Strengthening respeakers’ training in Spain: the research-practice connection

Nazaret Fresno
*The University of Texas Rio Grande Valley, nazaret.fresno@utrgv.edu*

Pablo Romero-Fresco

Follow this and additional works at: https://scholarworks.utrgv.edu/wls_fac

Part of the Modern Languages Commons

**Recommended Citation**


This Article is brought to you for free and open access by the College of Liberal Arts at ScholarWorks @ UTRGV. It has been accepted for inclusion in Writing and Language Studies Faculty Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.
Strengthening respeakers’ training in Spain: the research-practice connection

Nazaret Fresno*a and Pablo Romero-Frescob

aDepartment of Writing and Language Studies, The University of Texas Rio Grande Valley, Brownsville, United States;
bPablo Romero-Fresco. Universidade de Vigo/ University of Roehampton, Vigo/London, Spain/United Kingdom

*1 W. University Blvd, Brownsville, TX 78520, United States
nazaret.fresno@utrgv.edu
https://orcid.org/0000-0002-6702-159X

Abstract

After countless petitions and complaints from end users, live subtitling quality is slowly attracting the attention of broadcasters, regulators, the subtitling industry and scholars working in Media Accessibility. These stakeholders share an interest in providing better live subtitles, but their quality assessment is a thorny issue. Although quality studies are still scarce, the research undertaken so far has proven valuable in identifying the weaknesses of live subtitles in several countries. This article presents the main findings of the pilot project that preceded the first national quality assessment in Spain, which is currently underway. By focusing on this case study, we will argue that live subtitling quality research may fulfil an additional purpose: serving as a didactic tool in respeaking courses. In this paper we will outline the quality assessment method that we followed, discuss how its main results informed about the accuracy, speed and latency of our samples, and describe how these data may be brought to the classroom to fine tune respeakers’ training under a Performance Analysis approach.

Keywords: media accessibility, live subtitling, respeaking, training, quality.

1. Introduction

The presence of live subtitles1 on television has increased over the past twenty years, as more broadcasters have made their offer accessible to viewers with hearing loss and, in general, to all. Usually delivered in the same language as the programme that they accompany, live subtitles account for the acoustic information in live broadcasts for those who cannot access dialogue and soundtrack. As a response to users’ concerns and complaints, a number of

1 All throughout this paper we will use the term “live subtitles” to refer exclusively to “intralingual live subtitles”, that is, live subtitles delivered in the same language as the programme that they accompany.
countries have issued official guidelines or regulations aimed at describing the quality criteria that live subtitling should have (AENOR 2012; FCC 2014; ACMA 2016; CRTC 2019 or Ofcom 2019, among others). Nevertheless, as Romero-Fresco (2020) points out, quality has been approached very differently across countries. Some, such as the US and Australia, have defined broad unspecific quality requirements and have put in place complaint-driven approaches in which quality is monitored officially only after viewers have expressed dissatisfaction with the closed captioning provided in specific programs. These soft stances differ from hard approaches, such as those in place in the UK and Canada, which include more tangible quality criteria, mandatory assessments led by national regulators and descriptions of the methods that should be used when assessing live subtitling quality.

Live subtitling may be produced through several methods, but the most common ones are currently stenography (special keyboards that allow very fast typing) and respeaking, which Romero-Fresco (2011, 1) defines as follows:

“A technique in which a respeaker listens to the original sound of a (live) programme or event and respeaks it, including punctuation marks and some specific features for the deaf and hard of hearing audience, to a speech recognition software, which turns the recognized utterances into subtitles displayed on the screen with the shortest possible delay.”

Respeaking is the preferred method in Europe and in Spain, which will be the focus of this paper. At the time of writing this article, Spain is conducting its first official live subtitling quality assessment at the national level. Led by the Galician Observatory for Media Accessibility (GALMA) in collaboration with five television broadcasters and the Spanish Centre of Subtitling and Audio Description (CESyA), the QualiSpain project aims to define the strengths and weaknesses of the live subtitles currently delivered on Spanish television. By promoting a research-informed dialogue between academia, broadcasters, service providers and the national regulator, QualiSpain sets out to find potential solutions for those areas that could benefit from improvement. The findings obtained in this project should also be valuable to strengthen respeakers’ training in Spain. This article will tackle this research-training connection.

2. Respeakers’ training in Europe

In Europe, respeaking was introduced as a professional practice in the early 2000s and, at that time, respeakers were trained through in-house programmes offered by broadcasters that were in need of practitioners (Romero-Fresco 2020). Since then, respeaking has gained prominence and, as the number of live broadcasts keeps increasing, so does the demand for competent professionals. A recent survey undertaken by Robert, Schrijver, and Diels (2019a; 2019b) exploring the profile and background of live subtitlers from 27 countries showed that these professionals are on demand and that the in-house training model is still widely used among broadcasters and service providers. The industry can now hire practitioners with a background in audiovisual translation and/or subtitling, but in-house training is still available in the form of short courses focused on the use of speech recognition software, error correction and subtitling of live and hybrid contents (those that combine live sections with scripted segments that respeakers can use to prepare their subtitles beforehand). Depending on the depth with which each of those topics is covered, these in-house training programmes may last from a few hours to around three months (Robert, Schrijver, and Diels 2019a; 2019b).
As far as higher education is concerned, several universities have developed intralingual respeaking courses since 2005 (Eugeni 2008) and this media accessibility modality is currently being taught in around twelve universities across Europe (Romero-Fresco 2018; LTA 2019). These courses are mostly aimed at graduate students already trained in translation, interpreting and/or subtitling and, according to Robert, Schrijver, and Diels (2019b), they provide around 30 hours of instruction. The topics covered are normally respeaking theory, use of speech recognition software, and respeaking techniques and skills.

When it comes to pedagogical approaches to respeaking, Arumí Ribas and Romero-Fresco (2008) were the first to define a set of skills that professional intralingual respeakers should have. In the same article, these authors described the kind of practical exercises that could facilitate the acquisition of those skills in the classroom. Similar proposals followed in the next years by Eugeni (2009), Russello (2010) and, very recently, by the researchers involved in the European Commission-funded Live Text Access project, which tackles the need for harmonised training in real-time intralingual subtitling (LTA 2019). Even though they differ in several aspects, these pedagogical approaches coincide in that professionals should be able to master the following:

- **Linguistic skills**, which allow respeakers to comprehend the original message and to deliver it in the respoken subtitles.
- **Software-related skills**, which allow respeakers to properly dictate to the speech recognition software, and to change the subtitle format and position on screen.
- **Cognitive skills**, which allow respeakers to perform three main tasks:
  - Editing the original message in such a way that meaning does not get lost.
  - Splitting their attention when conducting several simultaneous activities.
  - Verbalising punctuation commands.

In order to work toward the acquisition of the aforementioned skills, intralingual respeaking instructors combine a variety of exercises that are typically used in interpreting courses (for instance, shadowing) and subtitling training (for instance, segmentation and condensation assignments), as well as some respeaking-specific practice (for instance, dictation assignments). In the following sections of this paper, we will argue that research exploring the quality of live subtitling may help trainers in this endeavour.

### 3. Live subtitling quality research as a didactic tool

Scholars working in live subtitling quality research analyse samples of live subtitles to trace their strengths and weaknesses, thus allowing to identify and promote best practices while facilitating informed interventions over those aspects that need to be reinforced. The final aim of these studies is to provide recommendations that will positively impact the respeakers’ work and, ultimately, the end users’ experience.

Live subtitling quality research is still scarce, but it has grown considerably in the last years. One of the first countries to require an official assessment of live subtitling quality was Canada. From 2012 to 2014, 265 programmes were analysed as part of an effort led by the Canadian Radio-television and Telecommunications Commission (CRTC) to improve the closed captions delivered in English (EBG 2014). In 2015, another large-scale quality study was completed in the UK under the lead of Ofcom and in collaboration with researchers at the University of Roehampton. This two-year project offered solutions to improve the quality of respoken subtitles on television, and led to the revision of the official guidelines in place in the
UK (Romero-Fresco 2016). Following in the footsteps of the Ofcom project, the ongoing QualiSpain study released its first tentative results in June 2019 (Fresno, Romero-Fresco, and Rico-Vázquez 2019). Although far less ambitious than the aforementioned projects, other studies have approached live subtitling quality by focusing on smaller corpora. In the US, for instance, the quality of live captioning has been explored in political debates, news programmes rebroadcast online and sport events (Fresno 2019; 2020; Fresno, Sepielak, and Krawczyk 2020). Additionally, several authors have analysed particular quality parameters of live subtitles, such as accuracy, speed, delay or display mode (Jensema, McCann, and Ramsey 1996; Eugeni 2009; Apone, Brooks, and O'Connell 2010; Luyckx et al. 2013), and more recently, scholars are expanding the focus of their analyses to include automatic subtitles in their research (Romero-Fresco and Fresno 2019; Rico-Vázquez 2021).

These kinds of studies inform about subtitling quality, but they also reveal interesting insights about the practitioners’ skills. Respeakers dealing with televised broadcasts need to multitask since they have to listen to the original message while respeaking to the software and monitoring their written output. Furthermore, they should do so with the least possible latency. Additionally, live programming is often fast-paced and may have multiple speakers interacting or even overlapping, which adds an extra layer of complexity. Given that respeaking is a cognitively demanding task (Szarkowska et al. 2016), it may be unrealistic to expect impeccable subtitling samples. Nevertheless, proficient respeakers should be able to minimise the errors in their subtitles, the severity of their mistakes, and find successful solutions in challenging situations. In this context, quality studies become valuable didactic tools since their live subtitling corpora may be used as source materials to undertake Performance Analyses (Svartik 1973).

Performance Analysis (Svartik 1973) originated in the field of Applied Linguistics as a response to Error Analysis, which was widely used in second language acquisition during the 60s and 70s (Richards 1980; Cherrington 2004). Error Analysis viewed the study of mistakes as a prominent aspect of the learning process, which had to be considered “essential within a pedagogical framework” (Granger 2002, 14). This idea has been embraced by scholars working in Translation Studies, where the analysis of poor solutions has been regarded as a useful didactic tool to train both translators and interpreters (Altman 1994; Kiraly 1995; Kalina 2000; Espunya 2014; Hurtado-Albir 2015). In the particular case of live subtitling, quality research is based on thousands of subtitles that have been marked for errors, as well as analysed for successful solutions. In this context, Error Analysis can be valuable for instructors in order to identify problem areas and work towards their reinforcement in the classroom. According to Corder (1975), three steps are needed to achieve this goal. First, a quantitative classification of errors that allows their systematic categorisation must be defined. Second, the frequency and severity of each type of error should be estimated so that each problem area can be prioritised. Third, the cause of each type of error should be investigated in order to undertake appropriate remedial measures. Quality studies such as QualiSpain constitute ideal source materials for Error Analysis since they compile subtitling mistakes and classify them according to their typology and severity. However, quality research may be more useful when approached from a Performance Analysis perspective that considers not only errors, but also those challenging aspects that professional respeakers solve in an effective manner.

For the QualiSpain pilot project, we assessed subtitling quality using the NER model, which is particularly well suited for training under a Performance Analysis approach. As we will further explain in the next section, NER classifies subtitling errors in two groups: those caused by poor decision-making on the respeakers’ part and those resulting from unsuccessful interactions.
between the respeaker and the speech recognition software. This distinction was indeed devised as an aid to training that would help instructors develop remedial actions to underpin specific skills. Additionally, since the accuracy of each subtitle is scored when using this instrument, trainers can easily identify areas of success. Training is, therefore, at the heart of this model. In fact, it is precisely this training orientation which has motivated several world-leading subtitling companies (AiMedia, Ericsson/Red Bee and Deluxe) to introduce NER as part of their in-house training programmes.

The next sections will describe the QualiSpain pilot project, its main results and their potential application to the training of intralingual respeakers following the aforementioned Performance Analysis approach.

4. QualiSpain pilot project: methodology

4.1. Materials

The corpus of this study included seventeen 10-minute samples of programmes aired by five Spanish broadcasters: TVE, Antena 3, Telecinco, La Sexta, and Cuatro. A total of 2,645 respoken subtitles belonging to news, chat shows, entertainment and sports programmes featuring football were assessed.

4.2. Quality assessment instrument

Following the recommendations in the Spanish Subtitling for the Deaf and Hard of Hearing standard (AENOR 2012), the samples included in this study were analysed with the NER model (Romero-Fresco and Martínez 2015). This instrument, which has been used as the reference method in many countries (Romero-Fresco 2016), was devised to assess the quality of live subtitles from a user’s perspective. This means that the analysis focuses on how viewers comprehend a subtitled programme and, to that end, NER compares the idea units (Chafe 1985) in the message uttered by the speakers to those provided in the respoken subtitles. Idea units, which could be understood as each piece of self-contained information within a sentence, may be independent (when they are essential to understand the message) or dependent (those that complement independent idea units and are not needed to transmit the main meaning of the sentence).

Under the NER model, discrepancies between the oral speech and the subtitles are considered errors only when meaning is lost, that is, when relevant information that was available to the hearing viewers is not included in the subtitles. When this is the case, NER classifies the errors according to their cause. If they were due to unsuccessful editing decisions (for instance, if respeakers omitted a whole sentence from their subtitles), they are considered editing errors. However, if they were caused by unsuccessful interactions between the respeakers and the speech recognition software, they are labelled as recognition errors. After this initial distinction, mistakes are further classified according to their severity as serious (when they provide an altered meaning that misinforms viewers), standard (when they omit independent idea units, thus compromising the viewers’ comprehension of the message) or minor (when they omit dependent idea units or involve petty mistakes, such as misspellings or punctuation issues).
As a result of applying the NER model to a set of respoken subtitles, a brief qualitative assessment is provided, together with a quantitative indication of the accuracy rate. For the subtitles to be considered of acceptable quality, they need to reach a minimum accuracy rate of 98% estimated according to the following formula, where \( N \) accounts for the number of words and punctuation marks in the subtitles, \( E \) shows editing errors and \( R \) reflects recognition errors:

\[
\text{Accuracy Rate} = \frac{N - E - R}{N} \times 100
\]

5. Results

This section will provide the results of our study, which will later be discussed and contextualised within live subtitling quality research in section 6.

5.1. Accuracy and reduction

The average accuracy rate found in this study was 97.72% (or 4/10 on a 10-point scale), which is close but below the quality threshold that the NER model deems as acceptable (98% or 5/10 accuracy rate). 12% of the programmes boasted an overall very good subtitling quality, 47% featured good or acceptable subtitles, and 41% showed substandard quality. Figure 1 shows the average accuracy rate per genre found in our study.

[FIGURE 1 here]

When it comes to reduction rate, the amount of text from the oral speech that was not included in the subtitles reached an average of 39%. Figure 2 shows the average reduction rate per genre.

[FIGURE 2 here]

As far as errors are concerned, 83% of the subtitling mistakes encountered in our samples were editing errors, whereas the remaining 17% were due to recognition issues. Figure 3 shows the distribution of editing and recognition errors per genre.

[FIGURE 3 here]

With regard to severity, 61% of the subtitling mistakes were minor, 36% were standard and 3% were serious errors. Figure 4 shows the percentage of minor, standard and serious errors per genre.

[FIGURE 4 here]

---

2 The NER model classifies subtitles according to their quality as follows: excellent (if they show an average accuracy rate over 99.5%), very good (99% - 99.5% accuracy rate), good (98.5% - 99% accuracy rate), acceptable (98% - 98.5% accuracy rate) and substandard (below 98% accuracy rate).
5.2. Speed

According to the Spanish standard (AENOR 2012), subtitles should not surpass 15 cps. In this study, the average subtitling speed was 14.49 cps, with 56% of all the subtitles delivered at over 15 cps, 44% of them exceeding 18 cps and 11% surpassing 25 cps. From all the samples assessed, news programmes showed the fastest subtitling speed (17.61 cps as average), followed by chat shows (14.27 cps), entertainment (12.81 cps), and sports (12.11 cps). Figure 5 shows the number of fast subtitles per genre.

[FIGURE 5 here]

5.3. Latency

The Spanish standard defines a delay up to 8 seconds as “reasonable” (AENOR 2012, 10). In our project we found an average latency of 8.6 seconds, with 59%, 38% and 18% of all the subtitles delayed over 8, 10 and 12 seconds, respectively. If analysed by genre, chat shows featured the largest delays (11.1 seconds as average), followed by sports (10 seconds), entertainment (9.2 seconds) and news programmes (5.8 seconds). Figure 6 shows the percentage of delayed subtitles per genre.

[FIGURE 6 here]

6. Discussion

Our results varied across genres. Therefore, to better interpret them and, especially to illustrate their application to respeakers’ training, it is worth making some distinctions per genre.

6.1. News

All the news programmes featured good or very good average quality. The average accuracy rate for this genre was 98.85% (7/10), very similar to that reported for news programmes in the Ofcom project (98.75%). Two factors contributed to these positive results in the case of the Spanish news: a low reduction rate and a limited number of subtitling errors.

With regard to reduction rate, the news samples showed an average of 19%, the lowest in this study. This figure is slightly higher than that obtained in the Ofcom project for news programs (13%), but it is substantially below the reduction rate found in the current research for chat shows (45%), entertainment (41%) and sports programmes (54%). When it comes to errors, news programmes showed the lowest percentage of subtitles with errors of any kind (22%). The vast majority of mistakes were due to poor editing (87%), rather than to misrecognitions (13%), and most of them were minor (78%), as opposed to standard (19%) and serious (3%).

The news segments in our corpus combined anchors’ scripted interventions with short pre-recorded videos and unscripted live connections with reporters. Respeakers were usually

---

3 The average subtitling speeds for the entire sample and for each genre were estimated following the traditional approach, that is, by dividing the total number of words in the subtitles by the exposure time of those subtitles.
provided with the presenters’ scripts and, sometimes, also with the videos beforehand. This allowed them to prepare their subtitles in advance, thus favouring better accuracy and fewer mistakes in those segments. In fact, only 10% of the subtitles accompanying anchors’ interventions showed errors of any kind. By contrast, the number of subtitles with one or more mistakes in live connections, which had to be subtitled in real time, increased to 47%.

The fact that pre-prepared subtitles were used in scripted sections had positive effects on accuracy, but it also led to an important drawback: it increased the subtitling rate. Pre-prepared subtitles showed verbatim or near-verbatim teleprompter extracts, whereas respeakers producing subtitles in real time often synthesised the speakers’ messages as a strategy to keep pace. This means that, in the same amount of time, pre-prepared subtitles displayed more words, which increased their speed over subtitles produced in real time. In this study, news were the genre with the fastest subtitling speed (17.61 cps as average). This figure exceeds the rate recommended by AENOR (2012) in the official Spanish standard (15 cps) and, according to Romero-Fresco (2015), samples with similar characteristics attract the viewers’ gaze between 60-70% of the viewing time, leaving 30-40% of the time to look at the images. Additionally, our news samples would possibly cause reading challenges to viewers due to the high number of subtitles delivered at fast speeds: 72% of all the subtitles from news programmes were presented at over 15 cps, 61% exceeded 18 cps and 16% surpassed 25 cps.

Finally, the fact that respeakers relied on scripted subtitles prepared beforehand for certain segments helped reduce latency to an average of 5.8 seconds. However, great differences were observed across the sections that composed the news samples. In those segments in which anchors read from the teleprompter, subtitles showed the lowest latency (3.9 seconds on average), followed by subtitles accompanying videos (5.1 seconds). However, the delay increased drastically in the segments with live pieces, which reached an average of 11.5 seconds and occasional latency peaks over 20 seconds. This illustrates that, while pre-prepared subtitles were in line with AENOR’s (2012) recommendations regarding latency, respeakers struggled to provide real time subtitles below the 8 seconds that the Spanish norm sets as acceptable. Actually, only 7% of all the subtitles produced for live pieces were within the acceptable lag.

6.2. Chat shows, entertainment and sports

The Ofcom project pointed at chat shows as particularly challenging for respeakers. In the UK, their average accuracy rate (97.9%) did not reach an acceptable quality level, as opposed to that of entertainment programmes, which obtained better scores (98.54% on average). In the present study, both genres showed poor results. Chat shows featured an average accuracy rate of 97.73% (4/10), with half of the programmes having good or acceptable accuracy levels and the other half showing substandard quality. In the case of entertainment samples, they showed an average accuracy rate of 96.54% (1/10), with half of the samples reaching acceptable scores and the other half presenting accuracy rates below 96% (0/10). When it comes to sports, a genre that was not explored in the Ofcom project, our samples showed an average accuracy rate of 97.15% (3/10), with only one programme reaching acceptable scores above the 98% quality threshold.

Since no scripts were available in advance for most of these programmes, their live subtitles were produced in real time. In order to cope with fast-paced speeches, respeakers provided heavily edited subtitles, which led to a substantial increase in the reduction rate for these genres: 45% for chat shows, 41% for entertainment programmes and 51% for sports as average. Although editing is not necessarily linked to loss of meaning, it may cause relevant information
to be inaccessible to the end user when performed inadequately. Correct editing was found in the present study, often to avoid unnecessary repetitions of the same idea. In the case of sports, the increased reduction had to do with the fact that our samples portrayed football games that featured some degree of redundancy between the visual information and the narrations provided by the announcers. Significant parts of their commentaries involved descriptions of the play-by-play, which the viewers could see on screen. A common subtitling strategy in these cases was to omit the play-by-play descriptions when the speech was rapidly-paced and would require the delivery of very fast subtitles. Part of the reduction in our chat shows, entertainment and sports samples involved the omission of redundant ideas or play-by-play descriptions and was therefore treated as correct edition. However, unfortunate editing decisions, which would negatively affect viewers’ comprehension, were also frequent in these samples. A closer look at the errors encountered for these three genres shed some light on this issue.

When it comes to chat shows, 35% of all the subtitles contained at least one error of any kind, which increased to 36% for sports and to 42% for entertainment programmes. As it happened in newscasts, most of the mistakes (80% for chat shows and 82% for entertainment and sports samples) were the result of poor editing strategies that caused relevant ideas to be missed, and only a small number of errors (20% for chat shows and 18% for entertainment and sports) were caused by recognition issues. Chat shows, entertainment and sports programmes still featured more minor errors as compared to standard and serious. However, while the percentage of serious errors remained very similar in all genres (2%-4%), that of standard and minor mistakes showed some interesting fluctuations. Namely, the proportion of minor errors decreased in chat shows, entertainment and sports samples as compared to newscasts (78% in news, 57% for chat shows and sports, and 53% for entertainment). Conversely, the ratio of standard mistakes more than doubled (19% for news, 39% for sports, 40% for chat shows and 45% for entertainment programmes). In practice, this translated into more omissions of independent idea units or full sentences, with 4-5/10 errors in chat shows, entertainment and sports programmes involving the loss of information necessary to understand the message.

This increase in the ratio of standard errors could be due to several factors. First, contrary to news programmes, in which anchors’ interventions followed a logically structured and clearly articulated script, most of the chat shows, entertainment and sports samples featured spontaneous speech that tended to be less organised and cohesive. Additionally, while newscasts usually showed one person speaking at a time, the other genres relied on several guests engaging in the same discussion, which often involved rapid changes of speaker, fast interactions and some degree of overlapping. This was perhaps more evident in the case of sports, since some of the commentators in our football games were journalists, but they shared the microphone with former football trainers or players. While the commentaries from all of them were spontaneous, those by some of the sportsmen could have posed particular challenges for respeakers because they were not always as articulate as the communication professionals. In some cases, the sportsmen interventions showed a poorer style peppered with repetitions, atypical collocations, occasional grammar mistakes and sentences being frequently interrupted to introduce new ideas. Since respeakers were delivering edited subtitles, reformulating this kind of imperfect speech probably required more time and effort. It was possibly the combination of the aforementioned factors with the fast speech rates found in chat shows, entertainment and sports samples which led respeakers to systematically omit information. While it probably allowed them to keep pace, it meant that a substantial amount of relevant information remained inaccessible to the end users.
With regard to speed, the average subtitling rate for chat shows was 14.27 cps, while that for entertainment programmes was 12.81 cps and that for sports reached 12.11 cps. These averages are below those recommended in the official Spanish standard (15 cps). However, a closer look at this parameter revealed that many subtitles were delivered at fast speeds: in chat shows, 55% of the subtitles exceeded 15 cps and 44% surpassed 18 cps; in entertainment samples, 44% of the subtitles were delivered at more than 15 cps and 28% showed speeds over 18 cps; in sports, 42% and 32% of the subtitles exceeded 15 cps and 18 cps, respectively. Therefore, even though the average subtitling speed would seem comfortable, nearly half of the subtitles in programmes belonging to these three genres could challenge the viewers’ reading skills and would leave very little time to look at the images (Romero-Fresco 2015).

Finally, the samples for these three genres showed high delays. In the case of chat shows, the average latency was 11.1 seconds, close to the average lag of live segments in the news (11.5 seconds). The average delay for sports reached 10 seconds and that in entertainment programmes was slightly lower, 9.2 seconds, but still above that recommended by the Spanish standard (8 seconds). The subtitles analysed for these three genres were seldom delayed below 5 seconds and their latency peaks surpassed 18 seconds. These lags are considerably higher than those reported in prior research exploring broadcasts with similar characteristics. For instance, the final US presidential debate had an average latency of 6.1 seconds (Fresno 2019) and the Ofcom project, which assessed live subtitling produced primarily by respeaking, concluded that chat shows and entertainment programmes in the UK were delayed 6 and 4.9 seconds, respectively (Romero-Fresco 2016). In the case of sports, Fresno, Sepielak, and Krawczyk (2020) found an average delay of 5.27 seconds for the Super Bowl 2018, which is substantially lower than that in the QualiSpain pilot project (10 seconds). Admittedly, the Spanish football samples could have posed more challenges to respeakers than those in the Super Bowl since they showed a more dynamic game pace, faster speech rates and more rapid changes of speakers with more frequent overlapping. Also, the subtitling production method could have played a role since respoken subtitles tend to show slightly higher delays than captions produced by stenography (Ofcom 2015). While these different characteristics may have influenced latency, it seems unlikely that they alone could explain such dissimilar results.

A more plausible explanation for this higher latency in Spain, which can also account for the low number of recognition errors in these samples, may have to do with the use of a subtitle validation process in the production of respoken subtitles. Often, after respeakers have dictated their subtitles to the speech recognition software, either a colleague or the respeakers themselves will read them to ensure quality (Romero-Fresco and Eugeni 2020). Normally, minor mistakes such as misspellings, punctuation errors or misrecognitions will be corrected before the subtitle is displayed on television for all viewers. This validation process is useful since it reduces the number of glaring mistakes, but it also increases the latency of the subtitles since all of them, regardless of whether or not they include errors, will be validated.

7. Transferring the QualiSpain findings to the classroom: a scaffolded training proposal

As mentioned in section 3, quality research can be used as a didactic tool to train respeakers in the making. We will now describe how the results of the QualiSpain pilot project may inform respeakers’ training by adopting a Performance Analysis approach that allows to look at problem areas and areas of success. This is used as a basis to present a scaffolded proposal for respeaking training that includes different levels, learning outcomes, exercises and materials.
Our analysis has helped to highlight a series of defining features of every genre that, on the one hand, explain some of the results obtained and, on the other, can be very useful in the training of respeakers. One of the main aspects of the news programmes analysed, for example, is their degree of scriptedness, which results in the key distinction between semi-live and live subtitling and accounts for some of the errors and successes identified in the study. In chat shows, the low degree of scriptedness is also an important consideration, as are the high speech rates and the interaction between several speakers, whereas in sports programmes, the redundancy between the images and the information included in the commentary has proved to be a very important element. With a view to transferring these findings to the classroom, the partners in the EU-funded ILSA project on live subtitling, led by the second author of this article, have produced a taxonomy to grade materials for live subtitling exercises (Pöchhacker and Alonso 2020). The guide is based on 3 levels (basic, intermediate and advanced) and five parameters: content (familiarity of subject matter, level of technicality, frequency of proper names, numbers and technical terms, etc.), language (lexical density, syntactic complexity [scriptedness], lexical frequency, enumerations & quotations, etc.), delivery (speech rate, intonation and unfamiliar accent), context (one speaker vs several speakers, noisy conditions, use of visual aids) and sound quality. Each parameter (except for sound quality) can be scored 1-3, which determines whether the audiovisual material to be used for respeaking training can be described as basic (a total score of 4-6), intermediate (7-9) or advanced (10-12). This guide allows trainers to gradually increase the difficulty of respeaking exercises as students progress throughout the course. Figure 7 shows this taxonomy.

[FIGURE 7 here]

Our findings also identified inappropriate editing as a shared problem area for chat shows, entertainment and sports programmes. These genres tend to feature fast speech rates, which often cause respeakers to make poor editing decisions and to omit important information (Dawson and Romero-Fresco 2021). The application of the NER model to our corpus also yielded numerous examples of correct editions, where respeakers managed to paraphrase and omit words from the audio without losing key information. This type of efficient editing can be tackled at two different stages in the training of respeakers. As suggested in Romero-Fresco (2011, 119), when starting their practice on dictation and respeaking, trainees can be tasked with identifying both their “optimum respeaking speed” (that is, the speed at which they are most comfortable and at which they can obtain their highest accuracy rate) and their “maximum respeaking speed” (the highest speed at which acceptable accuracy rates can be obtained). Once this is internalised, respeakers may find it easier to identify, even as they are respeaking, those programmes whose speech rates require them to reach (or go over) their maximum respeaking speed and where editing may be needed. Efficient editing can also be tackled at a second stage, when trainees are introduced to the notion of quality assessment through the NER model. As explained above, in order to apply the model trainees need to learn the difference between dependent and independent idea units, which can prove very useful to discern what type of information can be omitted in order to lose as little meaning as possible. The training of efficient editing can also benefit from incorporating another element that was revealed as important in the study presented in this article: the redundancy between images and commentary, where the latter can sometimes be omitted altogether from the subtitles. At an initial stage, trainees can be asked to watch sports programmes and weather reports and to read their transcripts in order to identify those segments that need not be included in the subtitles. Once this skill has been acquired “on paper”, they can attempt to put it into practice respeaking live programmes.
Finally, our study has also highlighted the key role played by the respeaking set-up chosen by every company with regard to error correction. In Spain or France, respeakers often validate their subtitles (and attempt to correct errors) before broadcasting them on air. This diminishes the number of recognition errors but increases delay. In countries such as the UK and the US, keeping delay to a minimum is the main priority, which means that (some) errors are corrected but only after they have been broadcast to the viewers. Either way, since error correction is time-consuming and can end up causing further delay and omission of information, respeakers must be able to ascertain, as they are respeaking live, what errors are worth correcting and what errors could be ignored. Here, the NER model may come in handy again, especially the distinction between minor, standard and serious errors. This proved useful in the Ofcom project, where professional respeakers in the UK were advised to correct serious and, whenever possible, standard errors, but not minor errors, which, by definition, do not impact negatively on comprehension (Romero-Fresco 2016). Although error correction has traditionally been absent from respeaking training (Romero-Fresco 2012), its importance in the research study included here has prompted its inclusion in the respeaking course created for the ILSA project. Likewise, until now the NER model has normally been introduced in the final stages of respeaking training (Arumí Ribas and Romero-Fresco 2008), but it may make more sense to introduce it at the beginning in order to help trainees acquire both editing and error correction skills. Once again, this shows the role played by the NER model as a bridge illustrating the connection and cross-fertilisation between research and training in the area of live subtitling.

Drawing on the findings of our Performance Analysis (errors and successes) and the work conducted within the framework of the ILSA project (especially the above-mentioned taxonomy), Table 1 presents a scaffolded proposal for respeaking training including different levels, learning outcomes, exercises and materials.

[TABLE 1 here]

8. Conclusions

This article has presented the main findings of the QualiSpain pilot project, which constitutes the initial step toward the first national assessment of live subtitling quality on Spanish television. Our pilot project analysed subtitles from news, chat shows, entertainment and sports programmes looking at their accuracy, latency and speed. Its main findings indicated that news samples featured an overall good quality, while those for the rest of the genres did not reach the minimum accuracy rate that the Spanish norm deems as acceptable. Our data revealed that recognition mistakes, most of which did not affect the viewers’ comprehension, were less frequent than edition errors, which often involved the loss of full sentences or independent idea units necessary to understand the message. Poor editing decisions caused the accuracy rates to drop, especially when respeakers had to produce real time subtitles for spontaneous discourse or for fast speakers interacting with each other. Possibly, respeakers skipped information as a way to keep pace and not increase latency. Even so, the delay found in our unscripted samples was unusually high and partly due to a validation process that involves the revision of all the subtitles produced. Fast subtitling speeds with the potential to challenge viewers’ reading skills were the last problem area identified in all the genres that we explored.

This paper has also discussed fruitful connections between respeaking professional practice, research and training. As part of our quality analysis, our corpus was marked for errors
according to the NER model, the quality assessment instrument recommended by the Spanish subtitling standard. Since training is central to this model, NER classifies mistakes according to their type and severity, which allowed us to trace the main problem areas for professional respeakers in Spain. Additionally, adopting a Performance Analysis approach, we also identified the areas of success. We have argued that this information, together with the work conducted within the ILSA project, can be valuable for respeakers’ training, be it delivered by higher education institutions, subtitling companies or broadcasters. This is illustrated by the scaffolded proposal for respeaking training presented in Table 1, which includes different levels, learning outcomes, exercises and materials for respeaking training.

Considering the strengths and weaknesses of live subtitling when designing respeaking courses should translate into future respeakers that are better equipped to deal with the current challenges, which should, in turn, lead to better quality in the services provided. As this article has argued, quality studies can help connect the dots between live subtitling research, training, professional practice and end users.

References

AENOR. 2012. “Subtitulado Para Personas Sordas y Personas Con Discapacidad Auditiva (UNE 153010).”


