carried out in the past decade all of which focused on increasing productivity, reducing translation costs and enhancing the quality of translation results through the introduction of technologies, in an attempt to meet market demands. The EU-BRIDGE[5] project aimed at developing automatic transcription and translation technology for the development of innovative multimedia captioning and translation services of audiovisual documents.

[3] TExSIS: http://www.lt3.ugent.be/en/ (accessed 24 November 2014).
[4] http://similis.org/linguaetmachina.www/index.php (accessed 21 January 2015).
[5] http://www.eu-bridge.eu/ (accessed 12 March 2015) *European Union grant agreement n°287658.*

The SUMAT-project[6] introduced statistical machine translation techniques into subtitle translation processes in order to develop an online subtitle translation service able to semi-automatize the subtitling of both freelance translators and subtitling companies. TransLectures[7] is another EU-funded project, aiming to develop innovative, cost-effective tools for the automatic transcription and translation of online educational videos. The SAVAS-project[8] aims at developing an Automatic Speech Recognition (ASR) technology for multilingual live subtitling, specifically tuned to the needs of the broadcasting and new media industries.

However, these projects mainly consider subtitling as translation mode and speech recognition and machine translation as technological support systems. Little research has been done into the specific features of off-screen dubbing. In a case study, Remael (2007) analyses the translation shifts of both subtitles and narration[9] due to language and/or ideology policies. Franco et al. (2010:83) discuss briefly that off-screen dubbing translators do have to take into account synchronicity constraints in order to align their translations with the visuals, even though voice-over creates the impression that synchrony does not matter because the viewer does not hear the original speaker. One ongoing study, as a part of the ALST-project (Matamala et al. 2012), links up with the present research as it investigates the application of machine translation and post-editing of off-screen dubbing for wildlife documentaries. In a preliminary analysis on the translations produced by MT engines, Ortiz-Boix mentions terminology as one of the challenges (forthcoming). She also carried out an experiment with twelve master students in translation who translated and post-edited an excerpt of off-screen dubbing for wildlife documentaries. The results showed generally that post-editing requires less temporal, technical and cognitive effort than the translation (Ortiz-Boix and Matamala, forthcoming).

In order to provide a useful contribution to audiovisual translation studies and given the challenges posed by documentary translation, especially in terms of the terminological challenges they present, this study narrows its focus to the impact of manual and automatic bilingual glossaries on the translator's workload and workflow for off-screen dubbing. Below

[6] http://cordis.europa.eu/project/rcn/191741_en.html (accessed 23 June 2015).
[7] https://translectures.eu// (accessed 23 June 2015).
[8] http://www.fp7-savas.eu/ (accessed 23 September 2015).
[9] The translation of the off-screen voices (= off-screen dubbing).

we present an experiment that involves professional translators who translate once with only reference resources (online dictionaries and internet information) and once with both reference resources and a bilingual, automatically or manually extracted glossary.

3. Experimental Set-Up

3.1 Organisation

Twelve professional English to Dutch translators, with an average of twenty years of experience (three to thirty-eight), participated in the experiment. Six were audiovisual translators, the other six had experience in different kinds of texts (e.g., financial, scientific, commercial, legal, technical). They were paid 100€ for the two sessions. We travelled to the working place of each participant with a laptop with *Inputlog*[10] installed, a logging software developed at the University of Antwerp.

This logging software served the purpose of observing the writing process of the candidates. Indeed, to establish whether a bilingual glossary reduces the translator's workload, we needed to observe the writing process as "writing fluency and flow reveal traces of underlying cognitive processes" (Leijten and Van Waes 2013: 360). Keystroke logging programs, in general, log and time stamp activity in order to reconstruct and describe text production processes. *Inputlog* was chosen because of its compatibility with Windows environments and its writing-oriented design (Leijten and Van Waes 2013:363). While the participants were translating, *Inputlog* registered the whole translation process. By means of this tool, it was possible to analyse the total translation time, pauses before terms and search for information. The features used for this study were the 'record' and the 'analyse' module. The record module logged all keystroke and mouse data in the Microsoft Word page used for the translation, together with a time stamp in milliseconds. The analyse module produced the general logging file and the summary analysis needed to identify the total process time and the pause time before terms, as these two data were crucial to determine the efficacy of a bilingual glossary. The total process time of the translations was provided in milliseconds by *Inputlog's* summary analysis. The pause time before terms could be deducted from the general analysis *Inputlog* supplies. It yields the very detailed analysis of the writing process needed for this type of study. Leijten and Van Waes (2013:364) explain that the output

[10] http://www.inputlog.net/ (accessed 20 October 2014).

features of the general analysis represent every log event, the cursor position, the document length, the start and end time of every event in milliseconds and are used to calculate the action and pause times. In order to determine the pause times before terms[11] needed for our experiment, we scrolled the log files manually selecting the rows belonging clearly to a pause time before terms, i.e. the event from the moment the participant entered a dictionary, the internet or the bilingual glossary until he/she entered the *Wordlog* document in which the translation was written down. If the participant surfed from one source to another without going to the *Wordlog* document in between (e.g., first to the dictionary then to the internet) this was considered one pause time. Figure 1 shows an example of the *Inputlog* general analysis output illustrating the output features. Every row represents one log event (second column 'id'). The first row of this example indicates that the candidate entered the TX-glossary, then used the CRTL-F function to look up the term 'mass' (vertical digits, row 1418 – 1421). The cursor position and the document length are represented ('positionFull' and 'doclengthFull') as well as the number of characters produced (charProduction). The next columns show the start time and the end time of every event in milliseconds. These data are used to calculate the action and - the most important information for this research - the pause time. An algorithm identifies the pause location and provides a classification ('pauseLocation' and 'pauseLocationFull') and in the last columns the mouse clicks are represented by x/y values.

[11] With 'terms' we mean 'labelled as such' in the manually extracted glossary about which we report in the section 3.2 'Glossaries'.

Figure 1: An example of the *Inputlog* general analysis output.

id	type	output	posit	docl	char	startTim	startClock	endTim	endClock	actio	pauseTi	pause	pauseLocat	x	y
1413	focus	Microsoft Excel - Hunt for the Higgs_TXtermenlijst	2636	5469	6107	2209828	00:36:49.82	2209828	00:36:49.82	0	0	10	TRANSITION		
1414	keyboard	LCTRL	2636	5469	6107	2209828	00:36:49.82	2210125	00:36:50.12	297	0	2	BEFORE WORDS		
1416	keyboard	LCTRL + F	2636	5469	6107	2210047	00:36:50.04	2210141	00:36:50.14	282	31	13	UNKNOWN		
1417	focus	Find and Replace	2636	5469	6107	2210125	00:36:50.12	2210125	00:36:50.12	0	0	10	TRANSITION		
1418	keyboard	m	2636	5469	6107	2210391	00:36:50.39	2210484	00:36:50.48	93	0	1	WITHIN WORDS		
1419	keyboard	a	2636	5469	6107	2210422	00:36:50.42	2210516	00:36:50.51	94	31	1	WITHIN WORDS		
1420	keyboard	s	2636	5469	6107	2210625	00:36:50.62	2210703	00:36:50.70	78	203	1	WITHIN WORDS		
1421	keyboard	s	2636	5469	6107	2210766	00:36:50.76	2210875	00:36:50.87	109	141	5	AFTER SENTENCES		
1422	keyboard	RETURN	2636	5469	6107	2212219	00:36:52.21	2212281	00:36:52.28	62	1453	4	BEFORE SENTENCES		
1423	mouse	Movement	2636	5469	6107	2212969	00:36:52.96	2213922	00:36:53.92	953	750	13	UNKNOWN	624	3
1424	mouse	LEFT Click	2636	5469	6107	2214016	00:36:54.01	2214078	00:36:54.07	62	94	2	BEFORE WOR	624	3
1425	focus	Microsoft Excel - Hunt for the Higgs_TXtermenlijst	2636	5469	6107	2214094	00:36:54.09	2214094	00:36:54.09	0	0	10	TRANSITION		
1426	mouse	Movement	2636	5469	6107	2214094	00:36:54.09	2214625	00:36:54.62	531	0	13	UNKNOWN	932	7
1427	focus	TASKBAR	2636	5469	6107	2214719	00:36:54.71	2214719	00:36:54.71	0	0	10	TRANSITION		
1428	mouse	LEFT Click	2636	5469	6107	2214719	00:36:54.71	2214797	00:36:54.79	78	0	2	BEFORE WOR	932	7
1429	focus	Microsoft Excel - Hunt for the Higgs_TXtermenlijst	2636	5469	6107	2214797	00:36:54.79	2214797	00:36:54.79	0	0	10	TRANSITION		
1828	focus	Large Hadron Collider - Wikipedia - Mozilla Firefox	3784	5652	6450	2299609	00:38:19.60	2299609	00:38:19.60	0	0	10	TRANSITION		
1829	mouse	Movement	3784	5652	6450	2299609	00:38:19.60	2300484	00:38:20.48	875	0	13	UNKNOWN	1028	
1830	mouse	LEFT Click	3784	5652	6450	2300641	00:38:20.64	2300844	00:38:20.84	203	157	2	BEFORE WOR	1028	
1831	mouse	Movement	3784	5652	6450	2300688	00:38:20.68	2300813	00:38:20.81	125	0	13	UNKNOWN	544	
1832	mouse	Movement	3784	5652	6450	2300859	00:38:20.85	2300859	00:38:20.85	0	46	13	UNKNOWN	544	
1833	keyboard	p	3784	5652	6450	2301375	00:38:21.37	2301453	00:38:21.45	78	516	1	WITHIN WORDS		
1834	keyboard	h	3784	5652	6450	2301609	00:38:21.60	2301750	00:38:21.75	141	234	1	WITHIN WORDS		
1835	keyboard	o	3784	5652	6450	2301688	00:38:21.68	2301813	00:38:21.81	125	79	1	WITHIN WORDS		
1836	keyboard	t	3784	5652	6450	2301813	00:38:21.81	2301906	00:38:21.90	93	125	1	WITHIN WORDS		
1837	keyboard	o	3784	5652	6450	2301875	00:38:21.87	2302047	00:38:22.04	172	62	1	WITHIN WORDS		
1838	keyboard	n	3784	5652	6450	2301969	00:38:21.96	2302109	00:38:22.10	140	94	5	AFTER SENTENCES		
1839	keyboard	RETURN	3784	5652	6450	2302453	00:38:22.45	2302547	00:38:22.54	94	484	4	BEFORE SENTENCES		
1840	mouse	Movement	3784	5652	6450	2303016	00:38:23.01	2304500	00:38:24.50	1484	563	13	UNKNOWN	448	3
1841	focus	photon - Google zoeken - Mozilla Firefox	3784	5652	6450	2303078	00:38:23.07	2303078	00:38:23.07	0	0	10	TRANSITION		

The corpus for the current experiment, used for both the creation of the bilingual glossaries and as a source text, was selected from a corpus of documentaries between 2005 and 2013 which VRT, the Flemish broadcasting channel, placed at our disposal. Of each episode (171 in total), the researcher was provided with the original English script and the translation into Dutch done by a professional, audiovisual translator. In terms of genre, the corpus was limited to natural science documentaries, as a pilot study (Hanoulle et al. 2015:9) demonstrated that these episodes contain enough recurring, domain-specific terminology to be considered for accurate, automatic term detection. The language combination of the corpus was English - Dutch, the major language pair on Flemish television and the target culture was Flanders, as the material was placed at our disposal by the Flemish broadcaster. In terms of audiovisual modes, only off-screen dubbing was considered. As mentioned before, every audiovisual mode (subtitling, lip-synch dubbing, voice-over, off-screen dubbing and audio description) faces specific constraints which might influence the translational choices (Franco 2000:240-241; Franco et al. 2010:83-93), hence the need to focus on one mode. In Flanders, the region of our target translators, the most frequent translation mode for documentaries is off-screen dubbing, often combined with subtitles. In the present study, we focus on off-

screen dubbing, the less investigated translation mode, as explained in the related research section above.

From the most recurring subjects, four episodes were selected to conduct the experiment: *The Earth Machine – Land* (geology), *Madagascar - Island of Marvels* (wildlife), *The Secret World of Pain* (human body) and *The Hunt for the Higgs* (astronomy). From the off-screen English dubbing scripts of these episodes, a selection of sentences counting a total of 1017 words was made. Each sentence contained one or more different terms. For a correct comprehension of the text, additional information was added where needed between brackets (e.g., the adjunct to which a relative pronoun refers).

The candidates were divided into two equal groups, each of which was composed of three audiovisual translators and three non-audiovisual translators. Two sessions involving the same source text were organized with an interval of four months in between, the time considered to be needed to 'forget' the first translation. In the first session, half of each group worked with the glossary and half worked without. The glossaries used in this experiment were a manually extracted glossary (called gold standard or GS) and an automatically extracted glossary (called TExSIS or TX). Detailed information on these is included in next section 3.2 'Glossaries'. In the second session, the condition was inverted to ensure an equal distribution of a possible residual memory effect (see method section). In the session where they translated with a glossary, one group used the gold standard (now referred to as the GS-group), the other group used the automatic glossary (the TX-group).

The participants did not make use of the video as the source text consisted of self-contained clips with no coherence between them. The terminology was clear from the verbal context and there was no need to match timing or style with the images and the intonation of the voice talent for the purpose of this study. Furthermore, even in daily practice, audiovisual translators do not always have access to the video, for example, in rush jobs and/or for copyright reasons. All the participants translated the same source text from English into Dutch. There was no set time limit, so all the candidates were able to complete the assignment.

3.2 Glossaries

The glossaries were extracted from the four episodes of the above-mentioned corpus. The manual glossary (GS) was created, labelling what is to be considered as a term on the conceptual basis of termhood and unithood, also defined by Kageuro and Umino as "the degree to which a linguistic unit is related to a domain-specific context" and "the degree of strength or stability of syntagmatic combinations or collocations" (1996:260-261). The automatic glossary was extracted by TExSIS, a hybrid system that combines a linguistic and statistical approach to bilingual terminology extraction (see also 'Related research'). Both glossaries provided the English term and its Dutch equivalent as found in the VRT translation. No other information typical for term records (e.g., definition, context, synonyms, etc.) was supplied. Table 1 shows an example of the terms TExSIS extracted from one episode (*The Earth Machine*). The first two rows of the example indicate that for the term 'crust', the system identified two translations, i.e. 'aardkorst' and 'korst'. For 'earth' the system extracted three different translations ('aarde', 'aardkorst' and 'aardoppervlak'). All the proposed translations are correct. However, for some terms, the system was not able to identify the equivalent term in Dutch (e.g., 'super-continent'). In such cases, the second column is empty.

Table 1: The first 34 candidate terms generated by TExSIS in '*The Earth Machine'*.

English	Dutch
crust	aardkorst
crust	korst
earth	aarde
earth	aardkorst
earth	aardoppervlak
super-continent	
lava	lava
lava	vulkaan
mantle	mantel
mantle	buitenmantel
planet	aarde
planet	planeet
planet	oceanen
tectonic	
surface	aardoppervlak
surface	grond
surface	oppervlakte
volcanoes	vulkanen

fault	breuklijn
fault	verschuiving
fault	breuk
miles	kilometer
miles	meter
plates	platen
crater	krater
crater	kraterwand
smoke-shrouded rim	
heat	hitte
heat	warmte
fourty-five miles (sic.)	
five-and-half miles	
storm-tossed seas	woelig water
never-ending movement	
tectonic plates	tektonische platen

Table 2 provides the number of term types (unique terms) per text in the Gold Standard only, in TExSIS only and, in the last column, the term types extracted by both glossaries. With 'term' we always mean 'labelled as such in the manual glossary'. The column 'only in TX' shows that a very small number of terms (only 1 for Earth Machine and 6 for The Higgs) was extracted by TX only and not by the GS. In other words, the GS glossary contains more terms than the TX glossary. Moreover, the total number of words in the first column in Table 2 illustrates the 'term density' in these texts. Table 3 shows the number of term tokens (total number of terms, repetitions included) per text in the GS-glossary and in the TX-glossary. For the accuracy of the term extraction, we refer to the pilot study (Hanoulle et al. 2015:6-8).

Table 2: Number of term types per text in the glossaries.

Episode (tot. n° words)	only in GS	only in TX	GS and TX
Earth Machine (234)	16	1	23
Madagascar (222)	31	0	17
Pain (319)	34	0	16
Higgs (242)	25	6	13

Table 3: Number of term tokens per text in the glossaries.

Episode (tot. n° words)	GS	TX
Earth Machine (234)	39	24
Madagascar (222)	48	17
Pain (319)	50	16
Higgs (242)	38	19

4. Data Analysis

4.1 Methods

The study design was organized as follows. The participants translated the same text twice, once with the glossary (half of the group with the GS, the other half with the TX glossary), and once without. As the translators might remember part of the text despite the four months' interval between the two sessions, half of them were asked to work with the glossary and the other half without in the first session. For the second session, this working condition was inverted. Hence, half of the candidates, those working with the glossary in the second session, had the advantage of the memory effect having already translated the text once, the other half had not (because they worked with the glossary in the first session). On the other hand, half of the candidates, those working without the glossary in the second session, also had the advantage of the memory effect having already translated the text once. This means that the advantage of the memory effect was equally distributed over both conditions: working with and without the glossary.

In order to implement this design statistically and to respond to research question one (i.e. Does the integration of a domain-specific, bilingual glossary in the translation process reduce the process and pause time before terms of professional translators?), all analyses were elaborated with the difference variables: the process time (and pause time) of the first translation minus the process time (and pause time) of the second translation, thereby creating process time differences and pause time differences. Subsequently, to examine whether translating with or without a glossary resulted in a significant difference, a comparison was made between group A, working without a glossary in the first session and with a glossary in

the second session, and group B, working first with and next without a glossary. If working with a glossary was more efficient, we expected group A to have longer process time differences (and pause time differences) compared to group B. In order to analyse whether group A has longer process time differences (and pause time differences) than group B, an independent sample t test was used. P-values below 0.05 were considered statistically significant.

In each group, there were 24 observations: 6 translators working on 4 texts. As we wanted to detect large effect sizes (Cohen's d = .80 as in Cohen 1988:35), a power analysis for this effect size, n = 24 and significance level $\alpha = .05$ shows that the power = .77 (Cohen 1988: 36). This means that we had a probability of 77 % to detect large effect sizes, which is considered important.

As a measure of the impact of glossaries on the translation quality and in order to provide an answer to research question two (i.e. Does the number of terminological errors decrease when translators work with a bilingual glossary?), all terminological errors were also assessed. A terminological error means that a term from the gold standard was translated with a term not corresponding to a correct translation in this context. A paired sample t test was used to analyse the statistical difference in term errors between the two working conditions of the GS group and the TX group. A bar graph presents the number of errors made by both groups translating with and without glossary. All statistical analyses were conducted in SPSS version 20.

Research question three (i.e. To which degree do professional translators use the glossary? Do they consult it more as they get used to it?) was addressed by means of a qualitative analysis of the candidates' search behaviour. The total pause time before terms was divided manually into two categories: pause time for bilingual glossaries (GS or TExSIS) and pause time for online dictionaries plus the internet (i.e. all other internet sources the participants consulted to gather information and find the correct translation).

Furthermore, by means of a line graph based on the pause time before terms per text, the evolution in the use of the glossary, the internet and dictionaries was examined towards the end of the assignment, as the candidates got used to the glossary. In addition, we investigated

the degree to which the candidates used the glossary, counting manually how many terms were double-checked (both in the glossary and in a dictionary or the internet), how many terms were looked up only in dictionaries or on the internet (despite the fact that they were listed in the glossary) and how many terms were checked only in the glossary. The average percentage was calculated for each category.

To conclude, a short retrospective survey was carried out among the candidates, inquiring into their experience of translating with a glossary and the set-up of the experiment. Four questions were asked:

1) Did you trust the glossary?
2) Do you think the glossary was beneficial for your translation?
3) In the second session, did you remember parts of the text/terms from the first session?
4) Do you think translating without video was a handicap? (this question was asked to the six audiovisual translators, not to the others)

The answer to question one was compared with the search behaviour for double-checking.

4.2 Results

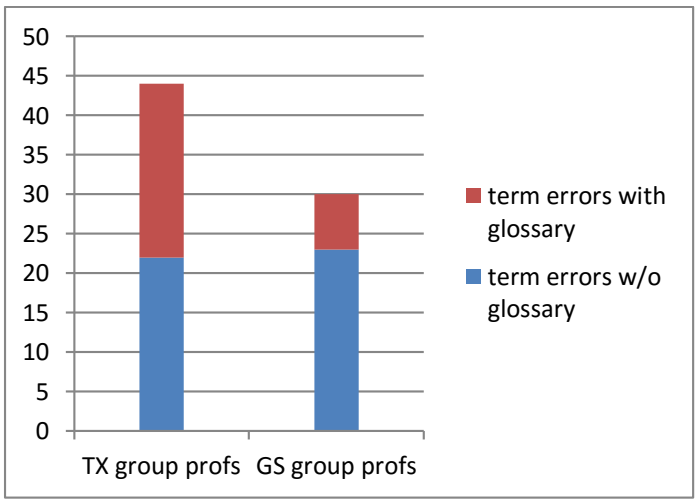
The purpose of the general question of this study was to determine the impact of domain-specific, bilingual glossaries on translators' workload and workflow. At first, we addressed the question as to whether the integration of a domain-specific, bilingual glossary in the translation process reduced professional translators' process and pause times before terms. It transpired that, the average process time differences and pause time differences were reduced significantly when they worked with the glossary. Group A, working without a glossary in the first session and with a glossary (GS or TX) in the second session, had longer process (or pause) time differences[12] than group B, working with a glossary in the first session, and without in the second session. The average reduction of the process time differences was 399.39 s (Confidence Interval of the difference 80.97 to 717.82 s, p. = .015) and the average reduction of the pause time differences for both groups was 113.52 s (C.I. of the difference 30.73 to 196.31 s, p. = .008). Next, we analysed the process time differences and pause time differences of the GS and the TX group separately. On average, for the TX group, both

[12] Differences of the process/pause time between the first and the second session.

process and pause time differences were significantly reduced when working first with a glossary compared to working first without. The average reduction of the process time differences was 469.75 s (C.I. of the difference 92.86 to 846.64 s, p. = .017) and the average reduction of the pause time differences was 124.75 s (C.I. of the difference 39.47 to 210.04 s, p. = .006). For the GS group, neither pause time differences nor process time differences were significantly reduced. The average reduction of the process time differences was 314.05 s (C.I. of difference -249.17 to 877.26 s, p. = .258) and the average reduction of the pause time differences was 102.61 s (C.I. of the difference -53.41 to 258.64 s, p. = .185).

Secondly, we examined whether the number of terminological errors decreased when working with a bilingual glossary. The analysis showed a significant difference between the number of term errors for the GS group when working with, or without, the glossary. The average difference was 2.67 (C.I. of the difference 0.50 to 4.83, p. = .025). However, no significant difference was noted in the TX group. The average difference was 0.00 (C.I. of the difference -2.89 to 2.89, p. = 1.00). In addition, Figure 2 provides the total number of term errors made in the translation of the whole source text (1017 words).

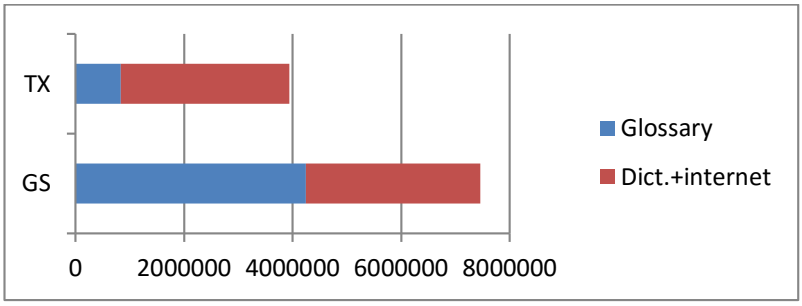
Figure 2: Number of term errors made when translating with and without a glossary.



Thirdly, we investigated the degree to which professional translators use a bilingual glossary and if they consult it more as they get used to it. Figure 3 shows the pause time before terms per source in milliseconds (glossary or internet and dictionaries). In the case of the gold standard group, the time for consulting the glossary was more than the pause time for consulting dictionaries or the internet, whereas the TExSIS group spent less than one fifth of

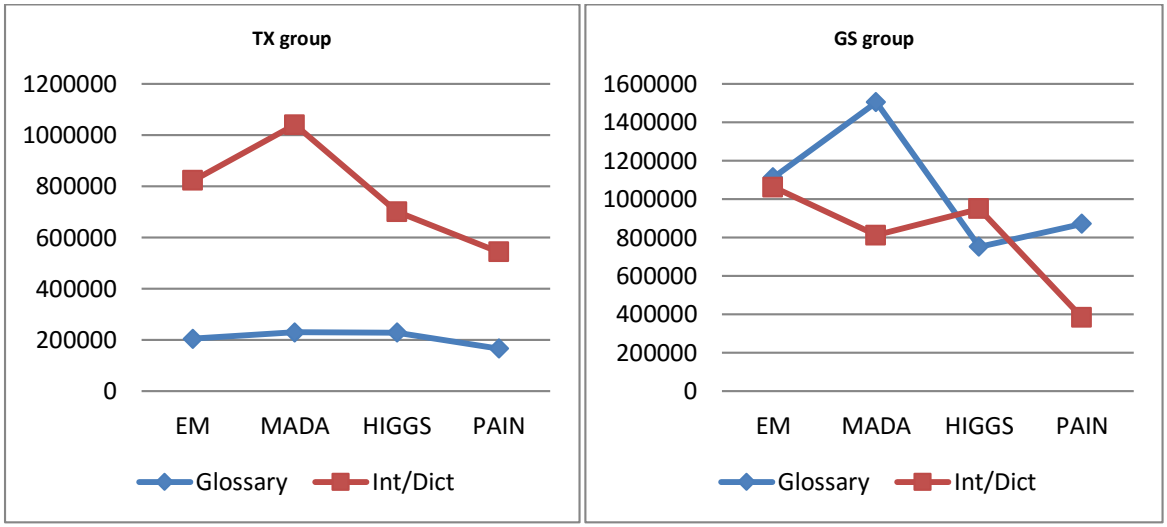
the pause time for consulting dictionaries or the internet to consult the glossary. In other words, the longer the glossary was, the more time the translators spent consulting it, the longer the total pause time and thus, the longer the total process time.

Figure 3: The use of the glossaries vs. the use of dictionaries and internet per group in milliseconds.



Narrowing the focus to the evolution in the use of the sources towards the end of the translation, we notice that the pause time before terms decreases for both the glossary and the internet or dictionaries, for the TX group as well as for the GS group. The line graphs in Figure 4 indicate that the candidates pause less before the terms as they proceed with the translation, but they do not use the glossary more towards the end of the assignment. Only the second text, Madagascar, seems to require longer pause times which can be explained by its high termhood. In this case, the domain-specific terminology consists mainly of names of endemic (and thus unknown) animals.

Figure 4: Evolution in the use of the glossary vs. the use of the internet and dictionaries through pause time before terms.



Regarding the search behaviour of the candidates, Table 4 shows that the GS group checks three quarters of the terms only in the glossary whereas the TX group checks one quarter of the terms only in the glossary. The figures in the first column indicate the average percentage of terms double-checked by the candidates (in both the glossary and the internet and/or dictionaries) compared to the total average number of terms they looked up. The second column shows the average percentage of terms the candidates checked only on the internet or in dictionaries, despite the fact they were listed in the glossary. The third column presents the average percentage of terms only checked in the glossary.

Table 4: Search behaviour per group, expressed in average percentages vs. total average n° of terms looked up.

	Average % terms double-checked	Average % terms only checked in int./dict.	Average % terms only checked in the glossary
GS group	13	10	77
TX group	3	69	28

In order to appreciate the level of difficulty of the terms, Table 5 provides the average % of terms looked up (in the glossary and/or in dictionaries or on the internet).

Table 5: Average % of terms looked up vs. the total average number of terms.

	Average % terms looked up vs. total average n° terms
GS group	48
TX group	38

The retrospective survey carried out among the candidates showed that five of them trusted the glossary, seven said they tended to double-check the terms on the internet or in dictionaries or even preferred to work without the glossary and search for the correct translation on the internet. However, when asked whether the glossary helped them in their translation, nine of the twelve candidates responded it did, mentioning time-saving and verification in case of doubt as their main reasons for using it. Questioned after the second session about what they remembered of the text from the first session, eleven candidates stated they remembered at best, the general content, no terms or at the most only one. The six audiovisual translators were also asked if they thought it was a handicap to work without video. Two of them replied they might have worked a little bit quicker with the video, but the product would have been the same. One mentioned that for one of the four texts, images would have clarified some of the content but not for the other texts. The other three candidates stated that for this assignment, it made no difference whether they had the video or not.

5. Discussion

The reduction in the process time differences and pause time differences of the GS and the TX glossary combined, addressed in research question one, supports the hypothesis that the integration of a bilingual glossary in the translation process of documentaries reduces the translators' workload and the number of terminological errors for professional translators, even though we might not have enough elements (24 observations) to carry out separate analyses on the use of the GS and the TX glossary. With little elements, the difference has to be very important to be statistically significant. In the case of the GS group, for instance, the positive result is not significant, whereas translating with the TX glossary shows a significant difference in both process and pause times. Regarding the number of terminological errors examined in research question two, the results indicate a significantly lower number of errors for the GS group. This might suggest that the GS glossary contains more terms with a high

termhood and/or that the GS group followed the glossary's suggestions more than the TX group.

The present study confirms the results of the pilot study with students, as both process and pause time of the relatively inexperienced translators were reduced significantly when they worked with a glossary, even with analyses carried out on the GS and the TX group separately. The number of terminological errors made by the GS group of the students was also reduced significantly when they translated with a glossary.

Regarding the degree to which professional translators use a glossary, the results seem to indicate that a longer glossary (in our case, the gold standard) encourages the translator to spend more time consulting it, which might be interpreted as 'more confidence'. Indeed, the GS group checked ¾ of the terms only in the glossary, whereas the TX group checked ¼ of the terms only in the glossary and more than ⅔ only on the internet. However, the GS group double-checked more terms than the TX group (13% vs. 3%). Moreover, in our experiment, there was no evolution towards consulting the glossary more and the other sources less as the candidates became more used to the glossary. A longer total pause time for the glossary was determined instead by the specificity of the terminology. Interestingly, both groups translated more than half of the terms without consulting any source, neither the glossary, nor the internet or dictionaries: the GS group looked up 48% of the terms, the TX group only 38%. One reason for this behaviour might be attributed to the nature of the text and the translation brief. The script of documentaries is written and translated for 'quick consumption' as the target audience has to understand the off-screen dubbing while listening to the commentator. This means that part of what is labelled as a term in documentary texts is standard language - although related to the domain-specific context — and thus, already known to the translators. A second reason for this behaviour might be the expertise of the candidates. All being experienced translators, they were probably familiar with some of the terminology. Indeed, the student candidates for the pilot study looked up more terms: the GS group looked up 62% and the TX group 53% of the terms.

A final observation in this discussion concerns the survey. Although five of the twelve candidates stated they trusted the glossary, two of them did not behave accordingly. They

double-checked terms as often as the 'not trusting' candidates or consulted the glossary hardly ever.

6. Conclusion

This article has provided evidence for the hypothesis that a bilingual glossary helps professional translators of documentaries reduce their working time. Indeed, the overall results of the manual and the automatic glossary appear to yield a significant decrease in both process and pause time. In addition, we have shown that the number of terminological errors is reduced when professional translators work with a glossary. These findings may constitute a useful sequel to Matamala's study about terminology in documentaries where she suggests that it would be highly interesting to find a mechanism to avoid duplication of efforts (2010:269). The analysis has also demonstrated that a longer glossary results in a longer pause time and thus, in a longer process time compared to translating with a shorter glossary. Nevertheless, with a longer glossary, translators make fewer term errors. Consequently, in order to optimize the use of a glossary to reduce the audiovisual translator's workload, a good balance between high termhood and number of terms, i.e. glossaries containing only the terms with the highest termhood, should be reached. Expanding the model to experiments with professional translators working with the GS glossary in one session and with the TX glossary in another session, could be a fruitful area for further research. Incorporating an analysis of the differences in extracted terminology between the GS and the TX glossary could prove useful to construct a model with the minimum requirements (e.g., parameters for high termhood) of a good bilingual glossary.

A final aspect that might influence the translators' working conditions is their attitude with regard to using a glossary. When they get used to the tool, they will probably double-check fewer terms, thus cutting down on both pause and process time. In all likelihood, translating regularly with a glossary will result in improved confidence in the tool and, consequently, in a less demanding workload and, hence, a better target text.

Acknowledgements

The subject of this research was inspired by Anna Matamala's project for the Spanish Ministerio de Economia y Competitividad, reference code FFI2012-31024, 'Accesibilidad lingüìstica y sensorial: tecnologìas para las voces superpuestas y la audiodescripciòn'.

References

Agost, Rosa (1999) *Traducción y doblaje: palabras, voces y imàgenes,* Madrid: Ariel.

Christensen, Tina P., and Anne Schjoldager (2010) 'Translation-memory (TM) research: what do we know and how do we know it?', *Journal of language and communication studies,* 44: 89-101. Available online at [http://download2.hermes.asb.dk/archive/2010/Hermes44.html] (accessed 21 August 2013).

Cohen, Jacob (1988) *Statistical power analysis for the behavioral sciences*. New York: Psychology Press.

Coombs, Jeremy (2014) 'How a large biotechnology company teamed with a translation service provider to define best practices', *Journal of commercial biotechnology*, 20 (1): 49-53.

Daille, Béatrice (1996) 'Study and implementation of combined techniques for automatic extraction of terminology', in Judith L. Klavans and Philip Resnik (eds), *The balancing art: combining symbolic and statistical approaches to language,* Cambridge MA: MIT Press, 49-66. Available online at [http://www.aclweb.org/anthology/W94-0104] (accessed 20 October 2013).

Daille, Béatrice (2000) 'Morphological Rule Induction for Terminology Acquisition', Actes, 18th *International Conference on Computational Linguistics (COLING 2000),* Saarbrucken, Germany, 215-221.

Espasa, Eva (2004) 'Myths about documentary translation', in Pilar Orero (ed.) *Topics in Audiovisual Translation,* Amsterdam/Philadelphia: John Benjamins, 183-197.

Franco, Eliana (2000) 'Documentary film translation: A specific practice?', in Andrew Chesterman, Natividad Gallardo San Salvador and Yves Gambier (eds) *Selected Contributions from the EST Congress Granada 1998: Translation in Context*, Amsterdam: John Benjamins, 233-242.

Franco Eliana, Anna Matamala and Pilar Orero (2010) *Voice-over translation: an overview***.** Bern**:** Peter Lang.

Frantzi, Katerina T. and Sophia Ananiadou (1996) 'Extracting nested collocations', in *Proceedings of the 16th conference on computational linguistics,* 1: 41-46. Available online at [http://dl.acm.org/citation.cfm?id=992639] (Accessed 10 January 2014).

Gregory, Michael and Susanne Carroll (1978) *Language and situation,* Language and Society Series. London: Routledge and Kegan Paul Ltd.

Hanoulle, Sabien, Véronique Hoste and Aline Remael (2015) 'The efficacy of terminology-extraction systems for the translation of documentaries', *Perspectives. Studies in Translatology*, 23(3): 1-16. Doi 10.1080/0907676X.2015.1010549

Kageura, Kya and Bin Umino (1996) 'Methods of automatic term recognition. A review', *Terminology,* 3: 259-289. Available online at [https://user.phil-fak.uni-duesseldorf.de/~rumpf/WS2006/IE/Pub/KagUmi96.pdf] (Accessed 25 March 2014).

Lagoudaki, Elina (2010) 'Translation memories survey 2006. Translation memory systems: Enlightening users' perspective' Paper presented at *ASLIB International Conference on Translating and the Computer*. 2006, London. Available online at [http://mt-archive.info/Aslib-2006-Lagoudaki.pdf] (Accessed 5 September 2013).

Leijten, Mariëlle and Luuk Van Waes (2013) 'Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes', *Written Communication* 30 (3): 358-392. Available online at [http://www.inputlog.net/publications.html] (Accessed 12 September 2014).

Macken, Lieve (2010) *Sub-sentential alignment of translational correspondences* (Doctoral dissertation). Available online at [https://biblio.ugent.be/input/download?func=downloadFile&recordOId=966953&fileOId=966963] (Accessed 02 may 2012).

Macken, Lieve, Els Lefever and Véronique Hoste (2013) 'TExSIS: Bilingual terminology extraction from parallel corpora using chunk-based alignment', *Terminology*, 19(1): 1-30.

Matamala, Anna (2010) 'Terminological challenges in the translation of science documentaries: a case-study', *Across Languages and Cultures,* 11: 255-272.

Matamala, Anna, Anna Fernàndez-Torné and Carla Ortiz-Boix (2012) 'Technology and AD: The TECNACC project', *Languages and the media 2012,* Berlin.

Nichols, Bill (2001) *Introduction to documentary*, Bloomington & Indianapolis: Indiana UP.

Ortiz-Boix, Carla (forthcoming) 'Post-Editing Wildlife Documentaries: Challenges and Possible Solutions'

Ortiz-Boix, Carla and Anna Matamala (forthcoming) 'Post-Editing Wildlife Documentary Films: a new possible scenario?'

Plantinga, Carl R. (1997) *Rhetoric and representation in nonfiction film*, Cambridge: Cambridge University Press.

Renov, Michael (ed.) (1993) *Theorizing documentary*, London: Routledge.

Vintar, Špela (2010) 'Bilingual term recognition revisited. The bag-of-equivalents term alignment approach', *Terminology,* 16(2): 141-158.

Ward, Paul (2005) *Documentary: the margins of reality*, London: Wallflower.

Warburton, Kara (2013) 'Processing terminology for the translation pipeline', *Terminology*, 19(1): 93-111.

Wong, Wilson, Wei Liu and Mohammed Bennamoun (2009) 'A probabilistic framework for automatic term recognition', *Intelligent data analysis,* 13: 499-539.

Filmography

Haken, Lucy, Laura Warner and Ruth Roberts (Producers), Nick Shoolingin-Jordan and Mike Slee (Directors) (2011) 'Earth machine – Land', [Television series episode], in Lucy Van Beeck (Series producer), *Earth machine*, UK: BBC/Discovery Channel.

Hornsby, Gaby (Producer & Director) (2012) 'Horizon - The Hunt for the Higgs', [Television series episode], in Aidan Laverty (Series editor), *Horizon,* UK: BBC2.

Learoyd, Sue (Producer & Director) (2011) 'Horizon - The Secret World of Pain – Season 47 – episode 11', [Television series episode], in Aidan Laverty (Series producer), *Horizon,* UK: BBC2.

Summerill, Mary (Producer) (2011) 'Madagascar – Island of marvels', [Television series episode], in Michael Gunton (Executive Producer), *Madagascar*, UK: BBC NHU/Animal Planet.