Language & Communication 49 (2016) 56–69

Contents lists available at ScienceDirect

Language & Communication

journal homepage: www.elsevier.com/locate/langcom

Accessing communication: The quality of live subtitles in the UK

Pablo Romero-Fresco [a,b,*]

[a] Department of Media, Culture and Language, University of Roehampton, SW15 5SL London, UK
[b] Universidade de Vigo, Campus Universitario, s/n, 36310 Vigo, Spain

ARTICLE INFO

Article history:
Available online 5 July 2016

Keywords:
Live subtitling
Respeaking
Subtitling for the deaf
NER model
Accuracy
Media accessibility

ABSTRACT

This article presents the results of the first official assessment on the quality of live TV subtitles in the UK as a form of access to audiovisual communication for deaf and hard-of-hearing viewers. Carried out in collaboration with the British governmental regulator Ofcom, this is the largest analysis ever conducted on the quality of live subtitles, covering 300 programmes and 78,000 subtitles. The results provide data on the key issues of accuracy, speed, reduction rate and latency of the subtitles. The article concludes with a number of suggestions to update the existing guidelines on live subtitles at an international level so that they can provide full access to audiovisual communication for viewers with hearing loss.

1. Introduction

Over the past 15 years, audiovisual translation has come of age as a discipline in its own right within the wider research field of translation and interpreting studies. Audiovisual translation scholars focus on film dubbing, subtitling and voice-over as forms of multimodal communication (Gambier, 2006; Pérez-González, 2014), involving the interaction between verbal and non-verbal signifiers and a complex combination of semiotic choices from different sign systems (Chaume, 2004). Also included within audiovisual translation is the subfield of media accessibility, which covers subtitling for the deaf and hard of hearing and audiodescription for the blind and partially sighted. The latter are examples not only of multimodal communication but also of Jakobson's (1959) intersemiotic translation or transposition from one system of signs into another: from aural to written in subtitling for the deaf and from visual to aural in audiodescription for the blind.

Widely regarded as an essential tool for hearing-impaired and foreign viewers to access audiovisual communication, subtitling for the deaf has recently been the subject of discussion amongst scholars, user associations and governmental regulators (Díaz Cintas and Baños, 2015). Although until now these discussions have typically revolved around the issue of quantity (i.e. the increase of national subtitling quotas and the extent to which they are met), the focus is gradually shifting to quality (MAA, 2014). In line with this trend, this article presents the results of the first official assessment of the quality of live subtitling in the UK and the largest analysis of live subtitling quality conducted to date.

In May 2013, Ofcom, the governmental communications regulator in the UK, published a consultation document in order to gather views from broadcasters, researchers, subtitle providers and user associations as to how to improve the quality of live subtitling on UK TV to benefit deaf and hard-of-hearing viewers (Ofcom, 2013). Following this consultation, Ofcom

* Department of Media, Culture and Language, University of Roehampton, SW15 5SL London, UK.
E-mail address: p.romero-fresco@roehampton.ac.uk.

http://dx.doi.org/10.1016/j.langcom.2016.06.001
0271-5309/

decided that broadcasters should be required to assess the speed, reduction rate, latency and accuracy of their live subtitles applying the NER model (Romero-Fresco and Martínez, 2015), which is currently being used as the standard method for the assessment of live subtitling quality in countries such as the UK, Spain, Switzerland, Germany, Belgium, Italy, Brasil and Australia. Ofcom also decided that the University of Roehampton (London) would review the measurements provided by broadcasters from a third party standpoint. This 2-year project is the first official assessment of the quality of live TV subtitles ever conducted in the UK and has triggered demands for similar assessments in countries such as Canada, Switzerland and Australia (Ai-Media, 2014; MAA, 2014). The present article describes the NER model as one of several methods to assess the quality of live subtitles, reviews similar studies conducted so far on the main features and quality of live subtitles and presents the results and conclusions of this project, whose ultimate aim is to enable a revision and update of the existing subtitling guidelines in the UK (Ofcom, 2012) and at an international level.

2. The assessment of live subtitling quality: in search of a model

New European legislation on media accessibility for viewers with hearing loss has led to increasing amounts of TV subtitles. Now that countries such as the UK, Switzerland and France are approaching 100% coverage on national TV, the focus has been placed on quality, with particular emphasis on live subtitles, which are characterized by the presence of errors and latency.

Unlike in the US, where most live subtitles for TV are produced by stenographers, in Europe TV subtitles are mostly made by respeakers, known in the US as voice writers or real-time voice writers (Keyes, 2005). Respeakers listen to the original audio of a live programme and repeat it or rephrase it simultaneously to a speech recognition software that turns the recognized utterances into TV subtitles (Romero-Fresco, 2011). Given the considerably shorter training time required for respeakers as compared to stenographers, respeaking has taken over as the preferred method for most European broadcasters, although stenographers are still employed for the fastest and most difficult programmes. Another difference between live subtitles in the US and in Europe is related to the reduction rate, also known as edition rate (Romero-Fresco and Martínez, 2015). Whereas live subtitles in the US are almost verbatim (Jensema et al., 1996), in Europe they vary from the near-verbatim UK subtitles to the more heavily edited subtitles in Spain or Switzerland (Romero-Fresco, 2009).

When it comes to assessing live subtitling quality, methods and approaches vary greatly across and even within countries, which up until now has made it difficult to compare results. What may be expected from these models of assessment is that they meet at least some of the following requirements: 1) they are functional and easy to apply, 2) they take into account not only the accuracy of the subtitles but also the comparison to the original speech, 3) they account for the possibility of reduced and yet accurate subtitles depending on the different national editing conventions, 4) they provide information not only about the accuracy of the subtitles but also about other relevant elements regarding quality such as delay, position, speed, character identification, etc., 5) they account for the fact that not all errors have the same origin or impact on the viewers' comprehension and 6) they provide an assessment of quality as well as an overall idea of aspects to be improved, in other words, food for thought as far as training is concerned.

Most models of assessment take as a starting point the basic principles of WER (word error rate), which have traditionally been applied to the analysis of accuracy in speech recognition. The US National Institute of Standards and Technology uses the formula shown in Fig. 1 to calculate word accuracy (Dumouchel et al., 2011):

In this model, N is the total number of words spoken by the user. D, S and I are the errors caused due to words deleted, substituted or inserted incorrectly by the speech recognition software. Designed as it is for the use of speech recognition, this model poses a significant problem when applied to subtitling in European countries, as it does not account for instances in which a subtitler edits the original text without changing or losing meaning, as shown in Fig. 2:

The example above features the omission of relatively unimportant asides ('you know', 'I mean', 'kind of'), which constitutes a useful strategy commonly applied by respeakers to catch their breath and keep up with the original speaker. Traditional WER methods would yield an accuracy rate of 52%, whereas a model suited to respeaking may consider this respoken subtitle as 100% accurate.

In order to solve this problem, the Centre de Recherche Informatique de Montréal (CRIM) adapted the above model to the specificity of live subtitling by respeaking (Dumouchel et al., 2011). Their model adds a further step: once the transcription of the audio and the subtitles has been aligned for the analysis, a human operator goes through the text and decides whether or not the deletions have caused loss of information. In this way, both verbatim and edited respeaking can be accounted for. Yet, a number of problems remain unsolved. First of all, the decision of when a deletion brings about loss of information is entirely subjective and may thus vary from person to person. Secondly, while requirements 1 to 3 above are met, 4, 5 and 6 are not. The accuracy rate obtained with this model may provide useful data, but the deletion figure can be ambiguous. Indeed, the

$$\text{Accuracy rate} = \frac{N - \text{Errors } (D + S + I)}{N} \times 100 = \%$$

Fig. 1. Formula used by the US National Institute of Standards and Technology to calculate word accuracy.