

Closed Captions *en español*. Live Captioning Quality in Spanish-language Newscasts in the U.S.

Nazaret Fresno¹, The University of Texas Rio Grande Valley

ABSTRACT

The broadcasting landscape in the United States includes a limited number of Spanish-speaking networks that serve the Hispanic community. As part of the accessibility regulations in place, the Federal Communications Commission issued the first closed captioning quality rules in 2014, which apply in the same manner to English and Spanish pre-recorded and live programming. Over the last few years, comprehensive studies assessing the quality of closed captioned broadcasts have begun to emerge, either led by or in collaboration with Media Accessibility researchers. These kinds of projects have shed some light on the strengths and weaknesses of current live closed captions, and have gathered valuable data that may be used to improve the training of practitioners. This article presents the main findings of the first study aimed at assessing the quality of the closed captions that accompany Spanish-language programming in the United States. By focusing on the national newscasts, it discusses the completeness, placement, accuracy and synchronicity of the closed captioning in Spanish currently delivered on television. Results show excellent completeness and placement, an average captioning speed aligned to the existing guidelines, and improvable delay and accuracy under the NER model.

KEYWORDS

Closed captioning, live closed captioning, subtitling for the deaf and hard of hearing, live subtitling, quality, accuracy, media accessibility.

1. Introduction

When the first station with a full programming in Spanish was created in the United States (U.S.) in 1955 (Wilkinson, 2016), the Hispanic broadcasting industry was a minor niche. However, the growth of the Hispanic population that took place between the 1980s and the 1990s catalysed the development and expansion of Spanish-speaking television networks (Retis, 2019). The linguistic landscape of the country led broadcasters to adapt their programming to appeal not only to their traditional monolingual viewers but also to second-generation Hispanics whose dominant language was English but had learnt Spanish at home. Today, the Hispanic community represents 19.1% of the total population of the U.S. and, with 63.7 million individuals, it constitutes the largest ethnic minority in the country (US Census Bureau, 2023). It is estimated that 13.3% of the households use Spanish to communicate (US Census Bureau, 2022). This means that over 42 million people in the U.S. speak Spanish at home, a figure that comes close to the entire population of Argentina or Spain (Instituto Nacional de Estadística, 2023; Instituto Nacional de Estadística y Censos de la República Argentina, 2022). In terms of their perceived language ability, over 16 million Hispanics report speaking English “less than very well” (US Census Bureau, 2022).

When it comes to TV viewing habits, 71% of Latinos consume some news in Spanish, but this percentage reaches 89% among foreign-born individuals living in the U.S. (PEW Research Center, 2018a). As far as national news is concerned, the most recent data available estimate that Univision and Telemundo, the two major Spanish-speaking networks in the U.S., had a combined audience of over three million viewers

¹ ORCID 0000-0002-6702-159X; E-Mail: nazaret.fresno@utrgv.edu



in 2018 (PEW Research Center, 2018b). This explains why Hispanic television is believed to have a competitive commercial niche today (Retis, 2019).

From a regulatory point of view, closed captioning in Spanish is subject to the same requirements as English closed captions. This means that since the Federal Communications Commission (FCC) issued its quality regulations in 2014 all the new programmes in Spanish must be closed captioned, as well as 75% of all the content created before this rule was adopted. According to the FCC (2014), if not exempt, the closed captions accompanying Spanish programmes need to be complete, properly placed, accurate and in synchrony with the images. In other words, the entire programme has to be captioned in such a way that relevant information, such as graphic elements or the speakers' mouths, are not obscured by the captions. The closed captioned text needs to be as verbatim as possible, and it should coincide with the dialogs to the fullest extent possible. Finally, closed captions should be delivered at a pace that allows their reading. These regulations apply to both pre-recorded and live programmes, although the FCC (2014) acknowledges that achieving completeness, placement, accuracy and synchronicity becomes more challenging in the case of live closed captions.

Given the relevance of the Spanish-language networks in the broadcasting industry, this article will explore the quality of the closed captioning delivered in the national news programmes broadcast in Spanish in the U.S.

2. Prior research on closed captioning quality and reception

Although comprehensive quality assessments are not plentiful, they have increased in the past years as a response to the concerns raised by the deaf and hard of hearing community about the quality of the closed captions accompanying live broadcasts. Most of these studies have traditionally focused on closed captions in English. For instance, the first large study on closed captioning accuracy was conducted in the U.S. and dates back to 2003, when Jordan et al. (2003) analysed the captions delivered in local and national news programmes. Using a methodology in which five coders rated a set of samples individually, the researchers explored, among other things, how closely the captions matched the audio track and if they were easy to understand. The research team concluded that only 25% of the local news featured clear closed captions, 32% included inaccurate captions that would be very difficult to comprehend, and 43% were somewhere in between. When it comes to national news, results were better: 82% of the captioned programmes were described as "clear", only 18% were "somewhat clear" and almost 65% of all the problems encountered were minor (Jordan et al., 2003, p. 11). A total of 73% of the samples they analysed contained substitutions, omissions or deletions, while typos and technical issues were present in 50% of the local news and 62% of the national newscasts. Also in the U.S., after several metrics to assess accuracy were created or applied to the captioning field, Apone et al. (2011) evaluated the accuracy of the live captions of 20 news programmes using the Weighted Word Error Rate (WWER)¹. The researchers concluded that 55% of their samples featured good accuracy and 10% of the programmes had poor closed captions that would hamper comprehension.

Two years later, in 2013, the UK regulator Ofcom led the largest study conducted to date, which assessed the quality of more than 78,000 live subtitles over a two-year

period. This project used the NER model, which was designed by Romero-Fresco and Martínez in 2015 and is described in section 3.2.2 of this article. As explained in Romero-Fresco (2016), the Ofcom project covered mainly respoken subtitles and news programmes proved to be the genre with the highest accuracy rate (98.75%)². Entertainment programmes were slightly below (98.54%), and talk shows did not make it to the minimum accuracy threshold (97.9%). As far as reduction rate³ is concerned, newscasts showed the lowest figure (13%), followed by entertainment programmes (23%) and talk shows (32%).

Focusing again on the U.S., where live closed captions have traditionally been produced using stenography technologies, several smaller-scale accuracy and quality studies have also been conducted using the NER model. When it comes to programmes in English, the closed captions delivered during the 2016 final presidential debate and during the 2018 Super Bowl were analysed for quality. The assessment of more than 9,400 captions from the political event revealed a 98.84% accuracy rate on average, and a reduction rate close to 6% (Fresno, 2019). In the case of the Super Bowl, results were even better, with an average accuracy rate of 99.42% and a reduction rate slightly over 5% (Fresno et al., 2021). More recently, Fresno (in press) studied the accuracy of the closed captioning of the national newscasts broadcast by four networks. The author found an average accuracy rate of 98.83% with a 6% reduction rate. Given that the present article focuses on closed captions in Spanish, the ongoing QualiSub project, led by Universidade de Vigo in collaboration with the main broadcasters in Spain, could also be of interest. The preliminary findings obtained so far point at an average accuracy rate of 98.85% for news programmes, with a 20% reduction rate (Fresno et al., 2019).

Although accuracy is often considered a reflection of quality, parameters such as the live captioning speed and delay should also be part of any quality assessment due to their influence on the viewers' comprehension. As far as subtitling speed is concerned, the news programmes in the Ofcom project were delivered at 152 wpm on average (Romero-Fresco, 2016), close to the 146 wpm (12.420 cps) found in the 2018 Super Bowl (Fresno et al., 2021), but below the average captioning speed reported by Fresno (2019) for the 2016 presidential debate (167 wpm or 14.454 cps). When it comes to news programmes in the U.S., the captioning speed in Fresno's (in press) samples was 190 wpm (17.772 cps). In the Spanish-speaking world, QualiSub found an average subtitling speed of 177 wpm (16.779 cps) for this genre (Fresno et al., 2019).

In terms of latency, the news programmes in the Ofcom project were delayed 5.2 seconds on average (Romero-Fresco, 2016). In the case of the 2016 political debate, captions lagged 6.1 seconds (Fresno, 2019), slightly below the 5.3 seconds of the Super Bowl 2018 (Fresno et al., 2021). When it comes to Spanish live subtitles, QualiSub news showed 6.6 seconds of latency on average (Fresno et al., 2019). Overall, the quality studies conducted to date have shown good or very good accuracy levels in different genres in the U.S., as well as manageable average speeds and latencies. Furthermore, news programmes feature good accuracy rates on average in the UK, the U.S. and Spain. Results also indicate that, as could be expected, reduction rate is lower in the U.S. due to their preference for verbatim captions, as opposed to the European tradition, more prone to edition.

Also relevant for the purposes of this paper are some studies that have explored the reception of subtitled materials. When it comes to subtitling rate, Jensema (1998) conducted an experiment with deaf and hard of hearing adults and concluded that materials presented at 145 wpm were read comfortably, but those exceeding 170 wpm compromised comprehension. Using a different methodology, Sandford (2015) found in the UK that his participants considered 171 wpm to be a convenient speed for scrolling subtitles and 177 wpm for block subtitles. Both Jensema and Sandford explored users' subjective preferences, which are not aligned to the results obtained in experimental research dealing with comprehension. For instance, Romero-Fresco (2012) pointed out that news pieces presented at 180 wpm were challenging for 53% of his deaf and hard of hearing respondents, with this percentage increasing up to 80% when the subtitles were delivered at 220 wpm. These findings are consistent with Burnham et al.'s (2008) results, who found that captioned documentary films were challenging for his deaf and hard of hearing participants with lower reading skills when presented at 130 wpm.

This effect of speed on comprehension has also been the focus of several researchers that have used eye-tracking technologies to explore the reception of subtitled content. Liao et al. (2021) pointed out that, when faced with very fast subtitling speeds, viewers spend less time fixating on individual words, and both word skipping and saccade length increase. Similarly, Kruger et al. (2022) argued that as speed increases, subtitle processing becomes incomplete and more superficial. These findings, together with those from additional eye-tracking studies, outline a complex scenario where a variety of different factors (subtitling speed, viewers' reading skills, programme complexity or the linguistic features of the subtitles) could potentially interact and influence our comprehension (de Linde & Kay, 1999; Jelinek Lewis, 2001; Miquel Iriarte, 2017; Szarkowska et al., 2011, 2016).

Finally, some eye-tracking research has been done in order to explore the viewers' attention distribution when watching subtitled programmes. Jensema et al. (2000) found that closed captions delivered faster caused the users to look at them longer. And, according to Romero-Fresco (2015), when block subtitles are presented at 150 wpm, viewers can use 50% of their viewing time to read the captions and the remaining 50% to look at the images. However, as subtitling speed increases, so does the time spent on the captioning area, which reduces the time available to see the images.

3. Live captioning quality in Spanish-language newscasts in the U.S.: methodology

3.1 Materials

In this study, we worked with 20 ten-minute samples of news programmes broadcast live by the two main Spanish-speaking networks: Telemundo and Univision. In total, 5349 closed captions were included in our samples. All the segments were extracted from national newscasts and were recorded from television in the southmost Texas area using a *Hauppauge WinTV-HVR-1955* TV tuner. Their closed captioning files were then extracted using *CC Extractor GUI 0.82* and the selected segments were assessed by the author of this article taking into account the quality criteria as defined by the FCC (2014): completeness, placement, synchronicity and accuracy. To ensure representativity and to counterbalance any effects that captioners' tiredness could

cause, a similar number of segments from the beginning, middle and end of the programmes were included in the analysis.

3.2 Method

3.2.1 Completeness, placement and synchronicity

In order to analyse the completeness and placement of the closed captions on screen, all the samples were watched to verify that they were captioned from beginning to end, and that no relevant information, such as graphic inserts or the speakers' mouths, was blocked by the captions.

In terms of synchronicity, both captioning speed and delay were taken into account. As pointed out by Fresno & Sepielak (2020), several methods have been used in Media Accessibility research to estimate subtitling speed. In order to allow comparisons with previous studies, this paper will report on the average captioning speed calculated following two methods:

(1) As the total number of characters (Ch) in a sample divided by the time of exposure (T) of those captions:

$$\text{Average Captioning Speed} = \frac{Ch}{T}$$

(2) By calculating the speed (SP) of each subtitle in a sample ($SP1... SPn$) and defining its average speed as the mean of those values:

$$\text{Average Captioning Speed} = \frac{SP1+SP2+\dots+SPn}{n}$$

As far as latency is concerned, the delay was estimated by synchronising the 5349 captions so that each of them was displayed on screen exactly when the speaker was saying those words. This process was done semi-automatically with the aid of subtitling software (Subtitle Edit 3.5.16) and forced alignment technologies (The Munich Automatic Segmentation System MAUS, developed by the Institute of Phonetics and Speech Processing Ludwig-Maximilians-Universität München, n.d.). These tools were used to automatically align the orthographic transcriptions of the texts to their audio recordings. In order to make sure that the MAUS-timed subtitles were correct, each video was played together with its MAUS-generated .srt file. Some instances of inaccurate alignment were identified, for instance in the case of pieces with noisy background. The timing of these subtitles was manually corrected to ensure proper spotting. Latency was then estimated by comparing the initial time of each subtitle in the original .srt file (which contained delayed subtitles) to the same subtitle time in the re-synchronised .srt file (which contained the correct spotting).

3.2.2 Accuracy

In order to assess accuracy, an adaptation of the NER model (Romero-Fresco & Martínez, 2015) was used. The original instrument estimates an accuracy rate for a given captioning sample as follows:

$$\frac{N - E - R}{N} \times 100$$

“N” reflects the number of words and punctuation marks contained in the subtitles. “E” are edition mistakes, which are caused by incorrect editing strategies used by the practitioner (for instance, the omissions of relevant information). “R” shows recognition errors, which are the result of unsuccessful interactions between the subtitlers and their technical equipment. After this initial categorisation, the NER model further classifies mistakes according to their severity in minor (those that do not have a negative effect on users’ comprehension), standard (those that are noticed and hamper the viewers’ comprehension), and serious (those that deliver misleading information to the user that cannot be identified as such). For instance, misspellings would be classified as minor errors, omissions of full sentences would be standard mistakes, and incorrect figures would be considered serious errors.

Although the NER model was devised to assess the accuracy of respoken subtitles, it has also been used to analyse closed captions produced by stenography technologies and automatic subtitles (Romero-Fresco & Fresno, 2023). Respoken and automatic subtitles rely on a speech recognition process and the distinction between edition and recognition mistakes is often quite clear. However, when it comes to stenography-produced closed captions, this classification is not always straightforward since the cause of some errors is difficult to identify⁴. In the current research, it was not possible to confirm the production method of the captions in our samples⁵. Some of them included what looked like mistakes commonly found in respoken samples (e.g., *construir un segmento de baya* [to build a section of the berry] instead of *construir un segmento de valla* [to build a section of the wall]). However, we also found atypical errors (e.g., *11 tener esta pared* [11 have this wall] instead of *vamos a tener esta pared* [we are going to have this wall]). This is why, following the same approach used in prior studies undertaken in the U.S. with the NER model (Fresno, 2019; Fresno et al., 2021), errors were classified according to their severity without distinguishing between edition and recognition.

The NER model was preferred over other accuracy instruments available because it accounts for correct edition, that is, instances in which the captions are not verbatim but keep the same meaning as the original text. As opposed to word-based accuracy metrics, such as the Word Error Rate (WER) or the Weighted Word Error Rate (WWER), which consider word substitutions or deletions as mistakes, the NER model focuses on meaning and, hence, treats different wording as an error only when captions omit relevant information or when they introduce an altered meaning as compared to the original message. Because it is not possible for live captioning practitioners to always provide fully verbatim closed captions (see Fresno, 2019; Fresno et al., 2021), an instrument that takes correct edition into account seems more appropriate to assess accuracy.

4. Live captioning quality in Spanish-language newscasts in the U.S.: results and discussion

This section will summarise the main findings of the study described in section 3 according to the four quality parameters defined by the FCC (2014): completeness, placement, synchronicity and accuracy.

4.1 Completeness and placement

All the samples included in this research were subtitled from beginning to end using roll-up scrolling captions displayed in two or three lines. Closed captions were located at the bottom of the image and appeared on screen without blocking graphic inserts or the speakers’ mouths. Therefore, in terms of completeness and placement, our closed captions proved impeccable.

4.2 Synchronicity

When it comes to synchrony, the captioning speed and the delay in our samples were analysed.

4.2.1 Speed

The average captioning speed estimated as the total number of characters divided by the time of exposure of the captions was 11.574 cps (or 121 wpm)⁶, with an average captioning speed of 11.863 cps in the programmes aired by Telemundo, and of 11.240 cps in those broadcast by Univision. Table 1 shows the average speed, range and standard deviation in the study.

	Average (cps)	Max. speed (cps)	Min. speed (cps)	Range	SD
Telemundo	11.863	52.727	1.740	50.987	10.694
Univision	11.240	51.111	0.460	50.651	12.011
All broadcasters	11.574	52.727	0.460	52.268	11.181

Table 1. Average speed, range and standard deviation.

It is interesting to note that 39% of our captions exceeded 15 cps, 27% surpassed 18 cps, and 23% were above 20 cps⁷. Figure 1 shows the percentage of fast closed captions per network.

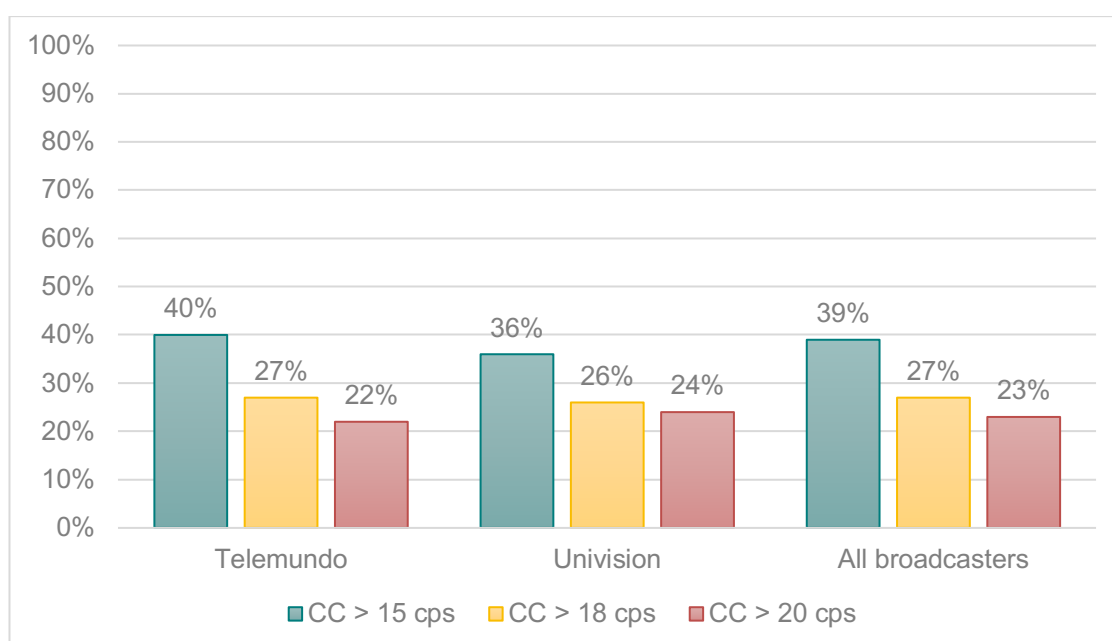


Figure 1. Percentage of fast captions per broadcaster.

The overall captioning speed, as well as the average per network (all of them between 11.240 and 11.863 cps) are aligned to the UNE 153010 standard, which regulates subtitling in Spain and recommends that subtitles for deaf and hard of hearing viewers do not exceed 15 cps (AENOR, 2012). Interestingly, the average captioning speed found in our samples across all broadcasters (11.574 cps or 121 wpm) falls behind that reported for news programmes in similar research projects, such as QualiSub (16.779 cps) or Ofcom (152 wpm) (Fresno et al., 2019; Romero-Fresco, 2016). Furthermore, the percentage of fast captions (39%, 27% and 23% delivered at over 15, 18 and 20 cps) is well below that revealed in news programmes in Spain (65%, 55% and 47%, respectively)⁸. These results can be better contextualised by looking at the speech rate and reduction rate of the samples.

The average speech rate found in the newscasts analysed in the present study across all broadcasters was 167 wpm, with 164 wpm in the segments aired by Telemundo and 170 wpm in those by Univision. This is lower than the data obtained in prior research exploring the news programmes aired in English in the U.S. (172 wpm) and the UK (175 wpm), and in Spanish in Spain (169 wpm) (Fresno, in press; Fresno et al., 2019; Romero-Fresco, 2016). In terms of reduction rate, our samples reached 26% (23% for Telemundo and 30% for Univision). Possibly, the unhurried oral delivery by the news anchors, together with the edition applied to the captions, which reduced the message approximately in one fourth, favoured a generally reasonable captioning pace. However, our samples still presented a number of fast captions delivered at speeds that prior research has signalled as potentially problematic in terms of attention distribution (Jensema et al., 2000; Romero-Fresco, 2015) and as challenging with regard to comprehension (Burnham et al., 2008; Jensema, 1998; Romero-Fresco, 2012).

Also noteworthy is the fact that, as seen in previous reception studies, the linguistic features of the subtitles and the complexity of the audiovisual materials could play a role in the interaction between captioning speed and comprehension (de Linde & Kay, 1999; Jelinek Lewis, 2001; Miquel Iriarte, 2017; Szarkowska et al., 2011, 2016). This idea has not been thoroughly researched so far in the subtitling domain, but given that newscasts often touch upon more complex issues than other genres and that it is very hard to keep captioning speed under a specific threshold while captioning live, the interaction between programme complexity, linguistic features of the closed captions, and captioning speed deserves further attention.

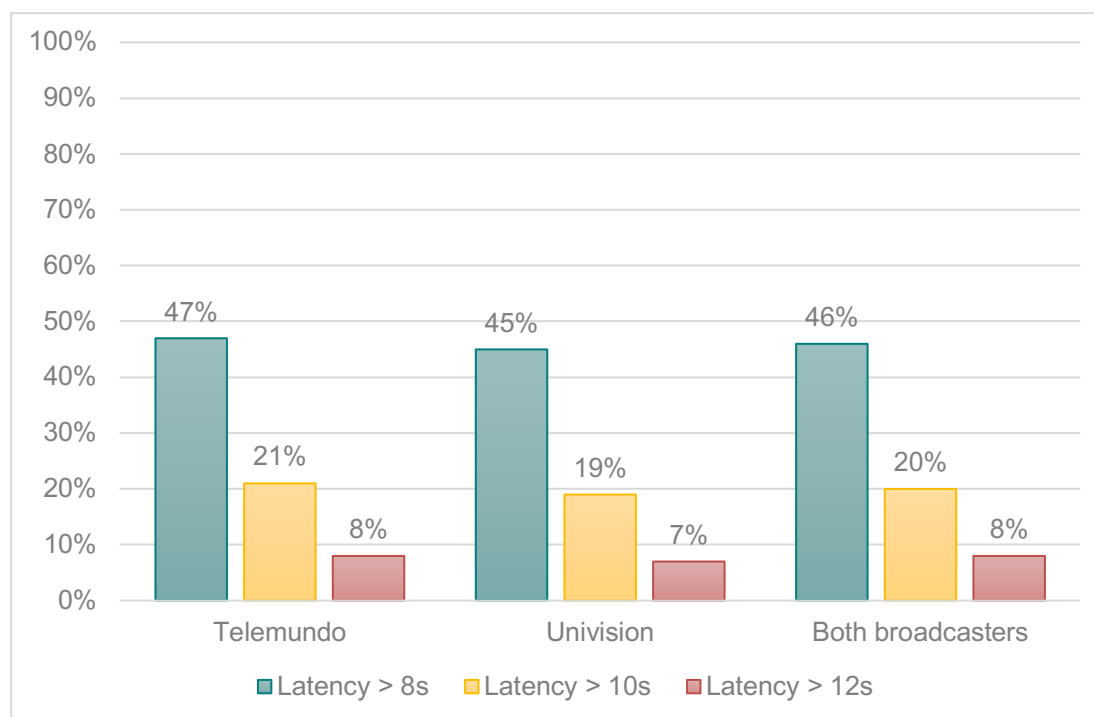
4.2.2 Latency

The closed captions in the current study had a delay of 8.2 seconds on average (8.2 seconds in the programmes aired by Telemundo and 8.3 seconds in those by Univision). Table 2 shows the average latency, range and standard deviation.

	Average latency	Max. latency	Min. latency	Range	SD
Telemundo	8.2	24.9	2	22.9	2.7
Univision	8.3	41	0.7	40.3	3.3
All broadcasters	8.2	41	0.7	40.3	3

Table 2. Average latency, range and standard deviation.

A total of 46% of all the closed captions in the corpus were delayed over 8 seconds, 20% surpassed 10 seconds, and 8% had a latency of 12 seconds or more⁹. In a few occasional instances, latency peaks close to 20 seconds were identified. Figure 2 shows the closed captioning delay per broadcaster.

**Figure 2. Delayed closed captions per broadcaster.**

To the best of our knowledge, this is the first study that analyses the latency in news programmes broadcast in the U.S. in English or Spanish. When compared to the results obtained for other genres, the average delay in our samples across all broadcasters (8.2 seconds) is higher than that found in the 2016 final presidential debate (6.1 seconds) and in the Super Bowl LII (5.3 seconds) (Fresno, 2019; Fresno et al., 2021).

If we look at the results of prior research undertaken with English programming in other countries, the latency reported for entertainment programmes and talk shows in the UK (4.9 and 6 seconds, respectively) was lower than the delay in our study. When it comes to news programmes, the Ofcom project revealed an average latency of 5.2 seconds (Romero-Fresco, 2016), and the tentative QualiSub results pointed at 6.6 seconds in Spain (Fresno et al., 2019). Respeakers in the UK and Spain often have access to the newscast script a while before its live broadcast, which allows them to prepare the subtitles of the scripted part beforehand. These pre-prepared subtitles are launched as the anchors read from the teleprompter and are combined with subtitles produced in real time for the unscripted segments (for instance, live connections with reporters). This hybrid method decreases latency and mistakes since the pre-prepared subtitles are typed and corrected before the programme begins. In the U.S., however, it is not customary to share the newscasts scripts. Although the hybrid mode approach could help explain the higher latency found in our samples as compared to the averages encountered in the UK and Spain, drilling down the QualiSub preliminary

data and distinguishing between pre-prepared subtitles and those created in real time may be helpful to make more sense of our results. This refined analysis revealed that in-studio segments, which had subtitles prepared beforehand, were delayed 5.1 seconds on average, whereas the lag of the pieces featuring connections with reporters, which were subtitled in real time, increased to 9.8 seconds. Therefore, the delay found in our U.S. samples is actually lower than the latency seen in Spain for news subtitles produced in real time. Furthermore, the average latency in the present study (8.2 seconds) is aligned to the 8 seconds that the subtitling for the deaf and hard of hearing standard in Spain considers acceptable (AENOR, 2012).

The quality studies undertaken so far have revealed greater latencies in Spanish live subtitles and captions than in their English counterparts. As explained before, this could be due to the use of different methods (e.g., stenography vs respeaking) and practices (e.g., use of scripts to prepare captions in advance, validation process, etc.). However, according to the FCC (2014), the closed captioning industry in the U.S. considers a 6-7 second lag to be acceptable in live programmes, although 4-5 is the target that captioners should aim at. In the case of our newscasts, these latency ranges seem too optimistic. In fact, none of the segments in our samples featured a latency below 5 seconds, and the captions in 85% of our samples were delayed more than 7 seconds. The limited quality studies that have been undertaken to date in the U.S. seem to confirm that the aforementioned industry estimations are attainable when it comes to English captions (Fresno, 2019; Fresno et al., 2021). However, the findings presented in this paper suggest that Spanish news programmes may not adjust to these latency recommendations so well. At this stage, more quality studies are needed to assess the delay in the Spanish live captions accompanying other genres in order to evaluate if their current latency is in line with the industry expectations. If it is not, strategies to decrease latency would need to be explored and, should they not be feasible, the industry may want to issue recommendations that adjust to what the current live captioning technology and captioners can deliver when working in Spanish.

4.3 Accuracy

Our samples featured an average accuracy rate of 97.04%, with the news from Telemundo averaging 97% and those from Univision reaching 97.10%. The reduction rate was 26% on average (23% for Telemundo and 30% for Univision). A comprehensive description of the accuracy and the issues encountered in these samples can be found elsewhere (Fresno, 2021). Figures 3 and 4 show the accuracy rate and the reduction rate per sample.

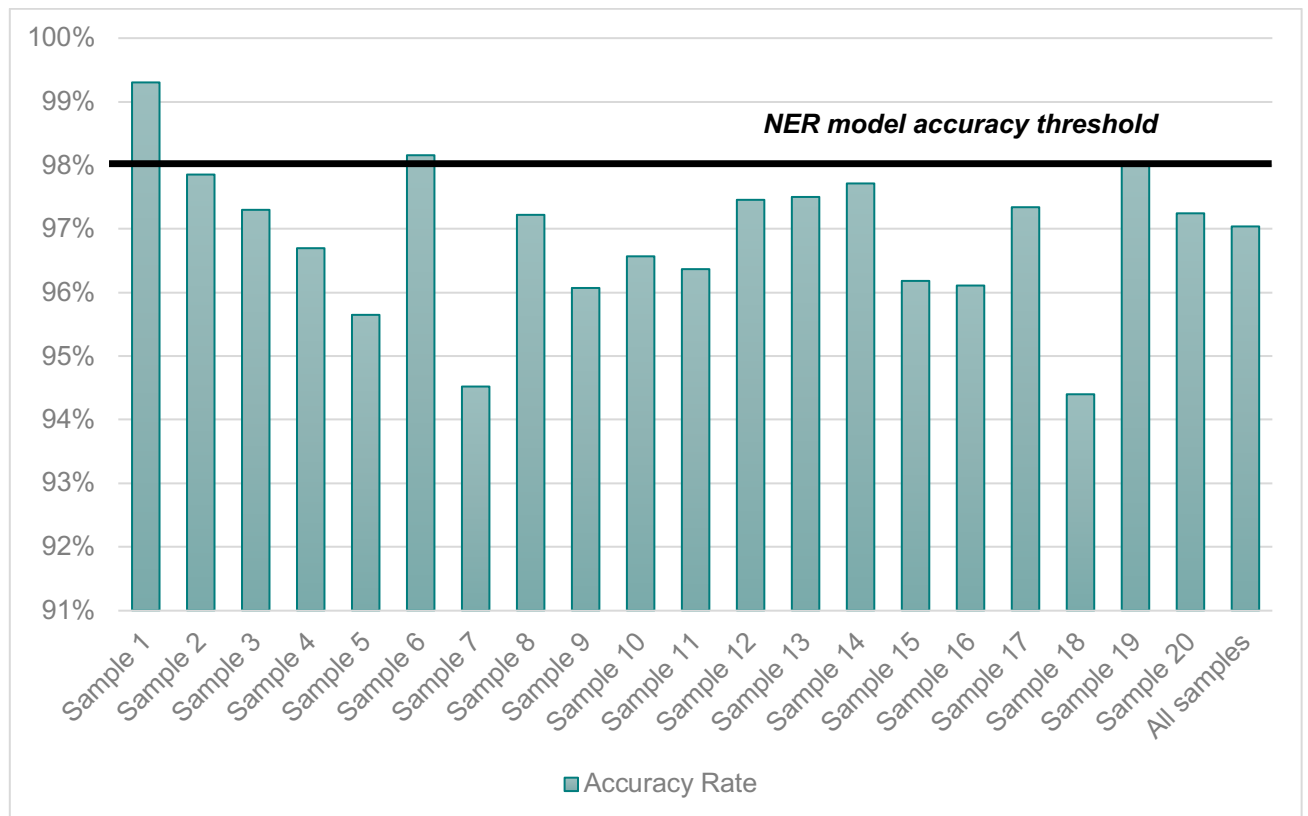


Figure 3. Accuracy rate per sample.

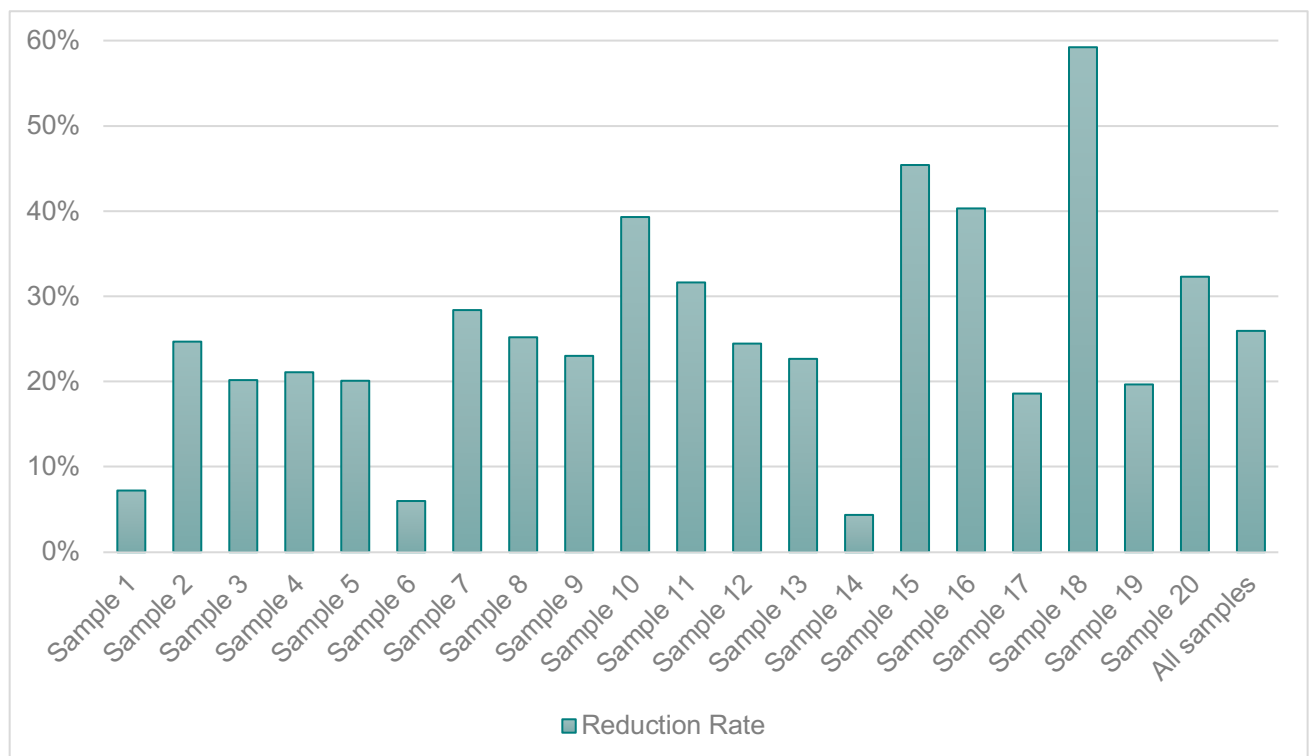


Figure 4. Reduction rate per sample.

The samples analysed in this research included three types of segments: in-studio pieces usually featuring the news anchor reading from a teleprompter, videos, which were narrated by a journalist and oftentimes included interviewees speaking one at a time, and connections with reporters, which covered breaking news or other important

stories. Given the different characteristics of these materials, they were analysed separately to identify potential accuracy variations across these groups.

In terms of accuracy rate, there was virtually no difference. Videos averaged 97.11% and were closely followed by in-studio segments (96.98%) and connections with reporters (96.90%). All the groups showed poor accuracy results below the minimum threshold that the NER model deems acceptable, which was the result of high reduction rates and elevated numbers of errors.

When it comes to reduction rate, it was high and, again, it showed minimal variation across groups. In-studio stories had the least reduction (25%), slightly below videos and stories broadcast by reporters (26% and 28%, respectively). According to these data, the reduction rates of all three groups, as well as that of the entire corpus (26%), were substantially higher than those previously reported in the U.S. (5% in the Super Bowl, and slightly over 6% in the 2016 final debate and in the news in English) (Fresno, in press, 2019; Fresno et al., 2021). Furthermore, in the case of the newscasts, while fifteen of the Spanish-language samples had a reduction rate that exceeded 20%, only one of their English counterparts showed comparable reduction levels (Fresno, in press).

Interestingly, the reduction rate in the present study was more aligned with the QualiSub findings, which showed a 30% reduction rate on average in the news segments that included live subtitles. This figure is similar to the average reduction rate in our in-studio stories (25%), videos (26%) and connections with reporters (28%), all of which featured captions in the same language (Spanish) that were also produced in real time.

Together with reduction rate, errors caused accuracy to drop below par. In terms of distribution, in-studio segments, videos and connections with reporters showed common trends. Minor errors were the most frequent in all three groups, followed by a limited number of standard mistakes and scarce serious errors. Videos showed slightly more standard and serious mistakes (20%) than in-studio segments (17%) and connections with reporters (13%). Figure 5 shows the error distribution per group.

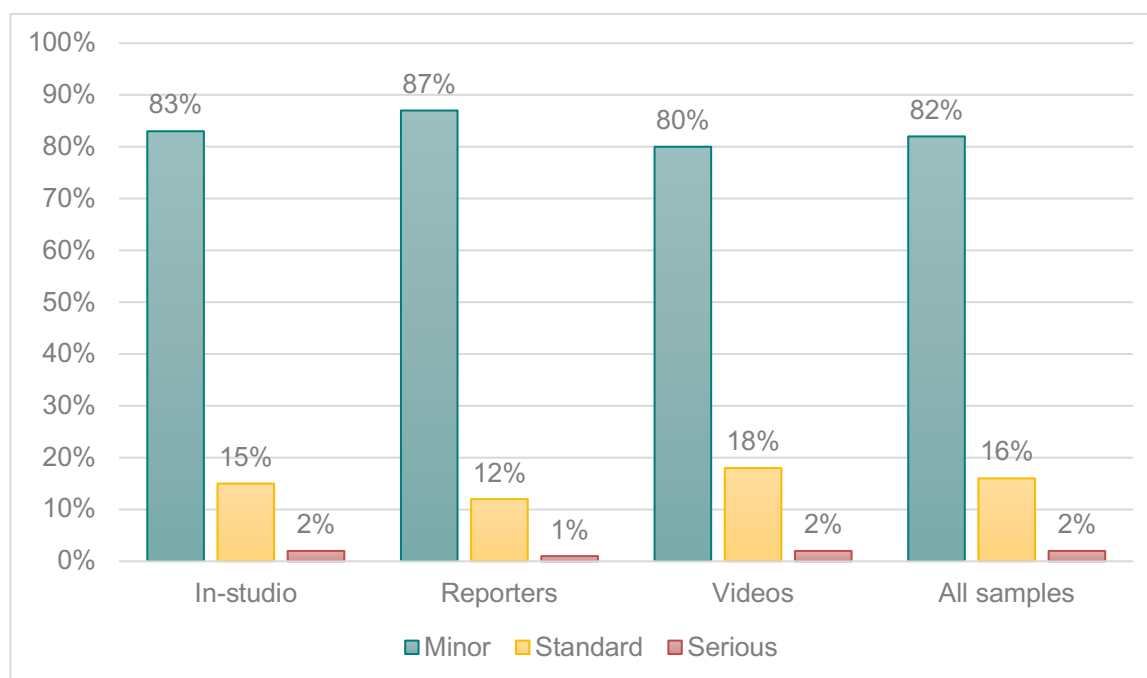


Figure 5. Error distribution per group.

When looking at the accuracy issues encountered, a similar pattern was identified for in-studio segments and videos. Most of the errors in these materials were omissions of relevant information and punctuation mistakes, followed by inaccuracies that seemed to be caused by unsuccessful interactions between the captioners and their technical equipment. A smaller number of orthographical and grammar inaccuracies was also found. Connections with reporters, however, followed a different trend, with the majority of the errors being punctuation mistakes. Omissions of relevant details were also frequent, followed by technical problems and, to a lesser extent, by grammar and spelling issues. Figure 6 shows the error typology per group.

For the sake of clarity, some examples of these errors are listed below:

Omission of relevant information: *un carro nuevo*. [a new car.], instead of *un carro nuevo sería genial*. [a new car would be great.]

Punctuation error: *por qué no podemos pasar* [why can't we get in], instead of *¿por qué no podemos pasar?* [why can't we get in?].

Unsuccessful interaction between the captioners and their technical equipment: *testigos que notes tifikaron* [meaningless clause] instead of *testigos que no testificaron* [witnesses who did not testify].

Orthography error: *trabajo 16 horas diarias* [I work 16 hours a day] instead of [trabajó 16 horas diarias] [he worked 16 hours a day].

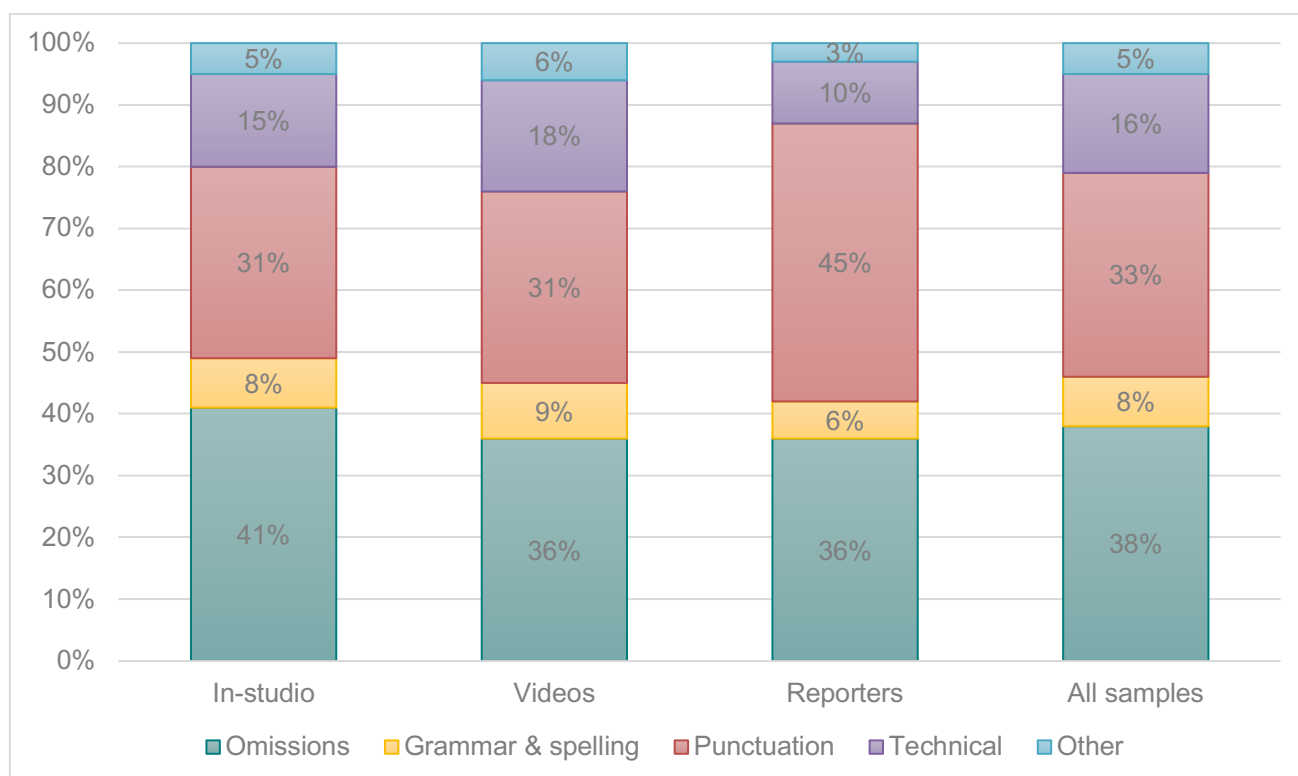


Figure 6. Error typology per group.

According to the aforementioned data, minor errors (82%) outnumbered standard and serious errors combined (18%), which means that most of the mistakes encountered would not hamper the viewers' comprehension of in-studio pieces, videos or connections with reporters. Nonetheless, the average accuracy rate of these three groups (96.98%, 97.11% and 96.90%) was below the 98% quality threshold that the NER model deems acceptable. This could be better explained in terms of error incidence, since 41% of the captions in the in-studio stories, 39% of those in videos and 46% in connections with reporters included at least one error, as opposed to, for instance, 13% of the captions accompanying the English newscasts in the U.S. or 32% of the QualiSub live subtitles (Fresno, in press; Fresno et al., 2019). In other words, while the captioners who worked in the segments included in the present research made good efforts to limit the errors that could affect the users' understanding of the news, they paid less attention to minor mistakes, which were plentiful and caused the accuracy rate to drop below the desired quality levels. Furthermore, these data suggest that connections with reporters posed the greatest challenge since they concentrated more mistakes overall with the potential to affect the viewers' comprehension. Interestingly, this difference could not be attributed to speech rate since the three groups showed similar averages (167 wpm for stories broadcast by reporters, 170 wpm for in-studio pieces and 166 wpm for videos).

5. Conclusions

This article has presented the main findings of the first quality assessment of the live closed captions produced for Spanish-language national newscasts in the U.S. The analysis of 200 minutes of captioned programmes featured exemplary closed captions in terms of placement and completeness. When it comes to synchronicity, the captions in our samples showed a manageable speed in line with the indications provided in the

subtitling national guidelines of several countries (AENOR, 2012; Ofcom, 2015), and in compliance with the FCC (2014) recommendation that closed captions should not be presented at a speed that is difficult to read. This study also revealed an average delay only slightly above the maximum recommended in the Spanish subtitling standard (AENOR, 2012). However, our closed captions did not adjust to the industry recommendations in terms of latency. Accuracy proved the main weakness of the captions assessed in the current project, especially in connections with reporters. The main issues identified included omissions of information and punctuation mistakes. Although most of the errors encountered were minor, they were more frequent than in news samples analysed in other studies, both in English and Spanish, which caused the average accuracy rate to drop below par.

The aforementioned findings, especially those concerned with delay and accuracy, warrant further research. For instance, it would be interesting to carry out studies that explore delay more extensively in the U.S. Analysing latency in a variety of genres in English and Spanish would help paint the big picture, allowing comparisons between both languages, and confirming if the higher latency found in the current assessment for Spanish news captions is also present in broadcasts belonging to other genres. If substantial differences were found between English and Spanish closed captioning, the industry may want to trace the reasons behind them. For instance, the production method may influence the average latency, as the Ofcom project found. According to the results reported in the UK, respoken subtitles were delayed 3 to 5 seconds more on average than those produced using stenography technologies. Additionally, other technical differences related to the transmission of closed captions could also influence the lag with which captions are displayed for viewers.

The comparative studies mentioned above could also offer a unique opportunity to the U.S. in terms of accuracy. Given that closed captions in both languages need to comply with the same regulations, quality studies can help identify weaknesses but also strengths in each case. Being aware of which areas would benefit from improvement could be helpful for practitioners, who could draw on research data to underpin their own performance. In addition, the strategies leading to successful results could be shared as best practices among captioners, which could translate into a positive cross-fertilisation of captioning in the two languages. Furthermore, quality studies such as the ones described here could contribute to the training of captioners willing to work for U.S. companies or broadcasters. By having detailed information on the errors more frequently encountered in current closed captioning samples, trainers would have a better idea of the areas that prove particularly challenging to practitioners, and would then be able to reinforce them in the classroom.

Finally, further research should explore how closed captions are being produced in the U.S. both in English and Spanish. The author of this paper has been unsuccessful when trying to find out basic aspects of current closed captioning on television, such as the production method used in specific broadcasts, if the captions produced in real time are corrected before being used in later reemissions of the same programmes, what quality control strategies are in place and so on. In the case of Spanish programming, it would also be interesting to investigate if captioners are based in the U.S. or elsewhere, since this could involve the use of different production methods and traditions (for instance, more edition). This kind of research, however, will not be possible without a much closer collaboration with broadcasters and the industry since

they alone can share key information that helps make full sense of the quality data. Studies conducted with the aforementioned support are key to provide solid data to inform future latency and accuracy recommendations that are both realistic in their expectations and truly meaningful for the end users.

Acknowledgements

This research has been conducted within the framework of the project “The Quality of Live Subtitling (QuaLiSub): A regional, national and international study” (Spanish Agencia Estatal de Investigación). AEI/10.13039/501100011033.

References

- Asociación Española de Normalización y Certificación (AENOR). (2012). *Subtitulado para personas sordas y personas con discapacidad auditiva* (UNE 153010). AENOR.
- Apone, T., Botkin, B., Brooks, M., & Goldberg, L. (2011). *Caption accuracy metrics project research into automated error ranking of real-time captions in live television news programs*. The WGBH National Center for Accessible Media. http://ncam.wgbh.org/file_download/131
- Burnham, D., Leigh, G., Noble, W., Jones, C., Tyler, M., Grebennikov, L., & Varley, A. (2008). Parameters in television captioning for deaf and hard-of-hearing adults: Effects of caption rate versus text reduction on comprehension. *Journal of Deaf Studies and Deaf Education*, 13(3), 391–404. <https://doi.org/10.1093/deafed/enn003>
- de Linde, Z., & Kay, N. (1999). Processing subtitles and film images: Hearing vs deaf viewers. *Translator*, 5(1), 45–60. <https://doi.org/10.1080/13556509.1999.10799033>
- Federal Communications Commission (FCC). (2014). *Report and Order, Declaratory Ruling, and Further Notice of Proposed Rulemaking*. https://apps.fcc.gov/edocs_public/attachmatch/FCC-14-12A1.pdf
- Fresno, N. (in press). Live captioning accuracy in English-language newscasts in the United States. *Universal Access in the Information Society*.
- Fresno, N. (2019). Of bad hombres and nasty women; the quality of the live closed captioning in the 2016 US final presidential debate. *Perspectives: Studies in Translation Theory and Practice*, 27(3), 350–366. <https://doi.org/10.1080/0907676X.2018.1526960>
- Fresno, N. (2021). Live captioning accuracy in Spanish-language newscasts in the United States. *Lecture Notes in Computer Science*, 12769, 255–266. https://doi.org/10.1007/978-3-030-78095-1_19
- Fresno, N., Romero-Fresco, P., & Rico-Vázquez, M. (2019). *The quality of live subtitling on Spanish television* [Conference presentation]. Media for All 8 Conference, Stockholm, Sweden.
- Fresno, N., & Sepielak, K. (2020). Subtitling speed in media accessibility research: Some methodological considerations. *Perspectives: Studies in Translation Theory and Practice*. <https://doi.org/10.1080/0907676X.2020.1761841>
- Fresno, N., Sepielak, K., & Krawczyk, M. (2021). Football for all: The quality of the live closed captioning in the Super Bowl LII. *Universal Access in the Information Society*, 20(4), 729–740. <https://doi.org/10.1007/s10209-020-00734-7>
- Instituto Nacional de Estadística y Censos de la República Argentina. (2022). *Resultados definitivos del censo 2022*. https://censo.gov.ar/index.php/datos_definitivos_total_pais/

Instituto Nacional de Estadística. (2023). *Población residente por fecha, sexo y edad (desde 1971)*. <https://www.ine.es/jaxiT3/Datos.htm?t=56934>

Institute of Phonetics and Speech Processing Ludwig-Maximilians-Universität München. (n.d.). *The Munich Automatic Segmentation System MAUS*. <https://www.bas.uni-muenchen.de/Bas/BasMAUS.html>

Jelinek Lewis, M. S. (2001). Television literacy: comprehension of program content using closed captions for the deaf. *Journal of Deaf Studies and Deaf Education*, 6(1), 43–53. <https://doi.org/10.1093/deafed/6.1.43>

Jensema, C. (1998). *Viewer reaction to different captioned television speeds*. *American Annals of the Deaf*, 43(4), 318–324. <https://doi.org/10.1353/aad.2012.0073>

Jensema, C., Danturthi, R. S., & Burch, R. (2000). Time spent viewing captions on television programs. *American Annals of the Deaf*, 145(5), 464–468. <https://doi.org/10.1353/aad.2012.0144>

Jordan, A. B., Albright, A., Branner, A., & Sullivan, J. (2003). *The state of closed captioning services in the United States. An assessment of quality, availability, and use*. The Annenberg Public Policy Center of the University of Pennsylvania. <https://dcmp.org/learn/static-assets/nadh136.pdf>

Kruger, J. L., Wisniewska, N., & Liao, S. (2022). Why subtitle speed matters: Evidence from word skipping and rereading. *Applied Psycholinguistics*, 43(1), 211–236. <https://doi.org/10.1017/S0142716421000503>

Liao, S., Yu, L., Reichle, E. D., & Kruger, J. L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading*, 25(5), 417–435. <https://doi.org/10.1080/10888438.2020.1823986>

Miquel Iriarte, M. (2017). *The reception of subtitling for the deaf and hard of hearing*. [Unpublished doctoral dissertation]. Universitat Autònoma de Barcelona.

Ofcom. (2015). *Measuring live subtitling quality results from the fourth sampling exercise*. https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/live-subtitling/sampling_results_4

PEW Research Center. (2018a). *Among U.S. Latinos, the Internet now rivals television as a source for news*. <https://www.pewresearch.org/fact-tank/2018/01/11/among-u-s-latinos-the-internet-now-rivals-television-as-a-source-for-news/>

PEW Research Center. (2018b). *Hispanic and Black news media fact sheet* [Fact sheet]. <https://www.journalism.org/fact-sheet/hispanic-and-black-news-media/>

Retis, J. (2019). *Hispanic media today: Serving bilingual and bicultural audiences in the digital age*. Democracy Fund. https://democracyfund.org/wp-content/uploads/2020/06/2019_DemocracyFund_HispanicMediaToday.pdf

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 and media accessibility at the crossroads* (pp. 111–131). Rodopi. https://doi.org/10.1163/9789401207812_008

Romero-Fresco, P. (2015). Final thoughts: Viewing speed in subtitling. In P. Romero-Fresco (Ed.), *The reception of subtitles for the deaf and hard of hearing in Europe* (pp. 335–342). Peter Lang.

Romero-Fresco, P. (2016). Accessing communication: The quality of live subtitles in the UK. *Language and Communication*, 49, 56–69. <https://doi.org/10.1016/j.langcom.2016.06.001>

Romero-Fresco, P., & Fresno, N. (2023). The accuracy of automatic and human live captions in English. *Linguistica Antverpiensia, New Series - Themes in Translation Studies*, 22, 114–133. <https://doi.org/https://doi.org/10.52034/lans-tts.v22i.774>

Romero-Fresco, P., & Martínez, J. (2015). Accuracy rate in live subtitling. The NER model. In J. Díaz-Cintas & R. Baños-Piñero (Eds.), *Audiovisual translation in a global context: Mapping an ever-changing landscape* (pp. 28–50). Palgrave MacMillan.

Sandford, J. (2015). The impact of subtitle display rate on enjoyment under normal television viewing conditions. *Inst. Eng. Technol.*, 7, 62–67. www.theiet.org/ibc

Szarkowska, A., Krejtz, I., Klyszejko, Z., & Wieczorek, A. (2011). Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American Annals of the Deaf*, 156(4), 363–378.

Szarkowska, A., Krejtz, K., Dutka, L., & Pilipczuk, O. (2016). Cognitive load in intralingual and interlingual respeaking - A preliminary study. *Poznan Studies in Contemporary Linguistics*, 52(2), 209–233. <https://doi.org/10.1515/psicl-2016-0008>

U.S. Census Bureau. (2022). *Selected Social Characteristics in the United States. American Community Survey, ACS 1-Year Estimates Data Profiles, Table DP02*. <https://data.census.gov/table/ACSDP1Y2022.DP02?q=DP02>.

U.S. Census Bureau. (2023). *Hispanic heritage month 2023*. <https://www.census.gov/newsroom/facts-for-features/2023/hispanic-heritage-month.html>

Wilkinson, K. T. (2016). *Spanish-language television in the United States: Fifty years of development* (1st ed.). Routledge.

Data availability statement

The data that support the findings of this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.10640932>.

Notes

¹ WWER (Weighted Error Rate) is an instrument devised to assess accuracy based on the WER (Word Error Rate). It classifies errors in specific categories, which are assigned a coefficient depending on how each error typology affects the viewers' comprehension.

² According to the NER model, subtitles need to reach 98% accuracy rate in order to be deemed as acceptable. Under the NER model, live subtitles are considered “excellent” (AR above 99.5%), “very good” (AR 99%-99.49%), “good” (AR 98.5%-98.99%), “acceptable” (AR 98%-98.49%) and “substandard” (AR below 98%).

³ Reduction rate is estimated by dividing the number of words in the captions by the number of words in the transcript of the same segment. Therefore, it indicates how shorter the closed captions are in comparison to the words that were actually spoken.

⁴ For instance, our samples included captions such as these three: *12 minutos después el avión* (...) [12 minutes later the plane (...)] instead of *2 minutos después el avión* [2 minutes later the plane]; *hay una nueva noticia* (...) [new news coming from (...)] instead of *hay una buena noticia* [good news coming from (...)]; *cuántos aspirantes demócratas* [how many Democratic candidates] instead of *cuál de los aspirantes demócratas* [which of the Democratic candidates]. In cases such as these ones, it is hard to know if the captioners dictated or typed the wrong information, or if they used the right words but these

were changed due to misrecognitions or to the fact that captioners used a mistaken combination of strokes.

⁵ Despite contacting both broadcasters several times, we obtained no satisfactory response and could not confirm how the closed captions for our particular samples had been produced.

⁶ If estimated as the mean of the speed of all our captions, the average captioning speed in our samples would reach 16.392 cps (or 171 wpm).

⁷ These percentages are cumulative because the captions delivered at 20 cps are also delivered at over 18 cps and they also exceed 15 cps.

⁸ These percentages are cumulative because the captions delivered at 20 cps are also delivered at over 18 cps and they also exceed 15 cps.

⁹ These percentages are cumulative because the captions delayed over 20 seconds are also delayed over 18 and 15 seconds.