

Antwerp, 25/08/2021

In relation to the publication G. M. Greco and A. Jankowska (eds.) (2019), *Quality in Media Accessibility*, special issue of the *Journal of Audiovisual Translation*, 2(2).

I the undersigned Anna Jankowska, as Editor-in-Chief of the *Journal of Audiovisual Translation*, declare that Gian Maria Greco acted as guest-editor, together with myself, for a special issue on “Quality in Media Accessibility” published in 2019 as issue 2, volume 2.

Anna Jankowska

Antwerp, 25/08/2021

In relation to the publication G. M. Greco and A. Jankowska (eds.) (2019), Quality in Media Accessibility, special issue of the Journal of Audiovisual Translation, 2(2).

I the undersigned Anna Jankowska, as guest-editor with Gian Maria Greco of the aforementioned special issue of the Journal of Audiovisual Translation, declare that Gian Maria Greco acted as the lead guest-editor, devised the idea of the special issue in the first place, wrote the call for papers, and performed the lead scientific responsibilities related to the role.

Anna Jankowska





Vol. 2 No. 2 (2019): Special Issue: Quality in Media Accessibility

DOI: <https://doi.org/10.47476/jat.v2i2> 

Published: 2019-12-31

Introduction

Framing Media Accessibility Quality

Gian Maria Greco, Anna Jankowska

1-10

 PDF

Research articles

Finding the Right Words

Investigating Machine-Generated Video Description Quality Using a Corpus-Based Approach

Sabine Braun, Kim Starr

11-35

 PDF

Feasibility, Quality and Assessment of Interlingual Live Subtitling

A Pilot Study

Hayley Dawson

36-56

 PDF

Usability as a Premise of Quality

First Steps Towards the Validation of the System Usability Scale (SUS) into Spanish

Irene Tor-Carroggio, Daniel Segura, Olga Soler-Vilageliu

57-71

 PDF

Translating Code-Switching on the Screen

Spanglish and L3-as Theme

Montse Corrius, Patrick Zabalbeascoa

72-91

 PDF

Museums as Disseminators of Niche Knowledge

Universality in Accessibility for All

Alessandra Rizzo

92-136

 PDF

Practice reports

Template Files:

The Holy Grail of Subtitling

Panayota Georgakopoulou

137-160

 PDF

Information

[For Readers](#)

[For Authors](#)

[For Librarians](#)

Framing Media Accessibility Quality

 **Gian Maria Greco** 

Autonomous University of Barcelona, Spain
University of Vigo, Spain

 **Anna Jankowska** 

Jagiellonian University in Kraków, Poland

Abstract

The article provides an initial general overview of the status of quality in media accessibility. After highlighting some of the reasons behind the importance of addressing quality in media accessibility, the article discusses some problems that undermine the potential for full maturation of research and practices on quality. Then, it presents some possible solutions and proposes to use “media accessibility quality” to refer to the overarching problem of quality in media accessibility. The article concludes by listing a set of actions that constitutes a first draft of an agenda for the future of media accessibility quality.

Key words: accessibility studies, audiovisual translation, media accessibility, media accessibility quality, quality.

Citation: Greco, G. M. & Jankowska, A. (2019). Framing media accessibility quality. *Journal of Audiovisual Translation*, 2(2), 1–10.

Copyright: ©2019 Greco & Jankowska. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. Introduction

Quality is one of the first topics collectively identified as a major issue in the field of translation studies. The official recognition of quality as a critical issue by the international translation community can be traced back to at least the third congress of the International Federation of Translators held in 1959, under the heading, “quality in translation”. In the book’s preface, summarising the preparatory discussions that led to the choice of the topic, Caillé (1963b) states:

After weighing the pros and cons for a long time, we fixed our choice on the theme of quality. The word opened such horizons that it lost its clarity. We knew it. However, it was right in the centre of the imperatives of the profession. It was a bit of a geometric locus – in abstracto – and had the merit of provoking the immediate reaction of translators as well as those who needed them. The term ‘quality’ was also the richest in extensions. It required the study of the entire mechanism of the profession, it involved looking into the training of translators, information, documentation, the exchanges of translators between countries, delivery times for a work, defending copyright, the remuneration for translated works, etc. (p. xiv)

If we look at the articles included in the over 550-page proceedings, we can identify many of the questions that are still under debate nowadays from the subjectivity of translation quality to the strive for the identification of objective measures, from fidelity to adaptation, from the impact of limited time and low pay on the work of professionals – and thus on the quality of the work – to non-professionalism (Cary & Jumpelt, 1963). Since then, quality has become one of the most hotly-discussed issues in translation studies, as testified by the massive number of monographs, edited books, journals’ special issues, articles, and conferences. A simple search on Google Scholar for “translation quality” gives 38.700 results.¹ Obviously, this search does not have any substantial bibliometric value, and thus does not accurately represent the status of the literature on the topic. It would need to be refined with more precision; for example, by excluding the publications where the search string occurs but that are not relevant while including those that discuss the topic but do not use that string. Still, the figure gives us a sense of just how significant a topic quality is in translation studies. Its prominent status has been reinforced by the advent of the information revolution, and it is now taking on a new life with the increase in technological developments, such as artificial intelligence, whose widespread adoption may trigger radical changes in our world.

Reaching a similar outlook is much more difficult for audiovisual translation (AVT) and media accessibility (MA). When searching for “quality + ‘audiovisual translation’” and “quality + ‘media accessibility’”, Google Scholar retrieves 4.720 and 1.960 entries, respectively.² Once more, the results are by no means representative of the situation and should be refined. A direct comparison with translation studies would also not be fair, given that MA and AVT are new

¹ Retrieved December 27, 2019.

² Retrieved December 27, 2019.

additions to the knowledge enterprise club. Nonetheless, given the low number of results, at least a bit more insight can be gained for the case of MA, being the one with the lowest number of entries and thus easier to analyse. Almost half of the entries ($n=928$) refer to works produced in the past five years (2015-2019), and only the 6% of them ($n=58$) are not relevant. We can see that of late, the topic of quality has become of increased significance among scholars. One more sign of the intensifying interest is the UMAQ conference held in Barcelona in June 2018, the first one ever explicitly devoted to quality in MA. It attracted so much interest that a second edition has been already scheduled to be held at the University of Vigo in September 2020.

This special issue of the *Journal of Audiovisual Translation* marks the first collective publication that deals specifically with quality in MA and AVT. Some of the articles included in this special issue discuss the topic of quality only marginally. The choice is not accidental. They accurately represent the breadth and complexity of the issue of quality as well as the current fragmented status of the debate, which we will further delve into in the following pages. With this special issue, the *Journal of Audiovisual Translation* has created a new section called *Practice Reports*. Once again, the choice is not accidental. Despite the complexity involved in specifying what constitutes quality, a multitude of MA actors – such as practitioners, companies, service providers, and users – deal with the issue of quality on a daily basis (Pedersen, 2017). Some of them have developed their own procedures, methods, models, and metrics, which are continuously tested and modified and are thus of great use in helping address quality.

Since the specific contribution of each article is clearly summarised in its own abstract, we will dedicate the following pages to a brief discussion of several of the general issues at hand. More specifically, in section two we will highlight the importance of addressing quality in MA. In section three, we will discuss some problems that undermine the potential for full maturation of research and practices on quality in MA, and we will discuss some possible solutions. Finally, in section four we conclude with a list some actions that constitute a first draft of an agenda for future research.

2. The Importance of Quality in Media Accessibility

Initially born, at least academically, as an area of AVT, MA has been steadily growing into an area of accessibility studies, becoming a driver for major social change. As proposed by Greco (2018), a prominent reason behind this process can be identified in the three shifts that MA has been experiencing over the past few years: a shift from particularist accounts to a universalist account of accessibility, from maker-centred to user-centred approaches, and from reactive to proactive models. Firstly, for many years MA was framed as the subdomain of AVT concerned with specific modalities related to persons with sensory disabilities. More recently, scholars have been increasingly embracing a universalist account that challenges traditional notions of AVT and MA, and sees MA as the area of accessibility studies that focuses on “access to media and non-media

objects, services and environments through media solutions, for any person who cannot or would not be able to, either partially or completely, access them in their original form” (Greco, 2019, p. 18). Secondly, though previously ignored or downplayed for years, the involvement of users has become a bedrock of MA. Users are now seen as bearers of unique knowledge that is essential for both the design and the evaluation of access solutions. And lastly, while MA solutions have long been regarded as an add-on and, as such, relegated to the end of the production process of an artefact, they are now at the centre of practices that promote their full integration into that very process, jumping from the ex-post to the ex-ante phase.

The change of focus from quantity to quality can be seen as connected to and influenced by the new position reached by MA as a consequence of the aforementioned shifts. For quite some time, research and industry have focused on the development and implementation of access services, while policies have focused on promoting the widespread adoption of those services and setting quotas (Romero-Fresco, 2016). However, over the past few years, the different actors, especially scholars, have increasingly started to turn their attention to the quality of those services, through the development of quality models (e.g., Pedersen, 2017; Romero-Fresco & Pérez, 2015; Romero-Fresco & Pöchhacker, 2018), the analysis of quality in production practices (e.g., Robert & Remael, 2016), and the discussion of the impact of machine translation on the quality of access services (e.g., Doherty & Kruger, 2018), just to name a few cases. As stated above, the rise of quality as a major concern can be connected to the three major shifts MA has been experiencing over the past few years. A case in point is the shift from maker-centred and expert-centred to user-centred approaches, which is evident in the recent increase of reception studies within MA (Di Giovanni & Gambier, 2018; Jankowska, 2019). In order to investigate the quality of MA services and products, identify critical issues, devise solution-oriented strategies, and define and test quality metrics, researchers have been conducting more and more tests with end-users using methods from other research fields, such as experimental psychology.

Given the situation briefly summarised so far, two questions stand out. What are the factors driving the move towards quality? And, why should quality matter in MA? A detailed answer to these questions would require more space than the one at our disposal. For the scope of this article, it is sufficient to draft some possible answers. As for the first question, let us sketch only some of the many factors that are guiding the rise in prominence of quality in MA. One factor is the recent change of pace in international, especially European, policies, with the publication of documents that either directly address or indirectly affect MA. Just consider the new Audio Visual Media Services Directive³ and the European Accessibility Act⁴. A second factor is the growing relevance of standards and their very practical influence on the daily life of individuals. Consider the increasing involvement of MA researchers in the work of standardisation organisations (Matamala & Orero, 2018) and the recent

³ <https://ec.europa.eu/digital-single-market/en/revision-audiovisual-media-services-directive-avmsd>

⁴ <https://ec.europa.eu/social/main.jsp?catId=1202>

release of a series of standards on accessibility, such the standard on “Accessibility requirements for ICT products and services”⁵ and one on “Design for All”⁶. A third factor can be identified in the social transformations brought about by information and communication technologies. They are radically changing the patterns of access to the world and making MA services more and more necessary instruments for guaranteeing access for all (Graham & Dutton, 2014; Greco, 2018). Finally, the second and third shifts mentioned previously have also significantly contributed to making quality a priority. For example, the search for ways to increase the quality of experience offered by MA services is strictly connected with the inquiry into the integration of those services in the production process (Fryer, 2018; Romero-Fresco, 2019).

In order to answer the second question however, we should look more closely at the first shift in MA, or better still, at its general standing. The shift from particularist accounts to a universalist account of MA mirrors and is a consequence of a more general shift concerning accessibility within the human rights framework. It is precisely in this shift that we can identify the core reason behind the importance of the topic of quality in MA. The most commonly accepted justification for human rights is that they are instruments for human dignity. In order to guarantee human rights, and thus human dignity, we need to provide access to the various material and immaterial goods that are the objects of human rights, that is, that are considered to be fundamental for a dignified life. This means that access is a necessary requirement for the fulfilment of human rights and the respect of human dignity (Greco, 2016). However, merely providing access services – i.e., quantity – is not sufficient. In a hypothetical world where access services are so widespread to pervade every aspect of life, human rights and human dignity could still be at risk if those access services do not provide an equitable experience to all. It is precisely in the experiential dimension that the importance of quality lies.

3. Quality in Media Accessibility: Some Problems

Despite an ever more dynamic situation, research on quality in MA is facing the risk to be undermined by a series of foundational and methodological problems that may hamper its growth to a mature stage. Openly addressing these problems should be a priority for scholars as well as for other stakeholders. If not tackled through a collective commitment in a timely manner, they may very well jeopardise the successful implementation of research results into practice, and ultimately run the risk of reinforcing old – or even producing new – inequalities.

⁵ https://standards.cen.eu/dyn/www/f?p=204:110:0:::FSP_PROJECT,FSP_ORG_ID:66602,855949&cs=1EDAD9D32175767F919BCC673EA996CFD

⁶ https://standards.cen.eu/dyn/www/f?p=204:110:0:::FSP_PROJECT,FSP_ORG_ID:62323,2301962&cs=1D28CFDC66E7CEF3CE441294CAA9FEABE

The most pressing problem is that of fragmentation. So far, attention has been spread out over a very diverse range of issues while lacking a cohesive framework for the investigation of quality in MA. To be clear, there is nothing wrong with focusing on minute aspects of a wider problem and trying to solve them. Actually, it could be quite a commendable strategy, if only put into the right structure. The Cartesian method of dividing a problem into smaller parts presupposes the existence of such a problem to divide. Without being contextualised in relation to a bigger problem, all the efforts put into addressing the small issues may risk being done in vain, or to have minimal effect, or to be even counter-productive. Until now, scholars have limited their efforts to address quality to only some modalities or contexts, and often exclusively on specific aspects of these. The general problem of what quality is in MA has been largely ignored. Actually, even the smaller problem of what quality is for those modalities or in those contexts has not been extensively addressed; and research has usually focused on some specific dimensions of quality for those modalities.

However, posing the overarching problem of quality in MA – as well as posing the general problem of quality for each MA service – and being aware of its complexities and implications is crucial for having a cohesive framework that could then support a solid inquiry into the small issues. Having such a unified framework for the investigation of quality in MA does not entail that all efforts should be focused on the formulation of a unified theory of quality in MA. Quality is an inherently vague term. Any attempt to formulate a unified theory of quality would be pointless or even detrimental, even in the case of a single MA service. However, a common framework would facilitate the correlation between the different aspects involved as well as encourage a harmonised discussion on quality in MA. It could also act as an ideal place where researchers might be able to coordinate their efforts in order to provide a more efficient and interconnected account of the various issues connected to quality.

A second problem is related to the relationship between MA and translation studies. In the context of MA, this means that a good portion of the research carried out so far is still too strictly based on, or even mirror with minimal adaptation, models and metrics from translation studies. More generally, research in MA still addresses the problem of quality adopting a translation-based and translation-centred approach. From a historical point of view, this is comprehensible. Academically, MA was born as a subfield of AVT, and thus of translation studies. The significant influence of translation studies should therefore not surprise. However, this view considers access problems as merely a case of a more general class of translation problems. As discussed in Greco (2019), this is rather controversial and could lead to a series of abuses and misuses of both translation and accessibility. AVT and MA do not overlap, but rather, intersect. This means that quality issues in MA cannot be reduced to AVT quality issues. The latter do not satisfactorily cover the problem of access, which on the contrary is the ultimate question in MA. Consider audio description. So far research on quality has focused on the most diverse aspects of a film as an audiovisual text, that is, it has approached audio description from a translation-based and translation-centred approach. Nontranslation-based aspects, such as delivery (e.g., choice of voice, pace, and intonation) and reproduction (e.g., sound mix)

have been deemed as secondary or even ignored. In MA, where the aim is to provide access to an equitable experience, these aspects are not secondary at all. The new position of MA as an area of accessibility studies, that only intersects with translation studies, requires us to start looking at the issue of quality from other points of view beyond translation (Greco, 2019; Greco & Jankowska, in press).

A third problem concerns the nature of quality. Despite its vagueness, or more properly as a way to deal with it, there is a widespread agreement to interpret quality as a multidimensional construct. According to this view, quality is a function of different variables, i.e., its dimensions. The multidimensionality of quality has received scarce attention in MA. Thus far, research has focused on one or a few dimensions of quality for some specific MA service, such as accuracy, segmentation, or synchronisation. As we said, such a focus on small problems is a commendable enterprise from a Cartesian point of view. However, if not explicitly clarified that the one under scrutiny is part of a set of dimensions, there is the risk of conducting research on some dimension of an MA service but then drawing conclusions at the general level of quality for that service. Such a risk is also faced by some of the various models produced so far for evaluating quality in the context of some MA service, which merely mention other dimensions. Consequently, they are not models for the evaluation of quality per se, but only for the evaluation of some dimensions. Nonetheless, they are often perceived as being models about (overall) quality, as a consequence of the classic synecdochal move of confusing a part for the whole. Moreover, the dimensions ignored or minimised are often the ones that do not involve forms of translation, which take us back to the problem of the dominion of translation-centred approaches to quality in MA.

A fourth problem is, *prima facie*, terminological. It concerns the lack of a specific noun for indicating the general topic of quality in the context of MA. Something similar to *translation quality* in the field of translation studies. We suggest to use *media accessibility quality* (MAQ). While seeming like a trivial issue, it is far from that; and it is not a mere terminological issue either. It connects especially with both the influence of translation-based approaches and the fragmentation of efforts. Adopting MAQ to refer to the general problem of quality in MA would help to keep present that whatever the specific topic at hand, it is part of a wider context. For example, in a paper that presents the research results on some dimension of quality in subtitling, just mentioning MAQ would prompt the readers to recall the general context within which they should frame both the research and its results. Starting to refer to MAQ may also facilitate a more systematic approach to the issue of quality in media accessibility, encouraging more coordinated efforts in research and practices. Most importantly, it would act as a reminder that, whatever the specific issue addressed at a certain point might be, it should be matched upon the problem of access, that is, the ultimate question of MA.

4. Conclusion: A First Draft of an Agenda for Media Accessibility Quality

Given the picture drawn in the previous pages, in this concluding section we will briefly suggest a series of lines of action that await MAQ in the course of time, and that we believe would benefit from a more systematic and coordinated discussion as per the terms discussed in the previous pages. Some of the actions mirror the problems analysed in the previous section. As in that case, the following list should not be considered exhaustive either. Our main goal is to raise awareness and encourage further work.

A first action concerns the very concept of quality. As we mentioned, quality is inherently vague. This is not ignored in MA. Yet, it should be acknowledged even more explicitly. Acknowledging – and reminding more often – the vague nature of quality could, for example, save us from falling into the trap of considering some model as *the* model.

A second action concerns the fact that dealing with quality means navigating the treacherous waters between the Scylla of objectivity and the Charybdis of relativity. This is an issue already extensively debated in other fields but only marginally in MA. It would mean, for example, trying to justify the choice of some specific dimensions. Is that choice completely arbitrary? Is it possible to find some mechanism that could confer some forms of objectivity?

This point leads to a third action: dealing with the lack of proper acknowledgement of the agency distribution involved in the definition of both quality and models for its evaluation in specific MA services. This means, for example, investigating the role of the human factor in the definition and assessment of quality, dealing with the differences among stakeholders, defining mechanisms for addressing the divergent perspective and expectations of the various stakeholders, and, especially, providing a justification for such choices.

A fourth action concerns the need make quality models and research results user-friendly (Romero-Fresco, in press). That is, finding ways to (a) make research results and models understandable by the different stakeholders, especially users; and (b) improve how we communicate research results to and how we engage with the other stakeholders. So far this has been done sporadically. We need to make it a regular activity.

A fifth action is to investigate the usefulness of adopting the distinction between *quality of service* and *quality of experience*, following the example of other fields.

A sixth action concerns the systematic extension of research on quality to all the different end users' groups – such as children, younger adults, the elderly, migrants – and not only to persons with disabilities. This means that future investigation on quality should also address the role played by intersectionality in both the definition and the assessment of quality in MA.

A seventh action is related to the pedagogy of MA (Greco, in press). More attention should be paid to (a) the investigation of the influence of current education and training programmes on how the question of quality is framed in MA, and (b) how the issue of quality should be presented and discussed in education and training curricula. An eighth action involves looking at how quality has been addressed in other fields. The development in time of new, ad hoc approaches, and thus methods and metrics, that respond to the specificities of MA is not a remote possibility, though one that would require substantial efforts. However, in line with the interdisciplinary nature of accessibility studies and thus of MA, a more practicable path would be to look at the theories, methods, and metrics produced in other fields that face the problem of quality and investigate if and how they can be applied, upon due modification, to the context of MAQ.

In his opening speech delivered at the congress of the International Federation of Translators we mentioned in the introductory section, Caillé (1963a) states: “in a world where translation is increasingly used, quality is a basic necessity and our touchstone” (p. 8). Paraphrasing his words, we believe that in a world where accessibility is increasingly becoming and used as an essential tool for human dignity, quality is indeed a basic necessity and should be the touchstone of all research and practices in accessibility studies, including MA.

Acknowledgements

We express our gratitude to the authors and the reviewers for their contributions to this special issue. Gian Maria Greco is the first author of this article. Teresa Canosa skilfully copyedited the final version. The research presented in this article as well as this special issue have received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 752659 and Mobility Plus Grant no 1311/MOB/IV/2015/0 of the Polish Ministry of Science and Higher Education for the years 2016-2019. Gian Maria Greco is a member of the TransMedia Catalonia research group, funded by the Catalan government under the SGR funding scheme (2017SGR113), and the GALMA research group, funded by the Galician government under the scheme Proxecto de Excelencia 2017.

References

- Caillé, P.-F. (1963a). Discours inaugural. In E. Cary & R. W. Jumpelt (Eds.), *Quality in translation. Proceedings of the IIIrd congress of the International Federation of Translators (FIT), Bad Godesberg Germany, 1959* (pp. 5-10). Oxford: Pergamon Press.
- Caillé, P.-F. (1963b). Préface. In E. Cary & R. W. Jumpelt (Eds.), *Quality in translation. Proceedings of the IIIrd congress of the International Federation of Translators (FIT), Bad Godesberg Germany, 1959* (pp. v-xviii). Oxford: Pergamon Press.

- Cary, E., & Jumpelt, R. W. (Eds.). (1963). *Quality in translation. Proceedings of the IIIrd congress of the International Federation of Translators (FIT), Bad Godesberg Germany, 1959*. Oxford: Pergamon Press.
- Di Giovanni, E., & Gambier, Y. (2018). Introduction. In E. Di Giovanni & Y. Gambier (Eds.), *Reception Studies and Audiovisual Translation* (pp. VII-XII). Amsterdam: John Benjamins.
- Doherty, S., & Kruger, J.-L. (2018). Assessing quality in human-and machine-generated subtitles and captions. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment* (pp. 179-197). Cham, Switzerland: Springer.
- Fryer, L. (2018). Staging the audio describer: An exploration of integrated audio description. *Disability Studies quarterly*, 38(3).
- Graham, M., & Dutton, W. H. (2014). *Society and the Internet. How networks of information and communication are changing our lives*. Oxford: Oxford University Press.
- Greco, G. M. (2016). On accessibility as a human right, with an application to media accessibility. In A. Matamala & P. Orero (Eds.), *Researching audio description. New approaches* (pp. 11-33). London: Palgrave Macmillan.
- Greco, G. M. (2018). The nature of accessibility studies. *Journal of Audiovisual Translation*, 1(1), 205-232.
- Greco, G. M. (2019). Accessibility studies: Abuses, misuses and the method of poietic design. In C. Stephanidis (Ed.), *HCI 2019* (pp. 15-27). Cham: Springer.
- Greco, G. M. (in press). Towards a pedagogy of accessibility. The need for critical learning spaces in media accessibility education and training. *Linguistica Antverpiensia*, 18.
- Greco, G. M., & Jankowska, A. (in press). Media accessibility within and beyond audiovisual translation. In Ł. Bogucki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation and media accessibility*. Cham: Palgrave Macmillan.
- Jankowska, A. (2019). Audiovisual media accessibility. In E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.), *The Bloomsbury companion to language industry studies* (pp. 231-259). London: Bloomsbury Academic Publishing.
- Matamala, A., & Orero, P. (2018). Standardising accessibility: transferring knowledge to society. *Journal of Audiovisual Translation*, 1(1), 139-154.
- Pedersen, J. (2017). The FAR model: Assessing quality in interlingual subtitling. *Journal of Specialised Translation*(28), 210-229.
- Robert, I., & Remael, A. (2016). Quality control in the subtitling industry: an exploratory survey study. *Meta: Journal des traducteurs/Meta: Translators' Journal*, 61(3), 578-605.
- Romero-Fresco, P. (2016). Accessing communication: The quality of live subtitles in the UK. *Language & Communication*, 49, 56-69.
- Romero-Fresco, P. (2019). *Accessible filmmaking: Integrating translation and accessibility into the filmmaking process*. London: Routledge.
- Romero-Fresco, P. (in press). Negotiating quality assessment in media accessibility: The case of live subtitling. *Universal access in the information society*.
- Romero-Fresco, P., & Pérez, J. M. (2015). Accuracy rate in live subtitling: The NER model. In R. Baños Piñero & J. Díaz Cintas (Eds.), *Audiovisual translation in a global context* (pp. 28-50). London: Palgrave Macmillan.
- Romero-Fresco, P., & Pöschhacker, F. (2018). Quality assessment in interlingual live subtitling: The NTR Model. *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, 16, 149-167.

Finding the Right Words: Investigating Machine-Generated Video Description Quality Using a Corpus-Based Approach

 Sabine Braun 

University of Surrey

 Kim Starr 

University of Surrey

Abstract

This paper examines first steps in identifying and compiling human-generated corpora for the purpose of determining the quality of computer-generated video descriptions. This is part of a study whose general ambition is to broaden the reach of accessible audiovisual content through semi-automation of its description for the benefit of both end-users (content consumers) and industry professionals (content creators). Working in parallel with machine-derived video and image description datasets created for the purposes of advancing computer vision research, such as Microsoft COCO (Lin et al., 2015) and TGIF (Li et al., 2016), we examine the usefulness of audio descriptive texts as a direct comparator. Cognisant of the limitations of this approach, we also explore alternative human-generated video description datasets including bespoke content description. Our research forms part of the MeMAD (Methods for Managing Audiovisual Data) project, funded by the EU *Horizon 2020* programme.

Key words: computer vision, machine learning, accessibility, audiovisual content, audio description, content description, content retrieval, video description, audiovisual translation, MeMAD.

Citation: Braun, S. & Starr, K. (2019). Finding the right words: Investigating machine-generated video description quality using a corpus-based approach. *Journal of Audiovisual Translation*, 2(2), 11–35.

Editor(s): G.M. Greco & A. Jankowska

Received: October 01, 2019


Accepted: November 19, 2019

Published: December 31, 2019

Funding: This publication is part of the EU funded project MeMAD, grant agreement number 780069.

Copyright: ©2019 Braun & Starr. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

 s.braun@surrey.ac.uk, <https://orcid.org/0000-0002-6187-3812>

 k.starr@surrey.ac.uk, <https://orcid.org/0000-0001-5236-1535>

1. Background

Audio description (AD) has established itself as a media service which facilitates access to audiovisual content for visually impaired audiences. Relying heavily on human resource, AD is currently an expensive part of the post-production process for traditional media companies, making it challenging to provide comprehensive media access (Sade, Naz, & Plaza, 2012, p. 270). The recent increase in user-generated audiovisual content has created a further challenge for media access. In parallel, research on automating the description of still images (“image captioning”) and video scenes (“video captioning”) has intensified and has begun to show moderate success (Krishna et al., 2017; Aafaq et al., 2019). The question of whether and to what extent these automated methods of description can be drawn upon to produce AD in order to reduce costs, and broaden media access without compromising quality, is therefore an economically and socially important question for research (Rohrbach et al., 2015b, p. 1).

Equally as important in this debate is the contribution to be derived from human video description analysis which has the potential to propel computer vision beyond standard object-and-action recognition tasks into the realm of multi-character, sequentially relayed narrative. However, since this is uncharted territory, new methods are required to bridge the human-computer void, bringing together the scientific world of algorithms and feature-extraction models with a humanities approach to cognition and understanding through a typically human lens.

The first step along this road is to form a deeper awareness of the distance that exists between human and machine, starting with a study of the main characteristics of automated image and video captions. Identifying patterns and behaviours that appear atypical in the context of human understanding allow us to isolate those areas of computer vision development which require most attention.

The current quality of image and video captioning and the absence of principles guiding the development of such captions raises ethical and quality issues from the customer perspective, as well as questions about the relevance and value of the “human voice” in this process. The “human” role is one aspect that the H2020 Project MeMAD (Methods for Managing Audiovisual Data: Combining Machine Efficiency and Human Accuracy)¹ is currently investigating. Our primary research focus in this project is to explore how our knowledge of human approaches to relevance and saliency in information selection can be used for modelling and improving the automation of video captioning in the fields of: (i) archive media retrieval; and (ii) AD for audiovisual media consumers. Whilst these practices overlap to some extent, the main driver for producing content descriptions for archival purposes is the likelihood of the re-use of the content internally or for re-sale to other media companies. Content descriptions for archival purposes therefore tend to be more “literal” or factual

¹ 2018–2020, Grant number 780069.

than AD, especially AD for filmic drama and movies, which is often “narrative” or figurative (Kruger, 2010; Ramos Caro, 2016). Whilst the (semi-)automation of content description is therefore likely to be a more achievable goal in the shorter term than a model for generating elaborate audio descriptions, the MeMAD project has adopted a two-pronged approach exploring content description for retrieval guided by work with broadcast archive journalists, as well as exploring how human knowledge can be used to support the (semi-)automation of video captioning in the context of enhancing traditional AD user experience. This paper focuses on the latter, while mindful of the former.

The initial expectation in the project was to harness human AD to inform the development of semi-automated solutions. A corpus-based approach was deemed appropriate, aimed at identifying patterns in human AD that are particularly relevant for the modelling of auto-generated descriptions. However, few AD corpora have been compiled to date, and even fewer are publicly available (Salway, 2007; Jimenez & Seibel, 2012; Rohrbach et al. 2015a; Matamala 2019). Preparations to compile our own corpus showed that differences in stylistic factors, density and granularity of available AD meant much current TV production content is of limited use to the audio extraction processes originally envisaged in the project. For example, while TV drama contains useful descriptions of narrative action which give insight into human meaning-making in story-telling, the extent of the AD is constrained by quick-fire direction (multiple short scenes and rapid shot-changes) and a shortage of audio hiatuses, such that the corresponding AD is minimal and largely a vehicle for announcing changes of location (“in the pub...”) or introducing new characters (“Bernadette and Tiffany arrive”). Other TV genres also proved problematic. Documentaries, for example, generally lack a clear narrative within the AD, which performs the task of overlaying supplementary factual information where this is visually relayed. By contrast film productions, due to their long-form narrative exposition, lend themselves to more elaborate and narratively sophisticated storytelling and AD scripting, with opportunities for the describers to paint an audio picture which does more than merely label the characters and their locations. This greater emphasis on explication in film storytelling is frequently matched by a richer lexicon and more complete descriptions than would be found in a standard television production. Lexically rich descriptions and contextualisation made feature-film AD a better candidate for inclusion in a corpus created specifically for our study. However, while AD has a perceived value in the context of informing machine-generated video descriptions, our pilot stage illustrated that extracting comprehensive visual information from AD can still prove problematic.

Irrespective of the differences between different audiovisual genres, in any material the absence of suitable hiatuses in the audio track, along with the “golden rule” of AD that prohibits interruptions to the original sound track (Hyks, 2005), often limits the extent to which any supplementary visual information can be inserted into the source material. In the context of human comprehension this is not problematic. AD is not a stand-alone text; its purpose is to facilitate meaning-making in conjunction with the primary audio track containing dialogue, narration, sound effects, and musical scoring (Braun, 2011). It capitalises on the human ability to assimilate texts and sensory

input by building mental models, establishing salience and relevance, and engaging skills of anticipation, inference and retrospective self-correction to retrieve the unsaid and the ultimately intended meaning (Braun, 2016; Fresno, Castellà, & Soler-Vilageliu, 2016; Vandaele, 2012). This, in turn, like any other language mediation activity, encompasses an element of interpretation and subjectivity. Unsurprisingly, therefore, rule-based methodologies for arriving at audio described outputs have largely eluded AD producers and researchers (Audetel/ITC, 2000; AENOR, 2005), as there is a lack of consensus between describers about what should be included and omitted (Vercauteren, 2007, p. 139; Yeung, 2007, p. 241; Ibanez, 2010, p. 144) and considerable variation between describers in the lexical breadth with which they choose to describe the selected elements (Matamala, 2019).

Computer vision algorithms, by contrast, currently lack complex inferential capacity. Large-scale captioned image and moving image datasets of the type used for machine learning are not sufficiently numerous, sizeable or broad-reaching to bridge this gap. For example, while most available datasets (COCO, TGIF, Visual Genome, Rohrbach's MPII-MD, Hollywood II) include still images or limited moving images, their application to training machines for the purposes of moving image description research is curtailed by the limited number of examples of each type of action or movement available. Whilst there are advances in parallel fields (e.g., task-driven facial recognition, emotion recognition, action detection etc.), the transferability of these different strands of research to narrative audiovisual content such as film is still a very challenging task.

What emerges from this is two-fold. On the one hand, existing training datasets for machine learning are not entirely relevant to the description of narrative audiovisual content. On the other hand, the highly idiosyncratic and individualistic nature of human AD suggests that it alone cannot provide sufficient data from which to elicit patterns that can inform and guide the automated production of human-like descriptions. In order to meet the requirements of the MeMAD project, namely, combining human knowledge of describing audiovisual content with machine learning and computer vision approaches, it became necessary to look elsewhere for human-produced descriptions of audiovisual content that can be used to identify patterns and strategies of human approaches. In short, the solution was to employ simpler human-produced "content descriptions" (non-interpretative) which more closely matched the types of description the machines are currently capable of producing (non-interpretive, observational, object/action oriented).

This paper outlines our approach to selecting and compiling appropriate corpora for this purpose and reports the outcomes of an initial comparison of human and machine-generated descriptions with regard to the quality of the descriptions. The final section will then discuss our findings and draw attention to the social and ethical implications that arise from these findings with regard to the automation of audiovisual content descriptions in the context of media accessibility.

2. Approaches to Analysing Video Captions

Addressing the first task, as outlined above, i.e. that of analysing auto-generated video captions and comparing them with human-generated descriptions in order to understand their structure and their current limitations led us to a corpus-based approach and the compilation of human descriptive corpora that are comparable with machine description outputs. For the reasons discussed above, this began with scrutiny of audio description texts. At first reckoning audio description appears the ideal candidate to fulfil the comparative brief as a linguistically and structurally sophisticated elaboration of the visual aspects of film material. Machine-generated video descriptions capture visual elements such as objects, characters, actions, locations and certain basic facial expressions, in a manner that is ostensibly similar to those selected by the human describer. However, the level of complexity in the narrative created by the audio describer far outweighs the lexically and syntactically naïve constructs currently produced by even the most advanced neural network model. Furthermore, the human being draws on cognitive skills to infer what cannot be explicitly included in the AD due to time limitations which are likely to be beyond reach in the field of computer vision for the foreseeable future. As pointed out above, an alternative, plainer version of human description was therefore deemed to be an important stepping stone in creating a multimedia corpus which promotes direct linguistic comparison between professional audio descriptions, human-generated content descriptions and machine-generated descriptions. In addition, the type of audiovisual material to be used for this comparison needed to be considered carefully. As pointed out above, the genre of feature films offers the most complete and elaborate AD but is likely to be too complex for the current state of video captioning. This section explains our approach to the comparative analysis, i.e. our solution for the selection of audiovisual material, and the approaches to, and benefits of, creating different corpora of human descriptions, i.e. an AD corpus and a corpus with a “plainer” content description.

2.1. Creating the MeMAD Video Corpus (MVC)

As stated above, feature films were selected for our study because of their professional quality audio description and narratively challenging content. Since large-scale “off the shelf” audio description corpora were not freely available, feature films which are already in the public domain and contain reliably accurate AD tracks, seemed a feasible alternative. Clearly, long-form and complex narrative of the type found in feature films is a giant leap for automated film captioning given the present state of the art, not least because concepts like sequencing and cohesion are absent. Nevertheless, a work-around for this problem was inspired by advances in automated visual storytelling (Huang et al., 2016) whereby short stories were devised by captioners using sets of five consecutive photos for the purposes of training the machine to orchestrate narrative. Our solution was to break down each of the feature films in our corpus into smaller, self-contained narrative units (somewhat similar to the short sequence photo experiment) with which, it was hypothesised, the machine might more successfully engage.

These took the form of stories-within-a-story (micro-narratives), containing clear, narratively significant beginning and end-points, and illustrating elements of crisis and resolution. However, the intention was that each “story-arc” would be treated in isolation for the most part, without recourse to the greater insights available in the storyline beyond the micro-narratives themselves. In total, 501 extracts were studied from across a body of 44 feature length films, with each extract representing one brief micro-narrative (story arc) of between 10 seconds and 2 minutes’ duration. Selection of an extract was dependent on there being a minimum of five separately identifiable images or actions across the duration, in order that the computer might detect change.

Mindful of the lack of sophistication in current machine-generated video descriptions, we selected examples of basic social interaction as the focus of our data mining exercise. Uniform parameters were applied to the selection of story arcs in order to standardise the dataset, and facilitate meaningful comparison and evaluation between human descriptions and those produced by machine learning techniques:

Table 1.

Story Arc Parameters

Category	Criteria	Observations
Source Text	Must contain audio description	Required to explore value of AD for informing computer-generated descriptions
Persons	1 or 2 principal characters	Incidental characters and small groups of people in the background of shots also permitted.
Actions	Minimum of 4 or 5 simple, common actions	e.g., sitting, running, talking, walking, hugging, kissing
Duration	20 seconds – 3 minutes	Limited duration story arcs should simplify sequence modelling
Storyline	Self-contained micro-narrative	e.g., initiating action/crisis, proposed solution, action based on solution, consequence, result
Objects	Unlimited	Although no limitation was put on the number of objects in an extract, only those objects regarded as key to the action were included in our annotations

A sample story arc, *Boy in a Field*, taken from the film *Little Miss Sunshine*, is illustrated in Figure 1. At the beginning of the extract a dispute arises between a teenage boy and his family. The dispute is subsequently resolved by the intervention of a young female family member. Screenshots of narratively key frames from the scene sit alongside a brief description of the action, provided in linear fashion:

Figure 1.

Sample Story Arc: Boy in a Field (Little Miss Sunshine)



On a family road trip, a teenage boy (Duane) discovers he can no longer follow his dream of becoming a fighter pilot. He demands the camper van the family are travelling in is stopped, and he jumps out. Refusing words of comfort from his mother, he runs into an empty field, and sits down alone, to contemplate his future.

Duane's young sister (Olive) offers to talk to him. She leaves the rest of the family back at the roadside and walks down a grassy slope towards her brother.

Olive crouches down behind Duane, and without speaking ...

... **puts** an arm around him, leaning her head tenderly on his shoulder.

Comforted by her presence and the knowledge that she truly understands his despair, Duane relinquishes his anger. They both rise ...

... **and** walk back towards the roadside where the rest of the family are waiting for them.

In a sentimental, reciprocal declaration of affection, Duane resumes his role as 'big brother', carrying his little sister up the sharp incline near the road.

2.2. Audio Description

The audio descriptions were captured and transcribed as text from the audio descriptive track delivered in parallel with the selected film productions comprising the MeMAD Video Corpus (MVC). As such, this material was produced by professional audio describers and their scripts represent interjections typical of the kind advocated by film production companies (i.e. dialogue-hiatus bound, narratively-driven, cognitively accessible). It was initially anticipated that such elaborate descriptions would provide information salient to the visual aspects of each film production against which the veracity and value of machine-derived descriptions created from the same source material might be assessed. However, not only is the process of arriving at relevant and timely audio descriptions highly complex as a cognitive and linguistic exercise, it is, by its nature, also an incomplete text covering a very specific sub-group of visual elements required to aid (primarily) sight-impaired audiences. In short, AD is applied to describe only those aspects of the film which the viewer cannot readily detect for themselves using the accompanying soundscape, whether dialogue, sound effects, non-verbal utterances or musical scoring. Visual cues for which simultaneous audio markers may be discovered either independently or in parallel with the on-screen action (e.g., dramatic music and the sound of a person screaming accompanying scenes of a burglary) and could therefore be regarded as redundant, are generally omitted from the AD. Such omissions represent a significant problem when considering AD in terms of a text through which to inform improvements to computer-generated video captions, given that the machine “sees” but does not simultaneously “hear” at present. For these reasons, it was concluded that AD did not provide the solution to training computers to deliver human-like video captions. AD does, however, represent a useful comparative text from which to determine the *narratively salient* visual cues from a human perspective in circumstances where these cannot be determined from the audio landscape. AD also contributes value in supplying data relating to the lexical characteristics of human description. Thus, as a professionally crafted corpus, movie AD can be said to comprise a high-quality body of material written in a style that is both lexically rich and narratively sophisticated. To this extent, the linguistic corpus derived via AD is reliable and considered (i.e. contains minimal errors either in comprehension of source materials or exposition in the AD output).

2.3. Content Descriptions

Having established that AD would not provide a one-stop-shop for sourcing linguistic material from which to extract comprehensive visual summarisations of film material, it was necessary to seek alternative annotations data in order to study human descriptive practices in comparison with machine video captioning. Our approach was inspired by our work with Finnish broadcaster YLE in the MeMAD consortium and by a consideration of archive retrieval approaches, metadata and ancillary texts (scripts, programme guides). Archive retrieval within the broadcasting industry is founded in metadata and the tagging of video programming, and this practice is generally referred to as “content description”. Industry moving-image annotations are search-focused (personality-

biased, relatively granular in nature, sales-oriented) and more prosaic than audio description, having less narrative interpretation and more overt labelling of key visual information.

As one strand of our study aimed at enhancing automated description services, the creation of a content descriptions corpus from the MVC, designed to inform computer-led video search and retrieval, appeared to be a reasonably attainable goal.

In order to safeguard objectivity as far as possible (bearing in mind that the points made about the subjectivity of AD apply to any form of human description/translation), the brief applied to building our human-generated content descriptions corpus (CD) was to generate a factual description of all discernible action occurring on screen while avoiding incursions into interpretation. Although the descriptions were kept brief, there was no need for them to fit around dialogue and other elements of the sound track. In practice, the standard applied to compiling content descriptions across the MVC was that the human annotator should identify actions and objects key to narrative, and describe those elements in relation to each other and the micro-narrative within which they were situated, without reference to events or themes derived from outside the current film extract.

As a result, the CD corpus can be regarded as a “ground truth” against which machine descriptions, governed by similar limitations inherent within the automation model, might be critically evaluated. Predictably, however, lexical variation within the AD is 29.66% greater when measured against the CD corpus (word-types) reflecting the more filmic, descriptive remit prevailing in most AD guidelines.

2.4. Training Data and Production of Captions for the MVC

A first-iteration corpus of captions (machine descriptions) was created by applying the MeMAD *DeepCaption* model (Sjöberg, Tavakoli, Xu, Mantecón, and Laaksonen, 2018), trained on image recognition using two large open access datasets, MS COCO (Lin et al., 2015) and TGIF (Li et al., 2016), to the MeMAD Video Corpus (MVC). Multiple captions were created for each of the 501 MVC clips, with one caption being generated by the machine at each computer-detected shot change. This means that the computer model is not applied to moving images per se, but operates on the basis of describing a single frame at a time (in our iteration, the middle frame of a shot), each of which is considered in isolation from the remaining imagery and any associated context. The quality of the resulting video captions is entirely dependent on the quality of the image descriptions contained in the training data and model feature extraction, since the captions are sourced from these datasets.

MS COCO comprises 2.5 million instances of objects in 328k images harvested from the social media website *Flickr*. Each image was annotated with one-sentence captions by five separate operatives (Chen et al., 2015), as shown in

Figure 2. TGIF consists of 100k short sequence animated images (GIFs) drawn from *Tumblr* and annotated with 120k natural language sentences. Both MSCOCO and TGIF harnessed the power of crowdsourcing to produce the annotations.

Figure 2.

Example of captioned image from MS COCO



(338317)

- i. There is a lot of foot traffic on this street during the day.
- ii. People walking down a sidewalk near a road and a building.
- iii. A street with various people walking by a building.
- iv. There are people that are walking on the street
- v. An image of a person walking down the street on her phone

2.5. Annotation Procedure








With regard to the methodological approach to the creation of CD, our “story-arc” annotators were drawn from a pool of doctoral students and post-doctoral researchers experienced in multimedia research. Each annotated extract was verified for accuracy by an alternate annotator to the one creating the original file. Further operatives ensured standardisation of annotations in terms of lexicon and terminology, and performed text normalisation and data cleansing tasks. Each film extract and the associated annotations were therefore checked by three independent operatives before being admitted to the final MeMAD Video Corpus.

The AD and the dialogues were transcribed from the original film tracks. The video captions were produced in electronic text format by our project partner, Aalto University Computing Department. The three corpora were aligned at clip level to allow for direct comparison of the different types of description/annotation at this level. Further detail about the corpus creation, annotation and alignment is given in Braun, Starr and Laaksonen (2020).

Using the story arc introduced earlier, Figure 3 shows an example of the descriptions/annotations. Corpus analysis software *SketchEngine* (Kilgarriff et al., 2014) was used to compute basic descriptive statistics, which will be presented in the next section.

Figure 3.

Sample MVC Annotation

Frame/Time codes	Audio Description (AD) / Dialogue	Content Description (CD)	Machine Description (MD)
02:100994/01:07:19.760 		Dwayne is sitting on the grass in a field, hugging his knees. He is sitting with his back to us.	a man is sitting in a field
02:101125/01:07:25 	He is sitting with his back to her, arms resting on his knees, gazing at the rocky soil at his feet, and doesn't turn as she comes near.	Olive walks towards Dwayne, who is sitting on the ground, staring at the grass. Sheryl, Frank and Richard are at the top of the slope, standing next to the van, looking down at them.	a man and a woman are talking to each other
02:101650/01:07:46.000 	Dressed in her red T-shirt, pink shorts and red cowboy boots, her long hair tied back, her huge glasses perched on her nose, Olive squats at Dwayne's side.	Once she has reached Dwayne, Olive slows down and bends her knees to sit next to Dwayne. Dwayne does not react.	a group of people are singing and dancing
02:101875/01:07:55.000 	She puts her arm around him and rests her head on his shoulder. His head turns slightly towards her.	Olive looks at Dwayne and then puts her arm around him, resting her head on his shoulder. Dwayne is trying not to cry.	a group of people are in a field
02:102325/01:08:13.000 	Dwayne: I'm OK... let's go.	Dwayne turns towards Olive. Dwayne reassures Olive that he is okay, and she looks at him and smiles.	a man is running
02:102475/01:08:19.000 	Olive stands up and Dwayne gets to his feet and goes with her to the bottom of the slope.	Olive and Dwayne stand up and slowly walk towards the bottom of the slope.	a man and a woman are walking in a field
02:102625/01:08:25.000 	Olive starts to climb, putting out her hand for support. Dwayne lifts her up underneath her arms and carries her to the top of the slope.	Olive climbs the slope but she wobbles. Dwayne helps her by carrying her up. Olive seems to be smiling.	a woman is walking down the road

2.6. Initial Corpus Comparison

Comparison of the three key corpora (machine descriptions, human-created content descriptions and audio descriptions) illustrates the fundamental differences between video descriptions produced as a result of basic machine learning, and those derived from human interaction with the same multimodal materials. Before turning to these, it should be noted that in terms of overall corpus size, the AD corpus is – as expected – smaller than the CD corpus, given the purpose and brief of the content descriptions (see above). The MD corpus is the largest, although the size is arbitrary and could easily be changed if the frequency/points at which the machine produces a caption is adjusted. As explained above, a caption was generated for the middle of each shot.

The number of unique words (*types*) represented in the MD corpus is considerably smaller – even in absolute terms, despite the larger size of the MD corpus – than that present in both of the human description modalities (MD: 580; CD: 3,061; AD: 3,969), illustrating at a glance the lexical poverty in the automated output. A similar pattern can be observed in relation to verbs (MD: 88; CD: 531; AD: 726) and adjectives (MD: 39; CD: 297; AD: 490).

In each case, the percentage of unique words appearing in the machine corpus as a percentage of the CD equivalents are: all words (19.72); verbs (16.57); adjectives (13.13). Whilst the same comparison in relation to uniqueness in the MD vs. AD corpus produces the following scores (%): words (14.68); verbs (12.12); adjectives (7.96).

The type-token ratio (TTR) of the three corpora (MD 0.008, CD 0.067, AD 0.158) supports this observation. As can perhaps be expected, the professionally created audio descriptions have the highest TTR, meaning that the lexical variation in this corpus is greater than in the other two. However, the TTR of the CD corpus is in the same order, whilst the TTR of the MD corpus is 20 times lower than that of the AD corpus and 8 times lower than that of the CD corpus. For comparison, TIWO, the AD corpus built by Salway (2007) based on AD of different TV genres, registers a TTR score of 0.044.²

These descriptive statistics paint an unequivocal picture of the overall shape and parameters of the machine corpus, which clearly falls short of human descriptions in all areas of lexical diversification. Indeed, not only is the size of the MD lexicon an average 17.2% of that created by human operatives (across AD and CD modalities), but adjectives comprise 10.9% of the CD corpus and 12.4% of the AD corpus, yet only 6.7% of the machine corpus (MD). It is perhaps not surprising that the human operative annotations deliver a description that is more creative, imaginative and entertainment-led than the machine currently produces, although this imbalance

² Due to the much larger size of the TIWO (over 300k words), this is only a rough indicator, as it is natural for the TTR to decrease with corpus size. However, the different genres may have had an impact.

might potentially be partially rectified in future machine iterations by changes to computer vision feature extraction.

Table 2.

Corpus Comparison

<i>Category</i>	<i>MD</i>	<i>MD</i>	<i>CD</i>	<i>CD</i>	<i>AD</i>	<i>AD</i>
	<i>Types</i>	<i>Tokens</i>	<i>Types</i>	<i>Tokens</i>	<i>Types</i>	<i>Tokens</i>
all words	580	70,315	3,061	43,829	3,969	25,039
type-token ratio (TTR)	0.008		0.067		0.158	
nouns	363	18,160	1,482	13,403	1,862	7,291
verbs	88	18,964	531	9,576	726	4,458
adjectives	39	460	297	1,448	490	1,221
adverbs	7	1,783	179	1,917	250	1,097
conjunctions	2	4,498	5	2,077	5	985
pronouns	14	1,938	21	3,477	21	2,888
prepositions	22	8,500	60	5,232	52	3,300

This quantitative overview serves to illustrate the differences between the corpora. Further insights come from our comparative qualitative analysis of the data for the purposes of identifying characteristic features and pattern deviations between machine- or human-led approaches. These insights will be outlined in the next section, which focusses on the assessment of the current quality of machine-generated descriptions.

3. Video Captions: Quality Assessment

Our initial quantitative analyses of the machine-generated descriptions, as exemplified in Table 2, show that at present, these descriptions hardly give insight into the essence of many of our micro-narratives. On the face of it, the computer algorithms often miss or mis-identify one or both of the main characters, key actions and the mood of a scene, they do not acknowledge repeated appearances of a character or object and, above all, they miss the intended meaning of our micro-narratives. As the application of automated image or video captions is relatively new territory to both human information retrieval and to human understanding in the context of media access, it is important to trace these observable phenomena back to source (their underlying problems). It is these issues which make current video captions appear trivial or naïve and which allow us to explore how human descriptive knowledge can potentially be applied to improve outcomes. We have therefore grouped the observed problems into three principal categories, each of which impacts the quality of outputs: methodological issues, where the problem is rooted in the nature

of the training data; computer vision problems, which result from current limitations in object detection/identification; and linguistic problems, which are related to how the output of computer vision algorithms is rendered into natural language. Each area will be discussed below.

3.1. Methodological Issues

A significant problem is the nature of the available training datasets. In the field of image recognition and description a number of large, comparatively high quality, annotated datasets are available when compared to other types of training data (e.g., in the business world). However, these captioned image datasets are not optimised in a way that serves linguistic studies. This can be illustrated with reference to one of the principal training datasets used to create the first iteration descriptions for our MVC corpus, MS COCO (Lin et al., 2015). As explained above, MS COCO is a meticulously designed and annotated large-scale dataset for visual object detection and captioning. Each still picture has been annotated with five captions, generated by five individual human operatives, describing the image content (Chen et al., 2015). The purpose of this exercise is to harvest visually pertinent information from which machines can learn the connections between the visual objects and actions, and the semantic labels given to them by the annotators. As with other data-related tasks of a similar scale, the MS COCO creators resorted to crowdsourcing service Amazon Mechanical Turk to collect the image captions (Chen et al., 2015). Although a widely accepted practice for manipulating datasets of this size, crowdsourcing annotations for training data in this manner introduces a number of factors which render the results from test data – in this case, our MVC corpus – less reliable, and demonstrably low in quality.




Firstly, the **type of work** undertaken is financially rewarded according to the number of units of material captioned, meaning that captions are produced spontaneously and rapidly, possibly without much thought being given to lexical variety or non-superficial observations. The protocols attaching to such image captioning tasks include word count and time limitations, which can have a significant impact on creativity, resulting in rigid syntax.

Secondly, in terms of **workers and their profiles**, Amazon Mechanical Turk and similar crowdsourcing services tend to attract college students from a computing background, leading to age and interest bias (Difallah, Filatova, and Ipeirotis, 2018). Research shows that the workers' profile has an impact on the quality of their work (Kazai, Kamps, and Milic-Frayling 2012) and that feedback can improve quality (Han, Roitero, & Gadiraju 2019). However, Chen et al. (2015) do not discuss the details of their approach to recruiting and working with the crowd workers, and the MS COCO captions suggest that at least some of the crowd workers are amateurs when it comes to the descriptive genre. The examples in Figure 4 illustrate the different skill levels. For instance, whilst caption 1.iii. sounds professional and forms a grammatically complete sentence with a verb in simple present, it includes an abstract value judgement ("beautiful"). Caption 1.iv. is factual but vague, not giving much detail about the actual objects in the room ("lots of furniture"). Similarly, in image 2, several captions refer

to the red sign, but lack the precise terminology (i.e. “no-entry sign”) that may be needed in the context of content description for archival purposes or AD.

Figure 4.

Examples of Captioned Images from MS COCO

	<p>1 (374628)</p> <ul style="list-style-type: none"> i. A kitchen made of mostly wood with a small desk with a laptop. ii. A full view of an open kitchen and dining area. iii. A beautiful, open kitchen and dining room area features an island in the center and wood cabinets and large windows. iv. A kitchen with wood floors and lots of furniture. v. A very spacious room with a kitchen and dining area.
	<p>2 (132394)</p> <ul style="list-style-type: none"> i. A red sign is on the gray sidewalk ii. A vandalized street sign on a side walk iii. A red cautionary sign with "know hope" in graffiti iv. A round red sign on the other side of a stop sign v. A red sign is at the corner of the street on the sidewalk
	<p>3 (290868)</p> <ul style="list-style-type: none"> i. A grandmother standing next to a child in a kitchen. ii. Baby trying to open wooden cabinets under the sink. iii. A woman and child stand in the kitchen. iv. An older woman is standing in the kitchen with a child. v. The little girl is trying hard to open the cabinets

The description **task** may also impact the quality of the results. The crowd workers for MS COCO were instructed to describe all “important parts” of the scene, using at least eight words, and not starting sentences with there is/are. An obvious problem is that crowd workers do not always follow the instructions. Albeit infrequently, they do use “there is/are” ($N=12817$, see e.g., Figure 2 above) and/or phrases such as “an image of”, “a full view of”, which are similarly redundant in this context. More importantly, the instruction rubric raises the highly relevant question: what are the “important parts” of any given image? Naturally, the answer is inextricably linked to matters of **relevance and saliency**. Considering image 1 in Figure 4 again, each caption highlights different objects, illustrating the differences in human perception and approach to simple tasks of this kind. In a video scene, whether it is important to mention the laptop or to highlight the mostly wooden outlay will depend on the context of the unfolding narrative.

Further issues inherent in this type of description are **accuracy, vagueness and lexical ambiguity**. Chen et al. (2015) explore recall (i.e. whether an entity that is present in an image is referred to

in the caption) and accuracy (i.e. whether the description is correct) for selected nouns, adjectives and verbs. Their results indicate high recall and accuracy rates for nouns denoting somewhat rare entities without many or any synonyms (e.g., “elephant”), but mixed rates for other more prosaic objects (e.g., “sidewalk”).

A more fundamental problem in our context is that although the aim of MS COCO was to present scenes, i.e. objects in context, it is still a database of **static images** without narrative coherence from one image to the next. As such, it can capture actions only to a limited extent and cannot provide examples of narrative cohesion (e.g., causal, temporal cohesion, links between characters, co-reference). As for actions, we clearly have the ability to identify visual actions in still images, especially in photos, using common knowledge of body movements, postures etc. Thus MS COCO has numerous instances of walking, playing, drinking, which can be detected from a single frame. In addition, it contains verbs denoting actions that would stretch over several frames in a video scene, e.g., opening (Ronchi & Perona, 2015), although these are considerably less frequent and occur in phrases such as “is trying to open”, suggesting uncertainty (see Figure 4, 3.ii and 3.v). Similarly, descriptions such as “he looks like he is falling”, although infrequent, indicate uncertainty in relation to such actions.

With regard to cohesion, **linkage of characters through actions** is limited and builds on a smaller set of verbs, mainly “talking”, but the frequent use of “talking” in our MD corpus is in itself problematic. It illustrates the point that human descriptions are narratively salient and relevant in a way that computer descriptions are generally not, at least consistently. When we see a man and a woman arguing about who does the washing up after dinner, narrative saliency may not to be found in the most common of computer captions, “A man and a woman are talking”. Adding a layer of emotional description may be possible if the computer determines facial expressions and therefore selects “A man and a woman are arguing”, although even then the saliency may not relate either to the household chore, or the argument, but instead indicate incompatibilities within the relationship. Most people would be able to detect this nuance by interpreting the dialogue in terms of the social setting, vocal tonality, facial expressions and body language. Meanwhile, the computer simply “sees” two people talking. As a measure of quality, the value for the viewer is to be found in the storytelling and not in the quasi-metadata description represented as a formulaic “man + woman + talk”.

Interestingly, while AD may assist in determining that a man and a woman are in the kitchen (the fact that they are arguing would be discernible to the viewer from voice tone and language), human content descriptions (CD) indicate everything that can be observed in the scene – two people, kitchen, washing up, angry faces, aggressive body language, arguing – falling short only on broader narrative interpretation which requires material from outside that specific scene (the failing relationship, perhaps). To this extent, and for this particular purpose, the CD corpus can be considered a more appropriate and quality-driven resource.

The lack of linkage of characters is one indicator of the dataset’s limitations with regard to creating a cohesive narrative. Another indicator is the lack of **temporal, causal or other links between individual actions**, i.e. the absence of relevant cohesive markers. While ‘and then’ occurs within the MVC corpus, instances can be traced back to split-screen images in the training data which prompted captioners to treat them in sequence, belying the superficially temporal implications of the phraseology. Finally, narrative coherence is constructed in the way human beings identify, recognise and refer to characters. MS COCO, however, does not include any support for this, for example, in the form of cohesive chains drawing on pronominalisation and other ways to create **co-reference**. The absence of co-reference markers is certainly one of the most noticeable features in the current MD corpus. Many examples in which a series of captions refer to the same characters read as shown in Figure 5. The story arc from which it is taken shows one man and one woman.

Figure 5.

Example of Machine Description from MVC Clip 200006

00:00:00.000 00:00:02.700 A man is talking and smiling and laughing
 00:00:02.700 00:00:04.533 A woman is smiling and talking to someone
 00:00:04.533 00:00:24.600 A man is dancing in a room with other people
 00:00:24.600 00:00:26.733 A woman is sitting on a couch and smiling
 00:00:26.733 00:00:28.266 A man is dancing in a room with a lot of people
 00:00:28.267 00:00:30.734 A man is walking through a door and then he falls down
 00:00:30.733 00:00:33.000 A woman is sitting on a couch and eating a sandwich
 00:00:33.000 00:00:34.600 A man is talking and smiling and laughing
 00:00:34.600 00:00:36.200 A man is sitting on a couch and talking
 00:00:36.200 00:00:40.967 A man is talking and smiling and laughing
 00:00:40.967 00:00:42.967 A woman is sitting on a bench and talking

Another difference is in the nature of the training dataset, i.e. a **mismatch between the content of the images in the training data and that of the MVC**. The images in MS COCO show simple everyday scenes of people walking, talking, eating, engaging in sports and so forth. The explicit aim of the MS COCO creators was to include non-iconic images, i.e. scenes without one person or object clearly standing out. In our corpus, which contains extracts from feature films, visual scenes are more deliberately composed, iconic and laden with narratively relevant *mise-en-scène*. They are also subject to editing techniques that manipulate visual content to include multiple shot changes, close-ups, panning and zooming techniques which render the material difficult for the machine to “read”.

Aside from the methods applied in relation to the purchase of training data captioning services from crowdsourced websites, and the differences in the nature of the visual material included in the training data and our MD corpus, other measures were taken during the application

of the training data to MD production which impacted results. In particular, the lexical poverty of outputs was increased by the elimination of tokens in the training data which occurred fewer than four times. These “long tail” words, being those which are uncommonly found in the corpus, are a regular feature of AD and human description adding nuance and colour. In this case, elimination from the training data before applying the *DeepCaption* model was a matter of computer processing expediency. Furthermore, topical bias is inherent in the types of data typically collected from *Flickr* and *Tumblr*, such that words like *laptop*, *microphone* and *surfboard* are over-represented in the test data results. Poor data cleansing within the training data also resulted in grammatical mistakes, lexical errors, and incomplete captions transferring across to the MVC machine descriptions. Finally, natural language processing as it has been applied to MD output, falls short of human descriptive requirements, being highly formulaic and syntactically repetitious in nature (“An X and a Y are + verb gerund”, as illustrated in the earlier examples). Taken together, these factors currently result in poor quality captions.

3.2. Computer Vision Problems

At the most fundamental level, visual storytelling relies on the successful identification of characters in order for the viewer to locate them successfully and consistently within the unfolding narrative. This is particularly the case for sight- and cognitively-impaired viewers, but also in the video retrieval scenario, where a certain character must be isolated from a vast wealth of video material. Separation between male and female protagonists where they are seen and not heard is generally helpful, notwithstanding issues of gender labelling and gender bias which are outside the scope of this study. Fully sighted human beings are capable of distinguishing between sexes featured in moving imagery in a traditional, binary sense with relative ease. The MD outputs from our computer model were unreliable in this regard, although the training data from which they were derived is unlikely to have a significant error rate. AD containing incorrect labelling of male and female characters would be unhelpful at best, and at worst represent a significant confound for audiences experiencing sight-impairment. Vocal gender profiling work will undoubtedly help to rectify this issue, compensating for unreliable computer vision feature extraction which is currently too rigid and rule-bound (e.g., a person with short hair is generally labelled as a man, irrespective of dress, mannerisms, voice and other cues implying gender).

Similarly, machine-based object detection remains unreliable to the extent that non-standard angles, changes of size/scale and rapid changes of light and shade can alter the description from “a car” to “a guitar” between one frame and the next. Equally curious, changes in the pixel structure of an image that cannot be detected by the human eye can change the descriptions in an unpredictable fashion.

Unusual or rare objects pose a further challenge as do facial expressions: laughing and grinning are difficult to distinguish in current models. More training data is needed to overcome these difficulties although, again, audio cues could assist once incorporated into the model.

3.3. Linguistic Considerations

As discussed above, the source of training data captions has resulted in MD lexical poverty in both variety and nuance. A study of verb usage in the MD corpus serves to illustrate this point:

Table 3.

MD Corpus: Verb Rankings

MD Corpus Verb Rank	Lemma	Frequency	MD Corpus Verb Rank	Lemma	Frequency
1	be	7806	24	live	51
2	talk	1686	25	wear	48
3	smile	1682	26	smoke	46
4	look	1657	27	run	42
5	dance	1119	28	make	38
6	walk	1087	29	eat	24
7	sit	1004	30	pour	20
8	kiss	328	31	blow	16
9	hold	302	32	take	15
10	play	238	33	swim	14
11	drive	230	34	do	14
12	fall	214	35	fly	13
13	stop	203	36	work	13
14	sing	179	37	wave	13
15	stand	134	38	move	13
16	jump	130	39	read	11
17	laugh	79	40	open	10
18	put	73	41	hug	9
19	turn	72	42	cut	8
20	lay	61	43	show	8
21	lie	55	44	crash	5
22	ride	52	45	type	5
23	drink	51	46	park	5

Eighty-eight verb lemmas can be found in the MD lexicon, only forty-six of which occur five or more times (see Table 3). The most commonly used verb lemma is “be” (frequency: 7806; relative frequency: 111.014.72/million), in contrast with the British National Corpus, which shows a relative frequency of around one-third of this rate (36762.66/m). In the MD corpus, 7549 instances of this lemma register in the third person singular (96.7%). Furthermore, 7508 of the 7549 instances of “is” in the MD corpus are to be found in concordance with a corresponding verb gerund (CQL search: [word="is" & word=".ing"]), e.g., “A woman is dancing”, “A man is talking”, and so forth. Parsing during the NLP phase of image processing might be improved to provide more syntactic variety in the rendering of these machine descriptions.

In addition, the top six verb lemmata are vastly over-represented in the MD outputs when compared to the MS COCO and TGIF training datasets (Table 4.), suggesting that feature extraction and other factors play a significant role.

Table 4.

MD Verbs: Comparative Statistics vs. Training Datasets

RANK	VERB LEMMA	MD <i>f</i>	MD verb/m	COCO <i>f</i>	COCO verb/m	TGIF <i>f</i>	TGIF verb/m
1	Be	7806	111014.72	154295	22188.44	90737	68149.11
2	Talk	1686	23977.81	3114	447.81	5914	4441.78
3	Smile	1682	23920.93	3913	562.71	3755	2820.24
4	Look	1657	23565.38	16902	2430.6	11071	8315.01
5	Dance	1119	15914.1	67	9.63	2392	1796.54
6	Walk	1087	15459.01	17921	2577.14	6480	4866.88
7	Sit	1004	14278.6	68705	9880.15	5076	3812.39
8	Kiss	328	4664.72	165	23.73	3242	2434.94
9	Hold	302	4294.96	30487	4384.19	5613	4215.71
10	Play	238	3384.77	15935	2291.54	4469	3356.5

An alternative source of information about the skewed nature of MD outputs are keywords. They provide score-based data regarding the uniqueness of the focus corpus in relation to a more generic and linguistically typical reference corpus. For this purpose, our comparison was made between the MD lexicon and that of the British National Corpus (BNC) which contains in excess of 96 million words, 6 million sentences, 1.5 million paragraphs and 700,000 unique items.

Analysis of keyness within the MD corpus illustrates the nature of lexical bias found within the captioned training data. In particular, the sources of imagery in the adopted datasets, which were derived from *Flickr* (in the case of MS COCO) and social media postings (TGIF), led to a preponderance of objects which were over-represented when compared with the more standard lexicon in the reference corpus (BNC). Technology and youth-relevant vocabulary scores highly in MD

keyness with *laptop*, *skateboard*, *trampoline* all ranking in “top 5” positions; *tv*, *microphone* and *piano* fall within the “top 20” items; and *surfboard*, *motorcycle*, *guitar*, and *skateboarding* rank in the “top 30”. These scores illustrate the youth and technology bias generally observed within social media postings and thus are over-represented in the training data. The over-represented nature of *hallway* (rank: 1; frequency 305; relative frequency: 4337.62/m) appears to derive from a particular phenomenon in the training data. Of the 305 occurrences in the MD corpus, 255 can be found in the concordance “walking down a hallway”, suggesting similar concordances occur in the training data. Indeed, while this phrase appears only five times in the COCO dataset, it can be found 65 times in the TGIF dataset (48.82/m). Clearly, the disparity in relative frequencies between the MD corpus and training data suggests that a level of bias is being introduced via the *DeepCaption* model, which requires further investigation. *Couch*, as the second ranked item in order of keyness, occurs 306 times in the MD corpus, with a relative frequency of 4351.85/m. A total of 296 of these MD occurrences feature in the concordance “sitting on a couch” (relative frequency: 4223.85/m) and “sitting on a couch and smiling” occurs 82 times (relative frequency: 1166.18/m). In the COCO dataset, “sitting on a couch” appears 872 times (relative frequency: 125.4/m), whereas in the TGIF dataset, it can be found 217 times (relative frequency: 162.98/m). Again, the imbalance between training data and MD corpus suggests that commonly occurring phrases become over-represented during the captioning process.

Table 5.

MD Corpus: Keyness Scores

Rank	Term	Score	(MD) corpus frequency	Reference (BNC) corpus frequency
1	hallway	920.81	305	417
2	couch	596.12	306	708
3	laptop	458.46	60	97
4	skateboard	355	42	77
5	trampoline	321.45	34	57
6	dance	286.34	1119	6132
7	smile	154.75	1682	17255
8	tv	118.73	14	77
9	singing	108.56	91	1228
10	shirtless	106.26	8	9

4. Conclusions: Quality Issues and the Future of AI for AD

The ongoing AI revolution has the potential to promote inclusive design, by personalising media products (James, 2019) and making them accessible for everyone, bridging language barriers as well as different physical and cognitive abilities. In the context of audiovisual accessibility, this is, however,

a long way off. The automatic generation of natural-language descriptions of video scenes still presents a non-trivial challenge for both the computer vision and the language-processing communities. This article has highlighted problems with object recognition, gender labelling, action interpretation and so forth. However, saliency, relevance and lack of narrative coherence emerge as fundamental issues (Huang et al., 2016). Currently, MDs not only fail to approach human levels of description, in terms of complex artificial cognition such as mental modelling, but they also fall at the first hurdle (e.g., mistaking a desk for a surfboard). Resolving basic computer vision (objects and actions) will therefore only solve some of the more immediate problems associated with complex narrative.

As was shown, the quality of MDs is currently affected by a lack of sufficient training data, especially moving imagery (Aafaq et al., 2019), and this further impacts the inter-relational linking of artefacts for narrative development, facial recognition, facial expression and emotion detection, amongst many other factors. One important finding of this exploratory research is reinforcement of the need for further relevant training datasets to be created, despite their limitations. What the present article has highlighted in this respect is that AD is not directly comparable with MD, and alternative human-derived datasets are more helpful for training the model. As discussed, CD appears to be a more reliable data source for the machine, but most importantly, the quality of future MDs is dependent upon a more syntactically flexible, lexically sophisticated, and coherent model for storytelling.

The most important ethical point that emerges from the data presented here is that poor-quality MDs cannot replace human AD as a service for sight-impaired audiences, as they do not meet legal requirements for the provision of meaningful description (Ofcom, 2017, Annex 4, p. 18). However, lower quality MDs may be acceptable for data retrieval purposes in commercial scenarios where certain film material lies outside the prime-resale category, i.e. as a means of increasing marginal profits by re-purposing those video assets considered less valuable and therefore not currently warranting human annotation.

Human approaches to audiovisual content description will continue to drive the agenda and establish priorities for future computer vision research. Semi-automation and post-editing afford further opportunities to enhance the machine's best efforts, with human-in-the-loop approaches being used to determine how human and machine intelligence can most productively and efficiently come together. Combining human and machine endeavours will also demonstrate to the human creators of AD that their involvement in developing semi-automated approaches will not mean that they are writing themselves out of their jobs, and indeed, is more likely to secure their involvement in the development of automated approaches where these can be useful. For instance, automation or semi-automation of AD carries enormous potential in the area of social media (YouTube; Facebook; Twitter images/gifs; Instagram) and in other multimodal information situations e.g., language learning and pedagogy more generally. The addition of AD to a range of social media content would serve the purpose of making AD more widely available generally.

This could be particularly beneficial if machines delivered AD where there is currently no alternative (low quality arguably being better than zero access). Taking this path will contribute to improving media accessibility for everyone while simultaneously invoking reflective practices and a mindful approach to the social, ethical and economic implications of automation in this area.

References

- AENOR (2005). Norma UNE 153020: *Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. [The Standard UNE 153020 Audio description for visually impaired people. Requirements for audio description and for the production of audio guides]. Madrid: AENOR. Retrieved from: <https://www.aenor.com/normas-y-libros/buscador-de-normas/UNE?c=N0032787>
- Aafaq, N., Milan, A., Liu, W., Gilani, S.Z., and Shah, M. (2019). *Video description: A survey of methods, datasets and evaluation metrics* (ACM Computing Surveys). Retrieved from: [arXiv:1806.00186v1](https://arxiv.org/abs/1806.00186v1) [cs.CV].
- Braun, S. (2011). Creating coherence in audio description. *Meta*, 56(3), 645–662.
- Braun, S. (2016). The importance of being relevant? A cognitive-pragmatic framework for conceptualising audiovisual translation. *Target*, 28(2), 302–313.
- Braun, S., Starr, K., and Laaksonen, J. (2020, forthcoming). Comparing human and automated approaches to visual storytelling. In S. Braun, and K. Starr (Eds.) *Innovation in audio description research*. London: Routledge.
- Chen, X., Fang, H., Lin, T-Y., Vedantam, R., Gupta, S., Dollár, P., and Zitnick, C. L. (2015). Microsoft COCO captions: Data collection and evaluation server. Retrieved from: [arXiv:1504.00325v2](https://arxiv.org/abs/1504.00325v2) [cs.CV].
- Difallah, D., Filatova, E., and Ipeirotis, P. (2018). Demographics and dynamics of mechanical turk workers. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, February 5–9th. Marina Del Rey, CA: ACM.
- Fresno, N., Castellà, J., and Soler-Vilageliu, O. (2016). What should I say? Tentative criteria to prioritize information in the audio description of film characters. In A. Matamala and P. Orero, (Eds.) *Researching audio description* (pp. 143–167). London: Palgrave.
- Han, L., Roitero, K., and Gadiraju, U. (2019). All those wasted hours: On task abandonment in crowdsourcing. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (pp. 321–329), 11–15th February. Melbourne, Australia: ACM.
- Huang, T-H., Ferraro, F., Mostafazadeh, N., Misra, I., Agrawal, A., Devlin, J., Girshick, R., He, X., Kohli, P., Batra, D., Zitnick, C. L., Parikh, D., Vanderwende, L., Galley, M., and Mitchell, M. (2016). Visual storytelling. *Proceedings of the NAACL-HLT*, June 12th–17th. San Diego, CA: Association for Computational Linguistics.
- Hyks, V. (2005). Audio description and translation: Two related but different skills. *Translating Today*, 4, 6–8.
- Ibanez, A. (2010). Evaluation criteria and film narrative. A frame to teaching relevance in audio description. *Perspectives: Studies in Translatology*, 18(3), 143–153.
- Independent Television Commission (2000). *Guidance on standards for audio description*. Retrieved from: https://www.audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf.

- James, L. (2019, 12 August). Getting personal with artificial intelligence and the cloud. *The Record*. Retrieved from: <https://www.technologyrecord.com/Article/getting-personal-with-artificial-intelligence-and-the-cloud-85390>.
- Jimenez, C., and Seibel, C. (2012). Multisemiotic and multimodal corpus analysis in audio description: TRACCE. In A., Remael, P. Orero, and M. Carroll (Eds.) *Audiovisual translation and media accessibility at the crossroads* (pp. 409–421). Amsterdam: Rodopi.
- Kazai, G., Kamps, J., and Milic-Frayling, N. (2012). The face of quality in crowdsourcing relevance labels: demographics, personality and labeling accuracy. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM'12*, October 29 – November 2nd. Maui, HI: ACM.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit J., Rychlý, P., and Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, (1), 7–36.
- Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L-J., Shamma, D., Bernstein, M. S., and Fei-Fei, L. (2017). Visual Genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 32–73.
- Kruger, J. L. (2010). Audio narration: Re-narrativising film. *Perspectives: Studies in Translatology*, 18, 231–249.
- Li, Y., Song, Y., Cao, L., Tetreault, J., Goldberg, L., Jaimes, A., and Luo, J. (2016). TGIF: A new dataset and benchmark on animated GIF description. *Proceedings of the IEEE computer society conference on computer vision and pattern recognition* (Vol. 2016 – December, 4641–4650). Las Vegas, NV: IEEE Computer Society.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girschick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2015). Microsoft COCO: Common objects in context. *Computer Vision, ECCV 2014*, 740–755.
- Matamala, A. (2019). The ADLAB PRO course materials: Structure, type, quantity and aims. Multipler event 6, Ljubljana, 3rd June. Retrieved from: <https://www.adlabpro.eu/results/multiplier-events/multiplier-event-6/>
- Ofcom (2017). Ofcom's code on television access services. Retrieved from: <https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/tv-access-services>
- Ramos Caro, M. (2016). Testing audio narration: The emotional impact of language in audio description. *Perspectives: Studies in Translatology*, 24(4), 606–634.
- Rohrbach, A., Rohrbach, M., Tandon, N., and Schiele, B. (2015a). A dataset for movie description. *Proceedings of CVPR*, Boston, Mass, June 8–10th. Retrieved from: <https://www.cv-foundation.org/openaccess/CVPR2015.py>
- Rohrbach, A., Torabi, A., Rohrbach, M., Tandon, N., Pal, C., Larochelle, H., Courville, A., and Schiele, B. (2015b). Movie description. *Proceedings of ICCV*, December 11–18th. Santiago, Chile. Retrieved from: [arXiv:1605.03705v1](https://arxiv.org/abs/1605.03705v1) [cs.CV].
- Ronchi, M. and Perona, P. (2015). *Describing common human visual actions in images*. Retrieved from: [arXiv:1506.02203v1](https://arxiv.org/abs/1506.02203v1) [cs.CV].
- Sade, J., Naz, K., and Plaza, M. (2012). Enhancing audio description: A value added approach. *Proceedings of ICCHP, Part 1, LNCS 7382* (pp. 270–277). Linz, Austria: ICCHP.
- Salway, A. (2007). A corpus-based analysis of audio description. In J. Díaz-Cintas, P. Orero, and A. Remael (Eds.) *Media for all: Subtitling for the deaf, audio description and sign language* (pp. 151–174). Amsterdam: Rodopi.

- Sjöberg, M., Tavakoli, H. R., Xu, Z., Mantecón, H. L., and Laaksonen, J. (2018). PicSOM Experiments in TRECVID 2018. *Proceedings of the TRECVID 2018 Workshop*. Gaithersburg, MD.
- Vandaele, J. (2012). What meets the eye. Cognitive narratology for audio description. *Perspectives: Studies in Translatology*, 20(1), 87–102.
- Vercauteren, G. (2007). Towards a European guideline for audio description. In J. Díaz-Cintas, P. Orero, and A. Remael (Eds.) *Media for all: Subtitling for the deaf, audio description and sign language* (pp. 139–149). Amsterdam: Rodopi.
- Yeung, J. (2007). Audio description in the Chinese world. In J. Díaz Cintas, P. Orero, and A. Remael (Eds.) *Media for all: Subtitling for the deaf, audio description and sign language* (pp. 231–244). Amsterdam: Rodopi.

Feasibility, Quality and Assessment of Interlingual Live Subtitling: A Pilot Study

 Hayley Dawson 

University of Roehampton

Abstract

Intralingual respeaking has been widely practiced since 2001 (Romero-Fresco, 2011); however, interlingual respeaking (from one language into another) is yet to take off. Interlingual respeaking is a hybrid form of subtitling and interpreting and calls upon skills used in both professions. To consolidate this mode of audiovisual translation (AVT) within media accessibility (MA), a programme must be created to train future interlingual respeakers. This paper presents the results of the first ever study on interlingual respeaking, in which 10 participants interlingually respoke three short videos using a language combination of English and Spanish. The main areas of research in this project are feasibility, quality and training. Before expanding training in this area, interlingual respeaking must be deemed feasible and an effective method of assessment must be in place to determine its quality. The NTR model is a quality assessment model for interlingual live subtitles, of which an accuracy rate of 98% or above indicates acceptable live subtitles. The average accuracy rate of the study is 97.37%, with the highest accuracy rate reaching the 98% threshold with 98.50%. The initial results point to interlingual respeaking as feasible providing a training programme is put in place to build upon existing task-specific skills and develop new ones to ensure interlingual live subtitles of good quality are produced.

Key words: interlingual live subtitling, quality, NTR model, training, respeaking, media accessibility (MA).

Citation: Dawson, H. (2019). Feasibility, quality and assessment of interlingual live subtitling: A pilot study. *Journal of Audiovisual Translation*, 2(2), 36–56.

Editor(s): G.M. Greco & A. Jankowska

Received: April 04, 2019

Accepted: November 19, 2019

Published: December 31, 2019

Copyright: ©2019 Dawson. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. Introduction

Respeakers use speech recognition software to repeat or paraphrase what is heard from an audiovisual text through a microphone while enunciating punctuation and adding special features, such as colours to identify the speakers. The software turns the spoken utterances into text on screen, which are cued as live subtitles (Romero-Fresco, 2011). Although intralingual respeaking has become an established practice within the industry, interlingual respeaking is yet to take off. Intralingual respeaking has been widely practiced since 2001, when it was first used in the UK as a method to provide live subtitles for the BBC World Snooker Championships (Romero-Fresco, 2011). Interlingual respeaking mirrors a similar process: a respeaker listens to an audiovisual text in its original language and respeaks it into another. The respeaker essentially simultaneously interprets what they hear, enunciates punctuation and endeavours to correct errors and add special features for a deaf and hard-of-hearing audience before cueing the subtitles. The shift in language adds a layer of complexity to interlingual respeaking. It is not widely practised and must be researched before better quality live subtitles can be produced.

Many aspects of intralingual live subtitling have been researched, such as sociolinguistic approaches to respeaking (Eugeni, 2008), training (Arumí Ribas & Romero-Fresco, 2008), (Romero-Fresco, 2012); and quality and audience reception (Romero-Fresco, 2012, 2015, 2016; Fresno, 2019). Such research has successfully informed the field of the challenges that lie ahead for intralingual respeaking. Broadcasters have increasingly used the NER model (Romero-Fresco & Martínez, 2015) due to a rise in focus on the quality of live subtitles. The model analyses the extent to which errors affect the coherence of the subtitled text or modify its content. As a next step towards improving quality in intralingual respeaking, respeaking certification has begun with LiRICS (Live Reporters International Certification). LiRICS aims to set and maintain high international standards in the respeaking profession, which is a step in the right direction to highlight respeaking as a qualified profession (Romero-Fresco et al., forthcoming).

Interlingual respeaking is considered a new discipline within the realm of AVT and MA. The ILSA project (Interlingual Live Subtitling for Access) is a large-scale project, which aims to design, develop, test and validate the first training course for interlingual respeaking. The project has become a starting point for research to take place and has explored the complex nature of the task including: research into understanding who live subtitlers are (Robert, Schrijver, & Diels, forthcoming), the task-specific skills required for interlingual respeaking (Pöchhacker & Remael, forthcoming) and what interlingual respeaking performance means for training in interlingual respeaking (Dawson & Romero-Fresco, forthcoming). SMART (Davitti, Sandrelli, & Romero-Fresco, 2018) is a similar pilot study carried out at the beginning of 2018 and aimed to compare the interlingual respeaking performance of trainees with previous experience in interpreting, subtitling and respeaking. The research conducted for this study (in February 2017) and the SMART pilot study (in January 2018) has already shed light on the feasibility and quality of interlingual respeaking and the task-specific skills required. Comparisons have been made with this pilot study

and the SMART study throughout this article. Although there is a small number of participants in this study, it is the first step towards launching and informing the practice of interlingual respeaking. The results have already successfully informed the main experiment of Intellectual Output 2 of the ILSA project.

The service of live subtitles has mainly been provided intralingually and is primarily for those with a hearing impairment but also caters for a hearing audience when sound cannot be used on the television such as in gyms, cafes and libraries. Over the past few years, an interest in interlingual respeaking has grown, which is prospective of revolutionising MA. Interlingual respeaking could heighten social impact as it not only caters for those with a hearing impairment. It also caters for foreign audiences, which demonstrates the potential to aid the integration of foreigners into society. Interlingual respeaking gives those access to a text, which, due to a language barrier, they would not usually be able to access in its original form. This mirrors the idea that MA can be an effective tool to foster human rights for all citizens, not only for those with disabilities (Greco, 2016). Shaping the training of interlingual respeakers is essential in implementing sturdy theory and techniques on how respeakers work in the UK and abroad. This is the key to contributing to the broadened scope of MA. Most importantly, it ensures that a DHOH (Deaf and Hard of Hearing) and foreign community can fully access media products and events in a different form and with live subtitles of good quality.

Training in respeaking has been included in AVT modules in universities across Europe (University of Antwerp, Autonomous University of Barcelona, University of Leeds, University of Roehampton, and Zurich University) and has typically focussed on intralingual respeaking with an introduction to interlingual respeaking included at the end of the module. Training in interlingual respeaking has now begun with the first known course, which was delivered online by the University of Vigo to seven students from January to June 2019. The course included three modules: simultaneous interpreting (English > Spanish), intralingual respeaking (Spanish > Spanish) and interlingual respeaking (English > Spanish); indicating that current professionals and students have started to take interest in the most recent mode of translation and interpreting.

This article, firstly, endeavours to present relevant data outlining the feasibility of interlingual respeaking and how quality has been measured using the current working NTR model (Romero-Fresco & Pöchhacker, 2017) (defined in section 3). Secondly, qualitative results on participants' perception of their performance, the task-specific skills and best-suited professional profile for an interlingual respeaker will be presented. Then, a brief evaluation on the effectiveness of the NTR model will be given. Finally, initial thoughts on training will be presented as a step towards producing interlingual live subtitles of good quality. The findings of this study aim to inform the design of a large-scale study of around 50 participants, which will seek to identify the task-specific skills and best-suited professional profile for an interlingual respeaker. The outcomes will eventually inform a training model for interlingual respeaking.

2. Methodology

The methodology of the pilot experiment took a train, practise and test approach. Participants filled out a pre-experiment questionnaire, a short session on respeaking was delivered, participants completed a respeaking test and filled out a post-experiment questionnaire. Finally, participants contributed further by answering some questions and making observations about their experience of respeaking. Quantitative data was collated in the form of analysis of the respoken subtitles and qualitative data in the form of questionnaires. Individual performance was recorded with Screencast recording software and analysed separately.

2.1. The Participants

This study took place in the language computer laboratory at the University of Roehampton. Participants received face to face training and the researcher was present to lead the training and experiment. Ten participants took part in the study. However, due to technical issues the data from participant 5 is not available and there is not enough data from participant 10 to carry out a meaningful analysis. Of the remaining eight participants, seven were female and one was male. Their average age was 32, the youngest being 23 and the oldest 48. There were seven native Spanish participants and one native English (participant 7). Two participants were professionals, one being a translation lecturer and another a speech-to-text interpreter. Six participants were postgraduate students in translation, two of whom also worked in translation and teaching while studying.

Table 1.

Details of Participants' Previous Experience

	Background
Participant 1	Subtitling, interpreting and intralingual respeaking
Participant 2	Subtitling and respeaking
Participant 3	Subtitling
Participant 4	Subtitling and interpreting
Participant 6	Subtitling and intralingual respeaking
Participant 7	Subtitling and interpreting
Participant 8	Interpreting and intralingual respeaking
Participant 9	Subtitling

2.2. Data Collection

Participants respoke three video clips interlingually: a narration, a speech and a news story. Due to two sets of language combinations (English into Spanish and Spanish into English) video clips of similar genres were made available in both Spanish and English. Only two participants respoke the news clip so the results have not been included in the quantitative analysis.

Table 2.

Details of Video Clips Used During the Pilot Experiment

Language combination	Genre	Description	Duration	Words per minute (wpm)
ES > EN	Narration	Wildlife documentary	2 mins 24 secs	73 wpm
ES > EN	Speech	Presidential speech	2 mins 24 secs	131 wpm
ES > EN	News	RTVE - Robot museum	1 min 58 secs	191 wpm
EN > ES	Narration	<i>Desperate Housewives</i> , opening scene	2 mins 33 secs	102 wpm
EN > ES	Speech	Presidential speech	2 mins 4 secs	101 wpm
EN > ES	News	BBC - <i>Can a robot do your job?</i>	2 mins 9 secs	173 wpm

The video clips used were chosen to represent a variety of content that would usually be respoken in a real-life scenario. The narration and speech videos had one speaker and the news videos had multiple speakers. The narration videos were chosen to allow participants to carry out interlingual respoking exercises with low speech rates and long pauses. The wildlife documentary clip covered animals which inhabit the Sahara and a clip from the opening scene of *Desperate Housewives* gave insight into a character's life before the show began. Although the genres are different, these clips were chosen as they were both delivered by a narrator and did not include specialist terminology. The two speech and news videos were chosen to replicate content in the form of a live event (speech) and live television (news). The English speech was of former President Barack Obama delivering his farewell speech for the American news channel NBC. The Spanish speech was of President Mariano Rajoy announcing former King Juan Carlos' abdication of the throne. The English news clip covered the public's views on robots carrying out professional jobs and the Spanish news clip covered the opening of a robot museum in Spain. It was hoped that slightly varying speech rates would allow for further analysis of participants' performance when respoking at different speeds, which has been considered during analysis.

Participants were given the option to watch each video before respoking it and attempt each one more than once. Only the first attempts were analysed for this pilot study. Participants were given one hour to work on the tasks and respeak each video into their native language through

a microphone attached to the speech recognition software, Dragon NaturallySpeaking. Since no subtitling software compatible with Dragon was available, participants were advised to place a reduced size window of DragonPad underneath the video. DragonPad is similar to a Word document and it allowed participants to make a text box underneath the video to simulate the effect of live subtitles. Screencast software captured each attempt, which recorded the mouse and keyboard movements on screen and audio from the microphone.

Participants completed a pre-experiment questionnaire before any training was given. The questionnaire was composed of the following sections: biographical information, language skills, training, competence, subtitles and respeaking. Closed questions served to determine the demographic of the sample and for participants to rate their own competence in subtitling and interpreting. In the training, competence and subtitles sections, most questions were multiple choice to reflect the limited options for response. The respeaking section was composed of open-ended questions allowing participants to express their current perceptions of respeaking and how they thought they might perform.

After the interlingual respeaking tests, participants completed a post-experiment questionnaire composed of the following sections: level of difficulty, expectations, performance and skills. Level of difficulty and performance required participants to rate their performance and share the most difficult elements of the exercises. Expectations and skills allowed participants to reflect in detail upon what happened during the exercises and note how they perceived their own performance. Participants' perceptions on the skills and best-suited profile for an interlingual respeaker were sought before and after the test.

3. Quality Assessment¹

Intralingual respeaking has been practised in the UK since the BBC tested it in April 2001 with the World Snooker Championship (Romero-Fresco, 2011). It is evident that the focus has indeed been on extending quantity rather than improving quality. The NER model considers the number of words in the respoken text (N), the number of edition errors caused by strategies applied by the respeaker (E) and the number of recognition errors that are usually caused by mispronunciations, mishearing or errors with the speech recognition technology (R). These errors can in turn be classified as minor, standard or serious. The threshold for a set of intralingual live subtitles to be considered acceptable is 98%. The need for human intervention is highlighted through the two additional elements of the model: correct editions (CE), which account for editing that has not caused a loss of information, and the final assessment, in which the evaluator can comment on issues such as speed, delay and flow of the subtitles.

¹ This section draws heavily upon *Quality assessment in interlingual live subtitling: The NTR model* (Romero-Fresco & Pöchhacker, 2017)

The NTR model (Romero-Fresco & Pöchhacker, 2017) considers the number of words in an interlingually respoken text (N), the translation errors (T) and the recognition errors (R) to calculate the accuracy rate. Thus, the NTR model uses a NER-based formula and accounts for the shift from intralingual to interlingual live subtitling by replacing edition errors (E) with translation errors (T). The latter are in turn subdivided into content (omissions, additions and substitutions) and form (correctness and style) errors. As in the NER model, errors are also classified according to three degrees of severity (in this case minor, major and critical) and the minimum accuracy rate required is 98%.

3.1. Applying the NTR Model

Respoken texts must meet an accuracy rate of 98% to be suitable for broadcast. The user must compare the original audiovisual text with the target text of subtitles to identify each translation and recognition error and effective edition:

- Translation errors must be identified with the error sub-types listed above and then penalised depending on their severity. Three categories of severity highlight the error as either recognisable (*minor*, -0.25), causing confusion or loss of information (*major*, -0.5), or introducing misleading information (*critical*, -1).
- Recognition errors must be identified and penalised depending on their severity as outlined above.
- Differing text that has condensed information or has introduced synonyms is not penalised, rather highlighted as Effective Editions (EE).

The NTR model was applied for the first time in February 2017 to calculate the accuracy rates for this pilot study. An extract of how the model has been applied to the analysis of a respoken text can be seen in Table 3. Examples of translation and recognition errors are highlighted in red and effective editions are highlighted in yellow. Once the criteria above had been applied to all changes between the original audiovisual text and the respoken subtitles, sums of translation and recognition errors were totalled and the NTR formula was applied to calculate the accuracy rate of the respoken text. An example of this can be seen in Table 4.

Table 3.

Extract of NTR Analysis of the Narration Text for Participant 2

Original text (transcribed audio)	Respeaking-based subtitles	Errors
<p>My name is Mary Alice Young. When you read this morning's paper you may come across an article about the unusual day I had last week. Normally there is never anything news worthy about my life, but that all changed last Thursday. Of course, everything seemed quite normal at first, I made breakfast for my family, I performed my chores, I completed my projects, I ran my errands.</p>	<p>Me llamo Mary Alice Young. Cuando lees (leas) (1) el periódico de esta mañana, quizá te encuentres un artículo sobre 10 en (el día) (2) inusual que tuve la semana pasada. Normalmente, no hay nada nuevo en mi vida, era (pero) (3) todo cambió el jueves (jueves) (4) pasado. No (todo) (5) era normal al principio, (hice) (6) el desayuno para mi familia, hice mis tareas domésticas, completé mis pequeños proyectos, y e hice mis recados. La verdad, me pasé</p>	<p>1. MinR (0.25): incorrect tense but it does not impact comprehension. 2. MajR (0.5): The target text becomes incoherent and as the error is strange the viewer would not be able to identify the source text. 3. MinR (0.25): This slightly deters from the meaning of the text. 4. MinR (0.25): The source text could still be recognised, given the idea has been mentioned previously.</p>

Table 4.

Extract of NTR Assessment of the Narration Text for Participant 2

Accuracy rate	
MinT: (4 x 0.25 = 1) (cont-omiss) x 1 (cont-sub) x 1 (form-corr) x 2 MajT: 0 CritT: 0 Total: 1	MinR: (5 x 0.25 = 1.25) MajR: (2 x 0.5 = 1) CritR: (1 x 1 = 1) Total: 3.25

	NTR accuracy rate
N = 243 (209 + 34)	$\frac{243 - 1 - 3.25}{243} \times 100 = 98.25\% (5/10)$

EE: 3

Assessment
<p>The quality of the subtitles is acceptable. The translation is good (only four translation errors, one regarding content and two regarding style) and perhaps too many recognition errors (10), of which two cause the viewers to lose information and another one introduces misleading information. Still, most errors are minor and therefore do not have significant impact on comprehension.</p>

4. Quantitative Results

An accuracy rate was calculated for each respoken text of the pilot study. Below, a breakdown of errors is provided for each participant and displayed as follows for translation errors: T (MinT – MajT – CritT); and for recognition errors: R (MinR – MajR – CritR). It should be noted that the participant who respoke from Spanish into English is participant 7.

“Video 1: narration” had an average score of 6.2 translation errors and 9.3 recognition errors per participant. The overall average accuracy rate is 97.35% (3/10). The most common errors were content omissions with an average of 9.5 per text, followed by content substitutions (2), form correctness (1.9), form style (0.4) and content addition errors (0.1).

Table 5.

Translation and Recognition Errors per Participant for Video Clip 1

	Translation Errors	Recognition Errors	Accuracy Rate
Participant 1	(1–0–0)	(8–5–0)	97.98% (4/10)
Participant 2	(4–0–0)	(5–2–1)	98.25% (5/10)
Participant 3	(4–0–0)	(8–6–0)	96.93% (2/10)
Participant 4	(3–0–0)	(10–3–0)	97.87% (7/10)
Participant 6	(4–1–1)	(4–4–0)	97.38% (3/10)
Participant 7	(2–4–0)	(3–5–0)	96.42% (1/10)
Participant 8	(9–3–1)	(0–1–0)	97.45% (3/10)
Participant 9	(8–2–2)	(3–1–0)	96.50% (1/10)
Average	(4.4–1.25–0.5)	(5.8–3.4–0.1)	97.35% (3/10)

“Video 2: speech” had an average score of 6.5 translation errors and 4.5 recognition errors. The overall average accuracy rate is 97.38% (3/10). This clip was deemed the easiest clip to respeak. The most frequently made error was content omissions with an average of 3.6 per text, followed by content substitutions (1.1), form style (0.8), form correctness (0.6) and content addition errors (0.1).

Table 6.

Translation and Recognition Errors per Participant for Video Clip 2

	Translation Errors	Recognition Errors	Accuracy Rate
Participant 1	(1–1–0)	(7–5–0)	97.86% (4/10)
Participant 2	(4–0–0)	(3–1–0)	98.75% (6/10)
Participant 3	(4–2–2)	(4–1–0)	97.17% (2/10)
Participant 4	(5–0–2)	(1–3–0)	97.86% (4/10)
Participant 6	(7–2–1)	(0–1–0)	97.58% (3/10)
Participant 7	(3–1–0)	(1–0–1)	96.09% (0/10)
Participant 8	(5–6–0)	(0–1–0)	96.99% (2/10)
Participant 9	(4–1–1)	(2–2–2)	96.73% (1/10)
Average	(4.1–1.6–0.8)	(2.3–1.8–0.4)	97.38% (3/10)

Participant 2 performed exceptionally well compared to the others and had a previous training background in subtitling and respeaking. Participants 1 and 4 performed very well in comparison to the others. Participant 1 had a background in subtitling, interpreting and respeaking and participant 4’s background was in subtitling and interpreting. Participants who scored between 95%–97% are not considered to have performed well, as the threshold for live subtitles of good quality is 98%.

Table 7.

Overall Individual Performance with Translation and Recognition Errors

	Translation errors	Recognition errors	Accuracy rate
Participant 1	(2-0-0)	(15-10-0)	97.92% (4/10)
Participant 2	(4-0-0)	(8-3-1)	98.50% (5/10)
Participant 3	(8-2-2)	(12-7-0)	97.05% (2/10)
Participant 4	(8-0-2)	(11-6-0)	97.86% (4/10)
Participant 6	(11-3-2)	(4-5-0)	97.48% (3/10)
Participant 7	(5-5-0)	(4-5-1)	96.26% (0/10)
Participant 8	(14-9-1)	(0-2-0)	97.22% (3/10)
Participant 9	(12-3-3)	(5-3-2)	96.61% (1/10)

The average accuracy rate for the SMART pilot study was 92.78% (0/10), compared to the average accuracy rate for this pilot study being 97.37% (2/10). It would be unfair to compare the two as the videos for this pilot study were approximately 2 minutes long, while those for the SMART study were around 7 minutes. SMART participants were trained for a total of 8 hours, whereas the participants in this study were trained for 2 hours. This may further highlight the impact that speed had on the accuracy rates. In this study, participants would not have experienced much fatigue and had breaks in between respeaking the videos, which could be a contributing factor to higher accuracy rates.

4.1. Translation and Recognition Errors

The most common translation errors were content omissions with an average of 6.6 errors per text, followed by content substitutions with an average of 1.6 errors, form correctness (1.3), form style (0.6) and content additions (0.1). This suggests omissions should become a focal point in interlingual respeaking training. A high number of omissions may indicate that participants either struggled to keep up with the speed of the text or did not know how to translate some parts of the text and resorted to making omissions. A strategy to keep up with the text is to edit or condense it. However, in the case of live subtitling and if editing is not performed correctly, it may cause loss of information for the viewer and in turn lowers the accuracy rate of the text.

An average of 5.8 minor omissions and 0.75 major omissions were made per text, suggesting that participants were omitting pieces of information but they managed to keep the main idea of sentences. As per the NER and NTR models, the loss of a dependent idea unit is the omission of part of a sentence containing information about the “where”, “what”, “when”, “who” and “why” piece of information and is usually scored as a minor translation error. In some instances,

the participant has missed a verb so the *what* piece of information from the sentence is missing. An omission of an independent idea unit is the omission of a whole sentence and is usually scored as a major translation error as it causes substantial loss of information. Most of the time, enough information has been displayed and the viewer can follow what is happening, however, they can still miss information that a hearing audience would have access to.

Results of the SMART study show that content omission and substitution errors were the most common errors, followed by form correctness, form style and finally content additions. Participants also found omissions and substitutions difficult to manage when respeaking. This highlights that managing speed, multitasking and dealing with dense information should be developed within a training programme. The results from the SMART study also point to interlingual respeaking as feasible and ambitious but not unattainable (Davitti, Sandrelli, & Romero-Fresco, 2018).

For this pilot study, participants made on average 6.3 recognition errors in each text, with 8.3 errors in the narration and 4.3 for the speech. For the SMART study, the average number of recognition errors was 5.4 errors per text and more omissions, substitutions and form correctness errors were made than recognition errors. Participants in the SMART study could have made fewer errors as they received more dictation and intralingual respeaking practice. Such practice would have allowed participants to train their voice profiles before beginning interlingual respeaking exercises.

Participants in this study made on average 4.2 minor errors, 1.4 major errors and 0.6 critical errors per text, suggesting they were able to manage the severity of errors. These results suggest more dictation practice is needed to train the user voice profile, dictate at a steady pace and volume and make regular pauses to release words on screen. Participant voices and on-screen actions can be observed in the Screencast clips. Some clips show that poor pace and dictation caused recognition errors. For future studies, more dictation practice is required to allow those who are good live translators not to underperform due to poor dictation or using an untrained voice profile. Participants did not pause enough to release the respoken words on screen. Therefore, they did not have the opportunity to monitor their respoken output and attempt to correct recognition errors live. Again, this points to the importance of building upon and developing new software skills within a training programme.

Participant 1 made the most recognition errors with an average of 12.5 per text and the least amount of translation errors, suggesting they focussed on translation but in turn compromised recognition. In contrast, participant 8 made on average 1 recognition error and 12 translation errors per text, demonstrating they focussed on dictation and compromised translation. There is no doubt that interlingual respeaking is a complex task. Trainers and students must find a happy medium between translation and recognition errors and explore strategies and techniques to overcome both types of error to produce interlingual live subtitles that meet the quality standard.

4.2. Translation vs. Recognition Errors

Those who achieved 97% or above have an average of 8.5 minor errors, 4 major errors and 0.7 critical errors (penalisation of -4.7). Those who achieved below 97% have an average of 13 minor errors, 8 major errors and 3 critical errors (penalisation of -10.25). This suggests the best respeakers from this experiment were able to keep critical errors to a minimum and managed to have half the number of major errors. For instance, in the case of omissions this means participants managed to omit part of the sentence instead of the full sentence on twice as many occasions. Translation errors can be controlled by the respeaker, whereas recognition errors tend to be uncontrollable when not caused by pronunciation or dictation issues, as they can also be caused by software or hardware malfunction. Therefore, to a certain extent, recognition errors can also be put down to luck.

The comparison between translation errors and recognition errors shows another interesting pattern. The four best respeakers collectively made more recognition errors (63) than translation errors (36), whereas the respeakers who didn't reach 97% collectively made more translation errors (64) than recognition errors (39). The best respeakers of this experiment have very good skills for live translation, as they have fewer and less severe translation errors. Most of the errors these participants have are recognition errors, which means that their scores could have been considerably higher had they received thorough software training. Of the bottom four respeakers, the number of translation errors (64) suggests that live translation was their main weakness, which could perhaps be remedied with extensive language and interpreting training.

5. Qualitative Results

After completing the exercises, participants were asked to rate their level of difficulty on a scale of 1–5 (1 *being easy* and 5 *being difficult*). Participants' self-perception of their performance does not appear to match their actual performance. Those who rated themselves as "satisfactory" scored higher than those who rated themselves as "poor", except for participant 3. This suggests participants were aware of their performance. For example, some reported they felt the software was giving them instant feedback in the form of recognition errors as they respoke. Some participants found this difficult to deal with while trying to perform well. The table below shows the average score of level of difficulty is 4, which shows that interlingual respeaking is perceived as a complex task. It must be noted that students would not have been aware of what score constitutes as poor, good etc. due to a lack of knowledge of the NTR model.

In the post-experiment questionnaire two participants stated the respeaking tasks were linguistically difficult, seven found the speed difficult and one deemed the content comprehension difficult. Other comments included: dealing with long sentences, fear of missing information, their feelings as a respeaker and monitoring their own output.

Table 8.

Participants' Self-Rated Performance Compared with Their Actual Performance

	Level of difficulty	Perception of overall performance	Actual overall performance
Participant 1	5	Satisfactory	97.92% (4/10)
Participant 2	4	Satisfactory	98.50% (5/10)
Participant 3	3	Satisfactory	97.05% (2/10)
Participant 4	4	Satisfactory	97.86% (4/10)
Participant 6	3	Satisfactory	97.48% (3/10)
Participant 7	4	Poor	96.26% (0/10)
Participant 8	4	Poor	97.22% (3/10)
Participant 9	5	Poor	96.61% (1/10)

In the post-experiment questionnaire, participants stated that an interlingual respeaker would likely come across the following challenges: the speed of the source text, remembering to enunciate punctuation, using the software, recognition errors, paying attention to the interpretation, subtitles appearing on screen, linguistic knowledge and multitasking. Participants noted linguistic knowledge and multitasking as the two main challenges for an interlingual respeaker, thus suggesting that an advanced level of the working languages is essential to produce interlingual live subtitles. Multitasking was highly regarded by participants as a skill. This perhaps indicates a requirement for experience of simultaneous interpreting to listen in one language while speaking in another, not forgetting the added element of working with software to correct errors and cue subtitles.

Participants identified various skills that interlingual respeakers would require to improve performance. Overall, these skills were noted in both questionnaires indicating participants correctly predicted the skills deemed necessary to perform well. The following required skills were noted: a strong level of comprehension in the source language, a strong level of expression in the target language, communication, speed, multitasking, listening, software knowledge, memory, segmentation and reformulation. The following qualities were noted as being useful for an interlingual respeaker: ability to work under pressure, ability to keep pace, and focus.

After completing the exercises, six out of eight participants identified that an interpreter would be the best-suited profile for an interlingual respeaker. Two added simultaneous interpreting as the specific mode of interpreting, one noted it should be an interpreter with training in respeaking and another noted an interpreter that can work with the software. Two suggested a translator would be ideal and nobody noted a subtitler as the best-suited professional profile. Some participants noted more than one best-suited profile. Although this experiment yields interesting results on the feasibility of interlingual respeaking, further research is needed to draw meaningful

conclusions. Experience in translation or subtitling along with experience of respeaking are clearly an advantage, but interpreting skills are expected to be the main feature of interlingual respeaking.

6. Training

Given the hybrid form of the task, a training model for interlingual respeaking should be centred around the skills necessary for subtitling, interpreting and intralingual respeaking. A practical proposal for intralingual respeaking was put forward by Arumí Ribas and Romero-Fresco (2008), who note that identifying skills is a fundamental first step for designing a respeaking course. A similar taxonomy of skills may be required to highlight the skills required for interlingual respeaking to inform training. Pöchhacker and Remael (forthcoming) describe interlingual respeaking as a three-step process including pre-process, peri-process and post-process tasks. Dividing skills as per the processes that make up interlingual respeaking would clarify why each skill is necessary.

Interlingual respeaking is about providing a service to heighten access. Therefore, it is important that training is made accessible for trainees. Trainees may be current undergraduate or postgraduate students aiming to learn a new skill, others may be already well-established translators, subtitlers and interpreting professionals. The demographic of participants that took part in this pilot study is likely to be representative of future interlingual respeaking trainees (current postgraduate students and language professionals). An online training programme would cater to the need for accessibility and allow students to work at their own pace. This would not only foster greater flexibility, but also empower students to take control of their own learning, which university students and already established professionals may be more inclined to participate in. From a social constructivist approach, Kiraly (2000) proposes that translator education should be a dynamic, interactive process. A process that is based on learner empowerment and encourages interpersonal collaboration in which teachers serve as guides, consultants and assistants. A training model based on this approach, as explained by Kiraly, would build a sense of responsibility toward their own learning and future profession.

A list of criteria must be drawn up to base the training model upon. Criteria will be guided by the results of both this pilot study, the large-scale study, questionnaire responses and by taking industry training requirements into account. It will also be informed by the NER and NTR models, which outline criteria for quality assessment models. A brief example of criterion is to ensure the model is adaptable and considers language combinations, fast paced change in working environments, industry needs, audience needs and the evolution of interlingual respeaking. Once results of the large-scale study have been examined, other concepts and models for interpreter and translator training must also be considered extensively. Gile's *Effort models of interpreting* (2015) outline a set of models that explain the difficulty of interpreting, which facilitate the development of strategies for better interpreting performance. A set of three main efforts are identified: listening and analysis, speech production and short-term memory. These theoretical concepts can be taught

to allow students to identify difficulties in interpreting and develop strategies to alleviate such difficulties. Due to the similarities between Simultaneous Interpreting and interlingual respeaking, skills such as listening, source text comprehension, speech production and short-term memory are required for both tasks. Therefore, existing research on interpreting pedagogy and the training of intralingual respeakers can both be applied to research and training in interlingual respeaking. A sequential model of translation (Gile, 2015) details the translator or interpreter's progression from the source language to the target language. The model focusses on the comprehension loop and the reformulation loop to verbalise a translation. Gile explains that although translators may have days to find translation solutions, simultaneous interpreters would only have a few seconds. The concepts of this model can be practically applied to interpretation; therefore, such concepts may also be applied to interlingual respeaking given the similarity of the tasks.

Data on translation and recognition errors from this pilot experiment is valuable for deciding on material and resources for further training and experiments. Participants struggled more with content errors than they did with form errors. Content errors are easier to make and refer to the broader content of the text, but form errors refer to the grammar and register of the text, which for many comes naturally. A training programme for interlingual respeaking should focus on the sub-types of translation errors, with an extra focus on managing speed to deal with omissions and managing the content of the text to deal with substitutions.

Audiovisual materials of different genres, topics, speed and content must be selected for a training course. If chosen carefully and set in order of difficulty, they could contribute to students' progress. Processing and reformulating information are necessary in between listening and speaking and are skills that can be built up over time and with each interlingual respeaking exercise. Starting off with videos with a reduced speech rate and long pauses would allow students to grasp the initial skill required to listen and speak at the same time. Then, allowing the speech rate to increase with each respeaking exercise would give students a sense of progress and would allow them to master recognition errors with speed, which are far easier to grasp than translation errors. A course could begin with documentaries, as the slow pace, long pauses and visual images on screen may aid beginners to focus on producing a live translation without feeling rushed. Slow speeches with non-specialised content could also be used, such as those that can be sourced from the EU speech repository for interpreting practice. Then, move on to working with more complex speeches, sports, the news and weather and eventually chat shows with a very high speech rate and multiple overlapping speakers.

This pilot study has identified areas to be targeted within training to master and combat recognition errors. Participants only had a brief intralingual dictation practice of 10 minutes to familiarise themselves with the speech recognition software. However, more extensive practice is required to train the user voice profile and allow students to monitor and correct recognition errors. A unit dedicated to dictation practice would give enough time for students to work on intralingual dictation, interlingual dictation (sight translation) and develop skills in reading in one language

and respeaking in another. This could be considered as an initial step to take before approaching exercises with audiovisual texts. Being able to manage the speech recognition software (including dealing with technical issues, training Dragon with vocabulary, creating macros and error correction) and required hardware (use of the microphone, knowledge of how much the computer can handle at the same time) are also important factors.

6.1. Applying the NTR Model

To continue some thoughts on training with regards to the NTR model, its primary function is to assess the quality of interlingual live subtitles. However, the model could also be used within a training programme to raise awareness of how live subtitles are assessed. The NTR model could allow students to assess each of their respoken texts, identify errors and categorise their severity. This may provide further understanding of how and why translation and recognition errors are made. It could give students a space to reflect on their performance with different types of audiovisual text and pave the way for students to develop techniques to manage the severity of errors or avoid them altogether. There may also be room for peer review within training to reduce the subjectivity of students analysing their own texts. Peer review may be an effective way of receiving feedback, which could be a turning point in training. Like the NER model, the NTR model allows for an overall qualitative assessment, which could prove insightful for critical peer review exercises between trainers and students alike.

A text that has reached an accuracy rate of 95% could be almost unintelligible but those not familiar with the model could misinterpret accuracy rates. An acceptable accuracy rate for both intralingual and interlingual live subtitles is 98%. To mitigate against any confusion, the NTR model offers a recalculation of the accuracy rate on a 10-point scale. This would serve as another benefit of using the NTR model within a training programme to highlight the actual quality of respoken texts from an early stage in a course and allow students to grasp the concept of accuracy rates quickly.

Subjectivity in scoring has been noted as a minor concern of the NTR model. When assessing translation errors and distinguishing between them, subjectivity could potentially threaten accuracy and consistency. However, previous testing of ten evaluators using the NTR model has already proven to be successful with a low average discrepancy of 0.3 on a scale from 1 to 10 (Romero-Fresco & Pöchhacker, 2017). The average discrepancy of the analysis of this pilot study is 0.38. The author of this article served as the first marker of the texts and a co-creator of the model served as the second. Most discrepancies were due to the content of translation errors, which could be due to the first marker translating out of their mother tongue. If, in an NTR analysis, a minor error has been scored as a major error it will have little impact on the final accuracy rate. For this pilot study such discrepancies minimally increased accuracy rates by an average of 0.12%, which changed the overall score by one point out of 10 only 12% of the time.

This pilot study was conducted in February 2017, so it is believed to be the first experimental study to have applied the NTR model for the analysis of interlingual respoken texts. A total of 18 texts were produced and analysed throughout this pilot study. The model was applied to a small volume of short texts of different types and proved to provide a simple and thorough method of assessing translation and recognition errors. The average number of words for each respoken exercise in this study was 261 words. Texts that achieved 97.50% or above took on average 21 minutes to analyse with the NTR model and texts that achieved 97.49% or less took on average 33 minutes to analyse due to the extra errors that had to be identified, categorised and scored. The simplicity and thoroughness of the NTR model allows for it to be flexible in its application to texts produced by different means i.e. by respoken and automatic speech recognition; as well as supporting training in interlingual respoken to analyse the accuracy of interlingual dictation (sight translation) to support the early training of interlingual respoken.

7. Conclusions

The results presented in this article demonstrate that interlingual respoken is feasible. Challenges could be overcome by developing task-specific skills through a training programme for interlingual respoken to build upon skills used in subtitling, interpreting and respoken. For subtitling, knowledge of SDH, segmentation, reformulation and edition is required. Developing short-term memory, speed and multitasking all highlight the requirement for elements of training to mirror simultaneous interpreting. Specific skills for respoken would be software related and should include the unlearning of skills, such as speaking in a pleasant voice due to the importance of dictation and enunciation. Live translation skills are essential for interlingual respoken as is the ability to dictate accurately to the speech recognition software.

Results from this pilot study have shown that an awareness of omissions and recognition errors should be incorporated into a training programme. Omissions have proved to have an impact on overall accuracy rates and cause loss of information for the viewer. Causes of recognition errors highlight dictation as an essential part of respoken training as many errors were due to participants over-dictating by pronouncing individual syllables of words and causing misspellings of short words. Managing the error severity is an essential aspect of interlingual respoken. Participants with good live translation skills can control the severity of translation errors by keeping major and especially critical errors to a minimum, even at the expense of increasing minor errors. The severity of recognition errors cannot always be controlled. This emphasises the need for extensive dictation practice and developed software skills to minimise such errors. Participants who are not strong live translators may find it difficult to reach the minimum accuracy threshold of 98%, even if their dictation is good.

The NTR model has proved to be an effective method of assessment for the quality of interlingual live subtitles. It has the potential to be incorporated within a training programme for students to

understand the differences between each sub-type translation error and become aware of what each of their translation choices mean within the overall quality of interlingual live subtitles. Measuring quality during training could prove to be a turning point in training by raising awareness of students' output and the impact it has on the quality of a text that a viewer receives.

More research is required to determine the best-suited professional profile for an interlingual respeaker. The qualitative data suggests an interpreter would be the best-suited. However, the quantitative data shows the highest performing participants did not have previous interpreting training, which may give some hope for subtitling skills to be maintained in training. At this stage, it would be fair to conclude that interlingual respeaking is feasible. This is providing a suitable training programme is produced to train interlingual respeakers to produce quality subtitles and ensure interlingual respeaking becomes a common practice to cater for a DHOH audience and to aid the integration of foreigners into society. Now that this article on the feasibility of interlingual respeaking has been completed, a forthcoming article will focus on the next stage of this project and present the results of a large-scale study. The large-scale study delivered short online training courses to mainly Spanish natives in the UK and Spain. Data, which is in the final stages of analysis, has been gathered for 50 participants from subtitling and interpreting backgrounds. It aims to identify the task-specific skills and best-suited professional profile for an interlingual respeaker. It is hoped that a final article based on the last stage of this project will present a research-informed training model based on the data of the large-scale study. Reporting on the individual stages of research aims to thoroughly document each one, which allows the results to be suitably shared and ensures accurate progression of interlingual respeaking as the most recent mode of AVT.

Acknowledgements

My thanks go to the students and academic staff for participating in this pilot study, without whom this research would not have been possible. I would also like to thank the organisers of the Understanding Media Accessibility Quality (UMAQ) conference held in Barcelona in 2018, where the first version of this paper was presented. My deepest gratitude goes to Pablo Romero-Fresco, Lucile Desblache and Aline Remael for their insightful comments and feedback on previous versions of this paper.

References

- Arumí Ribas, M., and Romero-Fresco, P. (2008). A practical proposal for the training of respeakers 1. *The Journal of Specialised Translation*, 10, 106–127. Retrieved from <https://www.jostrans.org/>
- Dawson & Romero-Fresco, (forthcoming). Towards research-informed training in interlingual respeaking: an empirical approach. *The Interpreter and Translator Trainer*, 14. Retrieved from <https://www.tandfonline.com/toc/ritt20/current>

- Davitti, E., Sandrelli, A., & Romero-Fresco, P. (2018). Interlingual respeaking: An experimental study comparing the performance of different subject groups. *Understanding Media Accessibility Quality (UMAQ)*, Barcelona, 4–5 June 2018.
- Eugeni, C. (2008). A Sociolinguistic approach to real-time subtitling: Respeaking vs. shadowing and simultaneous interpreting. In J. Cynthia, K. Bidoli, & E. Ochse (Eds.). *English in international deaf communication*. (pp. 357–382). Bern: Peter Lang.
- Fresno, N. (2019). Of bad *hombres* and nasty women; the quality of the live closed captioning in the 2016 US final presidential debate. *Perspectives*, 27(3), 350–366. Retrieved from <https://www.tandfonline.com/doi/abs/10.1080/0907676X.2018.1526960?journalCode=rmps20>
- Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.
- Gile, D. (2015). Effort models. In F. Pöchhacker (Ed.). *Routledge encyclopaedia of interpreting studies*. (pp. 135–137). London/New York, NY: Routledge.
- Greco, G. M. (2016). On accessibility as a human right, with an application to media accessibility. In A. Matamala & P. Orero (Eds.). *Researching audio description: New approaches*, (pp. 11–33). London: Palgrave Macmillan Limited.
- Kiraly, D. (2000). *A social constructivist approach to translator education: empowerment from theory to practice*. Manchester: St Jerome Publishing.
- Neves, J. (2004). Language awareness through training in subtitling. In P. Orero (Ed.). *Topics in audiovisual translation* (pp. 127–140). Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.
- Pöchhacker, F. & Remael, A. (forthcoming). New efforts? A competence-oriented task analysis of interlingual live subtitling. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 18.
- Robert, I., Schrijver, I., & Diels, E. (forthcoming). Live subtitlers: Who are they? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 18.
- Romero-Fresco, P. (2011). *Subtitling through speech recognition: Respeaking*. Manchester: St. Jerome.
- Romero-Fresco, P. (2012). Respeaking in translator training curricula. Present and future prospects. *The Interpreter and Translator Trainer*, 6(1), 91–112. Retrieved from <https://www.tandfonline.com/toc/ritt20/current>
- Romero-Fresco, P. (2012). Quality in live subtitling: The reception of respoken subtitles in the UK. In A. Remael, P. Orero, & M. Carroll (Eds.). *Audiovisual translation at the crossroads: Media for all 3* (pp. 111–131). Amsterdam/New York, NY: Rodopi.
- Romero-Fresco, P. (2015). *The reception of subtitles for the deaf and hard of hearing in Europe*. Bern/Berlin/Bruxelles/Frankfurt am Main/New York/Oxford/Wien: Peter Lang.
- Romero-Fresco, P. & Martínez, J. (2015). Accuracy rate in live subtitling: The NER model. In J. Díaz-Cintas & R. Baños (Eds.). *Audiovisual translation in a global context: Mapping an ever-changing landscape* (pp. 28–50). Hampshire/New York, NY: Palgrave Macmillan.
- Romero-Fresco, P. (2016). Accessing communication: The quality of live subtitles in the UK. *Language & Communication*, 49, 56–69. Retrieved from <https://www.sciencedirect.com/science/journal/02715309>
- Romero-Fresco, P. & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR model. *Linguistics Antverpiensia, New Series: Themes in Translation Studies*, 16, 149–167. Retrieved from <https://lans-tts.uantwerpen.be/index.php/LANS-TTS>

Romero-Fresco, P., Melchor-Couto, S., Dawson, H., Moores, Z., & Pedregosa, I. (forthcoming). Respeaking certification: Bringing together training, research and practice. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 18.

Szarkowska, A., Krejtz, K., Dutka, Ł., & Pilipczuk, O. (2018). Are interpreters better respeakers? *Interpreter & Translator Trainer*, 12(2), 207–226. Retrieved from <https://www.tandfonline.com/toc/ritt20/current>

Usability as a Premise of Quality: First Steps Towards the Validation of the System Usability Scale (SUS) into Spanish

 Irene Tor-Carroggio 

Universitat Autònoma de Barcelona

 Daniel Segura 

Universitat Autònoma de Barcelona

 Olga Soler-Vilageliu 

Universitat Autònoma de Barcelona

Abstract

Usability is a key factor when talking about the quality of a product. The System Usability Scale (SUS) is one of the most popular tools to measure usability due to its numerous advantages and, therefore, a very useful quality assessment tool. Originally designed in English, it is available in some other languages, such as Persian and Greek but no validated version in Spanish can be found yet. This paper bridges this gap by describing the process of statistically validating the SUS into Spanish. The results show that our translation of the SUS is reliable, although our modest sample of informants ($n = 50$) leaves room for improvement and future research. The validation of the SUS is framed within a European project that will use it for its testing phase.

Key words: usability evaluation, system usability scale, SUS, questionnaire translation, questionnaire validation, Spanish.

Citation: Tor-Carroggio, I., Segura, D. & Soler-Vilageliu, O. (2019). Usability as a premise of quality: First steps towards the validation of the System Usability Scale (SUS) into Spanish. *Journal of Audiovisual Translation*, 2(2), 57–71.

Editor(s): G.M. Greco & A. Jankowska

Received: April 24, 2019

Accepted: November 21, 2019


Published: December 31, 2019

Funding: This article has been funded by the EasyTV Project (GA761999). The authors are all members of the TransMedia Catalonia research group (2017SGR113).

Copyright: ©2019 Tor-Carroggio, Segura & Soler-Vilageliu. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

 Irene.Tor@uab.cat, <https://orcid.org/0000-0003-2924-014X>

 Daniel.Segura@uab.cat, <https://orcid.org/0000-0001-7669-0755>

 Olga.Soler@uab.cat, <https://orcid.org/0000-0001-9219-1913>

1. Introduction

Although often vaguely glossed as “ease of use”, usability is a difficult construct to define, probably because it is a non-functional requirement that “does not specify parts of the system functionality” (Lauesen & Younessi, 1988, p. 1). Yet, researchers such as Lauesen and Younessi (1988, p. 1–2) compiled the five factors that have traditionally been associated with any product tagged as “usable”, namely ease of learning, task efficiency, ease of remembering, understandability and subjective satisfaction. More recently, it has been formally defined under the ISO 9241-11: 2018(E) standard as “the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments”. User satisfaction with a product or service that successfully meets their requirements is one of the angles from which quality may be looked at (Elshaer, 2012). In fact, usability has directly been conceptualised as quality of use, although this seems to be changing into quality of experience (McNamara & Kirakowski, 2005). Be that as it may, usability is a key quality factor of many products, such as successful software systems (Winter & Deissenboeck, 2008). Therefore, it seems that the following conclusion can be drawn: any product that aims at ensuring quality, must be usable.

Given both its importance and implications, usability needs to be measurable. This means that usability requirements must be tangible enough to verify and trace them during the development of any product (Lauesen & Younessi, 1988, p. 1). Usability evaluation is often carried out through questionnaires. There are many usability evaluation questionnaires available nowadays, such as the Software Usability Measurement Inventory (SUMI) (Kirakowski & Corbett, 1993), the Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 1992), the Single Usability Metric (SUM) (Sauro & Kindlund, 2005), the Website Analysis and Measurement Inventory (WAMMI) (Claridge & Kirakowski, 2011) and the System Usability Scale (SUS) (Brooke, 1996). Baumgartner, Sonderegger and Sauer (2019) even developed a pictorial single-item scale for measuring perceived usability. The SUS is amongst the most popular ones and Brooke (2013, p. 29) reported that some publications have even referred to it as an “industry standard”, although it has never formally been through a standardisation process. It has a plethora of advantages, such as brevity and robustness, as well as being free of charge (Katsanos, Tselios, & Xenos, 2012, p. 302; Bangor, Kortum, & Miller, 2008). Despite its simplicity, Tullis and Stetson (2004) noted that the SUS, when compared to other usability questionnaires, yielded among the most reliable results across sample sizes. Klug (2017) underlines that it has also been successfully applied to a wide range of devices and systems, such as scholarly repositories, websites, medical technologies, decision aids and print materials. It has also been used to test landline telephones, non-web graphical user interfaces and automated telephone interfaces (Bangor et al., 2008), which proves its flexibility and lack of dependence on the system under study. Designed in the 80s by Brooke (1996), it consists of ten standardised questions, originally written in English, half of which are positive statements, while the rest are negative. Informants need to express how much they agree with the proposed statements selecting one of five options available, ranging from *strongly disagree*

to *strongly agree* on a Likert-type scale. Final scores for the SUS can range from 0 to 100 and 68 is considered the average score (Sauro, 2011b). Yet, the SUS has also some limitations that need to be taken into consideration, such as it not being a “diagnostic instrument”, meaning that “it cannot contribute to understanding the underlying reasons that explain participants’ perceptions about the quality of the user experience” (Katsanos et al., 2012, p. 302). Therefore, the SUS is no replacement for identifying usability problems (Sauro, 2011b). Grading the SUS is also a difficult process that can result in scoring errors if not carried out properly. Calculated grades should be normalised to obtain a percentage, which will be easier to understand especially by those not familiar with this scale and, therefore, prone to believe that a raw SUS score equals a percentage simply because it falls somewhere between 0 and 100 (Usability.gov, n.d.).

1.1. The EasyTV project

The SUS was selected as the usability testing tool for the EU-funded project EasyTV (www.easytvproject.eu). This project aims at making audio-visual content and media accessible to persons with sight and hearing loss and to the growing ageing population of Europe by developing new access services. Some examples of these new services are customised subtitles, and a crowdsourcing platform by means of which videos in sign language can be uploaded and shared. These access services are expected to grant equal and better access to audio-visual content in terms of both choice and quality.

The EasyTV project follows a user-centric approach, which means that the services developed need to be tested and approved by users at different stages of development (Matamala et al., 2017). This approach had previously been defended by, for instance, Lauesen and Younessi (1988, p. 2) in the case of usability testing. These researchers pointed out that “nobody can foresee the usability problems for a given user interface – not even usability experts (...) Only some kind of testing with real users can reveal the usability problems”. They also underlined that these kind of problems need to be identified during development. For the tests it was decided that we would be focusing on the service’s usability rather than their accessibility, which was taken for granted. It was thought that, since accessibility precedes usability, testing the latter would shed more light on the quality of the user experience. Also, since the “SUS data can help provide a more complete picture of the attitudes toward a website or system being tested when used in conjunction with usability test measures such as timed tasks and the number of tasks successfully completed” (Klug, 2017, n.p.), the EasyTV’s live tests will consist of users executing some specific tasks first and then answering the SUS.

The SUS was chosen as the testing tool, for several reasons. First, it was considered to be a simple and quick way to find out whether it was worth continuing the development of the access services. Second, as Manchón and Orero (2018) explained, its questions apply to the whole system under scrutiny, ignoring opinions about the content presented or shown by the system

(i.e., it gives a usability score about a subtitling platform, but not about the quality of the subtitles themselves). Third, it has been used in other similar European projects, such as HBB4ALL (RBB, 2016) and ImAc (Agulló, forthcoming). Fourth, since it is technology-independent, scores can be compared regardless of the technology used. Fifth, results are easy to share and to understand (Klug, 2017), and this was deemed important, since the project is interdisciplinary. Lastly, even though a large sample of participants will be drawn for the final tests, the number of users recruited for the intermediate tests was expected to be quite modest (five to ten users per service), hence the need to find a reliable tool that can deal with such modest samples.

When carrying out any sort of tests it is vital to make participants feel comfortable. This comfort can and should be provided by, for instance, administering the questionnaire in the participants' mother tongue, even if they can speak more than one language. This is necessary not only because it can ease the complexity of the activity, but also because it has been found that participants answering questionnaires in a language different than their native one are more likely to give higher non-respondent rates (Groves & Couper, 1998). Also, Choi and Pak (2005) noted that the language and culture of the participants in a questionnaire can affect their perception of questions and, therefore, have an impact on their answers. Thus, were the SUS to be used in the testing phases of the project, it would be necessary to have its translation statistically validated into Spanish. In fact, statistical validation is imperative to validate this kind of tests after translating them because the reliability of the original version does not necessarily transfer to its translations. The change of one term or the new way a question is phrased could result in an unpredicted alteration of the final score.

This paper aims to make the tests to be carried out in the EasyTV project easier by validating the SUS into Spanish so that it can be used as a solid and reliable usability testing tool. This article is divided into four sections: the first one reviews previous research regarding the SUS and questionnaire validation; the second one describes the methodology of the study; the third one presents the results obtained; and the last one draws some conclusions, as well as discusses some of the challenges we faced in our study. It also outlines new research paths worth exploring in the future.

2. Research Background

The SUS is a widely researched usability tool which many academic and industrial papers revolve around. One of the most comprehensive studies is that of Bangor et al. (2008), who analysed 206 studies in which the SUS had been applied. Several findings are presented in their paper: to begin with, out of the 2,324 individual surveys that those studies comprised, the average SUS score was 70.14 and the median 75. Moreover, fewer than 6% of the mean scores fell below 50, and there were no group scores below 30. Also, after having used the SUS with many different devices, graphical user interface for OS-based computer interfaces scored the highest with an average score

of 75.24. Fourth, unlike with sex, a significant relationship was found between age and SUS scores. Finally, the authors also explored what an acceptable SUS score was:

[...] This means that products which are at least passable have SUS scores above 70, with better products scoring in the high 70s to upper 80s. Truly superior products score better than 90. Products with scores of less than 70 should be considered candidates for increased scrutiny and continued improvement and should be judged to be marginal at best. (Bangor et al., 2008, n.p.)

Bangor, et al. (2009) conducted a study with almost 1,000 participants using a version of the SUS with an extra question, which asked them to rate the user-friendliness of a product as *worst imaginable, awful, poor, OK, good, excellent, or best imaginable*. Their results pointed out that the scale ratings were very similar to SUS scores and that, therefore, the inclusion of an additional scale may be of help. Finstad (2010) underlined the frustration respondents can feel with the 5-point Likert scale and proposed a 7-point Likert scale with the options *entirely disagree, mostly disagree, somewhat disagree, neither agree nor disagree, somewhat agree, mostly agree, and entirely agree*. Sauro (2011b) came up with very interesting conclusions too, such as SUS scores being able to predict customer loyalty, five seconds with a system being enough to generate a stable SUS score and positively and negatively worded items being more harmful than helpful. Klug (2017) collected some tips of advice that according to researchers could facilitate successful administration of the SUS, such as reminding users of the alternative nature of the SUS statements, alternating the order in which tools are tested, and participants filling in the questionnaire as soon as they finish testing the product so that they provide an accurate summation. This last point had also been recommended by Brooke (1996).

As for the SUS statistical properties, Lewis and Sauro (2018) report that, initially, 10 SUS items were selected from an initial pool of 50, based on the responses of 20 people who used the full set of items to rate two software systems, one of which was known to be easier to use than the other. The items selected for the SUS were those that provided the strongest discrimination between the systems. Since the original SUS just reported strong correlations between the selected values, some researchers have investigated its reliability (Bangor et al., 2008), validity (Bangor et al., 2008) and sensitivity (Lewis & Sauro, 2009).

2.1. The Concept of Validity and the Validation Process

Gandek and Ware (1998, p. 953) briefly defined “validity” as “the extent to which a score [of a questionnaire] means what is supposed to mean”. Arribas (2004, p. 27) differentiates three kinds of validity in a questionnaire: content validity (the extent to which the items in that questionnaire are indicative of what we want to measure), construct validity (the degree to which the questionnaire reflects the theory on the measured concept) and criteria validity (the comparison of each subject’s score on that questionnaire with a gold standard).

Thus, we can conclude that validity, as a broad concept used in the context of questionnaire design, refers to the ability of a questionnaire to accurately measure a variable representing an object of study. Under that premise, we define “validation” as the process followed to corroborate a questionnaire’s validity on an empirical basis. There can be many procedures to achieve that, hence the word validation referring to a specific methodology. It must be noted, however, that some translation scholars have already used the terms validity and validation before with very different meanings in mind. For example, Martín Casado and Sánchez-Reyes (1999, p. 149) defined validation as “the adaptation of the questionnaire to the receiving culture’s concepts”.

Regarding the topic of questionnaire translation and validation, extensive literature can be found. As an example, we can take the paper by Aguilar, de la Garza González, Miranda, and Villegas (2016) explaining their effort to validate the Computer System Usability Scale questionnaire (Lewis, 1995) in Spanish. Later, Aguilar and Villegas (2016) continued their research on questionnaires and validated an adaptation of the SUS that consisted of a positive version of the original questions. Also, some papers have been written about several translations of the SUS questionnaire and their respective validations. Greek (Katsanos et al., 2012), Persian (Dianat, Ghanbari, & AsghariJafarabadi, 2014), European Portuguese (Martins, Rosa, Queirós, Silva, & Rocha, 2015) Slovene (Blažica & Lewis, 2015) and American Sign Language (Huenerfauth, Patel, & Berke, 2017) are some of the languages the SUS has already been validated in. It has also been translated into Swedish and Finnish, but these were ad hoc translations that lacked psychometric evaluation (Blažica & Lewis, 2015, p. 112). Similarly, a German translation was produced by Rummel, Ruegenhagen, & Reinhardt (2013) and validated using backwards translation, i.e., translating the SUS in German back to English, but doing this alone does not completely validate the questionnaire (Blažica & Lewis, 2015, p. 112). Even though Sauro (2011a) reports the existence of unofficial Spanish translations, none of them seems to have been validated so far.

3. Methodology

The first step to validate the Spanish translation of the SUS was, obviously, to translate it. For that purpose, we decided to follow the instructions regarding questionnaire translation described by Tsang, Royse and Terkawi (2017). Their method was chosen because it is simple to follow, comprehensive and has proven to be effective in the past. Also, very similar procedures were followed in other articles focusing on questionnaire translation and validation, such as Domínguez, Balkrishnan, Ellzey, and Pandya’s (2006). The first step was to establish an expert committee, which Terkawi et al. (2017, n.p.) suggest as a preliminary step to “produce a prefinal version of the translation”. According to these researchers (2017, n.p.), the committee “should include experts who are familiar with the construct of interest, a methodologist, both the forward and backward translators, and if possible, developers of the original questionnaires”. Representatives of all the suggested categories were on the committee except for the developers of the original questionnaire. The committee agreed on, for example, the profile

of users that would participate in the validation. It was decided that the users in the study would be English speakers (both native and non-native) and native Spanish speakers, as respectively the control and the experimental group, from all age groups and of various educational backgrounds. The committee also chose the system to be tested in terms of usability (the EasyTV project website), as the results could also shed some light on what improvements could be made in it and therefore help the project, although that was not the main objective of our study. The main translation challenges were also discussed within this group and, with that feedback, the SUS questionnaire was translated into Spanish by two of the present study's researchers, who are Spanish professional translators. This team-based approach is said to be preferred among translation researchers, since it generates more translation options and provides more solid and less idiosyncratic translation evaluation (Forsyth, Stapleton, Lawrence, Levin, & Lewis, 2006, p. 4114). They worked separately and reached an agreement on the final draft, which was then sent to two different translators. These were briefed on the task and carried out a backwards translation into English. The researchers were available to answer questions and provide guidance, in case it was necessary. The objective of this exercise was to check that the translation did not convey any other meaning that was not present in the original questionnaire. Although the persons carrying out the backwards translation were professional translators, they were Spanish too, which obviously constituted a limitation in our research.

The following version of the SUS questionnaire was elaborated based on the input received from the two backwards translations. It was then posted on Google Forms, preceded by five tasks that the informants had to carry out in order to evaluate the usability of the EasyTV's website. The questionnaire also included some demographic questions about the informants' sex, age and mother tongue. The informants were also asked if they had already visited the website, since it was pointed out by Sauro (2011a) that previous experience with the website under study results in higher SUS scores. An important remark to be made is that instead of using the word *sistema* ("system"), which is the one used in the official version, we opted for substituting it with *página web* ("website"). Bangor et al. (2008) also reported that using "product" instead of "system" based on user feedback did not affect the results. It must also be highlighted that an English version of the questionnaire was also drafted so that it could be used in the control group.

For the pilot, four informants were recruited. After answering the questionnaire in Spanish, they also participated in a cognitive interview, similar to the one reported in Forsyth et al. (2006). This interview aimed at finding out how they understood each of the SUS questions and their feelings on the proposed tasks. The results of the pilot brought about several modifications. First, it was discovered that each informant had a different understanding of what the word "usability" means, so the researchers decided to eliminate this word in the title of the questionnaire to avoid biasing the subjects beforehand. The questions that contained this word were also rephrased. For example, the question "Did you find the website usable?" was substituted for "Did you feel you could use the website in an effective, efficient and satisfactory way?". It was also found out that the informants took less time to complete the tasks than what the researchers had anticipated,

so the information on how much time was needed was modified accordingly. Also, some of the tasks were changed or rewritten, as they were not clear enough to the informants or useful for the researchers. This is the case of, for example, a task that originally asked the respondents if they would be able to send an email to the project leader if needed. It was discovered that some of them did not even try to do it and just answered “yes”. Thus, it was decided that all tasks should look for specific information rather than just offering a yes/no type of answer. The following questions, extracted from the final questionnaire, indicate what kind of tasks informants had to perform in order to be able to answer them:

1. How many publications does Pilar Orero have (with other co-authors)?
 - 3
 - 4
 - 5
 - Don't know

2. How many members does the EasyTV consortium have?
 - 7
 - 8
 - 9
 - Don't know

3. What sex is the sign language interpreter in the “general overview”?
 - Male
 - Female
 - Don't know

4. How many languages are available in the EasyTV website?
 - 3
 - 4
 - 5
 - Don't know

5. How many objectives does EasyTV have?
 - 4
 - 5
 - 6
 - Don't know

6. How many categories can the news section be divided into?
 - 2
 - 3
 - 4
 - Don't know

Regarding the translation itself, it was revealed that some words like *inconsistencia* “inconsistency” led to confusion and they had to be changed (in this case to *inconsistente* “inconsistent”). This was not the first time that a word was found to be problematic, as the word “cumbersome” had already been found uncertain in previous studies in English and had, therefore, been changed for “awkward” (Bangor et al., 2008). Lastly, some commentaries about the questionnaire design were also made (for example, informants complained that the questions were too repetitive), but the researchers could do nothing about that, as the goal of the study was to validate the translation, and not to modify or improve the original SUS. The final version of the SUS that was approved was the following:

Table 1.

Items of the SUS and their Translation into Spanish

English version of SUS	Spanish version of SUS
1. I think that I would like to use this system frequently.	1. Creo que me gustaría usar esta página web frecuentemente.
2. I found the system unnecessarily complex.	2. Me ha parecido innecesariamente compleja esta página web.
3. I thought the system was easy to use.	3. La página web me ha parecido fácil de usar.
4. I think that I would need the support of a technical person to be able to use this system.	4. Creo que necesitaría la ayuda de una persona con conocimientos técnicos para usar esta página web.
5. I found the various functions in this system were well integrated.	5. Me ha parecido que las distintas funciones de esta página web están bien integradas.
6. I thought there was too much inconsistency in this system.	6. Creo que la página web es demasiado inconsistente.
7. I would imagine that most people would learn to use this system very quickly.	7. Imagino que la mayoría de la gente aprendería a usar esta página web de forma muy rápida.
8. I found the system very cumbersome to use.	8. La página web me ha parecido engorrosa.
9. I felt very confident using the system.	9. Tenía muy claro cómo usar esta página web todo el rato.
10. I needed to learn a lot of things before I could get going with this system.	10. Tuve que adquirir muchos conocimientos antes de poder usar esta página web.

Once the final version was agreed on, the approval of our university's ethics committee was requested and obtained in March 2019. The questionnaires, both in Spanish and in English were distributed between March and early April 2019. Although the users recruited for the EasyTV tests will be functionally diverse and elderly persons, there was no specific demographic characteristic requested for this study, since the SUS can be applied regardless of the user profile. The test results are described next.

4. Results

In this section, the results obtained during the tests will be analysed. We will try to find out if the questionnaire is reliable by using statistical methods and if the translation is valid by comparing the SUS scores of the English and Spanish versions. Also, we will analyse the usability of the EasyTV webpage on its own.

4.1 Spanish SUS Validation

A total of 50 informants answered the questionnaire in Spanish. Among those, 70% chose Spanish as their mother tongue, while 25% chose Catalan (2 users, or 5% of the total, chose "other"). The control group consisted of 19 informants, who performed the same exact tasks and answered the questionnaire in English. Because of the difficulty of finding native English speakers (only 10% of the respondents), non-native speakers with high language proficiency were allowed to answer. As the questionnaire's content was not difficult and the tasks were simple, we do not think that the control group results were affected by this choice. What could affect the results, however, were each participant's technological skills, as the system being tested was a website and the tasks to be done in the tests required some basic computer knowledge. Because of that, informants were asked how often they surfed the internet. All participants in the control group answered "Every day or almost every day", while 98% of the ones in the experimental group would give the same answer (1 user answered "3 times per week or more"). The participants in the control group were slightly younger on average than the ones in the experimental group (the average being 29.58 and 32.52, respectively). The questionnaires were distributed through Google Forms and the data for the statistical analysis was processed using the software IBM SPSS (version 23). Before performing any analyses, we corrected the SUS scale using the transformations to obtain positive scoring between 0 and 4 for all 10 items. This is important because questions 2, 4, 6, 8 and 10 are reversed questions.

We first tested the reliability of the Spanish version of SUS with two different reliability tests: Cronbach's alpha, a widely used measure of internal consistency, and Guttman lambda-2. Both Cronbach's alpha coefficient (.904) and lambda2 values (.911) are very high for the ten items, indicating that the questionnaire is highly consistent.

We also carried a factor analysis to identify possible subscales. Lewis and Sauro (2009) detected that English SUS can be subdivided into two subscales that measure Usability (questions 1, 2, 3, 5, 6, 7, 8, 9) and Learnability (questions 4 and 10). A factor analysis using common varimax rotation with Kaiser normalisation extracted two principal components, clearly identifying the suggested. Table 2 shows the components identified in the SUS scale. Questions 4 and 10 are clearly related to a different dimension than the rest.

Table 2.

Rotated Component Matrix (rotation converged in 3 iterations)

	Component 1	Component 2
Q7N	.795	
Q2N	.778	
Q1N	.775	
Q3N	.774	
Q9N	.769	
Q8N	.754	
Q5N	.745	
Q6N	.513	
Q10N		.894
Q4N		.814

Furthermore, we tested the reliability within both subscales. Cronbach's alpha for usability scale (8 items) is .907, and for learnability scale (2 items) equals .786.

According to these results, the Spanish version of SUS is a highly reliable scale. Within the scale we could identify two subscales, as defined by Lewis and Sauro (2009). Both subscales, usability and learnability, also show good consistency.

Next analyses consisted of comparing Spanish SUS scores with the control group scores, that is, the group of 19 people that answered the questionnaire in English. We ran a non-parametric test, the Independent-Samples Mann-Whitney U Test, which did not detect any differences in the scoring of all the questions across groups (all $p > .05$), meaning that the scores are not statistically different from one to another.

4.2 Evaluation of EasyTV website

The results obtained in our analyses indicate that our translation of the SUS questionnaire in Spanish is consistent and comparable to English SUS. Therefore, we can make use of the evaluation that our informants did on the EasyTV website' usability. The English SUS group qualified the website

with a mean score of 74.07, whereas the Spanish SUS group gave a rating of 69.6. The difference between these two scores cannot be interpreted as significant taking into account the number of respondents for each questionnaire, especially the one in English (<20). Both ratings are above the average score that Sauro (2011b) specified (68). Also, if we combine both scores we obtain a mean score of 71.83, which, since it is above 70, it is “at least passable” (Bangor et al., 2008). Nonetheless, it stands to reason that this score leaves room for further improvement so that the website’s usability can increase. In fact, future improvements could and should start by looking at the issues encountered by some informants. For example:

1. “The web is not user-friendly”.
2. “What I found the hardest to find were the publications”.
3. “Although there are several languages to choose from, not all the functions have the corresponding translation”.
4. “I think the website is easy to use and clear. I found all the information easily. My only comment perhaps is the images chosen for the front page – I think they are not very clear in terms of the message that the website wants to send. They are very cluttered when compared to the rest of the website.”

Finally, it should be borne in mind that the scores could change depending on the informants’ profile (for example, it could be less usable for users with functional diversity or of certain age), but, as it was mentioned in the methodology section, the demographic characteristics of the informants were not taken into account in this test. In general, though, we can conclude that the EasyTV project website is usable enough and the users were satisfied with it.

5. Conclusions

This study has its limitations. First, the backwards translation carried out during the process of coming up with the final Spanish version of the SUS was performed by two non-English speakers. Second, the samples obtained in the questionnaires (in both the English and the Spanish versions) were not randomly selected and their sizes also leave room for improvement. Third, some of the persons who filled in the questionnaire in English were not native speakers, which obviously can compromise the results. In fact, Finstad (2006) had found that non-native speakers had difficulty in understanding the word “cumbersome” as opposed to English native speakers.

Due to the limited sample of users, this article can be read as a first step towards a definitive validation of the SUS questionnaire in the Spanish language. The next logical step towards successful validation would be to conduct a study involving a higher number of informants. The vast majority of validation studies cited in the current article used samples ranging from 150 to 300 informants.

Consequently, we think that a similar number of informants for both the control group and the test group would yield statistically more powerful results.

However, even considering these limitations, the tests carried out show that the Spanish translation of the SUS questionnaire has been an (interim) success. The results of Cronbach's Alpha and Guttman Lambda tests demonstrate that the questionnaire is very reliable. This happens both if we consider the 10 questions as a whole and if we measure usability and learnability as two separate variables. Also, no differences were detected in the SUS scores in the English and Spanish versions. This demonstrates that the phrasing of the questions in the Spanish version did not lead to a different interpretation of their meaning or change the original questions' goal. We can conclude that, at least with our limited number of informants, the translation of the SUS questionnaire is good and precise enough as to be used in subsequent usability tests and, therefore, to guarantee their quality.

References

- Aguilar, M. I. H., de la Garza González, A., Miranda, M. P. S., & Villegas, A. A. G. (2016). Adaptación al español del Cuestionario de Usabilidad de Sistemas Informáticos CSUQ/Spanish language adaptation of the Computer Systems Usability Questionnaire CSUQ. *Revista Iberoamericana de las Ciencias Computacionales e Informática*, 4(8), 84–99.
- Aguilar, M. I. H., & Villegas, A. A. G. (2016). Análisis comparativo de la Escala de Usabilidad del Sistema (EUS) en dos versiones/Comparative analysis of the System Usability Scale (SUS) in two versions. *Revista Iberoamericana de las Ciencias Computacionales e Informática*, 5(10), 44–58.
- Agulló, B. (forthcoming). Technology for subtitling: a 360-degree turn. To appear in *Hermeneus*.
- Arribas, M. (2004). Diseño y validación de cuestionarios [Design and validation of questionnaires]. *Matronas Profesión*, 5(17), 23–29.
- Bangor, A., Kortum, P., & Miller, J. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594.
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114–123.
- Baumgartner, J., Sonderegger, A., & Sauer, J. (2019). No need to read: Developing a pictorial single-item scale for measuring perceived usability. *International Journal of Human-Computer Interaction*, 122, 78–89.
- Blažica, B., & Lewis, J. (2015). The SUS-SI. Intl. *Journal of Human-Computer Interaction*, 31, 112–117.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor and Francis.
- Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, 8(2), 29–40.
- Choi, B. C., & Pak, A. W. (2005). Peer reviewed: A catalog of biases in questionnaires. *Preventing chronic disease*, 2(1).

- Claridge, N., & Kirakowski, J. (2011). WAMMI: website analysis and measurement inventory questionnaire. Retrieved from <http://www.wammi.com/>
- Dianat, I., Ghanbari, Z., & AsghariJafarabadi, M. (2014). Psychometric properties of the Persian language version of the System Usability Scale. *Health Promotion Perspectives*, 4(1), 82–89.
- Dominguez, A., Balkrishnan, R., Ellzey, A., & Pandya, A. (2006). Melasma in Latina patients: Cross-cultural adaptation and validation of a quality-of-life questionnaire in Spanish language. *Journal of the American Academy of Dermatology*, 55(1), 59–66.
- Elshaer, I. (2012). What is the meaning of quality? Retrieved from <https://mpra.ub.uni-muenchen.de/57345>
- Finstad, K. (2006). The system usability scale and non-native English speakers. *Journal of Usability Studies*, 1(4), 185–188.
- Finstad, K. (2010). Response interpolation and scale sensitivity: Evidence against 5-point scales. *Journal of Usability Studies*, 5(3), 104–110. Retrieved from <http://uxpajournal.org/response-interpolation-and-scale-sensitivity-evidence-against-5-point-scales/>
- Forsyth, B., Stapleton, M., Lawrence, D., Levin, K., & Lewis, G. (2006). Methods for translating survey questionnaires. AAPOR – ASA Section on Survey Research Methods, 4114–4119. Retrieved from https://www.researchgate.net/publication/239856416_Methods_for_Translating_Survey_Questionnaires1
- Gandek, B., & Ware Jr, J. E. (1998). Methods for validating and norming translations of health status questionnaires: the IQOLA project approach. *Journal of Clinical Epidemiology*, 51(11), 953–959.
- Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238.
- Huenerfauth, M., Patel, K., & Berke, L. (2017). Design and psychometric evaluation of an American Sign Language translation of the System Usability Scale. *Proceedings of the 19th International ACM SIGACCESS Conference*.
- Groves, R. M., & Couper, M. P. (1998). *Non-response in household interview surveys*. New York: Wiley & Sons.
- ISO (International Organization for Standardization) (2018). *Ergonomics of human – system interaction. Part 11: Usability: Definitions and concepts* (ISO 9241-11:2018(E)).
- Katsanos, C., Tselios, N., & Xenos, M. (2012). Perceived usability evaluation of learning management systems: A first step towards standardization of the system usability scale in Greek. *Proceedings of the 2012 16th Panhellenic Conference on Informatics*.
- Kirakowski, J., & Corbett, M. (1993). SUMI: The software usability measurement inventory. *British Journal of Educational Technology*, 24(3), 210–212.
- Klug, B. (2017). An overview of the System Usability Scale in library website and system usability testing. *Journal of Library User Experience*, 1(6). Retrieved from <https://quod.lib.umich.edu/w/weave/12535642.0001.602?view=text;rgn=main>
- Lausen, S., & Younessi, H. (1988). Six styles for usability requirements. In *Proceedings of REFSQ'98*. Namur: Presses Universitaires de Namur.
- Lewis, J. (1992). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 36(16), 1259-1260. Los Angeles, CA: SAGE Publications.
- Lewis, J. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57–78.

- Lewis, J., & Sauro, J. (2009). The factor structure of the System Usability Scale. In M. Kurosu (Ed.), *Human Centered Design, HCII 2009* (pp. 94–103). Heidelberg, Germany: Springer-Verlag.
- Lewis, J., & Sauro, J. (2018). Item benchmarks for the System Usability Scale. *Journal of Usability Studies*, 13(3), 158–167
- Manchón, L. M., & Orero, P. (2018). Usability tests for personalised subtitles. *Translation Spaces*, 7(2), 263–284.
- Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., & Rocha, N. P. (2015). European Portuguese validation of the system usability scale (SUS). *Procedia Computer Science*, 67, 293–300.
- Matamala, A., Orero, P., Rovira-Esteva, S., Casas-Tost, H., Morales-Morante, F., Soler-Vilageliu, O., Agulló, B., Fidyka, A., Segura, A., & Tor-Carroggio, I. (2018). User-centric approaches in access services evaluation: Profiling the end user. *Proceedings of the Eleventh International Conference on Language Resources Evaluation (LREC 2018)*, 1–7.
- McNamara, N., & Kirakowski, J. (2005). Defining usability: quality of use or quality of experience? In *Proceedings of the International Professional Communication Conference* (pp. 200–204). Piscataway, NJ: IEEE.
- Rummel, B., Ruegenhagen, E., & Reinhardt, W. (2013). System Usability Scale [German trans.]. Retrieved from <http://www.sapdesignguild.org/resources/sus.asp>
- RBB. (2016). D3.4 – Pilot-A evaluations and recommendations. Retrieved from http://pagines.uab.cat/hbb4all/sites/pagines.uab.cat.hbb4all/files/d3.4-rbb_pilot-a-evaluation-and-recommendations_v1.00.pdf
- Sauro, J., & Kindlund, E. (2005). A method to standardize usability metrics into a single score. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 401–409. New York, NY: ACM Press.
- Sauro, J. (2011a). *A practical guide to the System Usability Scale: Background, benchmarks, & best practices*. Denver, CO: Measuring Usability LLC.
- Sauro, J. (2011b) SUSstified? Little-known system usability scale facts. *User Experience: The Magazine of the User Experience Professionals Association*, 10(3). Retrieved from <http://uxpamagazine.org/sustified/>
- Martín Casado, M., & Sánchez-Reyes, M.S. (1999). Más allá de la traducción: la validación de cuestionarios científico-técnicos inglés-español [Beyond translation: the validation of scientific-technical questionnaires English-Spanish]. *Livius (Revista de Estudios de Traducción)*, 14, 149–155.
- Terkawi, A. S., Tsang, S., Abolhair, A., Alsharif, M., Alswiti, M., Alsadoun, A., & Altirkawi, K. A. (2017). Development and validation of Arabic version of the short-form McGill pain questionnaire. *Saudi Journal of Anaesthesia*, 11(Suppl 1), S2.
- Tsang, S., Royse, C., & Sulieman, A. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia*, 11(1): S80–S89. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5463570/>
- Tullis, T., & Stetson, J. (2004). A comparison of questionnaires for assessing website usability. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.396.3677&rep=rep1&type=pdf>
- Usability.gov. (n.d.). *System Usability Scale (SUS)*. Retrieved from <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>
- Winter, S., & Deissenboeck, F. (2008). A comprehensive model of usability. In J. Gulliksen, M. B. Harning, P. Palanque, G. C. van der Veer, & J. Wesson (Eds.), *Engineering Interactive Systems. EIS 2007 Joint Working Conferences EHCI 2007, DSV-IS 2007, HCSE 2007, Salamanca, Spain, March 22–24, 2007. Selected Papers* (pp. 106–122). Berlin, Heidelberg: Springer.

Translating Code-Switching on the Screen: *Spanglish* and L3-as-Theme

 Montse Corrius 

Universitat de Vic- Universitat central de Catalunya

 Patrick Zabalbeascoa 

Universitat Pompeu Fabra

Abstract

This paper outlines the complexity, for the purpose of translation, of accounting for how languages alternate in multilingual films and the realisation that the number and importance of instances of L3 (forms of expression other than a text's main language) constitute a variable that can ultimately affect a translator's solutions. In particular, it focuses on an issue that has not received much scholarly attention so far, the fictional representation of code-switching (CS) in feature films, with examples drawn largely from *Spanglish* (2004), given that one of the main themes of the film is language diversity and its problems for interpersonal communication. The paper distinguishes different types of language shifts (alternations) as part of a film's plot or script, involving translation between characters (intratextual diegetic translation) or otherwise, in order to better characterise CS as a concept borrowed from sociolinguistics. We then include CS in a broader concept of language shifts and distinguish L3 as a translation problem from CS. A tripartite classification of films is proposed, according to the amount and importance of L3: anecdotal, recurrent, and L3-as-theme.

Key words: multilingualism, third language (L3), code-switching (CS), alternating languages, *Spanglish*, L3-as-theme.

Citation: Corrius, M. & Zabalbeascoa, P. (2019). Translating code-switching on the screen: *Spanglish* and L3-as-theme. *Journal of Audiovisual Translation*, 2(2), 72–91.

Editor(s): G.M. Greco & A. Jankowska

Received: September 19, 2019

Accepted: December 6, 2019

Published: December 31, 2019

Funding: This article is part of the research carried out in the MUFITAVI Project, funded by the Spanish Ministry of Finance and Competitiveness PGC2018-099823-B-I00).

Copyright: ©2019 Corrius & Zabalbeascoa. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

 montse.corrius@uvic.cat, <https://orcid.org/0000-0002-5032-1246>

 patrick.zabalbeascoa@upf.edu, <https://orcid.org/0000-0002-4027-5178>

1. Introduction

The importance and amount of multilingual films has increased over the last decades (Heiss, 2004, p. 208; de Higes-Andino, 2014, p. 211). Awareness of the variables involved needs to grow as well in order to provide insight, guidelines and justifications for greater quality in the translations of such films and TV shows. For the purpose of translation and translation studies, we use a notation system of L1 for the main language (not necessarily English) of any source text (ST), and L2 for the main language of any translation (TT); L3, then, signals any language in the source text other than L1 (L3ST), or any language in translated texts other than L2 (L3^{TT}). Thus, a source text comprised of L1+L3ST or any translation comprised of L2+L3^{TT} contains language variation by virtue of the presence of L3, or *third language* (Corrius, 2008). Androutsopoulos (2007) observes that in multilingual texts one language tends to predominate quite noticeably over the other(s). From the perspective of (audiovisual) translation, the ST and the TT¹ (AVT) have different predominant languages, L1 and L2, respectively, leading to the notion of *third language(s)* in ST and TT, for any other languages. L3 is less frequent in any text than L1 (or L2), and this is its defining feature; other L3 features—like comprehensibility, proximity, historical rivalry, stereotypes, and prejudices—are variables to be established anew for every case study (Corrius and Zabalbeascoa, 2011, p. 211). Although L3 was initially proposed to tag foreign languages it can also be used to locate any relevant use of language variation (dialects, sociolects, chronolects, idiolects) when used in fiction as a stylistic device to sound “foreign”. The important point to be made is that the *third language* is not a language (French, Chinese) but a sign of meaningful language variation; so, the notation system groups all multiple different languages under the same label of L3, with no need for L4, L5, etc. Different languages can be distinguished under L3 (L3Fr for French, L3Chi for Chinese, or L3a, L3b, and so on).

Viewers are usually meant to identify with a film’s in-group characters, who nearly always speak the same language as the audience. They tend to come from the same country (Bleichenbacher, 2008), as an important part of a shared identity, except of course, in history dramas, space operas, or fantasy films (e.g., *Gladiator*, 2000; *Star Wars*, 1977; *The Lord of the Rings*, 2001), where language is even more important because national background is not a factor of empathy with the audience. The out-group characters, by contrast, have strange, alien, or foreign identities, from the audience’s perspective, and the effect is reinforced when they use forms of expression the audience does not use or know, and this is true even for films set in remote or make-believe times and places. Alienation of out-group characters by use of L3 is a frequent practice for Hollywood productions. L1 is in line with the intended audience’s language, and this alignment tends to be used as a device for in-group “bonding”, especially when, as so often happens, the *third language* is used to stress the point that some of the characters come from “other” (out-group) cultural backgrounds

¹ TT / target text (translation) includes AVT as well as written translations and interpreting. Likewise, ST / source text includes audiovisuals as well as written and oral texts.

or when there is a need to show the audience that the cultural setting has changed. In AVT, quality may depend on how languages are assigned, birthplaces conveyed (and kept or changed) and possible discrepancies between the use and form of L2 and intended character portrayal. These factors are related to and depend on additional factors such as the mode of translation (dubbing, or subtitling, or a hybrid version of both), well-established AVT conventions and criteria, and, quite fundamentally for the main topic of this study, the AVT intended audience's language profile, their knowledge and tolerance of other languages, and, particularly, when L3ST happens to coincide with L2.

As the presence and importance of the *third language(s)* vary from film to film, we propose the following categories:

- *Anecdotal*, with only a few L3 words throughout the film, e.g., L3-French in *Moulin Rouge* (2001).
- *Recurrent*, where we can hear languages other than L1 in numerous parts of the film, e.g., L3 Spanish in *Land and Freedom* (1995).
- *L3-as-theme*, meaning that there is an intense and highly relevant presence of L3, language is referred to and talked about and metalingual elements or conversations might appear, too, e.g., *Inglourious Basterds* (2009).

L3-as-theme may involve a constant mixture of languages as part of the film's plot, and is featured heavily in films like *Um filme falado* (2003), which mixes Portuguese, French, English, Italian and Greek; or *Spanglish* (2004), with a fairly balanced mix of English and Spanish. In these films each character speaks their own language as an important part of their national identity. For Díaz-Cintas (2015), languages mark geographical and political borders and embody the characters' social, cultural and personal features. Linguistic uniqueness helps to develop the theme of characters' otherness. In Hatim and Mason's words (1994, p. 43), "[i]n situations where two or more codes coexist in a speech community, code-switching is not random and the translator or interpreter, like all language users, must be able to recognise the question of 'identity' involved". Language as a system of communication is a pervasive theme in both films. Needless to say, an awareness of the third language is relevant not only to language and linguistic diversity, but also to culture and identity. "We use variation in language to construct ourselves as social beings, to signal who we are, and who we are not and cannot be" (Lippi-Green, 1997, p. 63). L3 is usually one of the means by which a different culture is brought to the fore. Hence, third-language presence becomes more relevant as a signal of place, context, identity and cultural distance and not just a mere alternative means of conveying a verbal message.

2. Code-Switching, a Misleading Term?

Code-switching can be defined as “the use of several languages or dialects in the same conversation or sentence by bilingual people” (Gardner-Chloros, 2009, p. 4). CS is used in particular circumstances by bilingual people who alternate between languages in an unchanged setting (Bullock and Toribio, 2009). Thus, “code-switching has and creates communicative and social meaning” (Auer, 1998, p. 1). In fact, it even reflects social differences and tendencies within a given society: it identifies a social group, that is to say, it symbolises identities beyond the linguistic fact. Bilinguals may display different degrees of proficiency in each language in CS, and the actual features of CS may change in different contact settings. The reasons for using CS may vary, including, for example, a speaker’s need to fill certain lexical gaps, to express their ethnic identity or to achieve a particular discursive aim, among others (Bullock and Toribio, 2009).

The term “code-switching” was first coined by Hans Vogt in 1954—in the field of linguistics—as a form of bilingual behaviour. Later, it appeared in Haugen (1956), Diebold (1961), Jakobson (1961) and Álvarez Caccamo, (1990, 1998), among others. Other terms have been used to refer to this phenomenon or to certain aspects of it, such as: “transversion” (Clyne, 2003), “alternation” (Auer, 1984), “code-mixing” (Muysken, 2000), “insertion” or “congruent lexicalisation” (Pfaff, 1976), “interference” (Weinreich, 1953; Paradis, 1993; Treffers-Daller, 2005), “borrowing” (Thomason and Kaufman’s, 1988), “nonce-borrowing” (Poplack and Meechan, 1995).

Although “code-switching” has the widest currency among scholars, for Gardner-Chloros (2009, p. 11) the two parts of the term (“code” and “switching”) are misleading. Initially, “code-switching” referred to “a mechanism for the unambiguous transduction of signals between systems”; “code” originally had nothing to do with language and was borrowed from the field of communication technology. Nowadays, “code” is understood as a neutral term to refer to languages, dialects, styles/registers, etc., which may be used instead of the term “variety”. In AVT, Chaume (2001) has proposed various signifying codes, which include the linguistic code. Secondly, “switching” refers to the alternation between the different varieties that people speak, and in the early psycholinguistic studies carried out in the 1950s and 1960s, “psychologists assumed that something similar to flicking an electric switch went on when bilinguals changed languages” (Gardner-Chloros, 2009, p. 11). For a long time, scholars did not pay much attention to the study of code-switching, and research into it developed slowly and almost invisibly until the work done by Gumperz and his associates—Wilson and Blom—in the 1960s and early 1970s (Gardner-Chloros, 2009). In the last forty years, interest in the topic has grown and related research been carried out. CS as an object of study in AV fiction (as multilingual texts) is challenging because it deals with several languages, and CS comprises a broad range of contact phenomena (how closely languages interact), which makes it difficult to set the boundaries of a well-defined characterisation: its linguistic manifestation may vary from cases involving the insertion of a single word to other cases where languages alternate at a lively pace, involving larger segments of discourse.

Following Myers-Scotton, Monti (2014, p. 137) distinguishes three different types of CS in films: *turn-specific*, when code-switching is determined by changing languages in the turns of each character in a conversation; *intersentential*, across sentence boundaries within a given turn; and *intrasentential*, within the same sentence (from single-morpheme to clause level). CS in AV fiction can be used to establish the same types of relationships that are found in real multilingual communities (Monti, 2014, p. 137), though it is essential for researchers and translators not to forget the crucial distinction between real-life language use and the practices of scriptwriting.

3. Code-Switching Among Other L3 Alternations in Film Fiction

The studies and references mentioned above have outlined many of the formal features and variables of language variation, and even some of the functional and stylistic ones. According to Zabalbeascoa and Corrius (2014, p. 258–259) the most salient features of the third language include how L3 may be (in different degrees):

- real or constructed (e.g., Italian, Klingon);
- comprehensible or not (to the audience or to other characters);
- native or non-native (ranging from high level of speaker proficiency to high level of incompetence or farcical representation);
- source-text L3 may happen to be the same language as the main language of the translation (L2): $L3^{ST} = L2$;
- L3 may be part of a stereotype, either a stereotype of the language or a stereotype of some or all of its speakers, or certain aspects of its related culture(s);
- the various languages may be related (members of the same family) or have a history (e.g., rivalry, dominance) together that goes back more or less in time;
- L3 may be a familiar or recognisable language, in different degrees, or regarded as exotic and completely unfamiliar to the intended audience.

Our claim in this paper is that a full account of L3 features and variables can probably be reached by looking not only at the variables involved in each individual instance of L3 appearance in a feature film (as listed above), but that it is also necessary and fruitful to observe patterns of how L3 alternates with either L1 or L2, depending on whether we are dealing with ST films or their translated versions. By “alternate” we mean the dynamics of shifting (i.e., switching) from one language to another—any number of times—in any combination of L3 with L1 or L2, or even between different languages with L3 status. Formally, we could represent such language-shifting dynamics, including code-switching, with a “>” sign, for instance: $L1 > L3 > L1$, to show how L3 alternates with L1 by appearing between two L1 utterances, and $L2 > L3 > L2$ for the case of translation.

So, L3 variables are established by asking questions such as how many different L3-languages are there and which ones are they; and then by adding further details, as checklisted above.

And beyond that, the questions that arise from this approach include: Does it matter how many L1>L3 shifts or switches there are and where they occur in a given film (in terms of finding coherence and consistency in the TT)? Are there different types of shifts and switches and what are their variables? What kind of features can be associated with them? Do they have more to do with style, plot development or character portrayal?

We can establish (so far) at least five types of switches involving alternations of L1 (or L2) and L3 for the purpose of distinguishing code-switching in feature films from other interlinguistic alternations or uses of L3 by the scriptwriter:

- Intratextual (diegetic) translation; prototypically, one character is translating for another (e.g., acting as official or casual interpreter).
- Code-switching by bilingual characters (as defined by Gardner-Chloros, 2009, p. 11), involving one or more speakers and one or more interlocutors, or none (speaking to self).
- Background conversations, as spoken by extras, where L3 is distinguishable, even if the actual substance of the utterances is hard or impossible to discern.
- Foreign-language quotes or quote-like utterances, proverbs, sayings, etc. (e.g., to show erudition).
- Metalinguistic references or topics. Different kinds of bilingual wordplay will belong to this type, for example the bilingual (English-German) pun “Nice to *miet* you” which was used to in reference to car-rental companies in Majorca (Spain). The combination of the German word *mieten* (rent) and the English phrase “Nice to meet you” evokes the pleasure to rent a car.

As the presence and patterns of L3 vary from film to film, there are a number of factors that can also differ across audiovisual texts, as listed below. For a full appreciation of quality in AVT, it is necessary to see how these variables affect and function in the smaller units of AV fiction, i.e., conversations. We use the term “conversation” as used in the Trafilm Guide (2018, p. 6)² and as described by Zabalbeascoa and Corrius (2019): “the communicative context of an instance of L3. A Conversation may contain more than one L3-instance. All L3-instances belonging to the same [ST] Conversation (and their respective TT versions) share the same start time and duration”.

² Trafilm Project – “La traducción de filmes multilingües en España” (FFI2014-55952-P), (2015–2018) funded by the Spanish Ministry of Economy and Competitiveness. <http://trafilm.net/>

The types of alternations and their variables can be listed as:

- Number, distribution and duration of L3 appearances throughout a film.
- Number of L3 instances within a given scene or conversation.
- Position of L3 in a scene: start, middle, or end. This might be evidence that L3 is used solely for salutations, for example, but not necessarily, salutations can happen at any point in a scene.
- The presence of just one language as L3 in a given film (vs more; e.g., German, French, Italian, in *Inglourious Basterds* (2009)).
- The total number of characters involved (see 3.1, below).
- The number of switches (see 3.2, below).
- The number of L3 languages (see 3.3, below).
- The number of conversations that include L3 instance(s).

3.1. Characters

The number of characters involved has different combinations:

A. 1 character “self-switch”

When the same and only character changes the language because of self-translation, rephrasing, clarifying, diglossia, etc. For example, in *Butch Cassidy and the Sundance Kid* (1969), the three main American characters are trying to hold-up trains and banks in Bolivia, in one of the robberies, we can hear “Raise your hands! ¡Manos arriba!”, as Butch is translating his own words.

B. 2+ character “self-switch”

When two or more characters switch from one language to another because of self-translation, rephrasing, clarifying, diglossia, etc. Examples of this can be found in TV series like *Modern Family* (2009), with a lot of interaction between characters that are native speakers of either English or Spanish and with some knowledge of the other language, especially when one of such characters has to act as interpreter and gets help from others present, for example when Manny (Gloria’s first child) “also translates/interprets her mispronounced words” (Dore, 2019, p. 64), or when there is a conversation between two CS characters, and each one is strong in a different language.

C. 2+ languages

When every character speaks their own (different) language, in other words, they alternate monolingualism. For instance, *Um filme falado* (2003) narrates the story of a teacher who goes on a cruise with her daughter. Passengers are from different nationalities and everyone speaks their own language. During the Captain’s dinner, the captain (who speaks English) sits at a table

with three passengers: Delphine (French), Francesca (Italian), and Helena (Greek); each character contributes to the conversation in their own language with no diegetic interpreter or translation.

D. 2+ characters

When there is a total mix; i.e., both of the above (B and C) for 2+ combined. This is a complex situation and may appear in scenes with several people talking at once, and at least two conversations going on simultaneously.

E. 2+ languages + interpreter

Two or more characters speak their own language with the mediation of an interpreter, as in the interpreting scene in *Lost in Translation* (2003), or Example 1.

(1) *Spanglish* (2004) 2+ L-exchange + translator

Deborah: I rented a house here for the summer and now she needs to sleep in the house because of the bus schedule. (L1)

Man: *Ella dice que rentó una casa durante el verano y quiere que te quedes con ella.* (L3)

Flor: No, sorry. (L1)

Flor: *Porque tengo una hija y no puedo.* (L3)

Man: She can't because of her daughter. (L1)

In Example 1, a man in the street who can speak English and Spanish, probably a Chicano, acts as an interpreter for Flor and Deborah, who cannot understand each other's language.

3.2. Switches

The number of switches within a conversation may vary:

A. Single alternation

Often just one or two words in intra- and intersentential CS, but also turn-specific (Example 2), at the start or end of a conversation.

(2) *A Passage to India* (1984)

Waiter: *Memsahib* (L3 Hindi)

Mrs. Moore: Oh, thank you! (L1 English)

Waiter: Tea coming. (L1)

B. Sandwich alternation

A turn in a different language appears between two separate turns.

(3) Carla's Song (1996)

Conductor: Tickets, please. The tickets. (L1 English)

Carla: *No tengo.* (L3 Spanish)

Conductor: Could you show me the ticket, please? (L1)

C. Multiple alternation with single L3

When there is a constant back and forth of languages and turns between two or more characters but the third language does not change.

(4) *Land and Freedom* (1995)

Conductor: *Bienvenido a España.* (L3 Spanish)

Passenger: Welcome in Spain. (L1 English)

David: Thanks very much. (L1)

Conductor: *El sindicato ha tomado la línea y nosotros cobramos el billete de los que vienen a luchar contra los fascistas.* (L3)

Passenger: Trade Union is controlling all the line. (L1)

David: Thanks. (L1)

Conductor: *Gracias. Bienvenido.* (L3)

David: *Gracias.* (L3)

D. Multiple alternation with multiple L3s

There is an alternation of different languages between characters (more than two turns).

(5) *Raid on Rommel* (1999)

Signorina: *Porco cretino. E' l'ultima volta che mi fa questo gioco. Il generale sentirà per questo.* (L3a Italian)

Captain: *Italienische Huren beeindruckt mich nicht!* (L3b German)

Signorina: Why don't you speak English? (L1 English)

Captain: What seems to be the problem, *signorina*? (L1+ L3a)

Signorina: Why I'm not leaving on that plane to Tobruk? Why? (L1)

3.3. Third languages

The number of third languages present in an audiovisual text may also be different.

A. ONE third language (L3)

There are films that contain only one type of third language, this is very common and there are plenty of examples, for instance in *An American in Paris* (1951) or *Moulin Rouge* (2001) L3 is French, in *Spanglish* (2004) L3 is Spanish. Category A, (and B and C, below), may be complicated by various degrees of non-nativeness and L3 proficiency (e.g., broken English and broken Spanish in *Bread and Roses* (2000), or even pseudolanguages, farcical representations of real languages).

B. TWO different third languages. (L3a + L3b)

There are also many films which have two third languages, for example in *Raid on Rommel* L3a is Italian and L3b is German; in *Land and Freedom* (1995) L3a is Spanish and L3b is Catalan or in *Carla's Song* (1996) where L3a is Spanish and L3b French. Still, considering that accents can be used to mark otherness and thus they can be considered as L3, we have examples such as *Beauty and the Beast* (1991) where L3a is French and L3b is English with a French accent (Corrius, 2008).

C. MORE than two different third languages (L3a + L3b + L3c, etc.)

There are audiovisual texts that for some reason or other (signal otherness, create humour, represent different identities, etc.) use three or more languages; that would be the case of *Lost in Translation* (2003) with L1-English + Japanese, French, German and non-native English with a Japanese accent; or *Um filme falado* (2003) with L1-Portuguese + four different L3s: English, French, Italian and Greek.

4. Code-Switching in *Spanglish*

4.1. The Spanish in *Spanglish*, a Case of L3?

Spanglish (2004), a sample of L3-as-theme, stands out from most multilingual films that just sprinkle L3 here and there in token samples; CS and language mixing can be encountered throughout the film. And, unlike other L3-as-theme films (e.g., *One, Two, Three*, 1961, or *Bread and Roses*, 2000) it can be regarded as a precursor of TV shows in the 21st century, like *Modern Family* (2009), *Jane the Virgin* (2014), or *Narcos* (2015), all of which use constant alternations between L1-English and L3-Spanish, within the L3-as-theme category. The language alternations begin in *Spanglish* when an unmarried Mexican mother, Flor, decides to cross the border and migrate to the USA with her daughter Cristina. Once she arrives in Los Angeles, she finds a job as a housekeeper for the Clasky family. Because Flor does not speak English and the Claskys do not

speak Spanish, communicative problems arise on both sides. *Spanglish* characters can be grouped into three categories: the main character, Flor, who speaks only Spanish until almost the end of the film; the Claskys, who speak English only; the characters who are proficient in English and Spanish, Cristina, and her cousin Monica.

When the monolingual characters want to communicate with each other, an interpreter is needed to avoid miscommunication. At first this role is played by Monica, and later on, in different instances, by Cristina. For Delabastita (2002) and Díaz-Cintas (2015), multilingualism is sometimes used to create confusion or misunderstanding in order to construct humour. Authors like Chiaro (2007) and Dore (forthcoming) focus almost entirely on the third language as an element of humour. In AVT, quality may suffer if the L3 message is translated or otherwise made clear to the audience when there is a deliberate attempt in the script to keep the audience in the dark as to what is being said, for humour or for suspense or dramatic effect.

Spanglish might be said to be a metaphor of both the cultural clash and the (mis)communication issues between the USA and Mexico. Flor steps across the cultural divide, going from the Mexican to the US culture. Indeed, the film is about the journey many Mexicans embark on to the USA and presents the differences and similarities between these two cultures. It shows areas where the two cultures can find it very hard to understand and “feel for” each other. The audience can easily appreciate that Flor and her daughter are moving from their country to a foreign land, crossing a physical border, and a cultural one, with two distinct settings (landscape, furnishing, wardrobe and food) can be seen.

The title itself, *Spanglish*, refers to this constant mix of English and Spanish; it provides a label for the code-switching partaken in the film. However, in this particular case, “Spanglish” refers more to the mix of languages (typical of the way Mexican-Americans often speak) than to the CS used by each of the characters in the film. Spanglish might even be considered a language variety rather than a combination of two languages, a fusion creating a “new” language as spoken by its US–Spanish community. It includes the anglicising of Spanish words and syntax, translating phrasal verbs or using English words with a Spanish appearance (Mar-Molinero, 1997). Spanish is regarded as a heritage language in the USA, with its stereotypes and stigmas, and Spanglish may have a positive or non-evaluative meaning, referring to a sociolinguistic phenomenon that defines a certain speech community and social group which has its own culture. On the other hand, for Treffers-Daller (2012) and Boztepe (2003), CS can be seen as instantiations of interference. Sometimes Spanglish is seen as having a negative connotation, as a broken or “contaminated” English (by Spanish) or Spanish (by English), either as part of a learning curve for speakers, who still need to improve their proficiency, or as a historical or dialectal corruption of one of these languages, through excessive borrowing or interference from the other.

Code-switching in the film is more clearly portrayed between the two groups of people defined by their birthplaces and their native languages than within any individual group, as CS also refers to.

Following Monti's (2014) classification, we might say that, on the whole, *Spanglish* displays more *turn-specific code-switching* involving two or more characters in conversation. Towards the end of the film, though, Flor begins to alternate from Spanish into English when she converses with the Claskys. A few instances of *intersentential CS* can be encountered: for example, when Cristina is upset and says to Flor, "No, it'll never be all right. You're wrong. This is exactly what I was worried about. I will never be able to forgive you! *¡Nunca te perdonaré, nunca!*" or when Flor replies "Not a space between us. *¡Que sea la última vez, Cristina!*" Some plot development-related instances of *intrasentential CS* can be found, too. For instance, at the very beginning of the film we can see a note on the fridge, which reads, "*Cristina, pon queso en la tortilla y ponlo dentro del microwave por un minuto. Te quiero. Mami*"; or when Clasky's mother is being introduced to Flor and she says, "What's your name. *Llamo?* It's one of my five Spanish words". Mr Clasky also uses this type of code-switching a couple of times, "is *simpático* the word?", and a bit later, "Oh, boy, *engreído* is gonna be rough".

Yet, if we had to describe a character in the film as a Spanglish speaker, we would choose Flor as she is the one who most displays its traits in her speech. In the last 15 minutes of the film she speaks in clearly non-standard, non-native English. Flor is of a much lower social class than the Claskys, which is underscored by her use of Spanish. Here, the AV fiction is a reflection of certain sociocultural realities. According to Luna and Peracchio (2005, p. 761), in the USA, Spanish tends to be associated with a lower socioeconomic status and can activate feelings of inferiority, while "Hispanics tend to perceive English as the language of integration". English, of course, is the *de facto* official language of the USA and it "denotes the language spoken by the group that holds the political, cultural and economic power within a country" and it has more positive connotations (Luna and Peracchio, 2005, p. 761).

When we come across a film like *Spanglish* (2004), with such a large amount of L3 and CS (by number of instances and screen time), one realises that the distinction between "main" AV language (L1/L2) and "others" (L3) is not always necessarily clear-cut. A film like *Spanglish* (or TV shows like *Narcos*) might better be described as having two main languages, L1-English plus L1-Spanish in the ST. A possible criterion for establishing one language or another as the "main" language is how it matches up with the assumed language of the intended audience (for ST or TT). What matters is that without this language mix in L3-as-theme type films, like *Spanglish*, the plot would lose an essential component. Later in the 21st century, the trend becomes more established, with TV shows like *Modern Family* (2009), possibly with an increasingly diverse intended audience, including not only monolingual English-speakers but also native speakers of English with some knowledge of Spanish, and maybe even vice versa, in a similar case to the TV show *Narcos* (2015).

4.2. Rendering Code-Switching: When L3 Can Mix with L1

Cultural references, as well as code-switching, interference or other language phenomena integrated in a film script might be serious obstacles (i.e., strong restrictions according to Zabalbeascoa's (1999) *Model of priorities and restrictions for translation*) for achieving (the intended goals or priorities) of a successful translation. According to this model, translation quality is achieved when the priorities for a translation are clear and agreed upon by the stake-holders and intended users, and reflected in the TT. Restrictions, or constraints, are factors that make the job more difficult, or sometimes impossible, depending on their number and strength. In our analysis, the nature of the priorities and restrictions for the version of *Spanglish* to be distributed in Spain, explains why it is not dubbed (except for the unseen narrator utterances), contrary to common practice in Spain. In the version for Spain, L1 on-screen utterances (perceived as L3^{TT}-English for the TT viewers) are translated with L2-Spanish subtitles when the characters are on screen. Coincidentally, this is also a feature of the ST, in which L1-English subtitles are used to help the English-speaking ST audience understand the utterances in L3ST-Spanish. The few off-screen (unseen) utterances of Cristina-as-narrator are the only parts translated from L1-English into L2-Spanish and then dubbed (Example 6), also as a precursor to what happens in *Narcos* (2015) with its narrator.

(6) *Spanglish* (2004)

ST unseen narrator's voice

Narrator (Cristina): Holding out had helped though. She was no longer intimidated. Working for Anglos now posed no problems. It would just be a job. White America beckoned. She stepped across the cultural divide.

TT dubbed version

Narrator (Cristina): Su paciencia le había ayudado, ya no sentía cohibida, trabajar para los gringos ya no sería un problema, sería sólo un trabajo. La América blanca la reclamaba, y ella cruzó la frontera cultural.

Table 1 summarises the transfer of the ST languages (L1 and L3ST) to the TT for Spain. Basically, English (the main language of the ST and its intended audience) can also be heard in the TT, but in its AVT version for Spain, and because it is not dubbed, it becomes a foreign language (L3^{TT}) for the intended TT audience made up of native speakers of Spanish with no assumed knowledge of English beyond a basic level. For Cristina's unseen, off-screen narration her voice is dubbed from ST-L1-English into TT-L2-Spanish. Cristina's narration is used "to complement the information provided visually as well as some of the exchanges between the Mexican characters" (Sanz, 2011, p. 24). Interestingly, L3ST-Spanish is also retained (as L2) in the TT. "Spanish is maintained untranslated at the spoken level but conveyed through open subtitles in the original version of the film" (Monti, 2014, p. 155). The roles of the two languages have been swapped; in other words, L1, which is the language of the ST audience and thus the main language of the ST, becomes a foreign language for the TT

audience (L3^{TT}); and L3ST, which is a “foreign” language for the ST audience, becomes the main language of the AVT audience (L2). All of this is achieved largely because there is no distinction between two otherwise distinct varieties of Spanish, Iberian Spanish and Mexican Spanish (not to mention Chicano), which could well be treated as different languages in other films or AVT cases. For the version screened in Spain: English is L1 and L3^{TT}; Spanish is L3ST and L2.

Table 1.

Spanglish – Language Correspondences in AVT for Spain

ST audio	AVT audio
<i>L1 English</i>	<i>L2 Spanish</i>
<ul style="list-style-type: none"> • The Claskys and most secondary characters. • Cristina’s off-screen narration. 	<ul style="list-style-type: none"> • Mainly Flor and Cristina. • Cristina’s dubbed voice, as narrator.
<i>L3 Spanish</i>	<i>L3 English</i>
<ul style="list-style-type: none"> • Mostly Flor and Cristina (with English subtitles). 	<ul style="list-style-type: none"> • The Clasky’s • Most secondary characters (with Spanish subtitles).

A further feature needs to be pointed out in this discussion: the use of subtitles for L3 is a priority for the AVT because it was a priority for the ST, as an instance of intended equivalence, a concept borrowed from Zabalbeascoa (1994) and a part of his *P-R model*. Equivalence can be sought at multiple levels, not necessarily or exclusively lexicosemantic similarity (which is what people often understand equivalence to mean). In this case, a potentially important level, in the pursuit of quality would be to strive for an equivalent “combination of L3 + translated subtitles”, even if the roles of English and Spanish are reversed in ST and its AVT. The translation mode in *Spanglish* of combining subtitles with dubbing might be simply the result of an attempt to represent multilingualism. For de Higes-Andino (2014), multilingualism is portrayed or “marked” when dialogues are left untranslated, or, we might add, undubbed. Multilingualism, as manifested in code-switching, plays such an important part in *Spanglish* that we can assume that the obvious strategy for producing a version for a Spanish-speaking audience is to keep both English and Spanish audio in the AVT precisely because language variation is part of the plot (L3-as-theme). Each character is portrayed partly by their choice of language and occasion for using it. This is not always the case, though, not even in the English-Spanish language combination. In TV shows like *Modern Family* (2009) and *Dexter* (2006), the importance and degree of language variation is considerably reduced in translation, and this goes against any effort of increased quality, rather the contrary; it affects quality negatively, for the simple reason that some scenes rely heavily on the audience

being able to perceive multilingual dynamics. In *Bread and Roses* (2000) (quite analogous to *Spanglish* in its use of L1>L3 alternations, though more dramatic and political in nature), important subtle distinctions of how (well) each character could speak English and Spanish were practically wiped out in the dubbed version for Spain, with remarkable (negative) effects in the TT on the quality of character portrayal, the meaning of certain scenes, and the audience's ability to interpret exactly what was going on. Indeed, *Spanglish* seems oddly unique in this respect, and we can only speculate why. One reason might be that the lead actress, Paz Vega, is Spanish and very famous in Spain.

When AVT translators come across code-switching, they have to decide how to render L3 in the TT. Quality is not necessarily kept by simply leaving foreign languages untouched, especially when that negatively affects the portrayal of multilingualism, or CS, or scenes of miscommunication. As described by Corrius and Zabalbeascoa (2011, p. 120–121) there are five different types of solutions in translation. Quality may be searched for by means of creatively exploring potential solutions amongst a wide variety of different types and combinations provided first by broad exploratory descriptive studies that set out to collect samples of different ways of tackling a variety of problems posed by multilingual combinations and the reasons they are used for. We adapt them here, for a tripartite classification of strategies for rendering L3ST in AVT.

1. *Retaining one or more of the ST language(s):*

- a. If L3ST does not coincide with L2 (L3ST ≠ L2), L3 utterances are often left untranslated in the audio, and multilingualism is still noticeable in the translated version: ST [L1+L3] --> TT [L2+L3]. This is the case of the Spanish dubbed version of *An American in Paris* (1951) where the L3ST-French used in the ST has been retained in the TT for Spain;
- b. If L3ST happens to coincide with L2, the language of the intended audience for the AVT, and is left untranslated, then multilingualism can no longer be appreciated by the viewers as a feature of the film because L2 and L3^{TT} are the same language, rendering L3 undetectable. This results in poorer quality when certain important features of any film, such as the portrayal of a given character's "otherness", depend entirely on the noticeable presence of a third language and this is not compensated for either. For example, in *Carla's Song* (1996) the Spanish used as L3ST has been retained in the L2-Spanish version, resulting in L3 invisibility.
- c. L3^{TT} is the same language as L1, and L3ST happens to coincide with L2, which is so often the case for Spanish dubbed versions of US films and TV shows that include dialogues in L3ST-Spanish. In *Spanglish*, the effect of otherness is caused (in Spain) by English L1ST, which is perceived as L3 for the Spanish-language audience, not by L3ST-Spanish as it coincides with L2-Spanish.

2. *Avoiding any instances of L3^{TT}:* in this case, L1 and L3ST are equally rendered as L2. Again, this option gives rise to L3-invisibility, entailing that multilingualism (including CS) ceases to be a feature of the film. For instance, in the dubbed version for Spain of *Monsoon Wedding*

(2001) L3-Hindi, as a feature of the ST, is no longer perceivable in the TT, impoverishing the result when not compensated for.

3. *Using new languages in the TT*: even when the number of languages is retained in translation the actual languages may differ. Multilingualism is kept as a feature but the languages of the TT do not entirely coincide with those of the ST, so the effect caused in the TT might be different from the one caused on the ST audience. There are instances of this in *Butch Cassidy and the Sundance Kid* (1969) or in *Fawlty Towers* (1975), where Spanish L3ST is changed to French and Italian in the respective dubbed versions. In the former, there are several scenes where the main characters speak Spanish in the ST but French in the TT; and in the latter there are some humorous scenes, where Manuel (the Spanish character) speaks Spanish in the ST but Italian in the TT. It is interesting to note, too, that in the Spanish version even the character's name has been changed for an Italian one: Paolo.

In brief, the pattern of languages is not random in *Spanglish*, and its AVT for Spain does not erase the feature of multilingualism by dubbing everything into the same language (unlike *Bread and Roses*, 2000); rather, this version shows its quality by a creative strategy of systematic subtitling combined with a highly restrictive use of dubbing (for off-screen narrative) so that the multilingual nature of the film is not negatively affected, certainly not its CS. The marked multilingualism, through non-translation at the spoken level is quite significant in a country where dubbing is the most common modality of audiovisual transfer. In fact, as stated by Díaz-Cintas (2015), in a country where the translation of multilingual films tends towards standardisation with far less linguistic variation, the case of *Spanglish* has proven to be a model of translation practice, both in its quality and as a historical landmark. Retaining CS in the AVT gives “vital clues about the immigrant characters’ socio-linguistic hybrid identities and, as such, it is a very important feature of the filmic text” (Monti, 2014, p. 165).

5. Conclusions

In this brief study we have confirmed ideas coming from all of the references cited above that the use and presence of language combinations and CS is a much more complex issue than one might suspect on first approaching the topic. Studies like this are necessary to show the variety of problems posed, the number of features displayed and the range of possible solutions a translator might resort to, including the wealth of data already gathered, and that increased awareness of all this can help to improve the quality of AVT. The point we have insisted on here is that beyond any characterisation of L3 at the level of instances and types it is also essential to contextualise them as much as possible for an improved understanding of their functions and how they are rendered in translated versions, with a view to providing insight and guidelines for translators having to tackle this issue. We have provided a broader context for the notion of code-switching as borrowed from

sociolinguistics, and adapted it here for the purpose of analysing film scripts, and exploring ways of improving their translated versions. We have presented examples of CS while comparing it to related phenomena, such as intratextual translation and isolated uses of L3, including them all under the concept of alternations (or switches). A proposal is included for labelling multilingual films according to the presence and importance of L3: *anecdotal*, *recurrent* and *L3-as-theme*. We have also found the need to relate patterns of multilingualism and code-switching to functions of the plot or script, such as portraying cultural identity, be it stereotypical or otherwise. Another possible function of CS is to provide clues to show any changes in a cultural backdrop or setting.

Of course, much research still has to be carried out to show the full range of code-switching instances and patterns, their role in each film and how they are rendered in actual translated versions or what potential innovative solutions can be proposed for the future, making the viewing experiences of foreign audiences closer to that of the source text viewers.

References

- Alvarez Cáccamo, C. (1990). Rethinking conversational code-switching: Codes, speech varieties, and contextualization. In K. Hall, et al. (Eds.), *Proceedings of the sixteenth annual meeting of the berkeley linguistics society* (pp. 3–16). Berkeley: Berkeley Linguistics Society.
- Alvarez Cáccamo, C. (1998). From ‘switching code’ to ‘code-switching’. Towards a reconceptualisation of communicative codes. In P. Auer (Ed.), *Code-switching in conversation. Language, interaction and identity* (pp. 29–50). London and New York: Routledge.
- Androutopoulos, J. (2007). Bilingualism in the mass media and on the Internet. In M. Heller (Ed.), *Bilingualism: A social approach* (pp. 207–232). London: Palgrave.
- Auer, P. (1984). *Bilingual conversation*. Amsterdam: John Benjamins.
- Auer, P. (1998). Bilingual conversation revisited. In P. Auer (Ed.), *Code-switching in conversation. Language, interaction and identity* (pp. 1–24). London and New York: Routledge.
- Bleichenbacher, L. (2008). *Multilingualism in the movies*. Tübingen: Francke Verlag.
- Boztepe, E. (2003). Issues in code-switching: Competing theories and models. *Working Papers in TESOL and Applied Linguistics*, 3(2), 1–27.
- Brancato, C. et al. (Creators). (2015-2017). *Narcos* [Television series]. USA: Gaumont International.
- Brooks, J. L. (Director). (2004). *Spanglish* [Motion picture]. USA: Gracie Films.
- Bullock, B. E., & Toribio, A. J. (2009). *The Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press.
- Chaume, F. (2001). Más allá de la lingüística textual: cohesión y coherencia en los textos audiovisuales y sus implicaciones en traducción [Beyond Text Linguistics: cohesion and coherence in audiovisual texts]. In M. Duro (Ed.), *La traducción para el doblaje y la subtitulación* [Translation for Dubbing and Subtitling] (pp. 65–82). Madrid: Cátedra.
- Chiaro, D. (2007). Lost, found or retrieved in translation? Cross-language humour in multilingual films. In M. G. Scelfo & S. Patroni (Eds.), *Lingua, cultura ed ideologia nella traduzione di prodotti multimediali: Cinema, televisione, web* [Language, culture and ideology in the translation of multimedia products: cinema, television, web] (pp. 123–138). Roma: Aracne.

- Clesse, J. (Scriptwriter). (1975–1979). *Fawlty Towers* [Television series]. UK: BBC.
- Clyne, M. G. (2003). *Dynamics of language contact. English and immigrant languages*. Cambridge: Cambridge University Press.
- Coppola, S. (Director). (2003). *Lost in Translation* [Motion picture]. Japan and USA: Tohokushinsha Film.
- Corrius, M. (2008). *Translating multilingual audiovisual texts. Priorities and restrictions. Implications and applications* (Unpublished doctoral dissertation, Universitat Autònoma de Barcelona, Bellaterra).
- Corrius M., & Zabalbeascoa, P. (2011). Language variation in source texts and their translations: the case of L3 in film translation. *Target*, 23(1), 113–130.
- de Higes-Andino, I. (2014). The translation of multilingual films: Modes, strategies, constraints and manipulation in the Spanish translations of *It's a Free World*. *Linguistica Antverpiensia. New Series – Themes in Translation Studies*, 13, 211–231.
- Delabastita, D. (2002). A great feast of languages. Shakespeare's multilingual comedy in 'King Henry V' and the translator. *The Translator: Studies in Intercultural Communication*, 8(2), 303–340.
- Díaz-Cintas, J. (2015). Multilinguismo, traducción audiovisual y estereotipos: el caso de *Vicky Cristina Barcelona*. *Prosopopeya*, 9, 135–161.
- Diebold, A.R. (1961). Incipient bilingualism. *Language*, 37, 97–112.
- Dore, M. (2019). Multilingual humour in audiovisual translation. *Modern Family* dubbed in Italian. *European Journal of Humour Research*, 7(1), 52–70.
- Dore, M. (forthcoming). Multilingual humour in audiovisual translation. In M. Corrius, et al. (Eds.), *Translating audiovisuals in a kaleidoscope of languages*. Bern: Peter Lang.
- Gardner-Chloros, P. (2009). *Code-switching*. Cambridge: Cambridge University Press.
- Hathaway, H. (Director). (1999). *Raid on Rommel* [Motion picture]. USA: Universal Studios.
- Hatim, B., & Mason, I. (1994). *Discourse and the translator*. New York: Longman.
- Haugen (1956). *Bilingualism in the Americas: A Bibliography and Research Guide*. Alabama: University of Alabama Press / American Dialect Society.
- Heiss, C. (2004). Dubbing multilingual films: A new challenge? *Meta*, 49(1), 208–220.
- Jackson, P. (2001). *The Lord of the Rings* [Motion picture]. New Zealand: WingNut Films.
- Jakobson (1961).
- Lean, D. (Director). (1984). *A Passage to India* [Motion picture]. USA: Metro-Goldwyn-Mayer Studios.
- Lippi-Green, R. (1997). *English with an accent*. London and New York: Routledge.
- Lloyd, C., & Levitan, S. (Creators). (2009–2019). *Modern Family* [Television series]. USA: Lloyd-Levitan Productions.
- Loach, K. (Director). (1996). *Carla's Song* [Motion picture]. Spain: Alta Films.
- Loach, K. (Director). (2000). *Bread and Roses* [Motion picture]. Spain: Alta Films.
- Loach, K. (Director). (1995). *Land and Freedom* [Motion picture]. Spain: Messidor Films.
- Luhrmann, B. (Director). (2001). *Moulin Rouge* [Motion picture]. UK: Angel Studios.
- Lucas, G. (Director). (1977–1983). *Star Wars Trilogy*. USA: Lucasfilm.
- Luna, D., & Peracchio, L. A. (2005). Advertising to bilingual consumers: The impact of code-switching on persuasion. *Journal of Consumer Research*, 31, 760–765.
- Manos, J. (Creator). (2006). *Dexter* [Television series]. USA: Showtime.
- Mar-Molinero, C. (1997). *The Spanish speaking world*. London and New York: Routledge.
- Minelli, V. (Director). (1951). *An American in Paris*. USA: Metro-Goldwyn Mayer Studios.

- Monti, S. (2014). Code-switching in British and American films and their Italian dubbed version. *Linguistica Antverpiensia. New Series – Themes in Translation Studies*, 13, 135–168.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Nair, M. (Director). (2001) *Monsoon Wedding* [Motion picture]. USA: IFC Productions.
- Oliveira de, M. (Director). (2003). *Um filme falado* [Motion picture]. Portugal: Madragoa Films.
- Paradis, M. (1993). Linguistic, psycholinguistic and neurolinguistic aspects of — interference|| in bilingual speakers: The activation threshold hypothesis. *International Journal of Psycholinguistics*, 9, 133–145.
- Pfaff, C. (1976). Functional and syntactic constraints on syntactic variation in code-mixing. In B. Steever, et al (eds.), *Papers from the parasession on dyachronic syntax* (pp. 248–259). Chicago: Chicago Linguistic Society.
- Poplack, S., & Meechan, M. (1995). Patterns of language mixture: nominal structure in Wolof-French and Fongbe-French bilingual discourse. In L. Milroy, & P. Muysken (Eds.), *One speaker, two languages. Cross-disciplinary perspectives on code-switching* (199–232). Cambridge: Cambridge University Press.
- Roy Hill, G. (Director). (1969). *Butch Cassidy and the Sundance Kid* [Motion picture]. USA: 20th Century Fox.
- Sanz Ortega, E. (2011). Subtitling and the relevance of non-verbal information in Polyglot Films. *New Voices in Translation Studies*, 7, 19–34.
- Scott, R. (Director) (2000). *Gladiator* [Motion picture]. USA: Scott Free Productions.
- Tarantino, Q. (Director). (2009) *Inglourious basterds* [Motion picture]. USA: A Band Apart Productions.
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. Berkeley, Los Angeles, London: University of California Press.
- Trafilm project. (2015–2018). La traducción de filmes multilingües en España (FFI2014-55952-P) [Research Project]. Funded by the Spanish Ministry of Economy and Competitiveness. <http://trafilm.net/>
- Treffers-Daller, J., & Mougeon, R. (Eds.). (2005). The role of transfer in language variation and change: Evidence from contact varieties of French. *Bilingualism, Language and Cognition*, 8(2), 93–98.
- Treffers-Daller, J. (2012). Code-switching and transfer: an exploration of similarities and differences. In B. Bullock, & A. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 58–74). Cambridge: Cambridge University Press.
- Trousdale, F., & Wise, K. (Directors). (1991). *Beauty and the Beast* [Motion picture]. USA: Walt Disney Pictures.
- Urman, J. S. (Executive producer). (2014). *Jane the Virgin*. [Television series]. USA: Poppy Productions.
- Weinreich, U. (1953). *Languages in Contact*. The Hague: Mouton.
- Wilder, B. (1961). *One, two, three*. USA: The Mirisch Company.
- Zabalbeascoa, P. (1994). Awareness in translation. *Language Awareness*, 3-4(III), 161–174.
- Zabalbeascoa, P. (1999). Priorities and restrictions in translation. In J. Vandaele (Ed.), *Translation and the (re)location of meaning* (pp. 159–167). Leuven: CETRA.
- Zabalbeascoa, P., & Corrius, M. (2014). How Spanish in an American film is rendered in translation: dubbing *Butch Cassidy and the Sundance Kid* in Spain. *Perspectives Studies in Translation*, 22(2), 255–270.

Zabalbeascoa, P., & Corrius, M. (2019). Conversation as a unit of film analysis. Databases of L3 translation and audiovisual samples of multilingualism. In M. Pérez, & I. de Higes-Andino, (Eds.), *Monti*, 4(4), 57–85.

Zabalbeascoa, P., & Sokoli, S. (2018). *The Trafilm Guide: how the trafilm database works and how to feed in more metadata into the system, with basic trafilm concepts. Revised draft 21.*

<http://hdl.handle.net/10230/28223>

Museums as Disseminators of Niche Knowledge: Universality in Accessibility for All

 **Alessandra Rizzo** 

Università degli Studi di Palermo

Abstract

Accessibility has faced several challenges within audiovisual translation Studies and gained great opportunities for its establishment as a methodologically and theoretically well-founded discipline. Initially conceived as a set of services and practices that provides access to audiovisual media content for persons with sensory impairment, today accessibility can be viewed as a concept involving more and more universality thanks to its contribution to the dissemination of audiovisual products on the topic of marginalisation. Against this theoretical backdrop, accessibility is scrutinised from the perspective of aesthetics of migration and minorities within the field of the visual arts in museum settings. These aesthetic narrative forms act as modalities that encourage the diffusion of 'niche' knowledge, where processes of translation and interpretation provide access to all knowledge as counter discourse. Within this framework, the ways in which language is used can be considered the beginning of a type of 'local grammar' in English as lingua franca for interlingual translation and subtitling, both of which ensure access to knowledge for all citizens as a human rights principle and regardless of cultural and social differences. Accessibility is thus gaining momentum as an agent for the democratisation and transparency of information against media discourse distortions and oversimplifications.

Key words: media accessibility, counter information, access to niche knowledge, museum text types, the visual arts, accessibility as activism, universality in accessibility, interlingual museum translation, museum multilingual multimedia guide, migration and minorities.

Citation: Rizzo, A. (2019). Museums as disseminators of niche knowledge: Universality in accessibility for all. *Journal of Audiovisual Translation*, 2(2), 92–136.

Editor(s): G.M. Greco & A. Jankowska

Received: August 31, 2019

Accepted: December 16, 2019

Published: December 31, 2019

Copyright: ©2019 Rizzo. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. Introduction

As claimed in Pablo Romero-Fresco's study (2013), the last decades have seen an exponential growth in audiovisual translation (AVT) and accessibility services in Europe and all over the world. European countries, such as Spain, France, the UK and Italy, have undertaken important projects and initiatives (i.e., Spain, AENOR 2012; France, MFP 2012; UK, Ofcom 2016; Italy, ADLAB 2011–2014) that have contributed to increasing the interest in the inclusion of sensory impaired people within national and international cultural activities (Reviere, 2016). The creation of new legislation and accessibility guidelines has favoured the improvement of the quantity and quality of subtitling for the deaf and hard of hearing (SDH), and audio description (AD) for blind and partially sighted persons within media and artistic contexts, and also within creative frameworks, as far as the translation of foreign audio description scripts into source language (SL) scripts is concerned (Jankowska, 2015). Although media accessibility (MA) practices are still fragmented in numerous European countries (Reviere, 2016), there is a massive increase in attention to the rising awareness regarding accessibility as a result of collaborative exchanges between academia and the industry. Today, AD and SDH are considered as the most popular and common accessibility modes for the inclusion of sensory impaired persons. Thus, if, on the one hand, they remain the standard modes to access the media and arts (e.g., TV, cinema, museums, theatres, festivals, etc.) within AVT settings, on the other hand, a variant to AD and SDH as modes of accessibility relates to the concept of universal design that designates the degree to which public buildings and spaces can be accessed by people with disabilities (Díaz-Cintas, Orero, & Remael, 2007). Information and communication technology areas can also entail accessibility as the possibility for everyone to use and have access to technology, even though most attention is still predominantly given to users with disabilities (Díaz-Cintas, Orero, & Remael, 2007). Research into AD and SDH has thus confirmed their role as accessibility modes recognised within AVT, as well as their involvement with practices of translation as types of constrained translation (Mayoral, Dorothy, & Natividad, 1988) based on intersemiotic or intermodal translation transfers (Braun, 2008). The new millennium has welcomed accessibility as an AVT issue, and AVT modes and strategies as accessibility devices (Orero, 2005; Díaz-Cintas, Orero, & Remael, 2007; Díaz-Cintas, Matamala, & Neves, 2010; Remael, Orero, & Carroll, 2012; Bruti & Di Giovanni, 2012; Perego, 2012a, 2012b).

Accessibility entails diverse domains, all interested in providing services to citizens, especially sensory impaired persons. Nevertheless, in recent times, MA has embraced universal views that have extended the concept of access to all citizens and to social and political issues, while giving a crucial role to translation in its task of increasing and reinforcing universality for the most neglected social issues. Against this backdrop, translation, in its broad sense, which involves “the mediation of diffuse symbols, experiences, narratives and linguistic signs of varying lengths across modalities (words into image, lived experience into words), levels and varieties of language” (Baker, 2016, p. 7), is examined as a characterising concept in accessibility mechanisms. In other words, translation, by definition, goes hand in hand with accessibility. That accessibility is “a form of translation

and translation is a form of accessibility, uniting all population groups and ensuring that cultural events, in the broadest sense of the word, can be enjoyed by all” (Díaz-Cintas, Orero, & Remael, 2007, pp. 13–14), is a commonly accepted concept. Based on this fundamental belief, this study goes beyond the view of accessibility as limited to AD and SDH practices, and attempts to establish a conceptual link between accessibility and universality in relation to society and collectiveness. If accessibility is a form of translation, and translation is a type of accessibility device, then, it is the very significance of universality, which is rooted in the nature of translation, that makes it possible to extend the horizons of accessibility to the sphere of human rights (Greco, 2018). As a human rights concept, accessibility is open to any “social” user and to a variety of minorities – where the idea of minorities has implications within the spheres of sensory impairment, physical disability, social class, age, race, and language. In this regard, accessibility can be built, on the one hand, as a social potential that favours and stimulates knowledge dissemination, while assembling all citizens of the world (e.g., museums as spaces of social and multicultural encounters), and, on the other hand, as a universal concept in relation to its analogies with translation and interpreting processes as mechanisms of universal communication.

Going beyond the framework of the social model of disability, which recognises the consolidated significance of accessibility as the depositary of the needs for sensory impaired persons, motor, cognitive, and psychological disabilities, this study looks at accessibility as a service that gives access to niche knowledge, that is, a type of socially neglected content that can be and has to be “expanded to all potential audiences” (Di Giovanni, 2018) across diverse channels and forms (e.g., museums, the visual arts). By looking at niche knowledge as the whole of discursive forms on ontological truths which depict marginalised identities (Rizzo & Seago, 2018), I claim that accessibility involves universality as a concept that nourishes the dissemination of all types of knowledge for all citizens. This idea draws on Catalina Jiménez-Hurtado, Claudia Seibel and Silvia Soler Gallego’s studies on universal accessibility as a “powerful tool for facilitating access to knowledge” (2012, p. 1). In particular the diffusion of niche knowledge has acquired a significant role within the context of the arts and, in particular, through modes of translation as tools for museum accessibility, that is, “the ability to make any text using different codes or modes linguistically, cognitively and socially accessible to as many recipients as possible”, a principle which is, exactly, “inherent in the nature of translation” (2012, p. 2). Since the core of this investigation is the context of museums, access services to public and digital museum spaces concern the degree to which each individual/participant/user can, on the one hand, access physical environments (e.g., material or virtual) and, on the other hand, access the museum contents (to which attention is paid in this work), regardless of subjective abilities and personal interests.

Within the framework of Descriptive Translation Studies (DTS), which “correspond to a descriptive, empirical, interdisciplinary, target-oriented approach to the study of translation” (Rosa 2010/2016d, p. 1), this scrutiny is both a product- and target-oriented survey of Italian into English interlingual translations of catalogue texts and museum panels, and interlingual subtitles for digital museum videos. It is a product-oriented study, since its focus is on the comparative description of source texts

and their translations, and it is target-oriented, since it relies on the view of translations as “facts of target cultures” (Toury, 1995, p. 29). Target cultures are posited in a central dimension and are embedded within the reframing of accessibility entering the circuits of universal communication. On a theoretical level, this research aims at supporting the existing thesis that “accessibility” is not limited to the realm of individuals with sensory, motor, and cognitive impairments, but should also be used to address the access of minority social groups to knowledge and the society. At the same time, access is given to niche knowledge that contributes to the depiction of minority groups (e.g., migrants) who can eventually have the chance to acquire visibility.

By means of a corpus-based analysis, the main research question leading the study is descriptive and focuses on the features of interlingual museum translation and subtitling in the context of marginalisation, migration, and minorities. The second research question investigates whether the techniques observed in interlingual museum translation and subtitling in the context of marginalisation, migration, and minorities indicate that the translator(s) and subtitler(s) could have deliberately made the texts more accessible by using a simpler and clearer register, as far as museum text catalogues and wall text panels translation is concerned, and by transferring semantic content from the source language into the target one, as far as interlingual subtitling is concerned. Finally, a third research question, which is rooted in the idea that translations are “facts of target cultures”, interrogates whether the techniques observed in interlingual translations and subtitles in the context of marginalisation, migration, and minorities indicate that the translators and subtitlers could have deliberately made the issues of migration more visible as compared to the ST contents.

As declared by the International Council of Museums (ICOM), a leading international NGO museum organisation, a museum is a “non-profit, permanent institution in the service of society [...], which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study and enjoyment” (ICOM, online). Embedded in the context of aesthetics of marginalisation as part of the genre known as “aesthetics of subversion” (Mazzara, 2019), which refer to the depiction of migration issues in the wide artistic domain of the visual arts, a selection of bilingual museum panels and catalogue texts, and subtitles for digital museum videos has been made in order to identify them as spaces of re-narration of marginalised narratives which “speak out” and offer oppositional stories. In this sense, I look at the visual arts as instruments of accessibility that provide universal inclusion (e.g., the visual arts make niche knowledge accessible and give voice to minorities), thus, actively contributing to the spread of current stories of marginalisation through networked channels and public cultural spaces. In the context of museum exhibitions, websites and blogs, where niche knowledge is placed (the prime example being migration), interlingual translations for museum texts (i.e., catalogues and panels) and subtitles for audiovisual digital museums (i.e., subtitles for videos, interviews, and documentaries) are scrutinised as modes of universal accessibility able to reconfigure socio-political meanings in ways that depart from other representations in conventional media (e.g., the news bulletins). Accordingly, the translation of museum catalogues and panels into English,

and the English subtitling for museum videos, are viewed as political acts in which accessibility produces subversive counter information.

The first section of this study is dedicated to the relationship between universality and accessibility as a crucial assimilation for the construction of access practices within socio-political and cultural settings, where all citizens as human beings, and citizens of the world, have the right to take part in. In the second section, I argue that accessibility can be conceived as a space of activism and interventionism in its promotion of (audio)visual translation modalities for the spread of niche knowledge within the visual arts. In the central and final sections, I provide a corpus-based analysis centred on the qualitative study of selected interlingual translations and subtitles belonging to three diverse museum text types, which I place in the context of specialised exhibitions (i.e., marginalisation, migration, and minorities), where universal access is given to niche contents with the objective to disseminate socio-political narratives from non-hegemonic perspectives. Thus, the translated products are addressed to speakers of other languages, namely, people belonging to minority groups (e.g., non-native Italian speakers who live in Italy) and tourists. The final section, also the practical and comparative one, relies on the methodologies of Michael Halliday's Systemic Functional Linguistics (SFL) (1978, 2004) and Faber and Mairal Usón's approach to Lexical Semantic Analysis (LSA) within the framework of Lexical Grammar Model (LGM) (1999). The aim is, on the one hand, to explore how English as the dominant target language (TL) functions for the translation of specific niche fields and, on the other hand, which English semantic categories emerge in the transfer from Italian into English within the context of specific museum text types dealing with marginalisation, migration, and minorities. These methodologies have been applied to (a) the bilingual texts for the *Manifesta 12* catalogue (Palermo, 2018); (b) the bilingual museum panels for the museum exhibition *Towards the Museum of Trust and Dialogue for the Mediterranean – Lampedusa* (Lampedusa, 2016); and (c) the subtitles for museum multilingual multimedia guides for *eMMMMe – Porto M – Lampedusa* (Lampedusa, 2013-2014). The identification of the translation modes as instruments of accessibility has been carried out by taking into consideration Jiménez Hurtado, Seibel and Soler Gallego's (2012) taxonomy for the classification of target types and access modes for universal accessibility in the context of museums (i.e., this study investigates museums that are all set in Sicily).

2. Universality and Accessibility in the Visual Arts

Accessibility has for several years been a “legal and technical issue in various countries” in the attempt to guarantee persons with disabilities the possibility to have access to transport, facilities and culture. Accessibility has also become an important issue in the computer and telecommunications industries, the aim being “to optimise the user-friendliness of software, web sites and other applications” (Gambier, 2006, p. 4). Accessibility in the field of the distribution of audiovisual media has been relevant to serve the needs of user groups such as the deaf and blind persons. Today, the issue of accessibility has stopped being merely a question of supporting citizens

with special visual, auditory, motor and cognitive difficulties as too restrictive in light of the “digital divide [...] and the exclusion of certain sectors of society from access to information” (p. 4). Accessibility, in particular in the area of AVT, means that audiovisual services must be available to all users,

irrespective of issues such as where they live, their level of experience, their physical and mental capacity, or the configuration of their computer. Accessibility is not just an issue for the disabled: it does not only mean a barrier-free situation; it also means that services are available and that information is provided and easy to understand. (p. 4).

Nevertheless, accessibility, which is witnessing a surge of attention as an interdisciplinary research field, still remains a concept under scrutiny, open to different interpretations, ideas and resources and, the terminology itself, its practice and training, still require greater precision and demand for norms on national and international levels (Maszerowska, Matamala, Orero, & Reviere, 2014).

In recent research, significant importance has been given to accessibility as a human right (European Economic and Social Committee, 2014) of persons with disabilities and, more interestingly, “as a necessary instrument for the human rights of all, not only of persons with disabilities” (Greco, 2017, p. 94, my translation). The concept of accessibility for all, based on the idea that “all” is not limited to sensory impaired persons, but is extended to a variety of minorities in contemporary societies, relies on the belief that “accessibility is not a human right specifically concerned with persons with disabilities, but an instrument for the achievement, implementation and respect for human rights of all, above all of those people who are at risk of social exclusion” (p. 96, my translation). Since the Universal Declaration of Human Rights, the “human rights” concept has shaped “everyday language and culture, refashioning political rhetoric, and permeating literary works, movies, art and the media” (p. 1, my translation). The literature in the field demonstrates that accessibility has become a “strategic instrument in the agenda of many international actors for the achievement of some of the most pressing social priorities”, and has grown to such an extent that it is possible to talk about “accessibility studies”, in which the “shifts of accessibility” (Greco, 2018, p. 211) are the result of what can be referred to in terms of “accessibility revolution” (p. 210). The shift from an individualist perspective, where citizens are individuals who contribute to society improvement by performing work that is suitable for the individual’s physical features, to a universalist one, where citizens are first of all individuals who have the right to be an integral part of society regardless of their physical abilities, has indeed stimulated the growth of accessibility as a concept functioning and meaning equal opportunities for all at a societal level. Design for all, universal access and inclusive design are all different names of approaches that have largely described accessibility as a theoretical and practical system of application of principles and methods relating to social inclusivity, integration, equality and participation (Persson, Ahman, Arvei Yngling, & Gulliksen, 2015).

Against this backdrop, where light is shed upon accessibility as “more and more a priority and a widespread practice” (Di Giovanni, 2018, p. 189), the identification of the universalist account

of accessibility and its user-centred and proactive approaches (Greco, 2018, p. 211) testifies to the radical changes occurring in this research area, and also strengthens the relationship between universality and accessibility, where the universalist or holistic value implies that access is not limited to a specific group of people but concerns “*all human beings*” (p. 211). In this sense, universality and accessibility can be viewed as parallel concepts resulting in an achieved universal scope according to which accessibility is not something belonging exclusively to persons with disabilities. As already mentioned, universality also entails the sphere of accessibility in the computer-mediated dimension of knowledge, where universal access does no longer refer to computer-based applications for users with physical or cognitive disabilities, but embraces a universal vision of accessibility that regards society at large. For example, the numerous uploads of self-mediated narratives as digital videos and documentaries made accessible to a universal number of users, producers and consumers of computer-mediated knowledge are one of the responses to the universalist landscape in which accessibility offers universal services.

In the area of the visual arts, where greater emphasis is put on niche knowledge as counter discourse universally accessible through translation practices in English as the dominant language and means of global communication, universality goes beyond the well-established concept of MA as the set of services and devices used to provide access to audiovisual media content for people with sensory disabilities (Szarkowska, Krejtz, I., Krejtz, K., & Duchowski, 2013). It is still largely ignored that accessibility in its universalist approach represents a fertile terrain, where information on migration issues can circulate through the visual arts in opposition to anti-refugee media narratives, thus, effectively fostering a specific area in the human rights setting. In other words, MA “has been gradually pervading different aspects of our lives as well as a vast range of fields, giving rise to a plethora of fruitful new ideas, methods and models, and becoming an ever more key issue within a process that is reshaping the very fabric of society” (Greco, 2018, p. 206). Stories of migration provided by first-hand experiences and authentic narratives pervade different sectors of the cultural creative industries (i.e., the visual and performing arts, advertising, music, theatre and opera, festivals, etc.), both in digital and non-digital contexts, where participation, integration and description are the socially-engaged features of a new model of universal accessibility. “Social innovation” (Greco, 2018, p. 219) represents one of the leading processes that has taken place within accessibility mechanisms, while providing radical changes within its normative boundaries of epistemic and moral actions, where inclusion and intervention contribute to the spread and promotion of “social justice and epistemic justice” (p. 219).

In the artistic cases under scrutiny, accessibility has broadened its role as a service and practice for the transmission of meanings belonging to neglected areas in media cultures, and has called for the participation of a wider global audience, often disinformed or disinterested in learning about narratives on minorities confined to forms of niche knowledge. In this sense, accessibility can be said to be universal and to offer a service to ghettoised people, who can eventually have a voice and give public access to their stories. Accessibility is thus an open resource supporting the spread of underestimated information and social interaction of sensory-impaired persons and a universal

audience (e.g., immigrants, speakers of other languages). To some extent, it is transformed into a “positive, reversed trend not only for special audiences but an asset for all” (Di Giovanni, 2018, p. 189), through which any sort of information is made explicit. As already briefly stated in the introductory section, MA acts through (audio)visual translation, which is the field where it has grown as a discipline (Díaz-Cintas, 2007; Orero & Matamala, 2007) and, in particular, within the context of the cultural creative industries (i.e., public and (digital) museums), where participation and activism (e.g., websites and blogs, videos and documentaries are crowdfunded by citizens and subtitled by activist translators) transform citizens into financial supporters, artists into project creators and the migrants into the actors of their first-hand narratives in which untraditional content is made accessible and information is freely expressed.

This type of accessibility involves meaning-making processes in terms of production, consumption and reception of knowledge which can potentially empower the public with new “social” lens through which facts can be seen and understood differently. Interestingly, the dialogue between accessibility and social narratives within the space of the cultural creative industries is rooted in Greco’s belief according to which accessibility “has the potential to impact on a number of human rights for all individuals. For example, it affects the quality of life of the elderly, migrants and linguistic minorities, serving to grant access to culture, information and communication” (2016, p. 12). Narratives of migration, stories of marginalised subjectivities, among whom the disabled, homosexuals, political refugees, victims of the Second War World, and many others, have recently become accessible through the participative contribution of the visual and performing arts and the use of audiovisual translation modes in accessibility services. In their construction and depiction of minority experiences, the visual arts, in particular, exploit accessibility resources. In fact, the idea of accessibility within the arts is twofold: on the one hand, it involves the spread of niche knowledge and, on the other, it regards the diffusion of access practices addressed not only to the blind and visually impaired persons, or to the deaf and hard-of-hearing persons, but to any kind of minority group and individual, and to any kind of audience. This implies that accessibility is interpreted and described as the potential universal quality of being easily reached, entered, or used by people who have a disability of any sort, physical, material and geographical, and as the quality of giving universal access to any type of knowledge and information. In a way, accessibility is reflected in “the globalising digital world and technologies that are recontextualising many individual and collective social practices in relation to minority, minor, lesser-used and endangered language communities” (Folaron, 2015, p. 16).

Within the field of the visual arts, museums play the role of agents of social inclusion and cultural exchange, and their function is thus not limited to the acts of collecting, exhibiting and interpreting cultural heritage. The museum spaces that have as their primary focus the dissemination of narratives which deal with issues of race, unrepresented ethnic populations (e.g., immigrants), occupationally subordinated groups (e.g., women, black people), socially isolated individuals or economically depressed persons (e.g., unemployed, poor and old people), people discriminated for their sexual orientation, physical disability, mental health, physical diversities, and political perspectives,

are the museum settings where universal accessibility is rooted. The language that is adopted in these museum text types plays a strategic role in the development of discursive practices of inclusion, tolerance, social equality and, consequently, in pursuing ethical objectives, and the textualising processes that favour inclusiveness as a universal right rely on the linguistic potential of museum discourse in its relationship with human rights discourse. In fact, the construction of messages and discourses through language is paramount in the construction of inclusiveness and relevant social functions and, in this sense, it goes without saying that any translation activity is central within the space of accessibility that is created for the transmission of stories of marginalisation.

The universalist shift in accessibility has contributed to the broadening of interest in marginalised communities other than sensory impaired persons, and to increasing importance of the concrete experiences of artists and minorities who, as activists, act to spread narratives by means of translation in the visual arts setting. Indeed, more and more activists have mobilised and fought to counteract the media through networks of solidarity that crowdsource artistic interventions, while taking advantage of accessible services, such as translation, interpreting, subtitling and other forms of mediation, in order to bring inaccessible information and knowledge to the core of the political arena. Being conceptualised through new lens, accessibility in the context of the arts can be repositioned as a depository of “solidary activist communities” (Baker, 2016, p. 1) and viewed, to borrow Philip Rizk’s words, as the space of solidarity, where one can connect, communicate and translate in “more ways than just verbal translation” (Rizk, 2013, p. 237), with the intention of building “international solidarity networks that are nonetheless firmly rooted in the granular struggle of a particular place” (Selim, 2016, p. 84).

In this context, knowledge is offered against media barriers that both hinder disabled people’s autonomy and prevent the general public from being familiar with socio-cultural issues expressed through ontological narratives in the visual arts. Universal accessibility thus involves the inclusion of content in national and international cultural platforms, and the growing awareness that different groups and individuals in society exist, not only in terms of disability, but also in terms of age (e.g., children, young generations, elderly people), socio-economic background and culture, ethnic origins (e.g., the new generations of immigrants), intercultural and interlingual differences (e.g., dialects, slang, ELF). As remarked in the studies of Jiménez Hurtado, Seibel and Soler Gallego, “this awareness has encouraged the promotion of social policies committed to guaranteeing accessibility for all individuals” (2012, p. 2). Advances in universal access to any kind of knowledge coincide with the political role given to audiovisual collections, which are exhibited and displayed as agents for social inclusion, activism and for the democratisation of knowledge.

In the light of this, universality in accessibility epitomises discursive and non-discursive interventions which are made possible through a variety of accessible forms in public domains (e.g., outdoor and indoor installations, museum exhibitions, street art, documentaries) and within digital networked frameworks as competitive socio-cultural and political spaces. In these spaces,

niche contents in the shape of narratives, testimonies, ontological and public stories, paintings, visuals, etc. are meant to reverse commonly accepted views, often stereotyped, oversimplified, and manipulated, where minorities are recognised as aliens and masses, rather than individuals with their own rights, collective voices, and individual voices. Accessibility is finally viewed in opposition to the dangerous and outrageous role generated by the media gaze that relies on the reiteration of conventional formulaic expressions that are considered full of integrity, which benefit from institutionalised recognition. Accessibility turns out to be an active instrument which opens the dialogue with neglected topics by making the contents themselves accessible through accessible documentaries, accessible installations and exhibitions, accessible displays and videos in digital and non-digital museums. As Greco (2016) points out, MA, in relation to human rights, acts as a driving force for social change in relation to the fact that accessibility has potentially and intrinsically the capability to be turned into an instrument for change. By embracing the universalist perspective on MA as including persons with and without disabilities, MA can be employed as “a proactive principle for achieving human rights” (p. 23), thus, involving social, cultural and political cohesion and transparency, as well the promotion of education and literacy. This makes accessibility “inclusive and empathetic, as it makes everyone (whether or not they have disabilities) share the same need for accessing original content” (p. 10).

3. Activism in Accessible Museums: Translation(s)

In today’s technological and digital society and dynamic environment, where human exchanges are more audiovisual than ever before, the transmission of content information through the combination of audio and visual components has had an enormous reach, while changing the media’s traditional role in the way in which information is transmitted and in the type of message that is delivered. It goes without saying that the recent advances in technology and digitalisation have favoured the burgeoning of audiovisual communication, which has expanded and flourished while catering for the needs of a general public and a wider audience including sensory impaired persons, as already remarked in the previous sections. As a matter of fact, inclusion for all in the form of access to audiovisual media has become “an important debate in many countries around the globe, featuring prominently in legislation, academic exchange and broadcasters’ output” (Díaz-Cintas & Nikolić, 2018, p. 2).

Within interactive and multimodal communicative settings, the arts have been transformed into platforms for the promotion of inclusion and social integration, thus, offering universal access to knowledge. The art world has become the playground for universal accessibility and has overcome the mass media’s limits and boundaries in the spread of products and narratives on unspoken topics, while encouraging inclusion for all. This has permitted the establishment of a permanent dialogue between minorities/marginalised communities and the visual arts, having at their core accessibility as the quality of the audiovisual space that is made accessible to all, and that gives access to counter information. The arts have turned out to be interactive social agents, ultimately leading to inclusion

and universal accessibility. As one of the most accessible artistic spaces, to put it in Jiménez Hurtado and Soler Gallego's terms, museum accessibility based on translation and interpreting has contributed to "the creation of a truly universal knowledge society" (2015, p. 278). The dissemination of niche narratives in the context of the museums under scrutiny takes place through (a) audiovisual translation modes and (b) interlingual translation. Among the modes in the field of AVT for museums are audio descriptions for the blind and subtitling for the deaf, voice narration of panels, multimedia guides, multilingual multimedia guides (with subtitles), tactile and oral descriptions for visually-impaired and deaf and hard-of-hearing people. Other forms of translation include interlingual textual translation and textual adaptation, the choice between which depends on the target audience. Both types can be applied to labels and panels, catalogues and leaflets for speakers of other languages.

These intersemiotic and interlingual translation procedures make accessibility a space of "anti-establishment initiatives, and specific issues that exceed national and social boundaries" (Baker, 2019, p. 453), a space for activism (p. 453), potentially meant to subvert political establishment and the cultural forms it supports. Cases in point are the multilingual multimedia guide for *Porto M* (i.e., a proper video) and the interlingual translations for *Towards the Museum of Trust and Dialogue* and *Manifesta 12*, which can be looked at as forms of activism providing universal access to knowledge. Museum studies scholars seem to agree that museums are institutions created for the dissemination of knowledge and culture in the broadest sense, considering that the activities carried out by museums aim to make collections understandable to society and accessible to all. In this sense, accessibility is used in the broadest sense of the word. Museums have evolved and are no more perceived as "mere repositories of valuable objects that are studied by experts and visited by the upper levels of society. Museums have become places for social and cultural encounter and interaction" (Jiménez Hurtado & Soler Gallego, 2015, p. 280) and have been transformed into collaborative instruments "at the service of society and of all the cultural groups that compose it" (p. 280). The museums taken into account in this study are interactive, bilingual, multimodal and informative. In these museums, panels and labels, audiovisual texts, audio texts, images, paintings and displayed material objects, make contents, people and objects accessible through translation as the principle instrument of communication and knowledge dissemination.

Accessibility as a means of activism therefore does refer to the spheres of social activism, cultural activism, art activism and aesthetic activism – all terms that tend to highlight how artistic audiovisual initiatives can "challenge or undermine some aspects of the political establishment and/or the corporate culture that underpins it" (Buser & Arthurs, 2013, p. 2). Accessibility as cultural activism and activist intervention develops and stimulates the production of audiovisual artistic products which aim to challenge and stimulate "dominant interpretations and constructions of the world while presenting alternative socio-political and spatial imaginaries" (p. 2). The activist attitude shared by artists, citizens, linguists, and entire online communities, has posed accessibility as a challenge to mainstream values across diverse fields – political, economic, cultural, institutional, and socio-linguistic –, and audiovisual translation at the centre of all types of projects that support political

and aesthetic forms of intervention. Accessibility through activist modalities encourages the promotion of social counter narratives and relies on the power of translation and subtitling as the most challenging instruments for stimulating communication and dialogue across diverse platforms and media genres. In line with Luis Pérez González's concept of AVT as an "interventionist practice" (2014), accessibility can be looked at as an activist and interventionist practice that challenges traditional pre-arranged political and cultural orders. This process involves "[questioning] the current operation of global culture industries by providing a new model of content distribution and its organisation based on consumers' voluntary work" (Lee, 2011, p. 1132), and also "[challenging] the established global order by encouraging subtitlers to experiment with and develop innovative subtitling strategies that undermine restrictive conventions imposed by the industry" (Pérez González, 2013, p. 10). In this respect, activism in translation studies and, in particular, in the context of AVT, breaks down barriers in the media marketplace and makes unordinary knowledge accessible through collaborative modes. From this perspective, audiovisual translation products and translated museum panels as examples of accessible art forms give access to knowledge and are instrumental in enabling cultural participation and collaboration "to challenge the corporate and political order" (Baker, 2019, p. 459).

Accessibility of diverse types of content to a variety of visitors (i.e., sensory impaired people, the general public, tourists) strengthens its functional, active and political role in modern societies as the promoter and "[disseminator] of all types of cultures" (Jiménez-Hurtado, Seibel, & Soler Gallego, 2012, p. 4). As shown in the next sections, the museum exhibits taken into account in this study, are conceptually built for social purposes, convey units of meanings, and are spaces of communication and knowledge transmission (Santacana & Serrat Antolí, 2005). *Manifesta 12, Towards the Museum of Trust and Dialogue* and *Porto M* are interactive multimodal communicative contexts, where meanings transmit knowledge according to a textual macro-level (i.e., exhibitions/audiovisual products as genres) and to a textual micro-level (i.e., exhibitions, installations and audiovisual products in their relationship as specific text types that aim to fulfil contextually- and historically-related functions). Both levels are connected by the specific situational macro-context that is provided by the information transmitted through the accessible museums and artistic contexts. The interaction between the two levels, the macro and micro, influences the level of social discourse and stimulates the audience's responses in terms of raising cultural awareness. Against the Sicilian cultural backdrop, attention is paid to translation in its broad sense as a dynamic, collaborative and adaptable resource to the specific situational settings. Thus, translation and accessibility are turned into one single model of knowledge transmission within audiovisual and visual products addressed to both sensory-impaired people and foreign users or participants. As already stated, recent developments in MA (Greco, 2016) have testified to wider universalist descriptions as necessary requirements not only for persons with sensory impairments, but also for other types of participants who want to get involved in the socio-cultural activities within the European Union (Romero-Fresco, 2019). In particular, according to the latest international reports on subtitling, it is proved that this AVT practice is adopted by a heterogeneous target

for different reasons: deaf and hard-of-hearing people, persons having cognitive disabilities and linguistic problems, people who want to watch a movie in a foreign language, people who have difficulties in hearing the content due to environmental conditions (e.g., noisy contexts) or who have a certain type of hearing loss, people who cannot make use of the sound because it could be inappropriate in the context where they are (e.g., libraries, on a plane) (Romero-Fresco, 2019).

4. Corpus and Methodology

The corpus collected is made of interlingual translations of catalogue texts and wall panels, and of subtitles for multimedia guides created for temporary and permanent museum installations and exhibitions held in Sicily between 2013 and 2019. The methodological approaches applied to the museum text types under scrutiny draw upon Halliday-Ravelli's SFL applied to museum communicative frameworks and upon Faber and Mairal Usón's (1999) lexico-semantic model. The target texts (TTs) are considered types of marginalised narratives produced for the *Manifesta 12* event (Palermo, 2018), for the temporary museum exhibition *Towards the Museum of Trust and Dialogue* (Lampedusa, 2016) and for *Porto M* (2013-2014) as the only Lampedusa permanent museum on migration. The selected TTs are part of a large comparable corpus (which is not the focus of the present study as a whole), whose grammatical and lexico-semantic categories are specifically concerned with the representation of stories of marginalisation.

4.1. General Overview

This analysis provides a qualitative study of linguistic and thematic peculiarities contained in the selected TTs, which allow them to be categorised as sample cases (in a limited portion of the corpus) for each museum setting taken into account. This signifies that the selection of the TTs functioning as interlingual written translations and subtitles employed for comparative purposes has taken place according to linguistic features that entail the lexical component, syntactic structures and textual features which have made the translated narratives under scrutiny more accessible for the general public (who are familiar with the English language), and which can be used as samples of best practices. In fact, the selection of the TTs and subtitles in the selected subcorpus has been determined by (a) lexical, (b) syntactic and (c) textual features. The following chosen English texts are characterised by (a) a high number of terms specifically concerned with the field of marginalisation, migration and minorities (i.e., *sea, exchange, peoples, accompanying, narratives, victims, memory, desert, country, authentic, strangers, welcoming, images, drawings, refugees, floating*), (b) cases of lexical density that if, on the one hand, encourages redundancy, on the other, makes the TTs lexically productive (i.e., the repetition of content words such as *participatory, participating, monitored, controlled, sea, refugees, floating, drawings, narratives, etc.*), (c) the abundant use of verbalisations, passive and active structures, transitive and intransitive forms and, in particular, of relative clauses introduced by the relative pronoun *who* which provides explicitness and directness to message transfers (i.e., *attempts to create, is monitored and controlled,*

who are also controlled and guided, who died, shipwreck occurred, which is imposed in Eritrea, who were kept in detention and tortured, [...] over the years are exhibited, one takes), (d) the use of anaphoric and cataphoric systems of referencing that increase textual cohesion, and (d) the adoption of thematic sequences. In this sense, the selected TTs demonstrate that any pragmatic value is made directly accessible to the public through the dominance of verb phrases rather than by means of indirect constructions reinforced by noun phrases. Furthermore, the tendency to verbalisation in opposition to nominalisation gives emphasis on the explicit authorial voice in the translated texts, and makes the target products types of ontological discourse genres, not categorisable within the general label of “specialised discourse”, in which, on the contrary, nouns are commonly used instead of verbs to convey concepts relating to actions or processes (what in SFL is referred to as grammatical metaphor).

4.2. Manifesta 12, Towards the Museum of Trust and Dialogue for the Mediterranean, eMMMMe – Porto M – Lampedusa

The first museum product taken into account is *Manifesta 12* and its bilingual catalogue. *Manifesta* is Europe’s premier itinerant biennial exhibition of contemporary art. Being nomadic by definition, it changes its location every two years. Last year it took place in Palermo as the *Manifesta 12* edition. It “purposely strives to keep its distance from what are often seen as the dominant centres of artistic production, instead seeking fresh and fertile terrain for the mapping of a new cultural topography” (*Manifesta 12*, 2018). Its main focus is promoting innovation and reflecting on “emerging developments in contemporary art, set within a European context” (*Manifesta 12*, 2018). In doing so, *Manifesta* proposes forms of artistic expression that are connected with the needs of the host city. It is open to local, national and international audiences. The Palermo edition was particularly significant, since it was embedded in the context of coexistence and interaction, which is symbolised by the Sicilian space of reception, arrivals and cultural mixture. While exploring the prevailing historical and cultural identity of its temporary home, *Manifesta* maps out a new cultural topography featuring the work of local and international artists and architects. It provides the accessibility of a variety of information, which would never be commonly spread across mainstream platforms.

The *Manifesta* content is, thus, by definition, a type of counter discourse, and is diffused through video installations, performances, sculptures, depicting borders, migrations, postcolonial bodies, military experiments, farmers, drowning worlds, scenes of exchange, and collectives. In particular, *Manifesta 12* attracted 206,456 individual visits to the numerous exhibitions and installations, all as bilingual events (*Manifesta 12*, 2018). Furthermore, information about collateral events was collected in a bilingual catalogue that was created to guarantee access to these numerous events that were marginal compared to the rest of the artistic projects. This catalogue reports seventy-one collateral events running parallel to the *Manifesta 12* biennial events. Each of the 71 TTs featured in the catalogue comprises 10 descriptive passages organised under 10 headings. These include

the following areas: Opening; Duration; Open; Entrance; Address; Website; Social; Artist(s), Organiser(s), and Venue. All the bilingual catalogued texts (in Italian and English) are interdisciplinary and intercultural, have thematic and structural affinities, and regard both local and international communities.

In this study, attention is given to the description of the artistic performance *Enough about You* by Einat Amir (2018 [2016]). *Enough about You*, promoted by TRIAD (*Towards Regional Integration of Artistic Development*, London), is a project which gives access to the experiences occurring in significant participatory performances inspired by contemporary social topics, such as coexistence, interaction, dialogue, and respect. The project of the Israeli artist is an experiment based on the creation of interactions that blend the notions of authenticity and fabrication. Within the context of marginalisation and minorities, *Enough about You* depicts emotions in conflict laboratories as psychological experiments and participatory performances meaning to increase collaboration.

The second example is the exhibition *Towards the Museum of Trust and Dialogue for the Mediterranean*. It was a non-permanent exhibition and installation opened on the island of Lampedusa on June 3, 2016. On the website, it is declared: “Today more than ever it is important to understand the nature and dynamics of migration” (*Towards the Museum of Trust and Dialogue*, online). The museum exhibition aimed to exploit the arts in order to approach directly and empathetically the delicate and complex topic of migration, while reflecting on “the fragility and vulnerability, the strength and courage of those who leave their homes and set off in search of a better future” (*Towards the Museum of Trust and Dialogue*, online). It made available a variety of unedited documents and made visible the remaining personal belongings and everyday objects carried by the migrants. Narratives of individuals and collective voices were catalogued and displayed. Among these: drawings by Adal, an Eritrean refugee who depicted the torture in his home country; dead bodies of people found on the 2013 shipwrecks; personal items belonging to 52 people who died of asphyxiation in the hold of a boat; drawings of a Syrian girl, Sheradaze, elaborated in the camp of refugees in Idomeni. The museum exhibition also housed the “Wreck Room”, where visitors were offered a new multimedia itinerary of images and sounds where they could imagine and feel what the migrants experienced during sea crossings. Bilingual museum panels put emphasis on the migrants’ experiences and gave access to their stories from an individual perspective.

The last case that demonstrates how museums can provide access to topics on minorities and marginalisation is offered by *Porto M*, its exhibits and multilingual multimedia guide. *Porto M* is a permanent anti-institutional social museum located on the island of Lampedusa and created by the Askavusa Collective in 2013, where counter information is made explicit through language, visuals, and material objects. It aims to challenge mainstream institutions by producing counter narratives that depict the lives of the migrants before and after their passage across the Mediterranean Sea. The museum is chiefly conceived as an anti-museum, a non-standard

museum, containing no labels or panels, where the display of everyday objects represents migrants' material culture, which is rendered all the more immediate because of the routine nature of the items. In this type of museum, "museum translation", also viewed as the accessibility of first-hand experiences and authentic knowledge, is offered through multimedia guide tours, adapted guided tours and multilingual multimedia guides. The website of *Porto M* contains a blog page where a multilingual multimedia guide is uploaded as a short documentary or video, where the display of objects in combination with the narrating voice (also physically present), is a voluntary act of political intervention in the lives of marginalised people.

4.3. Methodology

Jiménez Hurtado, Seibel and Soler Gallego provide a taxonomy for the classification of target types and accessibility modes for universal accessibility in museums. As remarked in their study, where emphasis is put on translation and interpreting as accessibility tools of museum spaces, accessibility is offered to all types of audience through diverse modes of access. Table 1 only includes the audience types and modes of translation that are present in the subcorpus under scrutiny and takes into consideration Jiménez Hurtado, Seibel and Soler Gallego's "classification of museum accessibility resources according to visitor profile". This classification includes: "Children", "Teenagers", "Teachers and Students", "Families", "Speakers of other languages", "Visually-impaired people", "General (adults)", "Hearing-impaired people", "Mentally or intellectually disabled people", "Physically disabled people" (2012, pp. 4–5).

As already mentioned, the data selected are addressed to a general public (adults) and speakers of other languages (e.g., chiefly tourists, speakers of English as a second language or English as a lingua franca, immigrants), and access to information is given through auditory-visual and verbal-written channels (i.e., multilingual multimedia guides with subtitles) and the visual-verbal-written channel (i.e., interlingual translations). The type "Multilingual gallery printed texts (exhibit labels, text panels, gallery guides, text catalogues, educational materials)" has been added in bold to Table 1 and addresses speakers of other languages. The focus is on how niche knowledge is transmitted by making use of different modes and to what extent content dissemination changes according to the target language used.

Table 1.

Modes of Universal Accessibility Adapted from Jiménez-Hurtado, Seibel, & Soler Gallego (2012, p. 45)

General (adults)	Speakers of other languages
Multimedia guide: images, audio, video	Multilingual multimedia guide
Audio guide adapted to different tour durations and contents	Multilingual audio guide
Self-guided tour adapted to different tour durations and contents	Multilingual website and/or information leaflet
Guided tour adapted to different durations and contents	Multilingual gallery printed texts (exhibit labels, text panels, gallery guides, text catalogues, educational materials)
Online collection	
Online virtual tour	
Online multimedia resources: podcast, video, audio	

The comparative analysis of the Italian texts (both audiovisual and visual-textual) and English ones (i.e., selected text catalogues and wall text panels, and subtitles produced for niche content-oriented museum types) draws on SFL and LSA methodological models. Systemic Functional Linguistics has been applied to the printed written translations for the *Manifesta 12* catalogue. These have been used as a Translation Quality Assessment (TQA) model, where the three metafunction levels (i.e., ideational-representational, interpersonal-organisational, textual-interactional) have all been represented and analysed in order to offer an overall appreciation of the modalities by means of which SFL functions within translating contexts. Instead, the Lexical Semantic Analysis has been applied to the wall text panels for *Towards the Museum of Trust and Dialogue for the Mediterranean*, and to the subtitles for *Porto M* in order to demonstrate how and to what extent process types (i.e., verbal structures), which have been classified and identified at the level of SFL's ideational-representational metafunction model, can intervene to the activation of cognitive categories which, somehow, influence the reader-viewer's perception of texts and subtitles. On the one hand, SFL is useful to evaluate the quality of English in the translations as products which, in DTS, imply the description of individual translations or several translations of the same text. These translations are embedded in the target context (i.e., the context of target readers), and have a function, which is deconstructed by SFL as a method of TQA by means of which language is used to express social reality (i.e., contents are disseminated through the choice of certain words rather than others and selected from a lexical container at our disposal and, finally, combined together according to syntactic and textual criteria that speakers deliberately decide to use). On the other hand, LSA is useful for the identification of the prevailing processes (i.e., verb phrases) within a consciously selected syntactic structure in which each verb stimulates determinate cognitive areas. This permits us to discover what areas of experiences are activated in the viewer's or reader's

receptive contexts, and how affectivity is stimulated by means of language as social reality. LSA from the perspective of Faber and Mairal Usón's (1999) classification of lexical domains appraises the cognitive categories activated through the process of verb selection. Types of conceptual categories are thus activated by the most representative verbs that are present in the discursive forms used in wall text panels and subtitles for museums.

In this context, Halliday's SFL and Faber and Mairal Usón's lexico-semantic model have been applied to the subcorpus by relying on the concept of functional language, where the notions of "systemic" and "functional" are used to serve the application of a TQA model and the recognition of the semantic function of verbs and their lexical density for the production/creation of texts and visuals. The combination of the two methods aims to propose a rethinking of the relationship between the source and target texts in the context of museum translation, and the relationship between video transcripts and their subtitles, both aiming to provide access to counter information in museum environments, digital and non-digital.

Grammar in SFL refers to the resources used for creating meanings by means of wordings. The occurrence of language elements creates various interpretations according to three metafunctions: ideational (field), interpersonal (tenor) and textual (mode). Within this framework, functional language is a complex and dynamic semiotic system that is organised in four interrelated strata: semantics, lexicogrammar, phonology and phonetics. The focus is here on the first two strata, which relate to the experience of reality and how this experience is worded. Vocabulary (the conscious selection of lexis) and grammar (the deliberate use of certain structures instead of others) are not two distinct poles of the language system but part of a single continuum called lexicogrammar. The use of language for the construal of the experiential and interpersonal relationships (i.e., ontological narratives) transforms them into meanings which, in their turn, are put into words. Therefore, the notion of "function" in the context of SFL refers to the notion of a purposeful use that aims to achieve materialised goals. How language is used to make unheard stories of marginalisation accessible is discovered through SFL, which represents a TQA model for comparing and understanding the modalities of language use for the production of bilingual artistic panels and subtitles for the transmission of niche knowledge. In particular, at the level of the TQA model, the investigation is based on how SL metafunctions (experiential, interpersonal and textual) have been rendered in the TL ones, and whether variations have occurred on the basis of TL cultural needs. The semantic classifications of verbs as conceptual categories on which clauses depend are connected with the identification of processes as components in the experiential metafunction, where the selection of words to express meanings is essential to the type of message and the modalities by means of which a certain message is conveyed. In the SFL analysis, the main focus is on verbal structures, since the meanings of a sentence and the text that sentences form are arranged and patterned around verbs. Verbs are thus the most important categories for meaning production and transmission.

In the LSA method, here based on a parallelism with SFL, the belonging of a verb to a lexical domain is meaningful in relation to its semantic segmentation. In their scrutiny of 12,000 English verbs in a certain number of monolingual dictionaries, Faber and Mairal Usón (1999) pinpoint the following semantic verbal types: “be/happen”, “become”, “have”, “say”, “feel”, “do/make”, “know/think”, “move (go/come)”, “see/hear/taste/smell/touch”, “use”. This classification shares common features with Halliday’s system of transitivity and, in particular, in relation to the ideational metafunction and verb classification according to material, mental, relational, existential, behavioural and verbal processes. Faber and Mairal Usón’s identification of lexical domains consists of the following categories: EXISTENCE, CHANGE, POSSESSION, SPEECH, EMOTION, ACTION, COGNITION, MOVEMENT, PHYSICAL PERCEPTION, and MANIPULATION (p. 88). This implies that the selection of a specific lexical element is meant to satisfy specific semantic areas. The semantic classification is useful for two reasons. On the one hand, it helps classify verbs according to the lexical domain they belong to (e.g., migration, marginalisation) and, therefore, to relate these verbs to the same generic term, since they are members in the same lexical domain despite the fact that they may encode meanings from different perspectives (i.e., *escape*, *run*, *get away* [Action/Movement]). On the other hand, the lexical domain of a verb drives its argument structure, and the verbs belonging to the same lexical domain share the same structure of action. Verbs are categorised according to the selection of each superordinate within the subcorpus by means of which it is possible to decipher lexical density, and to evaluate how metafunctions are rendered within the lexico-grammatical system to address issues of translation quality and language use in specialist contexts (i.e., migration). Verbs are considered on the basis of their frequency in the texts, subsequently gathered according to the lexical domain in order to understand which superordinate dominates in the subcorpus. In other words, clauses have meanings built around verbs and every lexical field is defined according to its superordinate (Faber & Mairal Usón, 1999).

5. Analysis



By moving from the consideration that museums are institutions and cultural spaces meant to disseminate knowledge and culture in the broadest sense, it is possible to argue that it is fundamental in museum activities that collections are made understandable to society and its visitors in general. In order to make museum information knowable and understandable, it has to be accessible. Accessibility of niche information responds to the expectations of the target audience of migrant narratives, where linguistic and discursive features follow an established set of conventions (e.g., more nominalisations in Italian vs. verbalisations in English). The use of bilingual catalogues and panels, and subtitles for multilingual multimedia guides for the diffusion of niche knowledge, opens up new horizons of accessibility, not limited to local visitors, but extended to speakers of other languages. In the cases in point, English is the dominant language used for communicative reasons. In fact, the museum texts of the subcorpus, namely, catalogues, panels and subtitles, are types of “communicative events” which are hosted in multidisciplinary, multimodal, multilingual

and multifunctional museum settings (Jiménez Hurtado & Soler Gallego, 2015, pp. 280–281). In this communicative context, interactional processes (e.g., comments can be written and left in a room meant to welcome audiences’ feedback, and “likes” and “dislikes” can be posted, also adding comments, in websites and blogs) make accessibility more participative, and not limited to the reading of translated museum panels and subtitled videos.

According to Kavanagh’s studies (1991), a model of communication that may encourage conversational interaction with museum visitors requires that the layout and style of exhibit labels and panel texts take into account the conversational aspect of language use. This signifies that the exhibit texts should be rather informal (e.g., the use of the personal pronoun *you* is suggested, when it is possible, sentences are expected to be simplified and shortened in order to form more than one sentence if punctuation makes the reading less fluid: colons, for instance, make the sentence more difficult to follow). In brief, museum curators and translators have to have the audience in mind and be aware that the choice and combination of words are crucial to the transmission of textual meanings. Furthermore (audiovisual) translation scholars have recently begun cooperating with museum curators in order to make exhibitions accessible in the broadest sense of the word. According to Jiménez Hurtado and Soler Gallego’s classification of accessible translation and interpreting modalities in museum contexts, as shown in Table 2, situational contexts of art dissemination can be accessible through a variety of modes and involve different source text genres:

Table 2.

Classification of Accessible Translation and Interpreting Adapted from Jiménez Hurtado & Soler Gallego (2015, p. 279)

Accessible	Translation and Interpreting
  Modalities	AD SDH / Standard Subtitling Sign Language Interpreting (SLI)
Textual Adaptation/ Interlingual Translation Respeaking	Visual arts (painting, sculpture, installation/exhibition , film, video)
Source Text Types	Performing arts (theatre, dance, opera, music) Audiovisual programmes Architecture Natural site Video games

The selected museum panels, text catalogues and subtitles for multilingual multimedia guides, according to the classification above (Jiménez Hurtado & Soler Gallego, 2015), belong to the category

of visual arts (i.e., this type has been added to the classification in Table 2 for the purpose of this analysis), whereas the modes used to make ST types accessible are standard subtitling (i.e., it is used to refer to “subtitling for recipients without a hearing impairment”) and interlingual translation (i.e., it is added to the classification in Table 2 and used to refer to “translating for recipients of bilingual wall text panels and text catalogues).

5. 1. *Manifesta 12* – “Enough about You”

In *Manifesta 12*, in the case of which we focus on its bilingual catalogue, the study aims to trace the similarities and deviances within ST-TT pairs (Italian-English) and specifically draws on Louise J. Ravelli’s (2006) analysis of museum language according to Halliday’s SFL. In accordance with Halliday’s SFL model, Ravelli identifies three metafunctions of language for the production of meanings in museum texts: the representational meaning, which concerns how we engage with, understand and refer to the world we experience (Halliday’s ideational metafunction); the interactional meaning, which concerns the roles, the relation, and the attitude of text producers and receivers (Halliday’s interpersonal metafunction); and the organisational meaning, concerning how texts are shaped and structured to convey the previous two meanings (Halliday’s textual metafunction). The identification of each metafunction and its features leads to the development of a TQA model for museum texts, as remarked in Jiang’s studies (2010), which is based on how target readers perceive translated products. There are no absolute standards that measure translation quality unless appropriateness, which is a function given by different socio-cultural contexts, and which depends on the adaptation to each text type and the target readers’ expectations. TQA is thus based on the evaluation of whether a ST is relayed in the TT, and if the ST is relayed appropriately. The identification of deviances and similarities between STs, TTs and similar text types takes place on the basis of the first-order meanings belonging to the system of language as conceived in Halliday’s terms: the representational meaning is expressed by the participants, processes and circumstances in the clause; the interactional meaning is explicated through specific speech functions, tone of voice, pronouns of address, mood and modality; and the organisational meaning is given by the selection of themes and rhemes, the thematic progression and language complexity within the clause. The assessment procedures of quality description are divided into three phases: informativity, acceptability and intertextuality. The level of informativity is evaluated according to the shifts in the representational metafunction, the level of acceptability is evaluated according to the shifts in the interactional metafunction, and the level of intertextuality is evaluated according to the shifts in the organisational metafunctions and concerns the TT expectations and conventions in relation to museum TT types (Jiang, 2010). Appropriateness to museum TT types and readership is essential to translation quality assessments.

Deviances and similarities existing between STs and TTs are evaluated according to the lexicogrammatical systems of each language and generic comparison, and according to appropriateness to TT textual conventions connected with the audience’s expectations. As a case in point, *Enough about*

You is a performance that stimulates participation, human encounters, and interaction, which is also the general theme of the corpus which, all things considered, provides veritable narratives to encourage human relations. The descriptive target texts in the catalogue prove how the three ST meanings are realised in the TT ones as metafunctions that aim to fulfil specific purposes for a public who are familiar with the English language in an Italian setting.

Tables 3 and 4 investigate the interactional meaning (interpersonal in Halliday's terms) and witness what variants can be found in the TTs compared to the STs.

Table 3.

Enough About You: Interactional/Interpersonal Metafunction (Part 1)

Museum Communicative Frameworks	Source Language Text	Target Language Text
Interactional level 1	<i>Enough about You</i> è una performance collettiva nonché laboratorio d'analisi: il pubblico, monitorato e controllato , può osservare e sperimentare incontri «autentici» tra due estranei, in ugual modo controllati e guidati mediante una struttura sperimentale di audio-conversazione appositamente creata per la performance.	<i>Enough about You</i> is a participatory performance that attempts to create a monitored lab where the audience, who is monitored and controlled , can observe and experience «authentic» encounters between two strangers, who are also controlled and guided through an experimental audio-conversation structure, which is created specifically for this piece.

Table 4.

Enough About You: Interactional/Interpersonal Metafunction (Part 2)

Museum Communicative Frameworks	Source Language Text	Target Language Text
Interactional level 2	<p><i>Artista/i</i> Einat Amir realizza installazioni video e performance dal vivo. Il suo lavoro è dominato da un forte interesse per le dinamiche socio-politiche contemporanee.</p> <p><i>Organizzatore/i</i> Maria Teresa Setaro è curatrice e ricercatrice. Direttrice di TRIAD, è laureata in Arte Contemporanea e ha un forte interesse per le arti partecipative.</p> <p><i>Venue</i> Si prega di fare riferimento al sito web.</p>	<p><i>Artist(s)</i> Einat Amir works in video installation and live performance. Amir's work is infused by her strong interest in contemporary socio-political issues.</p> <p><i>Organiser(s)</i> Director of TRIAD, independent curator and researcher, Maria Teresa Setaro has an MA in Contemporary Arts and a specific interest in participatory arts.</p> <p><i>Venue</i> Please refer to the website.</p>

At the level of roles, the two texts, the ST and TT, share the same authorial perspectives, which are enhanced by the presence of statements (i.e., declarative passive sentences: *è dominato/is infused*). However, the TT, also due to the features of English syntactic constructions, presents aspects of readability and fluidity that render the text more accessible in terms of universal accessibility. In fact, thanks to syntactic devices and lexical reiterations, the TT is more cohesive and coherent than the ST. Connective textual elements contribute to the simplification of the TT and stress meaning-making processes that facilitate message-transfer mechanisms. Some examples of syntactic devices in the TT are the addition of passive structures and relative constructions (*monitorato e controllato vs. who is monitored and controlled; controllati e guidati vs. who are also controlled and guided; creata appositamente vs. which is created specifically for this piece*), processes of verbalisation that put emphasis on the action (*nonché laboratorio d'analisi vs. that attempts to create a monitored lab*), anaphoric references through the repetition of proper names and personal pronouns (*Il suo lavoro è dominato vs. Amir's work is infused*). Syntactic-grammatical devices make the relations between the participants in the event (and, therefore, in the clause) more explicit in the TT than in the ST, and also simplify the comprehension of the translated text in dialogue with the visual component in the catalogue.

At the level of generic structure, the organisational meaning (textual metafunction in Halliday’s terms), as shown in Tables 5 and 6, is given by the selection of a specific thematic progression, which in English functions through the lexical cohesion given by the repetition of lexical units according to grammatical criteria. A case in point is what is referred to as anaphoric referencing (i.e., *Einat Amir/Einat Amir; Il suo lavoro vs. Einat Amir’s work, Amir, her; un forte interesse vs. her strong interest*), accompanied by the economy principle, through which the English paragraph contains one clause vs. two clauses in Italian (*Director of TRIAD, independent curator and researcher, Maria Teresa Setaro has [...] vs. Maria Teresa Setaro è curatrice e ricercatrice. Direttrice di TRIAD [...]*). Besides, the English construction places more emphasis on the socio-cultural position of the participant as the Director of TRIAD (*Director of TRIAD vs. Maria Teresa Setaro*). This construction makes the educational and professional background as a general setting more accessible (since it occupies a thematic position) and shifts the reader’s attention from the person as such to objective societal roles. Also, as regards the number of words, the one sentence structure in the TT satisfies the principle of economy, which is preferred in the discourse structure of museum panels. The prevalence of a linear thematic progression in English is also given by the repetition of the relative pronouns in the embedded clauses (*audience, who is monitored and controlled; two strangers, who are also controlled and guided; an experimental audio-conversation structure, which is created specifically for this piece*).

Table 5.

Enough about You: Organisational/textual Metafunction (Part 1)

Museum Communicative Frameworks	Source Language Text	Target Language Text
Organisational level. Generic structure: interaction between language and purpose.	<i>Enough about You</i> è una performance collettiva nonché laboratorio d’analisi: il pubblico, monitorato e controllato , può osservare e sperimentare incontri «autentici» tra due estranei, in ugual modo controllati e guidati mediante una struttura sperimentale di audio-conversazione creata appositamente per la performance .	<i>Enough about You</i> is a participatory performance that attempts to create a monitored lab where the audience, who is monitored and controlled, can observe and experience «authentic» encounters between two strangers, who are also controlled and guided through an experimental audio-conversation structure, which is created specifically for this piece.

Table 6.

Enough about You: Interactional/Textual Metafunction (Part 2)

Museum Communicative Frameworks	Source Language Text	Target Language Text
Organisational level.	<i>Artista/i</i>	<i>Artist(s)</i>
Generic structure: interaction between language and purpose.	Einat Amir realizza installazioni video e performance dal vivo. Il suo lavoro è dominato da un forte interesse per le dinamiche socio-politiche contemporanee.	Einat Amir works in video installation and live performance. Amir's work is infused by her strong interest in contemporary socio-political issues.
	<i>Organizzatore/i</i>	<i>Organiser(s)</i>
	Maria Teresa Setaro è curatrice e ricercatrice. Direttrice di TRIAD, è laureata in Arte Contemporanea e ha un forte interesse per le arti partecipative.	Director of TRIAD, independent curator and researcher, Maria Teresa Setaro has an MA in Contemporary Arts and a specific interest in participatory arts.
	<i>Venue</i>	<i>Venue</i>
	Si prega di fare riferimento al sito web.	Please refer to the website.

The Italian structure, which coheres differently compared to the English one from a grammatical point of view, contains only a limited number of lexical reiterations – not always belonging to the same part of speech – as key terms in the corpus (i.e., *performance, controllato, controllati, sperimentare, sperimentale, participatory*). On the contrary, nominal and adjectival reiterations in the forms of nouns, proper names, personal pronouns, adjectives, past participles with adjectival functions, complements, etc. are used as standard grammatical categories and lexical phrases in museum text catalogues in English.

At the level of the representational meaning (ideational metafunction in Halliday's terms), as shown in Tables 7 and 8, English verbalisation processes vs. Italian nominalisations, are the most striking differences in terms of syntactic and lexical choices between STs and TTs. The Verbal Group Complex (VGC), introduced by the third person present tense *attempts*, emphasises the possible outcome of the main process *to create* and substitutes the nominalised, static structure, *nonché laboratorio d'analisi*. The English version slightly differs from the original meaning (*nonché laboratorio d'analisi* vs. *that attempts to create a monitored lab*) with the purpose of making explicit the role

of the participatory performance as a monitored lab through the use of a verbal structure that activates the scope of the performance itself. In the TT, the choice contrasts the function of *nonché* (meaning “as well as”, “and”) in the Italian clause, where the adverb *nonché* adds further roles to *Enough about You*, which is seen as a participatory performance and monitored lab.

Table 7.

Enough about You: Representational/Ideational Metafunction (Part 1)

Museum Communicative Frameworks	Source Language Text	Target Language Text
Representational level: how objects are experienced and contextualised.	<p><i>Enough about You</i> è una performance collettiva nonché laboratorio d’analisi: il pubblico, monitorato e controllato, può osservare e sperimentare incontri «autentici» tra due estranei, in ugual modo controllati e guidati mediante una struttura sperimentale di audio-conversazione creata appositamente per la performance.</p>	<p><i>Enough about You</i> is a participatory performance that attempts to create a monitored lab where the audience, who is monitored and controlled, can observe and experience «authentic» encounters between two strangers, who are also controlled and guided through an experimental audio-conversation structure, which is created specifically for this piece.</p>

Table 8.

Enough about You: Representational/Ideational Metafunction (Part 2)

Museum Communicative Frameworks	Source Language Text	Target Language Text
Representational level: how objects are experienced and contextualised.	<i>Artista/i</i> Einat Amir realizza installazioni video e performance dal vivo. Il suo lavoro è dominato da un forte interesse per le dinamiche socio-politiche contemporanee.	<i>Artist(s)</i> Einat Amir works in video installation and live performance. Amir's work is infused by her strong interest in contemporary socio-political issues.
	<i>Organizzatore/i</i> Maria Teresa Setaro è curatrice e ricercatrice. Direttrice di TRIAD, è laureata in Arte Contemporanea e ha un forte interesse per le arti partecipative.	<i>Organiser(s)</i> Director of TRIAD, independent curator and researcher, Maria Teresa Setaro has an MA in Contemporary Arts and a specific interest in participatory arts.
	<i>Venue</i> Si prega di fare riferimento al sito web.	<i>Venue</i> Please refer to the website.

As far as the processes are concerned, both texts, STs and TTs, contain three main processes, namely, relational, material and mental processes, as shown in Table 9.

Table 9.

Processes in Enough about You (2018)

Processes	Italian	English
Relational	è (2); has	is; has (1 implicit)
Material	sperimentare, realizzare, dominare, fare riferimento	create; monitor; control (2); experience; guide; create; work; infuse; refer
Mental	osservare	observe

Material processes express aspects of experiences focusing upon doing or happening, where the actor is the key participant, relational processes are concerned with being, possessing,

or becoming. Therefore, if material processes are processes of doing or happening in the transitive or intransitive clause, thus, as Halliday and Matthiessen affirm (1997), a material clause decodes doings and happenings, including actions, activities, and events. In other words, the material process is an arrangement of a process and participants involved, who require some input of energy to occur, possibly undertaking a change and expressing energy flow.

The English version presents some deviances produced by the translator's intention to make ST meanings more explicit and accessible to English readers through the verbalisation of parts of the clause (i.e., *that attempts to create, who is monitored and controlled, are also controlled and guided, is created*), and through the use of material processes that give relevance to actions instead of abstract concepts. The TTs verbalise the Italian nominalisations in the initial part of the clauses (as already shown above) and, besides, contain more material processes than the STs, where, in contrast, words such as *monitorato, controllato, controllati, guidati, creata* have adjectival functions. In the English texts, words such as *create, monitored, controlled, experience, controlled, guided, created* have all a verbal function that is given by the use of "to be" with which the lexical verbs form passive structures. Besides, the number of relational processes (i.e., to be, to have) is higher in the STs than the TTs. The functions of the relational processes are not always identical (e.g., *Maria Teresa Setaro è curatrice e ricercatrice* vs. *Director of TRIAD, independent curator and researcher, Maria Teresa Setaro has an MA [...]*). In fact, the English construction, in the case in which the verb "to have" has been chosen instead of the verb "to be", presents the facts as if the position as a curator and researcher is the result of the degree in Contemporary Arts. The Italian text, on the contrary, presents two different clauses, apparently disconnected (as already shown above). There are no connectives that create a link between them. The first clause does not seem to be the result of the second one. In brief, in terms of accessibility, the English text is more accessible due to the modalities by means of which the syntax is organised (i.e., an initial subordinate nominal group followed by an independent main clause, where *has* functions as a relational possessive attribute process). As remarked above, relational processes are processes of being, relation and identification. The Italian text presents one relational intensive attributive process: (i.e., *è una performance collettiva*), one case of relational possessive identifying process (i.e., *M.T. Setaro è curatrice*), and one case of relational possessive attributive process (i.e., *ha un forte interesse*). This makes the Italian texts more relational and less material, that is, more based on identification mechanisms than doing and active features. Thus, on the representational level, the STs and TTs diverge, since meaning transfers put emphasis on different aspects of reality. This implies that the STs and TTs have different systems of communication – the Italian texts rely on a High Context Culture setting in opposition to the English texts that rely on a Low Context Culture framework. As a result, the English texts present a satisfactory level of accessibility in relation to what Systemic Functional Linguistics demonstrates.

5.2. Towards the Museum of Trust and Dialogue for the Mediterranean

In *Towards the Museum of Trust and Dialogue for the Mediterranean*, where our focus is on bilingual museum panels explicitly on migration issues, the investigation is conducted by means of the LSA approach, which makes it possible to evaluate the semantic texture of migration narratives, and how these narratives are made accessible in English through the maintenance of standards at the level of discourse structure, style and semantics. On the ideational metafunction level (i.e., the transitivity system in SFL), the six processes or verbs are categorised as material, verbal, mental, relational, behavioural, and existential. As already specified in the methodological section, the classification of verb types leads to the identification of lexical domains, which are defined by a superordinate term that can include the meaning of one or more other verbs, as shown in Table 10. The English museum panels contain approximately the same set of lexical domains (where each lexical domain is defined by a superordinate word referred to as the “nuclear” term).

Table 10.

Lexical Domains and Superordinate Categorisation According to Faber and Mairal Usón’s LSA (1999)

EXISTENCE	Be, Happen
CHANGE	Become
POSSESSION	Have
SPEECH	Say
EMOTION	Feel
ACTION	Do, Make
COGNITION	Know, Think
MOVEMENT	Move
PHYSICAL PERCEPTION	See, hear, taste, smell, touch
MANIPULATION	Use

The lexical semantic analysis is applied to five selected panels on migration, which significantly shed light on TTs’ cultural deviances from the STs. The selected panels have the following titles in English: (a) *The RAI and the Museum of Trust and Dialogue*; (b) *What Stays*; (c) *The Drawings of Shahrazad*; (d) *Towards the Museum of Trust and Dialogue for the Mediterranean Sea - Lampedusa*; (e) *Adal-Eritrean Refugee*. The panels that have been transcribed in Tables 11, 12, 13, 14 are those that contain the narratives b, c, d, and e. The transcription of the narrative in wall text panel (a) has been left out for its length. Nevertheless, all the five panels have been used for the qualitative analysis.

Table 11.

Wall Text Panels at Towards the Museum of Trust and Dialogue for the Mediterranean

(b) Quello che resta (SL)	Oggetti personali di 52 persone morte soffocate nella stiva di un barcone durante il viaggio nel Mediterraneo. Si tratta dello stesso episodio raccontato nel film “Fuocoammare” di Gianfranco Rosi – Rai Cinema. Grazie da Palermo, Squadra Mobile Palermo.
What Stays (TL)	Personal belongings of 52 people who died suffocated in the hold of a wooden barge during the journey in the Mediterranean. This is the same story told in the documentary film “Fuocoammare” (Fire at Sea) by Gianfranco Rosi – Rai Cinema. Thanks to DDA Palermo, Palermo Flying Squad.

Table 12.

Wall Text Panels at Towards the Museum of Trust and Dialogue for the Mediterranean

(c) I disegni di Shahrazad (SL)	Guerra, fuga, viaggi e muri d’Europa disegnati dalla bambina siriana incontrata ad Idomeni da Pierfrancesco Citriniti e Diego Bianchi in un reportage di Gazebo, Rai Tre.
The Drawings of Shahrazad (TL)	War, escape, travel and walls of Europe are drawn by the young Syrian kid that Pierfrancesco Citriniti and Diego Bianchi met at Idomeni camp during one of the Gazebo reportage. Rai Tre.

Table 13.

Wall Text Panels at Towards the Museum of Trust and Dialogue for the Mediterranean

(d) Verso il Museo della Fiducia e del Dialogo (SL)	Frontiera della frontiera, comunità di donne e uomini ospitali, Lampedusa si rafforza quale simbolo di accoglienza. La mostra, organizzata dal Comune di Lampedusa e di Linosa, promuove valori di pace e di responsabilità sociale che l’isola esprime. Le opere che costituiscono il corpo della prima esposizione, transito verso la realizzazione del Museo della Fiducia e del Dialogo per il Mediterraneo”, possono essere considerate quali strumenti per orientarsi nel “Mare delle radici comuni”.
Towards the Museum of Trust and Dialogue (TL)	The real border of Europe, Lampedusa, is the symbol itself of welcoming migrants. The exhibition, by Lampedusa and Linosa municipalities, promotes the value of peace and social responsibility which are typical of the island. The works on show can be considered as tools to find our bearings in “the sea of our common roots”.

Table 14.

Wall Text Panels at Towards the Museum of Trust and Dialogue for the Mediterranean

(e) Adal- rifugiato eritreo (SL)	Adal è un rifugiato eritreo, fratello di una delle vittime del naufragio di Lampedusa del 3 ottobre 2013. Con i suoi disegni Adal ha raccontato le torture inflitte a chi cerca di sottrarsi alla schiavitù del servizio militare a vita imposta in Eritrea. Adal è scappato attraverso il deserto e il mare verso l'Europa. Ha raggiunto Malta che lo ha rimandato indietro in Eritrea insieme a 250 ragazzi come lui, imprigionati per tradimento e torturati. Adal è sopravvissuto ed è riuscito a scappare per la seconda volta. Ora vive nel nord Europa. Quei disegni Adal li ha fatti per spiegare al mondo cosa succede nel suo paese: "Sono libero e vivo in un paese libero, posso parlare, adesso. E posso mostrare al mondo quello che succede". I disegni tracciati da Adal in una intervista con il TG2 sono diventati prova d'accusa, acquisiti dalla Commissione d'inchiesta per i Diritti Umani delle Nazioni Unite nella relazione che condanna il regime eritreo per crimini contro l'umanità.
Adal- Eritrean Refugee (TL)	Adal is an Eritrean refugee and brother of one of the victims of the Lampedusa shipwreck occurred on 3 October 2013. Through his drawing Adal told of tortures inflicted on those who try to escape from the bondage of a perpetual military service which is imposed in Eritrea. Adal ran away across desert and sea to Europe. He reached Malta from where he was sent back to Eritrea together with 250 young people like him and there, charged with treason, they were kept in detention and tortured. Adal survived and successfully escaped for the second time. Now he lives in Northern Europe. Adal's drawings were made to explain what is happening in his country: "I am free and live in a free country. I can talk now. I can show the world what is happening there." The images drawn by Adal interviewed by TG2 have become damning evidence. They have been acquired by the UN Commission of inquiry on Human Rights in the reports that condemns crimes against humanity committed by the regime.

The comparative analysis between the original texts and their English renderings demonstrates that both texts are dominated by and organised through a discursive structure that is ruled by material processes, followed by the relational type. STs and TTs have however substantial variations of processes in numeric terms. Table 15 contains the English process types and their occurrences in the five panels, Table 16 indicates the lexico-semantic domains and their occurrences in the five panels, and Table 17 provides the lists of both the Italian and English processes according to the LSA approach.

Table 15.

English Process Types in Towards the Museum of Trust and Dialogue

English verbs – SFL	
Material	Promote, die, separate, offer, unite (2), bring, carry (2), float, survive, meet, acquire, condemn, give, provide, rebuild, divide, lie, come (2), choose, pay a tribute, drown, accompany, pick up, occur, serve, restore, participate, welcome, escape (2), begin, paint, draw, run away, keep, torture, impose, show, swallow, combine, reach, send back, seek, charge, make, boil down, live (2), happen (2), grow.
Relational	Become (2), be (18)
Verbal	Explain, talk, tell (3), ask, recount
Mental	Consider, mean (2), touch, question

Table 16.

English Lexico-Semantic Domains in Towards the Museum of Trust and Dialogue

English verbs – Lexical domains	
EXISTENCE	Be (18), happen (2), lie
CHANGE	Become (2), grow, die
SPEECH	Explain, talk, tell (3), ask, recount
ACTION & MOVEMENT	Promote, separate, offer, unite (2), bring, carry (2), float, survive, meet, acquire, condemn, give, provide, rebuild, welcome, divide, come (2), choose, pay a tribute, drown, accompany, pick up, occur, serve, restore, participate, escape (2), begin, paint, draw, keep, run away, torture, impose, show, swallow, combine, reach, send back, seek, charge, make, boil down, live (2)
COGNITION	Consider, mean (2), question
PHYSICAL PERCEPTION	Touch

Table 17.

English and Italian Processes Signalled Through Their Lexico-Semantic Domains

English processes	Italian processes
Separate (action and movement), unite (2) (action and movement), be (18) (existence), divide (action and movement), touch (physical perception), bring (action and movement), carry (2) (action and movement), float (action and movement), lie (existence), question (cognition), rebuild (action and movement), provide (action and movement), give (action and movement), welcome (action and movement), offer (action and movement), pay a tribute (action and movement), drown (action and movement), become (2) (change), choose (action and movement), come (2) (action and movement), mean (2) (cognition), accompany (action and movement), pick up (action and movement), ask (speech), recount (speech), serve (action and movement), restore (action and movement), participate (action and movement), begin (action and movement), tell (3) (speech), paint (action and movement), swallow (action and movement), combine (action and movement), seek (action and movement), boil down (action and movement), die (change), draw (action and movement), meet (action and movement), promote (action and movement), consider (cognition), occur (action and movement), escape (2) (action and movement), impose (action and movement), run away (action and movement), reach (action and movement), send back (action and movement), charge (action and movement), keep (action and movement), torture (action and movement), survive (action and movement), live (2) (action and movement), make (action and movement), explain (speech), happen (2) (existence), talk (speech), show (action and movement), acquire (action and movement), condemn (action and movement), grow (change)	Rafforzare (cognition), promuovere (action and movement), esprimere (speech), costituire (action and movement) considerare (cognition), orientarsi (cognition), trattarsi (cognition), essere (12) (existence), toccare (physical perception), raccontare (speech), sottrarsi (cognition), scappare (2) (action and movement), raggiungere (action and movement), rimandare (action and movement), sopravvivere (action and movement), vivere (action and movement), fare (action and movement), spiegare (speech), succedere (change), parlare (speech), mostrare (action and movement), diventare (2) (change), condannare (action and movement), rispondere (speech), offrire (3) (action and movement), partecipare (action and movement), raccogliere (action and movement), iniziare (2) (action and movement), separare (action and movement), unire (2) (action and movement), dividere (action and movement), intendere (cognition), crescere (change), trasportare (action and movement), galleggiare (action and movement), raccontare (speech), dipingere (action and movement), esistere (existence), interrogare (speech), ricostruire (action and movement), venire (2) (action and movement), cambiare (change), inghiottire (action and movement), limitarsi (cognition)

According to Faber and Mairal Usón's classification, the conceptual areas activated through the use of material processes in the TTs, as shown in Tables 15 and 16, belong to the lexical domain identified as "Action and Movement". The English panels contain 49 verbs that belong to the Action-Movement lexical domain (also recognisable as material processes), 21 verbs that belong to the Existence lexical domain (also recognisable as relational processes), 4 verbs belonging to the Change lexical domain (also recognisable as material processes), 7 verbs belonging to the Speech lexical domain (also recognisable as verbal processes), 4 verbs belonging to the Cognition lexical domain (also recognisable as mental processes) and 1 verb belonging to the Physical Perception lexical domain (also recognisable as mental processes). The Italian texts (the lexico-semantic domains are reported in Table 17) have a reduced number of material processes and the lexical domain identified as "Action and Movement" consists of 29 verbs. Furthermore, what clearly emerges in the STs is a more frequent use of nominalisations than verbalisations, as well as of verbs used as past participles, adjectives and complements, which hide the direct message transmitted by the English structure composed of subject + verb phrase. Some examples follow. In panel (b): *Oggetti personali di 52 persone **morte soffocate** nella stiva di un barcone durante il viaggio nel Mediterraneo* vs. *Personal belongings of 52 people who **died suffocated** in the hold of a wooden barge during the journey in the Mediterranean*; in panel (e): [...] *alla schiavitù del servizio militare a vita **imposta** in Eritrea* vs. [...] *the bondage of perpetual military service **which is imposed** in Eritrea; **imprigionati per tradimento** e torturati* vs. *and there, charged with treason, **they were kept in detention and tortured***; in panel (a): *L'ascolto e la ricerca, il diretto avvicinamento al luogo fisico e simbolico, dove l'incontro è possibile* vs. ***It is** a way of listening and of searching, a way of drawing close to the physical and symbolic place where encounter is possible; **alla ricerca di** un racconto* vs. ***to seek** a narrative*; in panel (c): *bambina siriana **incontrata** ad Idomeni da Piefrancesco Citriniti and Diego Bianchi* vs. *the young Syrian kid that Piefrancesco Citriniti and Diego Bianchi **met** at Idomeni camp*).

In relation to target responses, verbal structures are generally preferred in this type of narratives in English, and the verb is generally preceded by a relative pronoun (i.e., *who died suffocated* vs. *persone morte soffocate*). Nevertheless, although Italian narratives of migration, marginalisation and minorities are constructed with fewer processes or verbs compared to the English ones, material processes also have the highest frequency in comparison to the other process types in the Italian narratives. In some cases, ST lexically dense nominal constructions are omitted in English either because they may be irrelevant to providing access to meanings (i.e., *comunità di donne e uomini ospitali*) or because they may be confusing due to the reiterations of the same Italian words within the same noun phrase (i.e., *La frontiera della frontiera* vs. *The real border of Europe*). For instance, the target nominal phrase *The real border of Europe* demonstrates how accessibility through interlingual translation can develop further cognitive processes, such as that of explicitly identifying Lampedusa as *the real border*, as the door to Europe, which reminds the spectator of the real door built by Mario Paladino on the coast of Lampedusa as a physical symbol of access to the island and the new world for the migrants. Further on, the English phrasing puts

emphasis on the migrants themselves, who are introduced through an indirect verbal structure which contrasts with the Italian construction (*Lampedusa si rafforza quale simbolo di accoglienza* vs. [...] *Lampedusa, is itself the symbol of welcoming migrants*).

Nominal lexical additions are also present in English in order to strengthen meanings that can be obscure to readers. For example, in the ST noun phrase *nella stiva di un barcone*, which is transformed into the TT noun phrase *in the hold of a wooden barge*, the addition of the adjectival term *wooden* expands the semantic content, which, in its turn, sheds light on the material conditions in which the migrants' journey takes place. The relational process (belonging to the lexical domain of Existence) in the English clause *This is the same story told in [...]* substitutes for a mental process belonging to the lexical domain of Cognition, which appears in the Italian clause *Si tratta dello stesso episodio*. In the English clause, the relational process does not involve a relation between entities, where the carrier has got some qualities of the attribute, instead, a relation is established between entities in terms of identification (a relationship of identity), where the identified *this* is the identity of the identifier *the story*. This clause is very significant at the level of intertextuality, and relevant in terms of identification with what is narrated in the panel and with what the Italian filmmaker Francesco Rosi narrated in his documentary. The significant ST intertextual reference to Rosi's documentary *Fuocammare* is in fact lexically reinforced in English through the translation shift from a mental process in Italian to a relational identifying process in English, and also by the addition of the English title of the documentary, *Fire at Sea*, which speakers of other languages may be familiar with, since the film was also screened with English subtitles and entitled *Fire at Sea*.

5.3. eMMMMe – Porto M – Lampedusa

In the case of the *Porto M* multilingual multimedia guide, subtitling norms are not respected and subtitles function as spaces of re-narration of SL speech. Since the content of the video is not only expository of the anti-museum exhibits, but is imbued with political meanings, the subtitling activity is based upon the transfer of SL messages according to what is considered necessary or useful to learn about for a non-Italian audience. The TL subtitles are often manipulated in terms of linguistic constructions and lexical choices. Some words or phrases have been either omitted or reconstructed. The subtitles have been transformed in terms of technical and linguistic dimension into autonomous depositories of counter information on screen. Therefore, the technical dimension (i.e., the spatial and temporal considerations such as the number of words and lines in one subtitle, synchronization, etc.) and linguistic dimension (i.e., the segmentation of sentences, semantically self-contained subtitles, the respect of syntax in line breaks) are not arranged according to the norms of standard subtitling, but are the result of translation solutions, principally aiming at giving voice to migrant narratives that counter argue mainstream news reports.

A qualitative analysis based on the methods of SFL and LSA has been applied to the subtitles that scroll in the first 4.25 minutes (out of 11 minutes in total) in the multilingual multimedia guide for a general adult English-speaking public. The study is limited to the first part of the video, since it is the part that puts emphasis on the experiences of the migrants and gives indirect voice to their journeys. What is demonstrated is that materiality is the dominant lexical semantic area in both the Italian and English narratives. The material processes (in Italian 27, in English 30) and the lexical semantic domain of “Action and Movement” activate conceptual areas that stress the role of material objects and visual images as providing access to systems of exchange, and to the stories of marginalised groups, by privileging language and text as the essential ground for the mediation of experience and knowledge communication. The meanings transferred through the choice of processes and the activation of specific lexical categories bring social, cultural and material topics into meaningful relations and permit the audience to establish comparisons and recognitions. In this sense the English subtitling uploaded for the website of *Port M* can be viewed as an activist space of translation, a space where translation becomes interpretation, transformation, displacement and agency – all translation forms meant to provoke action and intervention and to produce a particular effect and result in terms of reversal of perspectives. The ST speaker/maker/producer becomes the TT translator/curator/circulator, who transmits TT viewers/users/consumers non-hegemonic narratives from a material and physical perspective, which strengthens the idea of migrants as human beings who can potentially subvert conditions of power. The “material” narration of migrant experiences (i.e., the journey itself and the quantity and quality of everyday objects that each migrant brings from his/her country) puts them in a state of activism, fight and hope. Ontological narratives become accessible through the subtitled multilingual multimedia guide which allows viewers to open up their own perspectives. Accessibility is thus conceived as potentially of help in challenging the established political and economic systems in order to affect the whole system and push for change.

Tables 18 and 19 provide the transcription of SL speech and TL subtitles according to the following criteria: the verb groups in bold indicate material processes, the words and phrases in italics signal Italian nominalisations vs. English verbalisations, the underscored words and phrases signal omissions in English translations, those both underscored and in Italics indicate English omissions and verbalisations, and those both underscored and in bold signal English additions.

Table 18.

Analytical Transcription of the Oral Speech and Subtitles in the Porto M Multilingual Multimedia Guide (Part 1)

Italian transcript – <i>Porto M</i>	English subtitles – <i>Porto M</i>
Questo è Porto M lo <i>spazio espositivo</i> con gli oggetti dei migranti	This is Porto M, <i>the space where the objects of the migrants [...] are exhibited</i>
che il Collettivo Askavusa ha raccolto negli anni.	that the Askavusa organization collected over the years.
Gli oggetti sono stati raccolti nelle varie <u>discariche</u> dell'isola.	All the objects have been collected from the boats cemetery.
Le imbarcazioni sono sempre state trattate come spazzatura.	Those objects are treated like rubbish,
Ci sono oggetti di diverso genere – del cibo, farmaci, cosmetici.	There are different kind of objects: kitchen-stuff, drugs and cosmetics.
<i>L'essenziale per affrontare</i> un viaggio.	The essential objects one <i>takes</i> for a journey.
Abbiamo raccolto migliaia e migliaia di oggetti che inizialmente abbiamo conservato nelle nostre case,	We collected thousands of objects. At the very beginnings we kept them inside our houses.
non c'era un'idea iniziale su cosa si voleva fare effettivamente con questi oggetti.	as we had no idea about what to do with all these items.
Come si può vedere gli oggetti sono semplicemente esposti .	As you can see the objects are simply exposed .
Non era quello il nostro intento, di chiudere gli oggetti come avviene nella forma museale,	We didn't have the intent of keeping them like you would do in a museum.
quindi datandoli , o descrivendoli o chiudendoli appunto in delle teghe.	We don't keep them inside showcase
	and there's no description nor dates relative to the objects.
Questi che vedete sono gli ultimi che provengono dai cimiteri delle barche,	Those objects are the last ones from the old boats cemetery.

Table 19.

Analytical Transcription of the Oral Speech and Subtitles in the Porto M Multilingual Multimedia Guide (Part 2)

Italian transcript – <i>Porto M</i>	English subtitles – <i>Porto M</i>
[...] a <u>gennaio</u> sono stati distrutti, e portati via come spazzatura.	All the boats were demolished and taken away from the island like rubbish.
Chi arriva sulle nostre coste, fa anche lo stesso <u>percorso</u> degli oggetti.	It is almost the same as what happens to those migrants that arrive here,
<u>Usati, sfruttati come profitto, no. Quindi gli oggetti nelle discariche,</u> i migranti nei vari centri.	<i>They become a source of profit</i> for those managing migrants' centres.
<u>Ancora una volta sfruttati e usati come profitto perché i barconi e gli oggetti che sono arrivati sono usati</u>	So the boats and the objects <i>are commodified</i>
come profitti per chi ha preso <u>milioni di appalti</u> per farli diventare spazzatura.	to benefit those companies that are in charge of demolishing them.
Lo stesso <u>percorso</u> che fanno anche i migranti,	The same happens to migrants
una volta <u>sfruttati</u> all'interno dei centri,	as <i>they are exploited</i> in the centres to be abandoned.
nei casi peggiori, <u>rimandati</u> nei propri paesi,	The worst that can happen to them is that <i>they are sent back</i> to their
o, nei casi migliori, <u>lasciati abbandonati</u> per strada, <u>nelle campagne</u> , senza alcun diritto.	countries. The "best" that can happen to them is that <i>they are abandoned</i> in the streets without any rights.

The cases of Italian nominalisations vs. English verbalisations include the following phrases: *spazio espositivo* vs. *are exhibited*; *l'essenziale per affrontare il viaggio* vs. *one takes on a journey*, *Usati, sfruttati come profitto, no* vs. *They become a source of profit*; *una volta sfruttati all'interno dei centri* vs. *as they are exploited in the centres*; *rimandati nei propri paesi* vs. *they are sent back to their countries*; *lasciati abbandonati per strada, nelle campagne, senza alcun diritto* vs. *they are abandoned in the streets without any rights*; *nei casi peggiori* vs. *The worst that can happen*; *nei casi migliori* vs. *The "best" that can happen*. Omissions or lexical variations in English have often taken place due to redundancy, impenetrability and inconsistency, or due to the fact that the translator may have evaluated certain terms as useless as far as the accessibility of niche contents is concerned. In some cases, the subtitler may have thought that expressions such as *Askavusa in lampedusano, nelle varie discariche* vs. *boats cemetery*; *milioni di appalti, a gennaio, nelle campagne* may be confusing and unnecessary, adding no further meanings to the core message to be conveyed in English. The concept of *discarica* ("landfill" in English), for instance, is not literally rendered in English. The metaphorical nominal expression, *boat cemetery*, also popular in Italian, has been chosen since it is a noun phrase the English audience may be familiar with. In fact, it has frequently appeared in English newspapers such as *The Guardian*, and in the Internet, and it is also commonly selected as a nominal phrase in British studies and artistic projects, such as those led by Maya Ramsay (*Countless*, 2016–2017) and Lucy Wood (*Boat Fragments*, 2011, and *TO6411*, 2013). Passive constructions have also been commonly used and appear to be frequently selected in these types of narratives, both in Italian and English, together with informal expressions, such as markers of explicitness and referencing systems that directly introduce personal pronouns and reduce the distance between writers and readers (i.e., *come avviene nella forma museale* vs. **you** *would do in a museum*; *quindi datandoli, o descrivendoli o chiudendoli appunto in delle teghe* vs. **We** *don't keep them inside showcase, and there's no description nor dates relative to the objects*). In search of clarity, the English subtitles move from impersonal general information to explicit formulations that give the viewer access to solutions, such as in the case of what procedures are adopted in that type of museum and what the Askavusa collective actually refuses to do (i.e., *We don't keep them inside showcase*). The referencing system in English also points out important elements in the sentences and strengthens the access to information, such as in the following cases that are present in two different subtitles: (i.e., **Those objects are the last ones from the old boats cemetery.**/**All the boats were demolished**). In these sample cases, the nominal groups a) *Those objects*, b) *the last ones*; c) *boats*, and d) *All the boats*, are embedded within a system of anaphoric references which strengthens the cohesion between the words and phrases *objects*, *ones*, *boats* and *All the boats*. The issue concerning the violation of human rights is what the English subtitler puts emphasis on. In fact, if, on the one hand, expressions such as *discariche* or *milioni di appalti* are omitted, thus, deliberately leaving out important questions about transparency and political corruption, on the other hand, emphasis is put on the notion of human rights through meaningful lexical and grammatical additions (i.e., the passive verbal constructions *to be abandoned*) that contribute to shedding light on the isolation and abandonment of the migrants left in the middle of nowhere with no rights at all.

6. Conclusions

Results show that marginalised narratives can be accessible through the intervention of English as the dominant language within the context of the visual arts, which act as instruments of niche knowledge dissemination. It has emerged that accessibility works as a form of translation and vice versa, and that the arts themselves are devices of accessibility. Museum narratives of marginalisation, migration, and minorities in their English renderings seem to respect and follow specific criteria that chiefly aim to satisfy the principle of directness vs. indirectness for clarity's sake and through the use of verb phrases rather than nominal ones. Pivotal to this objective is a number of linguistic norms that are traceable when scrutinising these artistic narratives from the perspective of translation practices, where the observed linguistic features can be regarded as the beginning of a local grammar adopted for interlingual translation and subtitling in museum contexts of marginalisation, migration, and minorities. The target audience's familiarity with certain linguistic and syntactic structures has influenced the translator's choices in relation to lexical choices and syntactic constructions used in the TTs under scrutiny. In fact, the spectator of the (audio)visual narratives taken into account has been identified as a heterogeneous adult public – not necessarily, and rarely, a native English speaker, but certainly a viewer and listener who uses English as a *lingua franca*. The identification of a type of local grammar, which, in this particular context, is the choice of a set of words employed in specialist domains (i.e., marginalisation, migration, minorities) for the construction and translation of text types, is structured around the conceptual activations of verbs. Each verb in the lexicon activates a determined context and sequence of events with definite semantic participants. This implies that each conceptual category involves a precise meaning area, whose activation depends on the selection of words belonging to the same lexical domain (Jiménez Hurtado & Soler Gallego, 2015).

It can be maintained that among the strategies involved in the process of translating/subtitling aesthetic narratives of marginalisation are lexical omissions and additions, passive constructions that state facts and events objectively, active sentences (subject + verb + object structures) in opposition to impersonal clauses, directness vs. indirectness strategies, and verbalisation instead of nominalisation processes. The preference for verbs instead of nouns, which is the core of these narratives, is strictly connected with the urgency to place them within a dynamic and active dimension. This helps disseminate representations of migrants' experiences of material facts, events and circumstances relating to their personal lives in order to deconstruct mass-media narratives and officially sanctioned news reports that depict these people as dangerous hordes.

The different levels of analysis have put emphasis, first of all, on the identification of the main processes within the SFL framework of Transitivity by means of which light has been shed on the translator's selected verbs (and their occurrences) as the principal semantic categories, which are meant to describe events, experiences and facts. Secondly, results have shown that the most frequent conceptual area activated by the lexico-semantic analysis is the "Action and Movement" domain, which involves the material verbal dimension (especially in *Towards*

the Museum of Trust and Dialogue and *Porto M*). The adoption of verb-based phrases rather than noun-based phrases has appeared to be fundamental and, in particular, the dominant presence of material processes (in the three subcorpora in English and Italian), which describe what the subject of the sentence is doing, or is expected to do, has carried out a great deal of information that serves to give a sense of completeness to the sentence. The LSA approach has led to the identification and specification of a local text grammar (i.e., a specific type for marginalised narratives), which involves the functional description of a set of lexical patterns that represent a function within the context of a certain text type. Such grammar has provided sets of recurrent lexical patterns, syntactically and semantically unvaried, which also share the same communicative functions, and have described small items looked at as sublanguages, which have transmitted specific contents and made use of recurrent linguistic patterns related to specific speech actions.

Although deviances between the two language and culture systems are present, TQA based on SFL and LSA methods has highlighted the general predominant presence of material processes and “Action and Movement” lexical domains in both STs and TTs. However, the target texts have been consistently characterised by a higher number of material processes compared to the Italian ones. Access to narratives of migration and marginalisation as counter information to conventional news media has taken place through interlingual translation and subtitling procedures, which have contributed to developing TTs aiming to adequately and successfully transmit niche knowledge across channels of international diffusion.

New views on accessibility as a universalist concept for the respect of human rights in contemporary societies can potentially improve the quality of life, social cohesion and inclusion of minorities, develop competences in different languages, and encourage the creation of a more equal society.

Acknowledgements

I am very grateful to one of the anonymous referees for the invaluable contribution in terms of suggestions, stimuli, and conceptual additions. The paper has deeply benefited from his/her profitable comments on a theoretical and practical level. I am also greatly indebted to Gian Maria Greco for helping me refine my critical thinking and for trusting my work from day one. This has made true collaboration and exchange feasible – something that is essential to delivering valuable academic research. Without his encouragement and that of Anna Jankowska, this paper would not have been possible.

References

- ADLAB. (2011–2014). Retrieved from <http://www.adlabproject.eu/home/>.
- AENOR. (2012). *Norma UNE 153010 Subtitulado para personas sordas y personas con discapacidad auditiva* [Subtitling for deaf and hard-of-hearing people]. Madrid: AENOR.

- Amir, E. (2018). *Enough about you*. Retrieved from <http://m12.manifesta.org/collateral-event/enough-about-you-a-performance-by-einat-amir/?lang=it>; <http://artis.art/grantees/einat-amir-enough-about-you/>
- Baker, M. (2016). Beyond the spectacle: Translation and solidarity in contemporary protest movement. In M. Baker (Ed.). *Translating dissent voices from and with the Egyptian revolution* (pp. 1–18). London & New York: Routledge.
- Baker, M. (2019). Audiovisual translation and activism. In L. Pérez González (Ed.). *The routledge handbook of audiovisual translation* (pp. 453–467). London & New York: Routledge.
- Braun, S. (2008). Audio description research: state of the art and beyond. *Translation Studies in the New Millennium*, 6, 14–30.
- Bruti, S., & Di Giovanni, E. (2012). *Audiovisual translation across Europe: An ever-changing landscape*. Peter Lang: Bern.
- Buser, M., & Arthurs, L. (2013). Connected communities: Cultural activism in the community, 1–11. Retrieved from <http://www.culturalactivism.org.uk/wp-content/uploads/2013/03/CULTURAL-ACTIVISM-BUSER-Update.3.pdf>
- Díaz-Cintas, J. (2007). Traducción audiovisual y accesibilidad [Audiovisual translation and accessibility]. In C. Jiménez Hurtado. *Traducción y accesibilidad. Subtitulación para sordos y audiodescripción para ciegos: Nuevas modalidades de traducción audiovisual* [Translation and accessibility. Subtitling for the deaf and audiodescription for the blind] (pp. 9–23). Bern: Peter Lang.
- Díaz-Cintas, J., Orero, P., & Remael, A. (2007). *Media for all: subtitling for the deaf, AD, and sign language*. Amsterdam: Rodopi.
- Díaz-Cintas, J., Matamala, A., & Neves, J. (2010). *New insights in audiovisual translation and media accessibility: Media for all 2*. Amsterdam: Rodopi.
- Díaz-Cintas, J. & Nikolić, K. (2018). Fast-forwarding with audiovisual translation. In J. Díaz-Cintas, & K. Nikolić, (Eds.). *Fast-forwarding with audiovisual translation*. (pp. 1–14). Bristol: Multilingual Matters.
- Di Giovanni, E. (2018). Audio description for live performances and audience participations. *Journal of Specialised Translation*, 29, 189–221. Retrieved from https://www.jostrans.org/issue29/art_digiovanni.php
- European Economic and Social Committee. (2014). Accessibility as a human right for persons with disabilities. Brussels.
- Faber, P., & Mairal Usón R. (1999). *Constructing a lexicon of English verbs*. Berlin: Mouton de Gruyter.
- Folaron, D. (2015). Translation and minority, lesser-used and lesser-translated languages. *Journal of Specialised Translation*, 24, 16–27. Retrieved from https://www.jostrans.org/issue24/art_folaron.php
- Gambier, Y. (2006). Multimodality and audiovisual translation. In M. Carroll, H. Gerzymisch-Arbogas, & S. Nauert (Eds.). *Audiovisual translation scenarios. Proceedings of the Marie Curie Euroconferences MuTra: Audiovisual translation scenarios*, Copenhagen 1–5 May 2006. Retrieved from https://www.euroconferences.info/proceedings/2006_Proceedings/2006_Gambier_Yves.pdf
- Greco, G. M. (2016). On accessibility as a human right, with an application to media accessibility. In A. Matamala & P. Orero (Eds.). *Researching AD. New approaches* (pp. 11–34). London: Palgrave Macmillan.

- Greco, G. M. (2017), L'accessibilità culturale come strumento per i diritti umani di tutti [Cultural accessibility as an instrument for the human rights of all]. In G. Cetorelli & M. R. Guido (Eds.), *Il patrimonio culturale per tutti [Cultural heritage for all]* (pp. 94–102). Rome, Italy: Direzione Generale Musei, Ministero dei Beni e delle Attività Culturali e del Turismo.
- Greco, G. M. (2018). The nature of accessibility studies. *Journal of Audiovisual Translation*, 1(1), 205–232.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Arnold: London.
- Halliday, M. A. K., with Matthiessen, C. (2004). *An introduction to functional grammar*. London: Edward Arnold. ICOM (International Council of Museums). Retrieved from <https://icom.museum/en/activities/standards-guidelines/museum-definition/>
- Jankowska, A. (2015). *Translating audio description scripts: translation as a new strategy of creating audio description*. Frankfurt am Main: Peter Lang.
- Jiang, C. (2010). Quality assessment for the translation of museum texts: application of a systemic functional model. *Perspectives: Studies in Translatology*, 18(2), 109–126.
- Jiménez Hurtado, C., Seibel C., & Soler Gallego, S. (2012). Museum for all. Translation and interpreting for multimodal spaces as a tool for universal accessibility. *MonTI. Monografías de Traducción e Interpretación*, 4, 1–24.
- Jiménez Hurtado, C., & Soler Gallego, S. (2015). Museum accessibility through translation: A corpus study of pictorial audio description. In J. Díaz-Cintas, & J. Neves (Eds.). *Audiovisual translation taking stock* (pp. 277–298). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Kavanagh, G. (1991). *Museum languages: Objects and texts*. Leicester & New York University Press.
- Lee, H. K. (2011). Participatory media fandom: A case study of anime fansubbing. *Media, Culture and Society*, 33(8), 1131–1147.
- Manifesta 12 – Palermo. (2018). Retrieved from <https://manifesta.org/biennials/about-the-biennials/>; <https://www.thethinkingtraveller.com/thinksicily/guide-to-sicily/general-information/manifesta-12-palermo-2018.aspx>
- MFP (2012). *Le sous-titrage de référence pour sourds et malentendants* [Subtitling for the deaf and hard-of-hearing people]. Paris: MFP.
- Maszerowska A., Matamala, A., Orero P., & Reviere, N. (2014). Introduction. From source text to target text. The art of AD. In A. Maszerowska, A. Matamala, P. Orero, & N. Reviere, (Eds.) *AD. New perspectives illustrated*. (pp. 1–10). Amsterdam and New York: John Benjamin.
- Mayoral, R., Dorothy, K., Natividad G. (1988). Concepts of constrained translation. Non-linguistic perspectives of translation. *Meta*, 33(3), 356–367.
- Mazzara, F. (2019). *Reframing migration: Lampedusa, border spectacle and the aesthetics of subversion*. Peter Lang: Bern.
- OFCOM. (2016). Retrieved from https://www.ofcom.org.uk/data/assets/pdf_file/0020/97040/Access-service-code-Jan-2017.pdf
- Orero, P. (2005). La inclusión de la accesibilidad en comunicación audiovisual dentro de los estudios de traducción audiovisual [The inclusion of accessibility in audiovisual communication within audiovisual translation studies]. *Quaderns. Revista de traducció*, 12, 173–185.
- Orero, P., & Matamala, A. (2007). Accessible opera: Overcoming linguistic and sensorial barriers. *Perspectives: Studies in Translatology*, 15, 262–277.
- Perego, E. (2012a). *Emerging topics in Translation: AD*. Trieste: EUT Edizioni Università di Trieste.
- Perego, E. (2012b). *Tradurre l'audiovisivo*. Roma: Carocci.

- Pérez-González, L. (2013). Co-creational subtitling in the digital media: Transformative and authorial practices. *International Journal of Cultural Studies*, 16(1), 3–21.
- Pérez-González, L. (2014). *Audiovisual translation. Theories, methods and issues*. London & New York: Routledge.
- Persson, H., Ahman, H., Yngling A. A., & Gulliksen, J. (2015). Universal design, inclusive design, accessible design, design for all: different concepts-one goal? On the concept of accessibility. *Universal Access in the Information Society*, 14,4, 505–526.
- Porto eMMMMe. (2013/2014). Askavusa Collective. Retrieved from <https://portomaremediterraneomigrazionimilitarizzazione.wordpress.com>
- Ramsay, M. Countless. (2016/2017). [Shipwrecked migrants' boat]. Retrieved from <https://www.mayaramsay.co.uk/work.php?s=countless-boats>
- Ravelli, L. J. (2006). *Museums texts. Communicative frameworks*. London and New York: Routledge.
- Remael, A., Orero, P., & Carroll, M. (2012). *Audiovisual translation and media accessibility at the crossroads, Media for all 3*. Amsterdam: Rodopi.
- Reviere, N. (2016). Audio description services in Europe: An update. *The Journal of Specialised Translation*, 26, 232–247. Retrieved from https://www.jostrans.org/issue26/art_reviere.pdf
- Rizk, P. (2013, May 20). Interview with Philip Rizk by Shuruq Harb. *ArtTerritories*. Retrieved from <https://monabaker.org/post/interview-with-philip-rizk-by-shuruq-harb/>
- Rizzo, A., & Seago, K. (2018). Translating the margin: Lost voices in aesthetic discourse. *InVerbis*, 1, 7–34.
- Romero-Fresco, P. (2019). *Accessible filmmaking. Integrating translation and accessibility into the filmmaking process*. London & New York: Routledge.
- Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation*, 20, 201–223. Retrieved from https://www.jostrans.org/issue20/art_romero.pdf
- Rosa, A. A. (2010/2016d). Descriptive translation studies – DTS (revised version). In Y. Gambier & L. van Doorslaer (Eds.). *Handbook of translation studies* (pp. 94–104). Amsterdam: John Benjamins. Retrieved from <https://repositorio.ul.pt/bitstream/10451/28112/1/Rosa-2016d-Descriptive%20Translation%20Studies%20-%20DTS%20%28revised%20version%29.pdf>
- Santacana, M., Serrat Antolí, J. & N. (cords.). (2005). *Museografía didáctica* [Museum teaching approach]. Barcelona: Ariel.
- Selim, S. (2016). Text and context. Translating in a state of emergency. In M. Baker (Ed.). *Translating dissent. Voices from and with the Egyptian revolution* (pp. 77–87). London/New York: Routledge.
- Szarkowska, A., Krejtz, I., Krejtz, K., & Duchowski, A. (2013). Harnessing the potential of eyetracking for media accessibility. In S. Grucza, M. Płużyczka, & J. Zajac (Eds.). *Translation studies and eye-tracking analysis* (pp. 153–183). Frankfurt am Main, Germany: Peter Lang.
- Toury, G. (1995). *Descriptive translation studies and beyond*. Amsterdam/Philadelphia: John Benjamins.
- Towards the Museum of Trust and Dialogue for the Mediterranean – Lampedusa (Verso il Museo della Fiducia e del Dialogo per il Mediterraneo–Lampedusa)*. (2016). Retrieved from <https://www.comitatotreottobre.it/museo-della-fiducia-e-del-dialogo-per-il-mediterraneo/>; <http://musei.beniculturali.it/progetti/lampedusa-verso-il-museo-della-fiducia-e-del-dialogo-per-il-mediterraneo>
- Wood, L. (2011). Boat Fragments. [Acrylic on wood]. Retrieved from <http://lucywood.net/ongoing/part2/boatfragments.html>.

Wood, L. (2014). TO6411 South Dock Marina Refugee Migrant Boat. [Refugee boat sailed solo from Lampedusa to London via Italy, France, Belgium, The Netherlands and UK. Time Capsule/Performance Piece]. Retrieved from https://artmap.com/profile/lucy-wood?print=do#_I5h0d.

Template Files: The Holy Grail of Subtitling

 Panayota Georgakopoulou 

Athena Consultancy

Abstract

The use of English template files in the creation of multilanguage subtitles from the same source audio assets, typically English, was one of the greatest innovations in the subtitling industry at the turn of the century. It streamlined processes, eliminated duplication of work, reduced direct costs, improved timelines and facilitated the quality control of large volumes of subtitle files whilst expanding the pool of available translators to complete the work (Georgakopoulou, 2006). Template files became the cornerstone for the globalisation of the subtitling industry and, almost two decades since their inception, they are still a topic of debate among language service providers and subtitlers. The present paper is a descriptive work, presenting a set of guidelines originally devised by the present author at the turn of the century, and used in practice for almost a decade in the production of multilingual subtitling work, enhanced and improved upon by the very subtitlers that worked with them. It aims at providing a reference point for debate to researchers in the subtitling field, so as to further the ongoing discussion on interlingual subtitling quality, practices and standards.

Key words: subtitling, templates, norms, quality, audiovisual localisation, language service providers.

Citation: Georgakopoulou, P. (2019). Template files: The Holy Grail of subtitling. *Journal of Audiovisual Translation*, 2(2), 137–160.

Editor(s): G.M. Greco & A. Jankowska

Received: September 19, 2019

Accepted: December 6, 2019

Published: December 31, 2019

Copyright: ©2019 Georgakopoulou. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. Introduction

Template files have dominated the globalised audiovisual localisation industry since the DVD era, starting in the mid '90s and used extensively as of the turn of the century. Being a young PhD researcher at the time of the DVD boom and eager to implement my learnings in DVD localisation, I decided to focus on subtitling templates, which I considered the Holy Grail of the subtitling industry (Georgakopoulou, 2005). “Holy Grail” because templates helped Language Service Providers (LSPs) better manage multilanguage subtitling production, by enabling efficient quality assurance across files in multiple languages, while also cutting down production times and cost, and at the same time expanding the pool of freelance translators available to work on such files by simplifying their training (Georgakopoulou 2006, 2009). “Holy Grail” also because of their elusive quality, as creating a template file suitable for translation into any given language would require substantial additional considerations to subtitling in a single language for a hearing audience.

In fact, the quality and suitability of subtitling templates for a given project or language has been at the heart of the debate between subtitlers and LSPs ever since their inception (O’Hagan, 2005, p. 10; Georgakopoulou, 2012, p. 81; Artegiani & Kapsaskis, 2014, pp. 420–421). Despite the intense dialogue between practitioners and agencies in the industry, rather surprisingly, there has been limited research on the topic of templates in the academic community. Kapsaskis (2011) and Nikolić (2015) wrote on the repercussions of the template phenomenon that dominated the industry in the first decade of the 21st century, whilst Artegiani and Kapsaskis (2014) also performed a qualitative analysis of the impact of template files on translation decisions, but not much else has been published to date. This is justified to an extent, due to the fact that template guidelines are typically part of internal company documentation and not publicly available. As template files have been serving the purpose of an intermediary file to aid in multilanguage subtitle production, they did not become a deliverable requested by content owners until recently. Hence, it is not surprising that not many such style guides on templates can be found online, with the notable exception of the guide supplied by Netflix (n.d., a).

After two decades of English-audio material dominating the multilanguage production chain at large LSPs, who have been catering to the need for centralisation, security and quality assurance that major content owners demand, the industry is now shifting to non-English language production, again for a multilingual audience. This is bound to raise questions once again regarding the appropriateness and suitability of workflows employed for such work, especially where these are tailored around the use of pivot English source template files for the translation in any and all languages required in a project, irrespective of its source audio language.

Having worked on the first published set of template subtitling guidelines (Georgakopoulou 2003), which were enriched and used for approximately a decade by professional subtitlers, I think it is appropriate to re-publicise them, accompanied by notes and examples collected over the years, so they can be referenced, discussed and questioned by researchers in the subtitling field,

as the industry is moving on to a new, exciting era that is bound to question a lot of established subtitling norms. This paper is, as a result, written in a descriptive manner, showcasing a representative example of an industry practice that still holds strong today. It is not meant to be an analysis of the practice of template writing, a justification of the rules it presents, or contrasted with recent research on subtitling norms and conventions, as it would be impossible to do so within the word limit of this article. Rather, the intention is for it to be used as a reference point in future discussions on current practices, so as to enable comparisons and the exploration of trends, as well as help shape new practices in an informed manner.

2. The History of Template Files

Template files first came about as a result of the need for multilanguage subtitle production for the same source material. The first traces of a template-based workflow in subtitling were found among subtitle files in Scandinavian languages, as I concluded from my PhD research (Georgakopoulou 2003, pp. 246–250): subtitling providers catering to Scandinavian audiences often used the Swedish file they created as the basis for the creation of files in other Scandinavian languages, largely retaining the subtitle segmentation and timing of the original Swedish files. The fact that subtitling traditions among Scandinavian countries were very similar certainly must have helped in this practice. The creation and use of spotting lists as templates by subtitling houses in workflows involving multiple subtitle stream creation was employed even before the DVD era, e.g., in corporate multilingual workflows (Bywood, 2019). Nikolić (2015) provides a useful analysis of the various types of template files in use in the industry today, many of which are created on the basis of the same workflow logic.

With the advent of the DVD, LSPs had to deal with the demands of content owners for large volume subtitling work in many languages simultaneously, reduced turnaround times for the completion of said projects, cost concerns, as well as security issues because of fears of piracy (Georgakopoulou, 2006). The template methodology served all these purposes, as it helped achieve shorter project turnaround times through the reuse of subtitle timing information from source template files and also reduced the creation cost for the language versions required per project. Templates also helped address the issue of multilanguage quality control (QC) of subtitle files from experts, typically linguists, who however did not necessarily speak all the target languages involved in a project, but who were fluent in the source English language and typically one or two other languages as well. The need for standardisation in multilanguage subtitle production, as well as security concerns were addressed with the centralisation of work in the hands of a handful select localisation service providers primarily located along the London – Los Angeles axis (Georgakopoulou, 2003, pp. 208–209; Carroll, 2004).

As no guidelines for the creation of template files existed at the dawn of the DVD era, LSPs naturally experimented with different styles of subtitling until each created their own internal guidelines.

Some used English Subtitle files for Deaf and Hard-of-hearing (SDH) as the basis, under the influence of ITC's (Independent Television Commission, 1999) subtitling guidelines, since the LSPs' headquarters were located in the UK and, as a result, often staffed by subtitlers trained in UK's hard-of-hearing subtitling standards. Eliminating hard-of-hearing elements from English subtitle files was the obvious thing to do to turn them into template files, and subtitle timings also needed to be adjusted so that sync would be with the dialogue rather than other sounds or shot changes, mimicking the style of translated subtitles that were the norm in traditionally subtitling countries.

Influenced by my PhD research and on the receiving end of subtitle template files for translation into Greek for a while, I embarked on drafting guidelines for subtitle template creation that our team of subtitlers would follow at the company I was working at the time, so that subtitle translation into other languages could be facilitated. I had read the *Code of Good Subtitling Practice* in Ivarsson's and Carroll's seminal book *Subtitling* (1998) and I was investigating dialogue reduction from audio to the subtitles in several European languages as part of my PhD, before the use of template files became the norm. I was interested in finding the golden mean so as to produce a template file that would be agnostic, a fair compromise among country-specific subtitling styles, which could easily be tailored to the norms and needs of each country if needed. I was also fortunate to be surrounded by colleagues from a good variety of language backgrounds and thus had the opportunity to debate template creation guidelines and their appropriateness for translation into Swedish, Norwegian, Portuguese, Dutch, Greek, Bulgarian and several other languages, thus indirectly taking into consideration a good representation of European subtitling country standards. The result was a set of guidelines that became part of my PhD thesis (Georgakopoulou, 2003), which were presented at conferences and later also published (Georgakopoulou 2010), but more importantly used in practice, amplified by the various subtitlers that worked with them, resulting in thousands of subtitle files distributed internationally.

3. English Template Guidelines

The present style guide builds on those earlier guidelines and includes further detail, especially as regards timing and reading speed issues, to exemplify the best practices followed for template file creation by a multilanguage subtitle service provider at the turn of the century. As traditional norms for subtitle (and thus template) file creation are questioned (Perego, del Missier, Porta, Mosconi, 2010; Krejtz, Szarkowska, & Krejtz, 2013; Szarkowska & Gerber-Morón, 2018), the present guide can provide topical information to serve as example in current discussions and comparative research on the topic.

An English template file is a subtitle file in English, meant to be used as the basis for translation into other languages, which may follow different linguistic conventions. As such, it seeks a compromise between different national styles of subtitling and is meant to aid the translator in focusing on the translation aspect of his/her job without being distracted by the timing of individual subtitles.

English template files are typically used in work involving English source audio assets, however they have also extensively been used as a pivot language in the case of non-English language pair combinations; this style guide addresses the creation of files from English source audio assets.

Although it is possible that template files are amended for timing at translation stage to best fit the chosen translation, this style guide assumes a strict template workflow, i.e., one in which translators are not allowed to manipulate the timings or the number of subtitles in the file they are translating, which are to remain identical in all language versions created out of a single template file, thus facilitating quality control and highlighting the importance of a well-crafted template file.

The style guide is meant to inform the practice of experienced subtitlers, i.e., people that have already mastered the art of timing and transcription of dialogue from audio, and who are familiar with various subtitling or captioning styles and guides, such as Ofcom's Code on Television Access Services (2017), the BBC Subtitle Guidelines (2018), the Captioning Key (DCMP, n.d.) and EBU-Tech 144-2004 (2004). As such, settings like number of lines per subtitle (typically two), number of characters per line (the file format required may impose its own limitations, such as 32 characters for closed captions and 37 for teletext, while 42 is most frequently used for open subtitles), font colour, style and size (white fonts without serifs are typically used, such as Arial), etc. are not explained or discussed further. The latter vary per project anyway, whereas the guidelines below are written so as to be valid irrespective of what these settings may be.

Finally, this style guide is written in a prescriptive manner but its rules are not absolute. It is understood that there may be occasions where the complex plot information one is required to convey in the subtitles overrides some of the rules presented here.

3.1. Timing the Subtitles

A template file is (usually) addressed to a hearing audience. Although there are situations involving the translation of SDH subtitles into other languages, in which SDH-styled templates need to be utilised, templates are most frequently used in contexts involving hearing audiences in multiple languages.

Accurate timing is crucial for the creation of high-quality subtitles, i.e., determining the exact moment when a subtitle should appear on screen (in-time) and disappear from view (out-time). Translation templates need to be timed in sync with the audio and, therefore, SDH conventions that bring subtitles in to shot changes, even if the character only speaks a second later, are not acceptable. Also, it is not the onset of sound that should trigger off a subtitle as in SDH files, but the onset of actual speech. If a character laughs, coughs or makes any other noise before s/he actually speaks, the subtitle should not appear until the first word is uttered. Hearing audiences can *hear* if subtitles are out of sync with the audio so, unless shot changes are an issue, as will be explained below,

a subtitle should come in on the same frame as a character starts speaking and come out when s/he has stopped speaking.

It is recommended that subtitles in template files are timed back-to-back to provide a smooth viewing experience, but also so as to give the subtitles ample reading time, which will be needed for the translation in other languages. Thus, subtitles in a sequence are usually contiguous, as are captions, unless there is a gap of over 20 frames in the dialogue. If the gap is less than 20 frames and does not cross a shot change, it is recommended that it is closed in order for the subtitles to be timed back-to-back, observing a set minimum subtitle separation which should be consistent throughout the file.

An exception to this rule would be if a shot change intervened between the two subtitles, e.g., if a subtitle ended on a shot change whereas the following subtitle begun 15 frames in. Such subtitles would typically be timed to audio, though for a translation template it may be advisable to close the gap between such subtitles as well, by extending the out-time of the first to cross the shot change and come up to one frame before the time-in of the second, so as to allow for more reading time.

When, on the other hand, the slow pace of dialogue allows for a gap between the subtitles or there is a dramatic pause that needs to be conveyed, it is recommended that the separation between subtitles is a minimum of 12 frames (half a second).

Finally, where no other subtitle follows immediately afterwards, a subtitle should not be taken off as soon as people have stopped speaking, as it may appear abrupt to the audience; it is best that there is at least a beat before the subtitle comes off. Delaying the time off of the subtitle in such cases also amplifies its reading time.

The proposed timing settings are the ones below:

- *Minimum duration of a subtitle: 1 second*
Minimum durations down to 12 frames have been observed in the industry, but rarely below.
- *Maximum duration of a subtitle: 7 seconds*
6 or 7 seconds is the most common industry setting, with the exception of songs that can stay on the screen for as long as the lyrics are being sung.
- *Minimum separation between subtitles: 1 frame*
2, 3 or even 4 frames are also very common; 4 frames is stipulated by the *Code of Good Subtitling Practice* (Ivarsson & Carroll, 1998).

3.1.1. Deviation from Sync

Ideally, subtitles will appear in perfect sync with the dialogue as outlined above. If pressed for time, however, it is possible to 'borrow' time before or after the speech so as to maximise reading time.

Where there is no previous subtitle, it is possible to bring in a subtitle a little early. This is a practice to be generally avoided, but when necessary it is recommended to bring the subtitle in ahead of speech by no more than 7 frames or it may be noticeable to the viewer and could create confusion as to the identity of the speaker.

Where there is no subtitle following the speech, it is possible to extend the out-time of a subtitle by up to a second for readability purposes.¹

In flowing speech, where subtitles are contiguous, the maximum deviation from sync should be 12 frames (forward or back). Such leniency should be applied only to gain a clear advantage in readability, i.e., that a subtitle can contain a complete sentence or clauses of a sentence, which would not otherwise be possible, and so awkward splits can be avoided.

3.1.2. Shot Changes

The aim when timing subtitles is to synchronise them with speech, however difficulties are encountered around shot changes. As a rule of thumb, shot changes are to be respected when timing subtitles, as “[s]ubtitles that are allowed to over-run shot changes can cause considerable perceptual confusion” (ITC, 1999, p. 12). The importance of shot changes though is generally lesser than that of ample reading time and appropriate subtitle breaks that make the reading of the subtitles effortless and faster for the viewer. Thus, cases of subtitles carried over shot changes are not rare in practice, especially as modern filmography makes more use of frequent shot changes.

Below are some handy rules so as to avoid the flashing effect of subtitles around shot changes:

- The one-second minimum duration of subtitles must always be respected.
- If a subtitle is to be carried over a shot change and its time-in or time-out is up to 6 frames either side of a shot change, it is advisable to 'snap' such time-in or time-out to the shot change.
- If a subtitle needs to be carried over a shot change so that it has enough time to be read, move its in-time and/or out-time so that it is at least 12 frames from the shot change. This is frequent in two-way conversations, where the audio is ahead of the shot change and the film cuts between the speakers.

¹ The Netflix English Template Guide (n.d., a) states a 12-frame maximum, or half a second.

- If a subtitle is to come out on a shot change and there is no subtitle following it, it is advisable to bring it out 2 frames before the shot change, so that it does not give viewers the impression of hanging on.
- If a subtitle is to come out on a shot change and the subtitle following it is more than 6 frames after the shot change, it is recommended to increase the gap between the subtitles, so that they don't flash (e.g., bring the first one out 2 frames before the shot change and the second in when the character starts speaking).

3.2. Editing – Helping the Translator

The definition of a template is that it is to be used as a base file for translation into other languages. Despite the fact that different percentages are quoted as to the exact degree of expansion when translating from English into other languages, it is generally agreed that considerable **text expansion** takes place when translating from English into most European languages.² As a result, a full two-liner in English loaded with information that cannot be omitted, such as nouns and uncommon or cultural-specific abbreviations cannot always be rendered easily in other languages as expansion will almost certainly have to be involved as part of the translation process. Full lines should thus be avoided in English template files, especially in two-liners. It is also recommended to always give a subtitle the maximum time possible, especially where many nouns are involved, as the latter are particularly difficult for translators to edit.

In the following example, the subtitle has enough time to be easily read in English, but the fact that it consists of a main and a secondary clause, however short, and includes an acronym could make it very tricky to find an adequate rendering in most languages. A higher duration should be given to such a subtitle to ensure it will have ample reading time when translated in other languages as well.

(1)

11:34:58:01 11:35:00:00

Reports say it's SOL.

Not every sound needs to be subtitled. Internationally recognisable features, such as exclamations and information that can be deduced from body and facial expressions, may easily be omitted. This is due to what is known as the “feedback effect” of film (Nedergaard-Larsen, 1993, p. 214) and is based on the fact that subtitles are added to the original production and, as such, do not need

² Indicative LSP sites quoting expansion percentages from English into other languages (and vice versa) are: Kwintessential (n.d.); Arancho Doc (2017); Andiamo (2019). There is guidance on this topic by W3C (Ishida, 2007), while IBM includes relevant instructions in its Knowledge Center (n.d.) regarding expansion allowances that should be made available when translating from English.

to reproduce the information that is already covered in the visuals or the soundtrack. (This same phenomenon can cause problems in the opposite scenario, i.e., when the visuals or soundtrack provide information that is contrasting to the target language audience's knowledge and culture and this dissonance needs to be mediated in the subtitles.) Easily omissible items from the audio to the subtitles are shown in the list below, which was inspired by Marleau's classification (1982, pp. 278–279) and Newmark's advice on redundancy and clichés (1995, p. 208):

- Repetitions;
- Names in appellative constructions;
- False starts / ungrammatical constructions;
- Certain internationally known words, such as “yes” and “no”;
- Expression followed by gestural language to denote salutation, politeness, affirmation, negation, surprise, telephone responses, etc.;
- Exclamations;
- Phatic phrases, such as “naturally”, “of course”, “understandably”, repeated implied superlatives (“basically”, “fundamentally”), prepositional phrases (“in view of the fact that”), rhetorical flourishes, sonorous phrases.

(Georgakopoulou, 2003, p. 216)

If it's irrelevant or obvious – leave it out! This simple rule of thumb goes back to the feedback effect of film (regarding the ‘obvious’ part) and constitutes a subtitler's core technique when it comes to editing down text. This is the very logic behind the cases below as well.

- There is no need to subtitle **stuttering and hesitation** for a hearing audience; hard-of-hearing audiences need such information, however hearing audiences have access to it in the dialogue track.
- No need to subtitle **exclamations** like “Oh!” and “Aah!” for the same reason. However, it would be advisable to subtitle “Oh dear” or “O Lord”. “Hey!” as an exclamation to attract someone's attention does not need to be subtitled, but when it is a substitute for “Hi” it is advisable to include it, e.g., “Hey, honey, I'm home!”.
- **Non-verbal utterances** should also be left out when subtitling. An exception would be the phrase “*Pretty, huh?*”, where the “huh” is necessary to indicate the type of question, i.e., “*isn't she pretty?*” rather than “*is she pretty?*”.
- **Names** should be subtitled if they are not well-known to the audience, for the first couple of times they are spoken, as it is not always obvious which words are names in other languages. It is good practice to introduce all names in the subtitles and leave them out in subsequent occasions.
- **Music.** Songs need not be subtitled unless specifically requested or if it is necessary to translate something sung in order for the audience to understand what comes after it. For example, if someone on screen is singing an entire song, leave it out. If someone bursts into a song in the middle of a conversation and only sings a line, this may need to be translated

if there is a reference in the dialogue to what has just been sung. Such plot-pertinent songs are typically sung to camera. It is recommended to use a symbol to indicate song lyrics, such as the music note if supported, or hash. Use a capital at the beginning of every song line, but do not use full stops or commas at the end of song lines, however commas can be used within the lyric lines if needed. Question marks and exclamation marks may also be used as needed.

3.2.1. Reading Speed

Reading speed is one of the most important settings in a subtitle file, as it determines the amount of text that can fit in the subtitles for the viewers to be able to read them. Thus, different reading speed settings are used depending on the audience, with subtitle files created for children having lower reading speeds than those addressed to adults. Different countries also follow different norms in terms of subtitle reading speed, and hence the amount of dialogue that is represented in the subtitles differs from country to country, with traditionally subtitling countries in Europe showing a strong preference for lower reading speeds than traditionally dubbing ones (Georgakopoulou, 2003, pp. 274–278).

Reading speed is traditionally measured in characters per second (cps), characters per minute (cpm) or words per minute (wpm), where a “word” in English is defined as 5 characters on average, including spaces and punctuation.³ The difference between the time-in and time-out of a subtitle is the subtitle’s duration. The latter, in combination with the reading speed setting applied to a specific file, determines how much text it is advisable for a subtitle to contain, so as to ensure there is enough time for the viewers to read it. Research has shown that reading text on the screen is considerably slower than reading text on printed paper, approximately 20–30% slower according to Dillon (1992).⁴

Reading time should be further maximised in template files, due to the expansion inherent when translating from English to other languages. It is thus recommended that the reading speed used in template files does not exceed 750 cpm (12.5 cps; 150 wpm). In particularly fast-paced

³ However it is an imprecise metric as not all words can be read in a time proportional to their length – word frequencies make a difference, as does syntax (discussed below) and also the video element that allows a viewer to spend more or less time on the subtitles depending on how much action is taking place on the screen. The use of this traditional subtitling metric is currently questioned in a relevant discussion among subtitle professionals on Twitter (Sokoli, 2019).

Between the two metrics, cps/cpm versus wpm, Martí Ferriol (2013) has shown that cps/cpm is more accurate across languages.

⁴ More recent research on print versus screen reading speeds would be interesting, as reading habits are changing over time with the effect of technology. Some studies on subtitle reading speeds listed in Section 4 indicate an increase in the tolerance of higher reading speeds among viewers.

scenes, reading speed can be increased exceptionally up to 900 cpm, i.e., 15 cps or 180 wpm, if needed.

However, keeping subtitles in sync may still be problematic, so it may also be necessary to edit the dialogue on behalf of the translator in order to allow for additional reading time. This technique is meant to help the translator, irrespective of whether the latter may well have to edit further anyway.

Breaks also affect the audience's reading speed, so that awkward line breaks and particularly subtitle breaks may disrupt the reading process (ITC, 1999, p. 8–9). The better a subtitle file is chunked into subtitle units and separate subtitle lines, on the basis of syntax and semantics, as explained in the section below, the greater the reading speed, as reading is facilitated (ITC, 1999, p. 8–9).

3.2.2. Syntax

Adherence to syntactic rules helps the audience process textual information easier. Thus it is better for the viewers but also for the translator if subtitles contain entire sentences. An ideal subtitle is sentence-long. When dealing with more sentences within a single subtitle, it is recommended they are placed in separate subtitle lines, unless they are too short. If a sentence needs to be split, it is recommended that this is done in up to three consecutive subtitles, as it is hard for audiences to retain information further back than that. In general, it is good practice to create fuller subtitles with a longer duration, as opposed to more, shorter subtitles with a short duration. The latter are harder to translate well, and they also require more of an effort by the viewer to read, according to Brondeel (1994, p. 28).

If sentences have to be split over two or more subtitles, it is important that this is done in a way that takes semantics into consideration, as this will increase the audience's reading speed, subtitle comprehension and overall film enjoyment.⁵ Such a split is called a **subtitle break**. As different languages use different sentence structures to English, awkward splits can cause problems, whereas subtitle breaks at the clause level facilitate both translation and readability. If subtitle breaks cannot be achieved at the clause level, then it is important to try and retain entire phrases, e.g., noun phrases, prepositional phrases, etc. within the same subtitle.

⁵ Keeping semantic units together has always been one of the cardinal rules of interlingual subtitling; one can read further on this in Ivarsson & Carroll (1998, pp. 76–78) and Díaz-Cintas & Remael (2007, pp. 172–180).

For example, the two subtitles below (Examples 2 and 3) may reflect the pauses the speaker makes in the audio, but for translation they are preferable as a single subtitle in order to form a complete sentence, or alternatively split after “fair” so that each subtitle contains an entire clause.

(2)

I’m going to the fair
and I’m going to buy

a strawberry ice-cream!

(3)

I’m going to the fair

and I’m going to buy
a strawberry ice-cream!

As one cannot possibly know what is appropriate syntactically in other languages, equivalence can be achieved more easily in larger syntactical structures, i.e., entire sentences, clauses and noun or verb phrases, which should preferably be kept within the same subtitle. Conjunctions usually provide a natural split for subtitle breaks. For example, the subtitles below are read and translated easier if split as in (5).

(4)

11:05:32:02 11:05:34:23
Announce a state of emergency
and arrest all members

11:05:34:24 11:05:38:10
of the Executive Council.
Break their chain of command.

(5)

11:05:32:02 11:05:36:06
Announce a state of emergency
and arrest all members of the Council.

11:05:36:07 11:05:38:10
Break their chain of command.

In the following example editing down helps retain a clearer syntactic structure across subtitles:

(6)

10:22:55:13 10:22:58:16

Although you're under 15
if you get more than 50

10:22:58:17 10:23:02:10

penalty points on your record,
you go to the regular courts.

(7)

10:22:55:13 10:23:00:09

Although you're under 15, if you
get more than 50 penalty points,

0:23:00:10 10:23:02:10

you're sent to the regular courts.

Though not as critical as subtitle breaks, **line breaks**, i.e., breaks within a subtitle from the top to the bottom line, can also ease reading and help viewers understand subtitles faster. Very much like subtitle breaks, line breaks should also be made at the highest syntactic node possible. Common mistakes to be avoided include separating articles or adjectives from the nouns they accompany, prepositions from the prepositional phrases they introduce, splitting proper noun phrases, etc. Punctuation is typically a good guide as to where to insert a line or subtitle break, as in most cases it will ensure semantic units are kept together. In the examples below the optimal line breaks are the ones in (9).

(8)

You'll all be surprised when
you see what I accomplished.

I really hope that the teacher will
not be late again this time.

It's incredible, I would have
never recognised him.

(9)

You'll all be surprised
when you see what I accomplished.

I really hope that the teacher
will not be late again this time.

It's incredible,
I would have never recognised him.

Simplifying syntax also helps viewers process a subtitle faster. Although a variety of syntactic structures may be possible in different languages, the most common word order is preferred as it is simpler and thus easier to read: *Subject-Verb-Object*. Canonical forms are generally shorter and simpler to read, argues Karamitroglou (1998), and hence they should be preferred in subtitles. One such example is the preference of active over passive voice, which also tends to be more concise and this helps reduce character count in the subtitles. This is a technique recommended specifically for interlingual subtitles, where translation takes place, and as such can be employed in the case of template files too, but not in SDH files where viewers are likely to lip-read.

(10) Original voice track:

One thing people never ought to be when they're buying used cars is in a hurry.

It is reported by many students that the course is particularly hard.

(11) English template file:

People never ought to be in a hurry
when buying used cars

Many students report
that the course is particularly hard.

3.3. Formatting the Subtitles

3.3.1. Dialogue

Subtitles ideally contain a single speaker's speech, but when two speakers need to be combined in a single subtitle, then each speaker's speech should be strictly placed on a separate subtitle line and preceded by a dialogue dash. Dialogue dashes are either consistently followed or not by a space, depending on the style chosen.

Dialogues can present a particularly problematic case in subtitling when the bulk of the information load rests on one of the two dialogue lines and expansion is not possible into a second line (as is the norm in some countries). The original needs to be restructured in such cases, as per the example below (Georgakopoulou, 2003, p. 212):

(12)

01:56:55:18 01:57:00:21

– Sunset? No! I mean, it's very late.

– What?

The possible alternatives are:

1. Splitting the subtitle in two subtitles.
2. Omitting the dialogue in the translation and translate the first line only.
3. Omitting 'Sunset' from line 1.

3.3.2. Punctuation

An old and widely used subtitling convention is the use of continuation and linking dots to indicate that a sentence continues from one subtitle to the next. While continuation dots are still used to a large extent in some countries, their use has been reduced over the years, and linking dots are a rare occurrence, while audiences are now largely conditioned not to rely on them.

In the benefits of saving space, and as viewers are able to deduct that a sentence has not finished by the proper use of sentence end punctuation marks, it is recommended that neither continuation nor linking dots are used in template files. They can be inserted, if needed, in the translation files.

Suspension dots (ellipsis) are however used to indicate speech trailing off, a pause in the speech, or speech picking up mid-sentence.

3.3.3. Italics

Italics are used to indicate speech that is coming from a source that is not visible on the screen. Common examples of the use of italics are for voice over speech, TV/radio speech, announcements at stations/airports, telephone responses where the person at the other end of the line is not visible, etc.

Italics should be applied to complete sentences only; if the speaker comes into vision at any point of the dialogue/scene, then italics are not needed.

Italics are also commonly used for:

- Single words or short phrases in a third language (L3) that have been transliterated rather than translated in the subtitles.⁶
- Song lyrics, when subtitled.
- Book, film, album and programme titles, though song titles are usually presented in quotes.

3.3.4. Positioning

Subtitles cover a significant part of the screen and as such it is best to have the largest load of text at the very bottom of the screen, where less action is taking place. Template subtitles are typically placed at the bottom of the screen and centred, with the exception of dialogues which are centred and left justified (in left-to-right languages). Thus, assuming syntax allows for it, it is best to keep the second line of a subtitle fuller than the first, and to place one-liners at the very bottom of the screen (where the second line would be placed in the case of a two-liner).

When there is principal photography on screen, however, e.g., newspaper text, or captions burnt-in on the screen, e.g., “London, 1941”, “John Smith: Producer”, then one must take care that the subtitle text does not cover the on-screen text. In such cases, subtitles should be raised either just above the on-screen text or to the top of the screen. Such subtitles are called **raised subtitles**. It is recommended to raise subtitles to the top of the screen in the case of template files. When doing so, the distribution of text in the subtitles should again follow the logic of placing the bulk of the text towards the edges of the screen where there is less action, i.e., the top line should be fuller than the bottom one in the case of raised subtitles.

⁶ For a detailed analysis on the use of L3 in subtitles, see Corrius & Zabalbeascoa (2011).

3.3.5. Forced Narratives

When text appears on the screen as part of principal photography, e.g., newspaper text, signs, letters, etc. or as captions burnt-in on the screen, e.g., date and time or location stamps, such text also needs to be subtitled and translated if plot pertinent. For example, you do not need to add a subtitle to translate the word “RESTAURANT” when a character is running past a restaurant on screen and this makes no difference to the plot. On the other hand, it is likely to be necessary to add a subtitle for the word “POLICE” if a character goes into a police station, or reads a note saying “Back in five minutes”.

Such subtitles are known as **forced narratives**. This is because forced subtitles will need to appear (to be “forced”) on a dub track in the target language as well, as such text would not normally be covered in the dubbed dialogue. It is important to mark forced narratives as such in a subtitle file, e.g., using a note or annotation or special commands in your subtitling software, as it may be necessary to also deliver a forced subtitle stream separately to the end client if one is needed for the dub track of the same film.

Forced narratives are typically timed to match the on-screen text, and also copy its punctuation and case style. It is recommended, however, that forced narratives which appear in between consecutive dialogue, e.g., the phrase “John Smith: Producer” appearing on screen in the middle of a character’s speech, are typed in ALL CAPS so they stand out from the dialogue that surrounds them. The exception to this rule is when there is a large block of text on the screen, such as the exposition or the epilogue of the film, in which case it is best to use sentence case for readability.

Forced narratives need only be subtitled once, even if they appear on the video recurrently (e.g., the name of a producer in an interview). When there is dialogue running consecutively over forced text and it is necessary to subtitle the forced text (if not, the dialogue always takes precedence), then it is allowable to go out of sync in order to include the forced narrative and give it a minimum duration. A forced narrative should never be combined with dialogue in the same subtitle.

Text in a third language (L3) in the dialogue is only meant to be subtitled in a template file (so that it is translated in all the target languages) if it is subtitled in the original version of the film, so as to follow the director’s intent and create the same effect to the target audience as for the source audience. Such subtitles are also considered forced narratives and need to be flagged as such.

3.3.6. Checklists

A template creation task is not complete until the following accompanying documents have been prepared: an italics list, a raised subtitles list, a forced subtitles list, and a names & notes list. All these lists are particularly useful when performing a QC of the ensuing translated files across multiple languages, even without knowledge of said target languages.

The first three types of lists have already been discussed. A **names & notes list** would include all proper names in a video, properly checked for spelling against the credits, script, official website, IMDb, Halliwell's, or other available source, and including information such as a person's sex if not obvious from the film, so this can be taken into account when translating. Information about song lyrics can also be included here, as well as notes regarding the relationship between film characters, especially in the case of series, where different episodes may be translated by different translators and, hence, it would be important in some languages to determine the familiarity level between characters so as to use appropriate politeness markers in the translation. Other translation notes are explanations designed to help the translator better understand the source dialogue, by providing useful information regarding the use of slang, dialect or cultural references in the subtitles that would be difficult for a non-native speaker to be familiar with. Frequently this is information that can also be found in the film script, if available.

4. Discussion

As already mentioned in the introduction, the motivation for writing this paper is to contribute to the ongoing discussion regarding quality and standards in subtitling by providing one additional source of reference, specifically for the creation of template files, which is a topic that remains under-researched to date.

Template files have, out of necessity, become a representation of a golden mean between varying subtitling practices and norms, favouring the ease of production and quality assurance of increasingly larger volumes of content, whilst trying to allow for regional variation to the extent possible. The language-specific subtitling guidelines published by Netflix (n.d., b) are a good representative example of this process in practise, by standardising aspects such as characters per line and subtitle reading speed, while allowing for regional treatment in formatting issues of subtitles, such as the use of dialogue dashes and italics. Despite their predominance in today's market, templates are still relatively unpopular among some groups of professional subtitlers and unions, mainly because of financial concerns, but also for reasons of quality (Artegiani & Kapsaskis, 2014; Nikolić, 2015), which are worth examining further, as specifications are being revisited.

With the developments in the entertainment industry and the Over-The-Top (OTT) distribution of content in any country and language, as well as the possibilities offered by online communication,

modern subtitling software and tools applied to subtitling research, such as eye-trackers, we are noticing today a rekindled discussion on subtitling norms and the questioning of long established practices. Some examples are: the study by Perego et al. (2010) on subtitle processing, which goes against the cardinal rule of subtitling for line segmentation; the analysis of the effects of crossing shot changes on subtitle re-reading by Krejtz et al. (2013), which shows that viewers are not induced to re-reading subtitles that go over shot changes, but also points to a higher number of gaze shifts from the subtitles to the image in such instances, a factor that would affect reading speed; Szarkowska's and Gerber-Morón's (2018) empirical work on the appropriateness of the famous six-second rule for subtitle readability; or their empirical verification of the rule about subtitle layout with respect to syntax (Gerber-Morón & Szarkowska, 2018).⁷ Such reception studies are extremely important in order to better define the needs of today's global audiences, which may have evolved and diverged from those catered for by national subtitling practices of the previous century as a result of technological innovation, exposure to and processing of vast amounts of audiovisual information, and conditioning to practices that promote homogenisation, such as subtitling workflows involving template files for translation purposes.

The most notable difference in the specifications laid out for template file creation in Section 3 above as compared to current practices by global LSPs providing subtitles in multiple languages, is the subtitle reading speed setting analysed in Section 3.2.1. The 1999 ITC standards for subtitling of pre-recorded programmes for adults in the UK recommended a reading speed of 140 wpm (or roughly 12 cps). This was later revised by Ofcom to 160–180 wpm (or roughly 13–15 cps) to reflect an increase in the audience's reading abilities, which was corroborated with a relevant study (2005),⁸ and it remains Ofcom's (2017) and the BBC's (2018) recommendation to date. This was also the standard loosely followed when drafting the present guide albeit for interlingual subtitles, which stipulates a reading speed of 150 wpm or 12.5 cps, with an allowable increase to 180 wpm or 15 cps in exceptional cases. This setting was verified in practice at the time, against locally-produced subtitle assets in a variety of European languages, and was shown to represent the average of the reading speeds applied in the sample files examined (Georgakopoulou, 2003, pp. 274–278). Such reading speeds are also in line with traditions in France, Spain and Germany (Mikul, 2014, p. 25)

⁷ Reception studies are also being carried out on the effects of innovative subtitling practices commonly used by fansubbers, which may also shed interesting light on today's audiences' reception capacities – see Künzli & Ehrensberger-Dow (2011), as quoted in van Tonder (2015).

⁸ The study performed by Ofcom was on deaf and hard-of-hearing participants, and intralingual subtitles. Nevertheless, its results could also be loosely applied to hearing interlingual audiences, who also “rely more heavily on subtitles” – “speed is much more of an issue because if they cannot read them fast enough they will literally lose the plot” (Ofcom, 2005). The comments of the participants and the results of the study regarding the issue of increasing subtitling speeds are of particular interest to those wishing to investigate this issue further.

as well as in Scandinavian countries (Pedersen, 2011, p. 133).⁹ Still, a lot of subtitling today, especially when produced in a centralised manner with a template workflow, is much closer to American standards of verbatim captions, with reading speeds above 200 wpm or 17 cps being the norm rather than an exception. For example, the Netflix Partner Help Centre specifies a reading speed of 20 cps (approx. 240 wpm) for adult programmes and 17 cps for children when it comes to English templates and SDH files (n.d., a), and 17 and 13 cps respectively for most other languages for interlingual subtitles (n.d., b).

This is a striking difference to the reading speed recommended for use in English template files in the previous decade. It is also the cause for a lot of debate. The Danish Subtitlers' Association, for instance, recently published a set of subtitling guidelines for Denmark, which are endorsed by the most prominent players in the Danish subtitling industry (Forum for Billedmedieoversættere, 2019); they specify 10–15 cps (120–180 wpm) as the reading speed standard in the country and encourage everyone involved in Danish subtitling production to respect it. On the other hand, the BBC R&D published a white paper (Sandford, 2015) on the impact of subtitle display rate on the enjoyment of television programmes, that speaks for higher reading speeds in SDH viewers, a tolerance as high as 242 wpm before the viewers started having difficulty to follow the programme and concludes that speed is not an issue if the subtitles follow the rhythm of natural speech. A similar finding was reported by Szarkowska and Gerber-Morón for hearing viewers (2018), who show that audiences are able to keep up with higher reading speeds, such as 20 cps, and even report re-reading of slower paced subtitles at 12 cps. Finally, Jean-Louis Kruger (2019), at a keynote speech he delivered at the 8th Media for All conference, spoke about studies performed on the topic of visual perception and skilled reading, and how audiences consume subtitled audiovisual material. He concluded that appropriate subtitle presentation speed is much more complex than simply counting words: word frequencies and word length also make a difference in viewers' reading speed, for instance, but other filmic elements, such as sound, are also factors in subtitle processing. Clearly reading speed is a topic that lends itself to research, thus new reception studies would be welcome in order to shed more light on the debate and help foster audience-friendly subtitling standards.

⁹ Romero-Fresco (2009) provides an excellent analysis of reading speed for SDH viewers and how this compares to the respeaking rate used in the UK for live intralingual subtitles. His analysis includes a section on reading speed research, including the findings of D'Ydewalle, Rensbergen and Pollet (1987) on the subtitle reading speed of hearing adults and the history of the famous "six second rule" (Romero-Fresco, 2009, p. 114).

5. Conclusion

This paper is a detailed presentation of a subtitling style guide used for the creation of English template files and practiced for approximately a decade, one of the few to be published to date. It is meant to serve as an example of best practice followed by a multilanguage subtitle service provider at the turn of the century and to be compared to the style guides in use today, so that it contributes to the ongoing debate about subtitling quality and practice, and highlights areas where research is necessary. The most important such area has been identified as the reading speed settings in use today and how well they cater to the needs of local audiences. More research on this topic would be most welcome.

Acknowledgements

This work would not have been completed without the contribution of multiple subtitlers I had the privilege to work with at the European Captioning Institute and who I learned from. They helped refine the present rules on the practice of English template file creation and produced some of the best subtitle template files I have come across. This paper is dedicated to them – thank you all!

References

- Andiamo (2019). How Much will your Text Expand or Contract? Retrieved from <https://www.andiamo.co.uk/resources/expansion-and-contraction-factors/>
- Arancho Doc (2017). Text Expansion & Contraction Chart. Uploaded 2017, July. Retrieved from <http://www.aranchodoc.com/wp-content/uploads/2017/07/Text-Expansion-Contraction-Chart3.png>
- Artegiani, I., & Kapsaskis, D. (2014). Template files: Asset or anathema? A qualitative analysis of the subtitles of *The Sorpanos*. *Perspectives: Studies in Translation Theory and Practice*, 22(3), 419–436. <https://doi.org/10.1080/0907676X.2013.833642>
- BBC (2018, May). Subtitle guidelines. Version 1.1.7. Retrieved from <http://bbc.github.io/subtitle-guidelines/>
- Brondeel, H. (1994). Teaching subtitling routines. *Meta* 39(1), 26–33.
- Bywood, L. (2019, May 13). [personal communication with Author of the paper].
- Carroll, M. (2004, September 14). Subtitling – Changing standards or new media? *Globalization Insider*, XIII(3.3). Retrieved from <https://www.translationdirectory.com/article422.htm>
- Corrius, M., & Zabalbeascoa, P. (2011). Language variation in source texts and their translations: The case of L3 in film translation. *Target* 23(1), 113–130. Retrieved from https://www.researchgate.net/publication/233638429_Language_variation_in_source_texts_and_their_translations_The_case_of_L3_in_film_translation

- D'Ydewalle, G., Van Rensbergen, J. & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In J. K. O'Reagan & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 313–321). Amsterdam / New York: Elsevier Science Publishers.
- Described and Captioned Media Program (n.d.). *Captioning Key*. Retrieved from http://captioningkey.org/about_c.html
- Díaz-Cintas, J., & Remael, A. (2007). *Audiovisual translation: Subtitling*. Abingdon: Routledge. Retrieved from www.routledge.com/Audiovisual-Translation-Subtitling/Diaz-Cintas-Remael/p/book/9781900650953
- Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics*, 35(10), 1297–1326. Retrieved from <https://www.ischool.utexas.edu/~adillon/Journals/Reading.htm>
- EBU Tech I44-2004 (2004, June). *EBU report: Access services, includes recommendations*. Retrieved from <https://tech.ebu.ch/docs/i/i044.pdf>
- Forum for Billedmedieoversættelse (2019, January 11th) *Guidelines for subtitling in Denmark* (Version 1). Retrieved from <https://undertekstning.dk/english>
- Gerber-Morón, O., & Szarkowska, A. (2018). Line breaks in subtitling: An eye tracking study on viewer preferences. *Journal of Eye Movement Research*, 11(3). <http://dx.doi.org/10.16910/jemr.11.3.2>
- Georgakopoulou, P. (2003). Reduction levels in subtitling. DVD subtitling: A compromise of trends. *Theses from the University of Surrey, Paper 242*. Retrieved from <http://epubs.surrey.ac.uk/id/eprint/602>
- _____ (2005). Subtitling: A quest for the Holy Grail? In B. Mayor (Ed.), *Readings booklet, Book 5 – Getting to work: Putting grammar into professional practice* (pp. 46–56). Milton Keynes: Open University.
- _____ (2006, July). Subtitling and globalisation. *The Journal of Specialised Translation*, 6, 115–120. Retrieved from http://www.jostrans.org/issue06/art_georgakopoulou.php
- _____ (2009). Subtitling for the DVD industry. In J. Díaz-Cintas, & G. Anderman (Eds.), *Audiovisual translation: language transfer on screen* (pp. 21–35). London: Palgrave Macmillan. https://doi.org/10.1057/9780230234581_2
- _____ (2010). *Reduction levels in subtitling. DVD subtitling: A convergence of trends*. Saarbrücken: Lambert Academic Publishing.
- _____ (2012). Challenges for the audiovisual industry in the digital age: The ever-changing needs of subtitle production. *The Journal of Specialised Translation*, 17, 78–103. Retrieved from http://www.jostrans.org/issue17/art_georgakopoulou.php
- IBM Knowledge Center (n.d.). Text translation design. Retrieved from https://www.ibm.com/support/knowledgecenter/ssw_ibm_i_72/nls/rbagstextdatatransdesign.htm
- Ishida, R. (2007). Text size in translation. *W3C Internationalization*. Retrieved from <https://www.w3.org/International/articles/article-text-size.en>
- Ivarsson, J., & Carroll, M. (1998). *Subtitling*. Simrishamn: TransEdit.
- Independent Television Commission (1999). *ITC Guidance on standards for subtitling*. February 1999. London, UK. Retrieved from http://sign-dialog.de/wp-content/richtlinien_199902_england.pdf

- Kapsaskis, D. (2011, July). Professional identity and training of translators in the context of globalisation: the example of subtitling. *The Journal of Specialised Translation*, 16, 162–184. Retrieved from http://www.jostrans.org/issue16/art_kapsaskis.php
- Karamitroglou, F. (1998, April). A proposed set of subtitling standards in Europe. *Translation Journal*, 2(2). Retrieved from <http://translationjournal.net/journal/04stndrd.htm>
- Krejtz, I., Szarkowska, A., & Krejtz, K. (2013). The effects of shot changes on eye movements in subtitling. *Journal on Eye Movement Research*, 6(5), 1–12.
- Kruger, J.-L. (2019, June 17–19). Towards an understanding of AVT as multimodal mediation. *The 8th Media for All conference – Complex Understadnings*, Stockholm, Sweden. Retrieved from <https://www.tolk.su.se/english/media-for-all-8/programme/live-streaming/opening-welcome-1.441494>
- Künzli, A., & Ehrensberger-Dow, M. (2011). Innovative subtitling – A reception study. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 187–200). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/btl.94>
- Kwintessential (n.d.). Text contraction and expansion in translation. Retrieved from <https://www.kwintessential.co.uk/resources/expansion-retraction>
- Marleau, L. (1982). “Les Sous-Titres... Un Mal Necessaire”. *Meta* 27(3), 271–285.
- Martí Ferriol, J. L. (2013). Subtitle Reading Speeds in Different Languages: The Case of *Lethal Weapon*. *Quaderns. Revista de Traducció* 20, 201–210. Retrieved from https://ddd.uab.cat/pub/quaderns/quaderns_a2013n20/quaderns_a2013n20p201.pdf
- Mikul, C. (2014). Caption quality: International approaches to standards and measurement. Sydney: Media Access Australia. Retrieved from http://www.mediaaccess.org.au/sites/default/files/files/MAA_CaptionQuality-Whitepaper.pdf
- Nedergaard-Larsen, B. (1993). Culture-bound problems in subtitling. *Perspectives Studies in Translatology*, 1(2), 207–241. Denmark: Museum Tusulanum Press. <https://doi.org/10.1080/0907676X.1993.9961214>
- Netflix Partner Help Centre (n.d., a). English template timed text style guide. Retrieved from <https://partnerhelp.netflixstudios.com/hc/en-us/articles/219375728-English-Template-Timed-Text-Style-Guide>
- _____ (n.d., b). Timed text style guides. Retrieved from <https://partnerhelp.netflixstudios.com/hc/en-us/categories/202282037-SPECIFICATIONS-GUIDES>
- Newmark, P. (1988). *A Textbook of Translation*. London: Prentice Hall.
- Nikolić, K. (2015). The pros and cons of using templates in subtitling. In R. B. Piñero, & J. Díaz-Cintas (Eds.), *Audiovisual translation in a global context: Mapping an ever-changing landscape* (pp. 192–202). Hampshire: Palgrave Macmillan. https://doi.org/10.1057/9781137552891_11
- O’Hagan, M. (2005). Multidimensional translation: A game plan for audiovisual translation in the age of GILT. In H. Gerzymisch-Arbogast, & S. Nauert (Eds.), *MuTra 2005 – Challenges of multidimensional translation. Conference proceedings*. Saarbrücken. Retrieved from http://www.euroconferences.info/proceedings/2005_Proceedings/2005_O'Hagan_Minako.pdf
- Ofcom (2005, January 6). Subtitling – An issue of speed? Retrieved from <https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/subtitling-speed>

- _____ (2017). *Code on Television Access Services*. 30 January 2017. Retrieved from https://www.ofcom.org.uk/data/assets/pdf_file/0020/97040/Access-service-code-Jan-2017.pdf
- Pedersen, J. (2011). *Subtitling norms for television: An exploration focusing on extralinguistic cultural elements*. Amsterdam: John Benjamins.
- Perego E., del Missier, F., Porta, M., Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, 13(3), 243–272. <https://doi.org/10.1080/15213269.2010.502873>
- Romero-Fresco, P. (2009). More haste, less speed: Edited versus verbatim respoken subtitles. *VIAL-Vigo International Journal of Applied Linguistics*, 6(1), p. 109–133. Retrieved from <https://vialjournal.webs.uvigo.es/pdf/Vial-2009-Article6.pdf>
- Sandford, J. (2015). The impact of subtitle display rate on enjoyment under normal television viewing conditions. *IBC2015 Conference*. BBC Research & Development, White Paper WHP 306. Retrieved from <http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP306.pdf>
- Sokoli, S. (2019, May 21). On #subtitling quality: When will we stop measuring readability just by counting characters per second? Message posted to <https://twitter.com/StavroulaSokoli/status/1130940807735857152>
- Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PLoS ONE*, 13(6), e0199331. <https://doi.org/10.1371/journal.pone.0199331>
- van Tonder, L. C. (2015). Fansubbing and creative subtitling – How amateur approaches can help improve professional products. *South African Translators' Institute – Second Triennial Conference: The Changing Face of Translation and Interpreting*. Kempton Park, South Africa. Retrieved from https://www.academia.edu/16564660/Fansubbing_and_Creative_Subtitling_How_Amateur_Approaches_can_Help_Improve_Professional_Products