A Netflix Original Closed Captioning Study:

How Netflix Closed Captions Make Audiovisual Content Accessible to Deaf Audiences

Anna Gomizelj

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements
for the Master of Arts

Department of Communication
Faculty of Arts
University of Ottawa

Abstract

Netflix is currently the world's largest subscription-based streaming platform, with 221.8 million subscribers worldwide (Maglio, 2022). Part of Netflix's enormous global appeal is its Netflix Original brand of films and TV shows – content it produces specifically for broadcast on its streaming platform. To make its content accessible to deaf and hard-of-hearing audiences, Netflix subcontracts the creation of closed captioning to vendors, instructing them to follow the Timed Text Style Guide (TTSG), which it makes freely available online.

My study examines how closed captions for Netflix Original content endeavour to make audiovisual content accessible to deaf audiences, and I demonstrate how the platonic ideal of "equal access" is out of reach due to the limitations of timed text. The objective of my study is to highlight and critique the transformations of meaning that occur when captions translate sound and spoken dialogue into timed text. Drawing on D'Acci's circuit model of media studies (2004) my thesis links the sociohistorical conditions from which captioning techniques and technologies were developed, the conditions of caption production, and the way in which the needs of deaf audiences are articulated in the TTSG. I explore how these three forces affect the content of closed captions. To this end, I engage in a close reading of the TTSG and a selection of closed captions for Netflix Original series and films, borrowing from Berman's (2000) theories regarding the deforming tendencies of translation to describe the changes that result from the intralingual and intersemiotic translation involved in captioning (Jakobson, 2004). My study is informed and inspired by my personal experience as a professional captioner.

Keywords: closed captioning; Netflix; deaf audiences; hard of hearing; SDH; media accessibility; equal access; streaming platforms

Acknowledgements

I would like to express my gratitude for the support, encouragement and, above all, patience of my friends and family over the past few years. Thank you to my thesis supervisor, Kyle Conway, for his invaluable guidance, suggestions, and insight. Thank you also to Maria Cherba and Philippe Ross for serving on my committee. A big thank you to Aline Saba for helping me with Spanish translations and delighting in *One Day at a Time* with me.

Finally, two enormous thank-yous. Firstly, to Humphrey for keeping me company while I wrote, cat-earing important pages of books and articles and keeping my spirits lifted. And finally, to Daniel Meltzer, my Excel wizard husband, for steadfastly supporting and encouraging me, for mastering all manner of coffee-making techniques and always keeping me supplied with a fresh cup, for indulging my ranting and raving about closed captioning, and for enthusiastically watching hours upon hours of captioned film and TV with me. Long may it all continue (except the thesis part).

Contents

Chapter 1: Introduction

In the summer of 2022, with the release of season four of *Stranger Things,* Netflix closed captioning entered the cultural zeitgeist. Filled with highly descriptive cues, like [tentacles undulating moistly], [menacing industrial synth playing], and [dissonant gibbering], the closed captions were celebrated by fans, who turned them into memes and took to social media to laud the captioner for how vividly they conveyed the grotesqueness and intensity of the show's soundscape (Janz, 2022; VanArendonk, 2022). This spotlight cast on Netflix's closed captioning increased public awareness of the accessibility service and called attention to the creative labour involved in producing captions. Within the subtitling industry, captioning has long been misrepresented as the verbatim transcription of dialogue coupled with straightforward descriptions of sound (Zarate, 2021; Zdenek, 2015a). Captioning is, rather, an interpretive and transformative practice, whereby the captioner identifies the most meaningful and plot-relevant elements of a movie or TV show's soundscape and reconstitutes them as timed text events to make audiovisual content accessible to deaf and hard-of-hearing audiences (Zdenek, 2015a).

It is ethically imperative for information communication technologies to be made accessible to deaf individuals to ensure they are fully able to participate in society (Díaz-Cintas et al., 2007; Ellis & Kent, 2015). Remael defines accessibility as "the degree to which a product, service, environment, concept or even person can be used, reached, understood or accessed for a specific purpose" (2012, p. 95). More specifically, "media accessibility" refers to practices that make information and entertainment disseminated via audiovisual media and the Internet accessible to all" (2012, p. 95). Many proponents of captioning describe its goal as providing "equal access" to the content of TV shows and films (Zdenek, 2015a). The Described and Captioned Media Program (DCMP) affirms this goal in their influential captioning style guide,

the Captioning Key, stating equal access requires that "the meaning and intention of the material" is "completely preserved" (DCMP, 2022). I will argue that this platonic ideal of accessibility is out of reach due to the limitations of the form of captions, as well as the conditions of their production. "Access to captioned content," according to Zdenek, "will never, strictly speaking, be the same as access to the sonic landscape. Rather the captioned text will always be inflected by the captioners' interpretive powers and the different affordances of sound and writing" (2015a, p. 6). Nevertheless, providing deaf audiences with equal access to audiovisual content is a goal captioners try to achieve. My research investigates how the captioning practices and approaches outlined in style guides both help and hinder captioners in their efforts to make audiovisual content accessible to deaf audiences.

While captioners do strive to preserve the meaning and intention of the material, this is not always possible, as captioners have to adhere to studios' style guides, negotiate reading speed limits, and respect character, line-length, and timing restrictions (Udo & Fels, 2010; Zdenek, 2015a). Captioning is typically outsourced to third party vendors and freelancers, who have limited or no contact with the creative teams involved in content production (Udo & Fels, 2010; Zdenek, 2015a). Captioning technologies and practices were initially developed by deaf people and by educators at schools for the deaf, and thus captioners were either members of the deaf community themselves or they were in regular dialogue with their audiences (Downey, 2008). As captioning mandates came into effect in the late twentieth century and the captioning industry exploded, deaf individuals were cut out of the production process as studios outsourced captioning work at scale (Downey, 2008). Thus, captioners today are doubly removed – first, from the content producers whose message they need to convey and secondly, from the deaf audience for whom they facilitate access.

Studios articulate their expectations for captions in style guides that they distribute to vendors. These guides are comprehensive instruction manuals, dictating stylistic requirements, technical considerations, and orthographical instructions for captions (Zdenek, 2015a). They help studios manage the diffuse network of vendors and freelancers, as they ensure a measure of consistency in caption appearance and style between episodes, seasons, and series. Some guides can be very rigid and specific, providing frame-specific timing instructions and character-per-second reading speed limitations (Zarate, 2021). Studios often structure their quality assurance processes around style guides, evaluating how well captions reflect and adhere to the guides (Zdenek, 2015a). This can lead captioners to prioritize adherence to the style guide above all else, even if situations arise where deviating from its instructions would better preserve the content's meaning and intention.

To demonstrate how captioners' pursuit of making audiovisual content accessible is complicated in practice by the formal constraints of captions and the conditions of their production, I elected to analyze a selection of Netflix's closed captioning and its style guide, the Timed Text Style Guide (TTSG). Netflix presents a unique opportunity for research, as streaming platforms can potentially usher in captioning innovation. Whereas broadcast television could only offer a single caption file for each program it aired, streaming platforms can offer dozens of subtitle and caption options, as well as the ability to customize their appearance. As more households "cut the cord" and transition from watching traditional broadcast television to watching television online (Fitzgerald, 2021), it is an opportune time to evaluate how streaming services make their content accessible for deaf and hard-of-hearing viewers. I specifically focus my lens on the captioning of Netflix Original programming – shows and movies that are produced by Netflix specifically for the platform. In addition to being the largest subscriber-

based streaming service, with 221.8 million subscribers worldwide (Maglio, 2022), Netflix is also an industry pioneer, having launched its platform in 2007 (Netflix, 2022). It began producing Original-branded content in 2013 and rapidly expanded over the next decade, now offering hundreds of Originals in dozens of languages (Netflix, 2022).

Because of Netflix's enormous scale, it relies on an expansive global network of vendors and freelancers for subtitling and captioning services. To streamline the sharing of information with these external parties, Netflix makes the TTSG, its comprehensive set of guidelines for captioning and subtitling in 47 languages, freely available online, as part of its Partner Help Center (Netflix, n.d.-j). This presents a unique opportunity for research, as most studios' captioning style guides are confidential and publicly inaccessible. A study of Netflix closed captions can thus be done with reference to the specific instructions that dictate reading speed, timing, formatting, and stylistic requirements to see how they are applied in the captions themselves. I am focusing specifically on Netflix Originals because the captions for Originals are required to follow the TTSG and are subjected to quality assurance processes to verify their compliance (Netflix, n.d.-e). The captions for the rest of Netflix's non-Original catalogue are not always obliged to follow the TTSG and they are not scrutinized the same way. Netflix echoes the DCMP's vision for captions as the purveyors of equal access to content, stating that users "want an effortless experience" which means the ability "to read subtitles or closed captions without being reminded that they are reading" (Netflix, n.d.-k). My research will explore how captions try to achieve this ideal of "effortlessness" and the areas in which this goal comes into conflict with the rules in the TTSG and the limits of the timed text medium.

Thesis Structure

I begin my research in Chapter 2 with a definition of closed captioning, explaining differences in terminology and approach between the European audiovisual translation industry and the American captioning industry. I establish that deaf people are the primary intended audience of captioning and explain the distinctions between the audiological condition of deafness, the status "hard of hearing," and the shared cultural identity of capital-D "Deaf." I defend my decision to use the term "deaf" to refer collectively to individuals who identify as deaf, hard of hearing or Deaf in the interest of inclusivity. I argue in favour of using the term "captioning" exclusively, rather than subtitling for the deaf and hard of hearing or intralingual subtitling in the interest of clarity and to emphasize the distinction between dialogue-focused subtitles and the mix of dialogue and descriptions of sound that constitute captions. Finally, I borrow from Jakobson's (2004) theory of translation to classify captioning as a practice of intralingual and intersemiotic translation. This theoretical framework helps me identify and articulate how meaning is affected by captions' transformation of speech and sound into timed text.

In Chapter 3, I describe my study's research design and methodology, explaining how I employ D'Acci's circuit model of media studies (2004) to map how the content of Netflix Originals closed captions is influenced by the socio-historical context from which captioning technologies and mandates emerged, the conditions and limitations of caption production, and how the needs of deaf audiences are articulated by regulating bodies and the Netflix TTSG. I recount how I organized my study around the DCMP and FCC's principles of quality captioning, and how I borrowed from Berman's (2000) list of the twelve deforming tendencies of translation to describe the way content and meaning is transformed as a result of the intralingual and intersemiotic translation involved in captioning. I also outline how my study is informed by my

personal experience working in the captioning industry. My working knowledge of captioning processes and workflows guides my decision to focus on three specific captioning practices: editing dialogue to accommodate timing and reading speed restrictions; the use of orthographical conventions and speaker identifiers to create clear and consistent captions; and the ways foreign dialogue is represented in captions. I identify the specific Netflix Original series and films I selected for my study and describe how I found specific caption extracts to analyze.

In Chapter 4, I dive into the sociohistorical context from which closed captioning emerged. I describe how captioning technologies were developed and how advocates and public broadcasters lobbied the US government to mandate captioning when private broadcasters refused to voluntarily participate in captioning initiatives. I emphasize that deaf individuals were hands-on in the early days of captioning, but as a result of the rapid expansion of the industry due to captioning mandates, they were excluded from the production process as the industry turned to vendors, freelance labour and outsourcing to meet captioning quotas. I also track the evolution of the DCMP, a publicly funded non-profit organization founded by deaf activists and educators, which along with the FCC, constitute the two most influential bodies that dictate standards for quality captioning. I conclude the chapter by situating captioning as a media accessibility service.

In Chapter 5, I draw from my professional experience to describe the structure of the captioning industry, the current captioning production process and the industry's hiring practices. I explain how the industry has become reliant on style guides, using them to train editors, to communicate studios' visions of media accessibility for deaf audiences, and to evaluate caption quality. I explain how Netflix uses its TTSG to this end.

In Chapter 6, I examine and evaluate the TTSG in accordance with the principles of quality captioning articulated by the DCMP and the FCC, focusing in particular on the conflict that arises between the TTSG's emphasis on the importance of accuracy and the strict timing rules and target reading speeds outlined within.

Next, in Chapter 7, I look at captioning practices in action, using excerpts of closed captions from a variety of Netflix Original series and films to illustrate how the directives in the TTSG are interpreted by captioners. I analyze the practice of editing dialogue to meet timing and reading speed requirements, examining the strategies captioners employ and revealing the losses that result from dialogue truncation. I identify how the TTSG employs orthographical conventions to reinforce clarity and consistency in captions and I assess how the rules regarding speaker identifiers are implemented by captioners to the detriment of caption clarity. Lastly, I examine how the TTSG instructs captioners to handle foreign language dialogue in captions, discussing the assumptions Netflix makes about its deaf audiences when it categorizes foreign dialogue along the binary of dialogue that is "meant to be understood" versus "not meant to be understood."

Finally, in Chapter 8, I summarize my findings and offer some recommendations for Netflix and the captioning industry, suggesting the need to move away from rigid adherence to style guides and the need to involve deaf people in the captioning process.

Chapter 2: What Is Captioning?

In this chapter, I provide a definition for closed captioning, and I explain that the deaf audience for whom captions are designed includes individuals with varying degrees of hearing loss who relate to deafness differently. I recount terminological debates that demonstrate differences between the European and North American captioning industries and reflect on the differences and similarities between captioning and subtitling. I conclude by arguing that captioning is a form of intralingual and intersemiotic translation (Jakobson, 2004) due to the transformations of meaning inherent in the practice.

Closed Captioning

Udo and Fels describe captioning as a system designed for deaf and hard of hearing people that consists of "the verbatim translation of spoken dialogue from television and film," which is "overlaid or scrolling on the original video on screen" (2010, p. 207). Udo and Fels' definition skews logocentric, as it neglects to mention the importance of captioning sound as well as speech. To this end, Butler defines captioning as "text that is placed on screen to linguistically represent audio messages" (2019, p. 535). Both of these definitions, however, overlook another critical reality of captioning – that it cannot provide access to all dialogue or to all sounds (Zdenek, 2015a).

Captions must work within the constraints of a multimodal audiovisual text, ensuring that captions can be read and comprehended in concert with the visuals onscreen. Captions are restricted by the amount of text that can be displayed on screen, how much dialogue is present in a scene, how quickly audiences read, how layered the soundtrack is and how the content is edited (Zarate, 2021; Zdenek, 2015a). Zdenek stresses that "definitions should avoid reducing captioning to sound only and stress instead how captioning is about meaning" (2015a, p. 34).

The captioner needs to decide which sounds are the most meaningful to the audience – which sounds inform the plot, describe the atmosphere, and reflect characterization – and these sounds need to be prioritized in the caption track (Zdenek, 2015a). The formatting and structure of captions are typically prescribed by captioning style guides and studio's formatting requests, but the work of captioners is deciding which nouns, verbs, adjectives, and adverbs they will use to describe and transcribe speech, sounds, and music.

The Audience for Captions

A study of captions needs to take into consideration the primary audience of captions: deaf, Deaf, and hard-of-hearing viewers. The same caption track is intended to make audiovisual content accessible to this diverse group made up of people who have differing degrees of hearing and experiences of deafness, and as a result, self-identify and relate to their disability differently (Downey, 2008; Leigh, 2008; NAD, n.d.; Neves, 2008). When writing about deafness, it is important to acknowledge the distinction between "deaf" and "Deaf." Lowercase "deaf" refers to "the audiological condition of not hearing" (Padden & Humphries, 1988, p. 2) and may encompass individuals with varying degrees of hearing loss. Uppercase "Deaf" refers to the cultural identity of deafness that is based around the shared language and culture of sign language (NAD, n.d.). Hard-of-hearing is generally understood to mean "less deaf" (Wilson, 2001 in Leigh, 2009) and includes many people who have acquired hearing loss later in life, post-spoken language acquisition (Downey, 2008). As a result, this group skews older (Downey, 2008). Hard-of-hearing people "may have residual hearing and can therefore share the experience of sound and of the world of hearers to different degrees" (Downey, 2008 p. 129).

Audiological and sociological definitions of deafness may resonate differently with deaf individuals depending on context because "those who select their labels and those who label

others may perceive the criteria for each label differently" (Leigh, 2009, p. 9). Leigh cautions that "the focus on differentiating categories can limit or deny the possibility of similarities between categories, as well as potentially create rationales for exclusion" (2009, p. 9). For example, deaf individuals who were socialized and educated in hearing environments and who communicate orally may feel ostracized from a conception of deaf identity that is structured around knowledge of sign language and vice versa. Lowercase "deaf" is thus the term I have adopted to refer to the general audience of closed captioning, to encompass all individuals with hearing loss who use captions when watching audiovisual content.

The number of differences between deaf identities has "led to an uneven demography, geography and sociology of deafness" (Downey, 2008, p. 24). Closed captions thus cater to a wide range of viewers with "distinct profiles and needs" and as a result, providing one set of captions for all groups is an imperfect access solution (Neves, 2008, p. 131). A study of captions needs to consider how the captioner perceives the needs of deaf audiences and assess whether their choices provide the same level of access to profoundly deaf and hard-of-hearing individuals. The way the captioner conceives of their deaf audience is built on their own personal assumptions of deafness, for better or worse. It is thus important to study the choices captioners make, as they will reflect captioners' impressions and understandings of which sounds and sound qualities are meaningful to deaf audiences (Zdenek, 2015a).

The Debate Surrounding Terminology

The term "closed captioning" was coined by Malcolm Norwood, a deaf academic who was instrumental in developing the technology that first enabled captions to be encoded and hidden in a television broadcast signal (Downey, 2008). Captions are "closed" when they are

able to be turned on and off by the viewer or they are "open" if they are hard-coded or burned-in on the image of the video and thus cannot be turned off (Downey, 2008; Szarkowska, 2020).

In the United States and Canada, where captioning was developed by deaf educators and public broadcasters in a mostly monolingual mass media landscape, captioning is typically approached through the lens of disability studies and media accessibility (Downey, 2008; Ellcessor, 2012; Neves, 2008). Caption studies is interdisciplinary, with links to the fields of communication, educational research, legal studies, information studies, and science and technology (Downey, 2008). In Europe, however, captioning is typically situated under the umbrella of audiovisual translation (AVT) (Neves, 2008; Zarate, 2021).

A particular point of contention amongst captioning and media scholars is whether captioning can or should be considered a type of subtitling (Neves, 2008; Zarate, 2021; Zdenek, 2015a). Translation scholar Díaz-Cintas defines subtitling as an AVT practice that involves "rendering in writing, usually at the bottom of the screen, the translation into a target language of the original dialogue exchanges uttered by different speakers, as well as other verbal information that appears written on-screen or is transmitted in the soundtrack" (2020, p. 150). This definition of subtitling which definitively categorizes it as an act of interlingual translation is reflective of the term's use in Europe's thriving AVT industry. Because subtitling and interlingual translation are so synonymous, some North American scholars bristle at the use of "subtitling for the deaf and hard of hearing" (SDH), the European term for captioning.

Some Canadian and American captioning scholars oppose the use of the term SDH because they believe equating captioning with subtitling erases the critical differences between dialogue-only subtitling and the full spectrum of audio content and nonspeech information that is included in captions (Clark, 2006; Zdenek, 2015a). Neves, however, believes that the

terminological discrepancy is a dialectal difference of no greater significance than the difference between the use of "lorry" and "truck" in British versus American English (2008, p. 130). She opines that the debate around the proper terminology is "rather sterile and quite worthless" because SDH specifically indicates its intended audience within its name (2008, p.130). Clark, a Canadian captioning advocate, rebuts this point, arguing that referring to captioning as "subtitling" is especially clumsy because it requires constant clarification between "subtitling for the deaf and hard of hearing" and "translation subtitling," while also impeding the accurate classification and indexing of research related to captioning (2006). Zdenek cautions that framing captioning as a subtype of subtitling risks reducing captioning "to speech within an economy organized around translation" (2015a, p. 109). Indeed, the American hearing public currently tends to use "subtitling" colloquially as a catchall term for both subtitling and captioning (see Dietz, 2021). Zdenek (2015a) worries that this conflation may encourage video producers to pass off speech-only subtitling as sufficiently accessible for deaf audiences. This type of corner-cutting is reflected in some current practices, such as international DVDs and in-flight entertainment systems that include subtitle tracks but no captions (Butler, 2019; Nović, 2020) and video hosting platforms that offer accessibility to deaf audiences via barebones automatic speech recognition-generated transcriptions of dialogue (Zdenek, 2015a).

Some media scholars use the terms "interlingual subtitling" and "intralingual subtitling" to distinguish between subtitling and captioning (De Linde & Kay, 2006). The use of "intralingual subtitling" as a stand-in for captioning exemplifies the clumsiness Clark (2006) critiqued. For example, subtitles that clarify dialogue that is difficult to understand are also technically intralingual, but they are not specifically designed for deaf audiences. It is a less suitable term than SDH, as not only does it erase the target audience of captioning, but it also

neglects the reality that captioning may be interlingual. Interlingual captioning is increasingly common on Netflix and in the internationalized world of streaming platforms. For example, the French-language show *Lupin* (Degeorges, Franck & Jaubert, 2021-present) is available on Netflix in Canada with an array of subtitling and captioning choices, as outlined in Table 1 below.

Table 1

English and French Subtitle and Caption Options for Lupin, Season 1 in Canada

Timed Text File	Intended Audience	Content Included
English	Hearing audiences that do not understand French	An English translation of dialogue and any relevant onscreen text
English [CC]	Deaf and hard-of-hearing audiences who understand English	An English translation of dialogue based on the English dubbed version and any relevant onscreen text, plus information about the audio
French [CC]	Deaf and hard-of-hearing audiences who understand French	A lightly-edited transcription of the original French dialogue plus information about the audio

Note. From "Degeorges, I., Franck, N. & Jaubert, M. (Producers). (2021-present). *Lupin* [TV series]. Gaumont; Netflix. https://www.netflix.com/"

Thus, for a show like *Lupin*, a reference to the "intralingual subtitles" would only indicate the French caption track. Netflix's top-three most-watched shows in 2021 were Korean, French, and Spanish-language shows (Spencer, 2021) and all three have interlingual English captions available in Canada.

For these reasons, I will exclusively use the term "subtitling" in reference to the practice of subtitle translation and "captioning" rather than SDH or intralingual subtitling to refer to the textual representations of audio that have been designed to provide access to audiovisual content for deaf audiences. Instead of using the term "subtitling" as a blanket term for captions and interlingual subtitles, I will use the more general term "timed text," which is the term Netflix

uses in reference to both. Netflix is itself not consistent in its terminology, using SDH in the TTSG, while offering "CC," the acronym for closed captioning, on its platform (English TTSG, I.).

Captioning is Intralingual & Intersemiotic Translation

Jan Ivarsson writes that "subtitling, when it is done to high standards of excellence, includes so many of the elements essential to art and above all, demands so much skill, imagination and creative talent that it *is* indeed an art" (Ivarsson, 1992, p.8). I would argue the same can be said about captioning, but as Neves (2008) and Clark (2006) note, the work of captioning is not held in this same regard; within AVT, captioning is instead often framed as straightforward transcription or adaptation (Neves, 2008; Zdenek, 2015a). While I agree that captioning and subtitling are distinct disciplines, the two modalities do share some fundamental similarities, as both accompany audiovisual content and both convert spoken dialogue into written text, introducing reading to the cinematic experience (Ascheid, 1997, DeLinde & Kay, 2006, Díaz-Cintas, 2020). Both also negotiate the same spatial and temporal constraints of timed text files (Díaz-Cintas, 2020). DeLinde & Kay note that in both subtitling and captioning "language is being transferred between distinct linguistic systems, between two separate languages and/or between different modes of a single language, while functioning interdependently with another, visual, semiotic system" (2006, p.1). This transfer of language can be understood as a type of translation.

While subtitling involves interlingual translation from a source to a target language (Jakobson, 2004), captioning involves a combination of intralingual and intersemiotic translation (Martínez-Martínez, 2022; Taylor, 2020; Zarate, 2021). Jakobson describes intralingual translation as a practice of rewording that entails "the interpretation of verbal signs by means of

other signs of the same language" (2004, p. 139). Captioners engage in intralingual translation when they reconstitute spoken dialogue as text, as it is often not possible to faithfully render the full and exact content of speech in timed text events. Dialogue may be condensed, simplified, or truncated due to the limitations of the form.

Jakobson refers to intersemiotic translation as "transmutation" which involves "the interpretation of verbal signs by means of nonverbal sign systems" (2004, p. 139). Eco (2001) notes that Jakobson's definition is so focused on the interpretation of verbal signs that it overlooks the plethora of possible transmutations between other types of semiotic systems. Translation scholars and semioticians have since widened the concept to account for "the proliferation of different forms of multimodal texts in today's digital environment, where semiotic resources 'coexist, cooperate, and get translated' on a regular basis" (Kourdis 2015, p. 311; O'Halloran, Tan & Wignell, 2016). Audiovisual texts are multimodal, combining sounds, images, speech, written text, and music to produce meaning (Díaz-Cintas, 2020, p. 13). Hearing audiences who have access to both the text's visual and auditory channels are able to process the audio and situate it within a visual context (Taylor, 2020). Deaf audiences, however, rely predominantly on the visual channel, and thus captions need to interpret the soundscape, translate the relevant sounds into text, and situate them within their context (Taylor, 2020). To describe and qualify sounds and dialogue, captions include nonspeech information (Zdenek, 2015a). Nonspeech information, which includes speaker identifiers, language identifiers, sound effects, paralanguage, manner of speaking identifiers, music, and channel identifiers, all function to translate, clarify, contextualize, and describe sound and speech for deaf audiences (Zdenek, 2015a, p. 139).

Jakobson identifies “equivalence in difference” as the cardinal problem of language and translation (2004, p. 139). When a translator “recodes and transmits a message received from another source” the result is “two equivalent messages in two different codes” (Jakobson, 2004, p. 139). The transformation of meaning that occurs within captioning – how the source and the captions differ and how the captioner strives to create an ostensibly equivalent representation of the audio in pursuit of making content accessible to deaf audiences – is the focus of my research. Captions are the product of individual captioners’ creative labour, as they identify which sounds are significant and how they should be represented, but they are affected by the physical constraints of captioning. The limitations of representing aural content in timed, written text can lead to losses as well as gains in access. Zdenek notes captions occasionally provide superior access to content, as they provide clarifications and can offer audiences information that they would not otherwise have access to, such as the name of a speaker or the spelling of a unique term (Zdenek, 2015a, p. 5-6). However, captions also embody a loss of information that results from the translation of the aural into the written. The goal of providing equal access is complicated by the fact that “access to captioned content will never, strictly speaking, be the same as access to the sonic landscape” because captions “will always be inflected by the captioners’ interpretive powers and the different affordances of sound and writing” (Zdenek, 2015a, p. 6).

In addition to the transformations that result from the translation of sound into text, the conditions of caption production affect captioning strategies and captioners’ pursuit of equal access. Captioners usually work independently from the creative production teams, isolated from both the deaf audience they serve and the content producers (Udo & Fels, 2010; Zdenek, 2015a), whose “meaning and intention” they need to reflect in their work. Studio style guides are thus

distributed to vendors and captioners to provide them with a set of instructions on how to make content accessible in accordance with industry standards and the studios' conception of accessibility (Zarate, 2021; Zdenek, 2015a). Style guides have the most direct influence on the strategies captioners employ and the way they caption content, as they prescribe specific instructions regarding reading speed, timing, and style (Zarate, 2021; Zdenek, 2015a). The reliance on style guides is a result of the scale of captioning operations – due to decades of advocacy and lobbying, the captioning of television became mandatory with the passage of the Telecommunications Act of 1996 and the Twenty-First Century Communications and Video Accessibility Act of 2010 (Downey, 2008; FCC, 2019). While deaf and hard-of-hearing individuals were once intimately involved in the development of captioning processes and technologies, they were excluded as a result of the industry boom triggered by captioning mandates. Studios and networks had to ensure all of their programming was compliant with the laws and was broadcast with closed captioning. This led to the widespread practice of outsourcing captions to vendors and freelancers, who are typically hearing individuals (Downey, 2008). During the captioning process, the deaf user is largely an imagined entity to the captioner – experience or familiarity with deafness or deaf individuals and disability training are not prerequisites for most captioning positions (Downey, 2008; Udo & Fels, 2010; Zarate, 2021).

Thus, all together, the conditions of caption production, the social and political pressures that shaped captioning technology and legislation, and the captioner's conception of the needs of the deaf audience influence how content is captioned. D'Acci's circuit model of media studies (2004) is an apt model to illustrate how these forces interconnect in the context of Netflix Original content's closed captioning and its pursuit of providing equal access to deaf audiences.

Chapter 3: Framework and Methodology

In this chapter, I will describe how I employ D'Acci's circuit model of media studies (2004) to structure my study, as I examine how the socio-historical context from which captions emerged, the conditions of caption production and the way the needs of deaf audiences are perceived affect the content of Netflix Originals' closed captions. I describe how my professional experience as a captioner guided and informed my study and the selection of Netflix Original content I chose to close read. I draw on Berman's (2000) theories of the deforming tendencies of translation to describe the way the content and meaning of captions is transformed as a result of the creative choices captioners make.

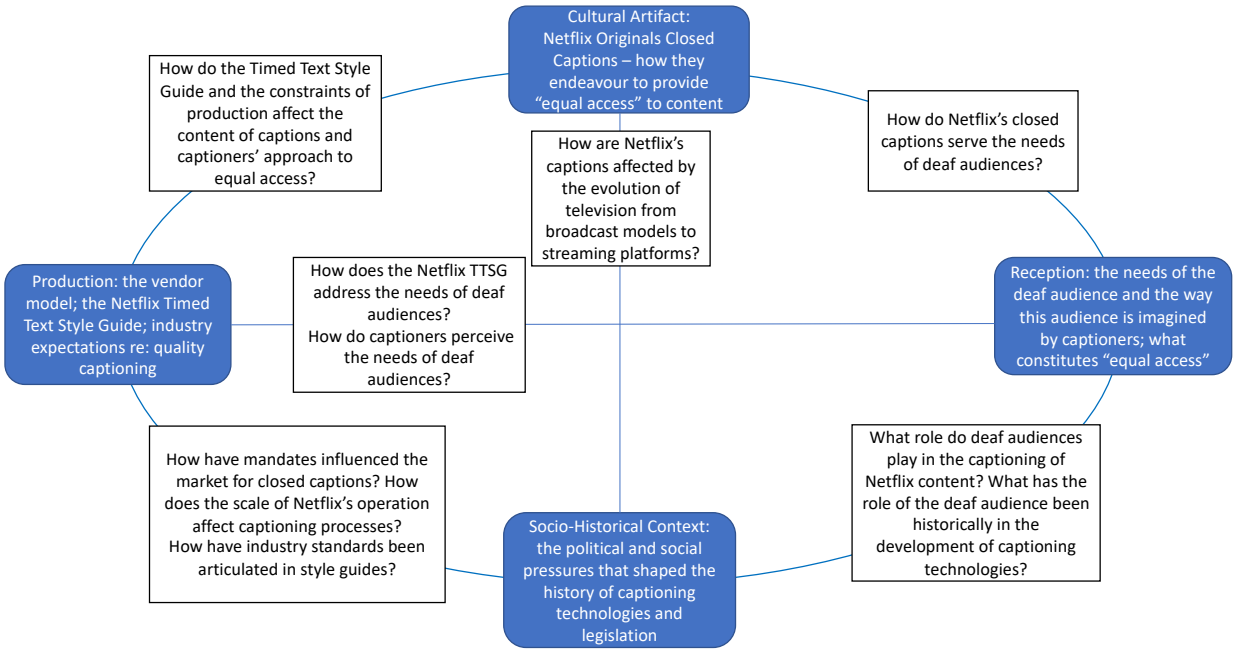
The Circuit Model

D'Acci's circuit model of media studies arises from her observation that cultural studies' approaches to television tended to be too text-centered, neglecting the forces of production, the conditions and intricacies of reception, and the specificities of the televisual form (2004, p. 422). Her model builds on Stuart Hall's (1980) encoding/decoding model, which presents a three-stage process by which media messages are produced, disseminated, and received. D'Acci proposes instead four discursive sites which are all interconnected via articulations: the cultural artifact, its production, its reception, and the socio-historical context in which the artifact is situated. She highlights the central role of the researcher within the model, as "it is the researcher who produces his or her version of the object and its articulations" (D'Acci, 2004, p. 432). For D'Acci, articulation is "the process by which an 'element' gets linked into a particular discourse to become a 'moment' in that discourse—to take up its meaning in that particular discursive practice" (p. 435). The researcher identifies the points of articulation, interrogating how the sites converge.

Though initially conceived to facilitate the study of television production, the model can be adapted for a study of closed captions. The strength of the model is that it represents each point on the circuit as "a convergence of discursive practices" – according to D'Acci, this means that "each site or moment of the circuit (seen as distinct for analytical purposes) mobilizes conjunctures of economic, cultural, social, and subjective discourses" (p. 433). The model is especially useful for contextualizing production, as it allows the researcher to see the economic imperatives of industry in relation to cultural artifacts, reception, and their sociohistorical context. As Zdenek observes, "While the practice of captioning will present a number of simple scenarios for the captioner, the subjectivity of the captioner and the ideological pressures that shape the production of closed captions will always be close the surface of the captioned text" (2015a, p. 6). D'Acci's model presents the opportunity to bring those ideological pressures to the surface.

Figure 1

Circuit Model of Netflix Originals Closed Captioning



Note. The four blue boxes indicate the four discursive sites of the circuit: the cultural artifact, the conditions of its reception, the socio-historical context from which the artifact emerged and the conditions of its production. The lines linking the sites represent the articulation between the sites, with the white boxes containing questions that interrogate and reveal how the sites on either side of the line link and converge. Adapted from "D'Acci, J. (2004). Cultural studies, television studies, and the crisis in the humanities. In J. Olsson & L. Spigel (Eds.), *Television after TV: essays on a medium in transition* (pp. 418-445). Duke University Press."

When applied to my study of how Netflix Originals closed captions strive to provide equal access to audiovisual content, the captions themselves are the cultural artifact. The locus of production encompasses the structure of the captioning industry, which includes the dependence on vendors and freelance captioners, as well as the reliance on style guides to train, direct and evaluate caption quality. The sociohistorical context includes the history and development of closed captioning technology, with a focus on the deaf individuals, educators, non-profit organizations, and public broadcasters that put pressure on networks and on government to make television closed captioning mandatory (Downey, 2008; Neves, 2018). It also includes the FCC and DCMP, who define what constitutes quality captioning, establish standards for the captioning industry, and advance the idea of equal access. The site of reception reflects the needs of the deaf audience captions are intended to serve. The deaf audience, however, exists mostly in an imagined form within the model, with captioners imagining the needs of the deaf viewer from their own conception of deafness and from the instructions in style guides.

To set the stage for my study and explore the articulations between the four sites, I start with an overview of the history of closed captioning, from the early innovations of deaf individuals and organizations through the development of closed captioning for television that

was facilitated by deaf educators and public broadcasters. Their efforts concretized and enshrined deaf audiences' right to media accessibility through the passage of the Telecommunications Act and the CVAA (Downey, 2008). From there, I recount the captioning production process, explaining how captioning work is delegated to vendors and freelance captioners, and how Netflix organizes the captioning of its Original branded content. As a professional captioner with more than eight years of experience producing captions for a major post-production studio, I am able to identify specific ways in which the demands and scale of Netflix captioning influence the production process. Then, to understand how the captioning industry at large and how Netflix articulates the needs of deaf audiences in the TTSG, I explore the principles of quality captioning as defined by the FCC and DCMP and explore how they are embodied in the TTSG and some of Netflix's captioning practices.

How to Study Netflix Originals' Captions

My research is focused on captions for Netflix Original content rather than those for other offerings in Netflix's catalogue because the captions for Originals must follow the TTSG and they receive more scrutiny from Netflix's quality control and production teams (Netflix, n.d.-e). More resources, attention and care are typically directed towards Netflix's higher-profile titles, and occasionally, the content producers are involved in the subtitling and captioning, making the goal of preserving the meaning and intention of the material more attainable.

Professional Experience

My research is inspired and informed by my professional experience as an English editor at a major post-production studio. My research and my critique of the TTSG draws on my own experiences interpreting guidelines and captioning content, and I bring insight into the creative decisions captioners make, how captioners interpret instructions, and how they understand their

mission of making audiovisual content accessible. By shining a light on the context in which captions are created for Netflix by the vast network of freelance captioners who are armed with a style guide and little formal training, I challenge the public perception that captions are authoritative companion texts authored by Netflix.

Zdenek comments that "ethnographic studies from inside captioning firms are needed to explore the decision-making processes, organizational values, client demands, and economic constraints that shape the production of closed captions" (2015a, p. 96). Such studies are difficult to conduct because most captioning firms are not interested in transparency. Post-production studios and localization vendors typically have employees sign non-disclosure agreements restricting how much information employees can share to safeguard security and client relationships, as well as keep proprietary information internal (Romero-Fresco, 2021).

I have worked for a major localization and captioning vendor for the past eight years. Because I am bound by a non-disclosure agreement, I cannot engage in a proper ethnographic study of my workplace, nor can I disclose any proprietary information or discuss the specifics of projects I have been involved with. In my captioning career, I worked firstly as a freelance contractor, independently subtitling and captioning English content for a variety of clients, following a variety of client guidelines. I then worked in-house as a full-time editor where I worked more closely with project and account coordinators. In addition to captioning and subtitling tasks, my responsibilities included training editors to follow Netflix's TTSG, creating internal training manuals and style guides, and providing feedback and support to freelance editors, responding to them when they had difficulties, questions, or queries. Working both as a freelance contractor and as an in-house editor afforded me a unique perspective, as it familiarized me with the ins and outs of the global processes involved in captioning and

subtitling. It also made me attuned to the types of issues captioners encountered most frequently and it exposed me to the client quality assurance process. Freelance editors are often shielded from these experiences, as they mostly work in isolation on content that is assigned to them. My experiences and knowledge of the project management pipeline, the process of captioning, and of following various style guides inform my critiques of Netflix's captions and the style guide. However, in my analysis of captions and the TTSG, I do not disclose any proprietary or confidential information, nor do I discuss any elements of Netflix captioning that are not publicly referenced or explained on Netflix's Partner Support website.

Timed Text Style Guide Analysis

My analysis of the TTSG and of Netflix captions assesses how Netflix understands its mandate to make content accessible and how well the captions negotiate providing equal access to deaf audiences. To this end, the first step I take is a close reading of the TTSG to take stock of its recommendations and evaluate how Netflix transforms the captioning industry's standards and expectations into concrete, specific instructions for captioners. As I will discuss in Chapter 6, the FCC and DCMP's principles of quality captioning – accuracy, completeness, synchrony, readability, placement, consistency, clarity, and equal access (DCMP, 2022; FCC, 2021) – are a logical point of entry when trying to determine the rationale behind the opaque instructions of the TTSG.

For the scope of this project, I narrowed my focus to three specific captioning practices: editing text to meet reading speed requirements, the use of orthographical conventions and speaker identifiers, and the way foreign dialogue is captioned. I chose these three areas to concentrate on because I have observed that, despite being presented as straightforward concepts in the TTSG, captioners frequently struggle with these practices, and their application

demonstrates how some of the FCC and DCMP's principles come into conflict. These practices are also straightforward to isolate and identify within caption files – for example, editing practices are commonly employed in fast-paced content; speaker identifiers and orthographical conventions are present in all caption files; and foreign dialogue is more likely to be found in content that is set abroad or features characters of different cultural backgrounds. I ruled out other potential practices, such as the use of sound cues, the way music is captioned, or how dialects are represented, because these practices are much more varied and context-dependent, necessitating a more longitudinal study across multiple episodes, seasons, and series to demonstrate how the TTSG's directives are applied. The three practices I chose lend themselves well to individual case studies that are emblematic of how the TTSG's guidelines are applied.

How to Analyze Netflix Captions

In preparation for my analysis, I downloaded XML caption files from Netflix for an assortment of twenty-five Netflix Original series and films that launched from 2015 to 2021: nine comedy series/stand-up comedy specials, ten dramas and six unscripted shows (three reality shows, three documentary series). I strove for a balanced sample with a variety of genres, however some of my specific content choices were purposive, informed by the insight I have gained in my captioning career, as I selected content that I knew would demonstrate the three captioning practices I sought to analyze. For example, I selected fast-paced comedy series and dramas that feature layered soundscapes and quick dialogue, since they present more captioning challenges than slower-paced content with fewer speakers, which is typically straightforward to caption. I also purposefully chose some series and films that featured characters who spoke foreign languages so I could see how context affected captioners' decisions to transcribe or cue the dialogue.

Once I had downloaded my corpus of caption files, I used the open-source software Subtitle Edit to convert the XML data into Excel spreadsheets complete with timecodes. This allowed me to read through the caption files and to be able to refer back to specific scenes in the videos on Netflix. The Excel format also gave me the ability to perform targeted searches on the caption files; for example, I could search "speak" to find examples of [speaking language] cues, and I could search for quotation marks to isolate instances where they were used. This helped guide my research and made it possible for me to find the specific examples I refer to in my analysis. In the end, I narrowed my focus further and used excerpts or examples from *Aziz Ansari Live at Madison Square Garden* (Ansari, 2015), *Chef's Table* (Fried et al, 2018)*, Gilmore Girls: A Year in the Life* (Sherman-Palladino, 2016)*, The Irishman* (Scorsese, 2019), *One Day at a Time* (Calderón Kellett et al., 2017b), *Tiger King* (Chaiklin et al., 2020)*, Stranger Things* (The Duffer Brothers, et al., 2016), and *Wet Hot American Summer: First Day of Camp* (Showalter & Wain, 2015).

Captioning as a "Trial of/for the Foreign"

In my analysis of Netflix captions, I illustrate some of the losses and gains inherent in the intralingual and intersemiotic translation of the captioning process. To evaluate, articulate and classify the different ways meaning is transformed within the act of captioning, I borrow from Berman's twelve deforming tendencies of interlingual translation (2000) and recontextualize them within captioning. Berman describes translation as a "trial of the foreign" as well as a "trial for the foreign" (2000, p. 284). As a trial of the foreign, the translation is received in its foreignness: translation "establishes a relationship between the Self-Same and the Foreign by aiming to open up the work in its utter foreignness" (2000, p. 284). At the same time, translation is a trial for the foreign as "the foreign work is uprooted from its own language ground" and

exiled from its original context, which may "reveal the foreign work's most original kernel, its most deeply buried, most self-same, but equally the most 'distant' from itself: (2000, p. 284). Berman is thus interested in exploring what he calls the "deforming tendencies" of translation; he enumerates twelve ways in which translation transforms the meaning of the original text.

Many of these tendencies are present in the practice of editing dialogue to improve readability. When captions are displayed at a pace that is too fast to be comfortably read, the captioner must edit content by reducing, deleting, or condensing the original dialogue (English TTSG, I.1). Captioners have to choose how to edit the dialogue and consider which elements can be omitted without significantly compromising the meaning and intent of the content. The result is that edited captions produce an ennobled and quantitatively impoverished version of dialogue Berman, 2000), cutting back on discourse markers, repetitions, and redundancy to fit within the spatial and temporal constraints of captions. The flourishes and fillers of spoken English are often excised, leaving behind a more formal version of the dialogue, presenting a translation in the captions that is markedly different from the source. Furthermore, edits often follow similar patterns, with captioners electing to remove the same types of words and phrases, while also simplifying spoken English to make it more compatible with written English. The result is what Berman calls "the destruction of linguistic patterns," which creates a "patchwork" text (p. 293). Zdenek refers to these types of formalizing reductions as the accuracy and accessibility balance (2015a, p. 8). My analysis evaluates how the practice of editing affects the way the meaning and intent is conveyed in the captions and assesses the types of transformations the text undergoes in the name of creating an accessible viewing experience.

Next, I assess how the TTSG's rules surrounding orthographical conventions and speaker identifiers facilitate equal access through the creation of clear and consistent captions. While

Netflix is diligent in tracking consistency across subtitling through the use of their proprietary term base manager, it does not have any formal content-based consistency tracking mechanisms for captioning. Instead, Netflix focuses on enforcing its orthographical and stylistic conventions to ensure speech and sound are represented clearly and consistently. My interrogation of some of these conventions demonstrates Zdenek's observation that in transforming sound into text, captions clarify content, occasionally providing superior access to the material (2015a, p. 8). I will explore how the practice of using speaker identifiers toes the line between providing inadequate versus potentially superior access to the deaf audience. The TTSG instructs captioners to avoid using the name of a character that has not yet been introduced, using instead a generic identifier like [man 1], [woman 1], et cetera (English TTSG, I.16). However, the TTSG does not account for the fact that in many cases, hearing audiences are introduced to the characters' voices before their names are mentioned in the dialogue. I explore how the TTSG's speaker identifier rule prioritizes consistency over clarity, ultimately offering inferior rather than superior access to the content.

In the case of foreign language dialogue, the TTSG instructs captioners to include [speaking language] cues when dialogue "is not meant to be understood" (English TTSG, I.8). This practice paradoxically provides both a clarification, by introducing information about the specific language spoken, while also providing an impoverished version of it. When cues are used instead of transcriptions of the foreign dialogue, access to the original rhythm of the dialogue is denied to the deaf audience (Berman, 2000). I examine how foreign dialogue is captioned based on instructions in the TTSG, the creative intent of the content producer, and the captioner's understanding of how to best serve the needs of deaf audiences. Netflix captions typically assume that the deaf viewer is monolingual and thus they frequently default to using

cues in place of transcriptions of foreign languages, an impoverishing practice which sometimes adds bulk to the caption track. For example, replacing *"Bonjour!"* with [greeting in French] is significantly longer but also less meaningful.

Chapter 4: The History of Closed Captioning

An important avenue to explore when studying closed captioning is the history of the technology and how it came into being, who was involved in its development, how the production process evolved, and the debates and issues that have shaped the industry. In this chapter, I examine the political and social pressures that shaped captioning initiatives and technologies and I ask what role has the deaf audience played in the development in captioning technologies? I offer an overview of the history and development of captioning technology from its early days to the Internet era, revealing how market forces pushed deaf individuals, who were innovators in the industry, out of captioning production. I also trace the evolution of the DCMP, the most influential publicly funded non-profit organization that has been authoring captioning style guides and advancing principles of quality captioning for decades. This chapter concludes with an assessment of how the captioning advocacy movement crystalized captioning as a media accessibility service and explores how disability scholars interpret accessibility for deaf audiences.

Captioning Film

While audiovisual media was largely inaccessible to deaf audiences in mid-century America, the opposite was true in the 1920s. The silent film era of early cinema was the "golden age" of accessible cinema for deaf audiences (Downey, 2008, p. 20). Silent film actors conveyed the story through gesticulation and heightened displays of emotion, while any dialogue or important expository information was written either on paper programs or on intertitles (Downey, 2008). Intertitles were cards containing stylized text that were photographed and inserted into film strips (Downey, 2008). Early cinema was exciting for deaf audiences, as unlike

radio or theatre, it was a public form of entertainment that incidentally happened to be accessible to them (Downey, 2008).

While the advent of sound films in 1927 marked a technological revolution and paradigm shift for cinema, it also marked the abrupt abandonment of deaf audiences. As silent films were phased out, the "'talkie' quickly became the true 'silent film' for the deaf" (Norwood, 1988). Only foreign-language films presented with English subtitles offered deaf audiences a degree of accessibility (Zdenek, 2015a). Deaf audiences needed to lobby, advocate, and innovate solutions to make films accessible – a process which would repeat with the development of broadcast television and Internet video streaming (Downey, 2008; Zdenek, 2015a).

Initial experiments by deaf activists and educators to make sound film accessible involved dual projectors that simultaneously broadcast captions on a second screen alongside the films (Downey, 2008). In 1947, Emerson Romero, a Cuban-American deaf actor who was forced out of the film industry in the transition to talkies, developed the first captioning technique by splicing cards with dialogue into film reels, mimicking the use of intertitles in silent films (Schlanger, 2020). This crude technique did not catch on widely due to its lengthening effect on the runtime of films, but his reels were circulated among deaf educators and community groups (Downey, 2008). From this grew a grassroots effort by educators at schools for the deaf to make film more accessible by etching captions directly onto prints of films (Boatner, 1980). In 1949, the initiative was formalized and incorporated into the non-profit Captioned Films for the Deaf (CFD) (DCMP, n.d.).

The CFD secured federal funding from the Department of Health, Education and Welfare (HEW) to expand their captioning operation and to circulate captioned films to schools for the deaf and to deaf social clubs (Boatner, 1981; DCMP, n.d.; Public Law 85-905). The CFD's

mandate was to provide “enriched educational and cultural experiences through which deaf persons can be brought into better touch with the realities of their environment” (Public Law 85-905). Though the etching technique was a drastic improvement over the intertitle method, captioning films was nevertheless time-consuming and labour-intensive. Captioners were supplied with timecoded continuity documents that contained descriptions of each scene and transcriptions of the dialogue, making it possible for deaf individuals to create captions, as it could be done without viewing the films (Downey, 2008; Boatner, 1981). Though the CFD was initially established to make Hollywood films accessible to deaf audiences, once it became a publicly funded agency of the HEW, the CFD expanded into captioning more educational content for use in classrooms (DCMP, n.d.; Downey, 2008). The agency continued to evolve throughout the 1970s and 80s, becoming Captioned Films/Videos for the Deaf in 1984 when they introduced captioned videocassettes, expanding their scope again, towards the end of the twentieth century, when they became the Described and Captioned Media Program in 2006 after adding audio description services (DCMP, n.d.). While the CFV continued to circulate captioned films and videocassettes for deaf audiences throughout the 1970s and 1980s, the film industry was largely unwilling to caption theatrical release movies (Downey, 2008). At the same time, deaf activists and educators began to turn their attention to television.

Captioning Television

Captioning experiments and initiatives for American television began in the 1970s under the remit of public broadcasting (Downey, 2008). Television became the most popular form of mass media in the 1960s, but its early critics were concerned about the quality of content made for television in a free market environment. They argued the medium’s reliance on advertising for revenue and thus its need to attract a mass audience “made it structurally incapable of serving

the broader cultural, informational, and education functions of a democratic mass communication system" (Hoynes, 1994, p. 1). In response to this pressure, Congress created the Corporation for Public Broadcasting, which in turn founded the Public Broadcasting System (PBS) to broadcast and produce non-commercial programming through its member stations (Downey, 2008; Hoynes, 1994). Public broadcasters were instrumental in the development of closed captioning, as the project of making audiovisual content accessible to deaf audiences aligned with their social and educational agenda (Ellcessor, 2012).

The first experiment in captioning broadcast television occurred in 1972, when Julia Child's *The French Chef* aired with open captions across the United States on PBS (DCMP, n.d.). The following year, *ABC News* became the first regularly scheduled program to be aired with open captions, appearing on 190 PBS stations (Downey, 2008). Though public reception to the captioned programs was generally positive, other networks were resistant to caption more programming, fearing it would cut into revenue (Downey, 2008). Network executives believed that the percentage of the public that was deaf or hard of hearing was smaller than the percentage of people who would be bothered by the presence of captions (De Linde & Kay, 2006; Downey, 2008). Thus, PBS took up the mantle of increasing caption access as part of its public service mission, continuing to offer open captions on a selection of programming while also devoting resources to the development of a closed captioning system that would keep captions out of sight from the hearing public (Downey, 2008).

In the late 1960s, the National Bureau of Standards realized that text could be encoded within the lines of an NTSC television broadcast signal (Downey, 2008). Television executives saw potential to use this technology for closed captioning and subsequently tested it at the 1971 National Conference on Television for the Hearing Impaired, presenting a close-captioned

episode of *Mod Squad* (Norwood, 1988; Downey, 2008). In 1976, after years of successful tests of the closed captioning technology, PBS successfully lobbied the FCC to pass regulation that would allow captions to be encoded on Line 21 of the vertical blanking interface of the television broadcast signal (Downey, 2008). The FCC was wary, however, and debated whether there would be a sufficiently large audience for captioning to justify the endeavour, and more urgently, whether audiences and networks would be willing or able to pay the associated costs: the public would need to purchase special decoders for their televisions to view the captions and networks would need to pay to commission the creation and encoding of captions for their programming (Downey, 2008).

The major networks' aversion to investing in captioning led PBS to propose the creation of a centralized, publicly funded non-profit dedicated specifically to captioning (Downey, 2008). The National Captioning Institute (NCI) was established in 1979 under the purview of the HEW to "produce captions, manage the development and marketing of decoders, as well as research future possibilities in captioning" (Downey, 2008, p. 94). Though the creation of the NCI led ABC and NBC to participate in closed captioning starting in 1980, captioning faced a "chicken and egg problem" (Ball in Downey, 2008, p. 98) – "people would not buy caption decoders without more captioned programming, but programmers refused to caption more content without a larger audience of decoder-owning viewers" (Downey, 2008, p. 98). Captioning advocates thus began lobbying the government for more comprehensive legislation to force the expansion of closed captioning. Having witnessed the networks' dismissal of deaf audiences as too small a market to cater to, they were forced to "rhetorically expand" the size of the audience that could benefit from closed captioning (Downey, 2008, p. 225). Advocates thus argued that in addition to improving the literacy skills of deaf Americans, closed captioning would benefit hearing children

learning to read as well as immigrants learning English (Downey, 2008). Legislating closed captioning necessitated framing the technology as "not only a broadcast right for the few, but a broadcast enhancement for the many" (Ruggiero, 1986).

Congress resolved the "chicken and egg" dilemma by passing the Television Decoder Circuitry Act in 1989 which made it mandatory for new television sets to have decoders built in, meaning captions could be accessed by anyone, without purchasing special technology (Downey, 2008). To resolve the lack of captioning issue, Congress included a provision in the Telecommunications Act of 1996 that gave networks a deadline of 2006 to have all their content captioned (Downey, 2008). The FCC oversaw the implementation of the measure and they decided that program distributors rather than producers would be responsible for meeting the caption quotas, reasoning that it would be easier for consumers to direct their feedback to networks and networks would themselves delegate the responsibility of providing captions to program producers in their contracts (Downey, 2008, p. 237).

Internet Captioning

In 2010, President Obama signed the Twenty-First Century Communications and Video Accessibility Act (CVAA) to extend the accessibility laws of the 1980s and 1990s into digital, broadband, and mobile technological innovation (FCC, 2019). With regard to captioning, it "requires video programming that is closed captioned on TV to be closed captioned when distributed on the Internet" (FCC, 2019). This requirement does not apply to programming, like Netflix Original content, that is only broadcast on the Internet (FCC, 2019) and thus Netflix did not immediately offer captioning on its streaming platform. In 2010, the NAD filed a class-action lawsuit against Netflix, alleging that Netflix's website constituted a "place of public accommodation" and thus its lack of captioning violated the 1990 Americans with Disabilities

Act (NAD v. Netflix, 2012). The ADA is a national mandate that seeks to eliminate discrimination against individuals with disabilities from employers, public and governmental entities, and privately-owned places of public accommodation (Podlas, 2018). Although the lawsuit was dismissed, it drew public attention to Netflix's lack of accessibility services, and thus, despite no legislation or legal prerequisite, Netflix signed a consent decree in 2012, committing to captioning 100% of its content within seven days of launch by 2016 (NAD v. Netflix, 2012). Maintaining positive public relations by providing accessibility services is a strong motivator for companies in the age of the Internet, as "an ADA lawsuit, even if unsuccessful in the courtroom may be successful in real life, by bringing attention to the needs of underserved consumers and motivating businesses to voluntarily implement accessibility changes" (Podlas, 2018, p. 264).

Shining a light on the way captioning technology was developed and how captioning advocates successfully lobbied the US government for captioning legislation provides context for the current conditions of caption production and the way in which the needs of deaf audiences have been considered. As I will describe in Chapter 5, since the Telecommunications Act and CVAA came into effect, the captioning industry began relying on vendors, freelance labour, and outsourcing to accommodate the volume of content that needed to be captioned. As a result, captioners are removed both from the content creators and from the deaf audiences that they seek to serve, complicating their pursuit of providing equal access to the content. This is in marked contrast with the early days of captioning when deaf individuals and educators were highly involved in developing captioning techniques and in the captioning of film. Captioning mandates increased the volume of content that needed to be captioned, and thus the industry expanded with a view to keeping costs minimal. The reluctance of content producers and broadcasters to budget

for captioning services is a trend that informs current captioning practices. The tendency to de-emphasize the benefits of captioning for deaf audiences to make captioning appear more commercially attractive continues today, such as in discussions about the potential to harness caption metadata for search engine optimization or the benefits of captioning for hearing, nondisabled audiences (Ellcessor, 2012; Gernsbacher, 2015).

This historical overview of captioning also establishes the importance of the FCC and DCMP as arbiters of caption quality. The DCMP, with its origins going back to 1950s grassroots film captioning initiatives, authored the Captioning Key, a comprehensive style guide that informs many current studio style guides, including Netflix's TTSG. The DCMP and FCC together articulate principles of quality captioning, which I will discuss in greater detail in Chapter 6. These regulating bodies stand in for deaf audiences, articulating their needs and describing how captioners should endeavour to make content accessible to them. Otherwise, the needs of deaf audiences are most directly considered when they threaten legal action against broadcasters for providing substandard access (Podlas, 2018).

Media Accessibility

In their fight for closed captioning, deaf advocates and educators repeatedly faced the same obstacles, locked in a cycle that indicates "a pervasive unwillingness to consider people with disabilities as a central audience" (Ellcessor, 2012). When networks and government agencies argued that the deaf audience was too small to be a market worth catering to, advocates refocused their campaigns on the educational benefits for hearing children learning to read and to immigrants learning English (Downey, 2008).

The power of the Telecommunications Act and the CVAA, combined with the influence of the Americans with Disabilities Act forced streaming platforms to provide captioning, lest

they face negative media coverage and public outrage (Neves, 2018). To implement widespread captioning, advocates could not outright argue for equal access to media or the right to enjoy film and television. Rather, the argument had to be contorted to appease ableist, paternalistic attitudes surrounding the needs of deaf audiences. Captioning advocates had to underscore the educational value of captions and its potential to improve deaf individuals' literacy skills (Downey, 2008). That deaf people should have equal access to film and TV purely for their own enjoyment was an insufficient argument to gain buy-in from studios, broadcasters, and legislators. Despite arguments that captioning could benefit children and English language-learners, the success of developing widespread captioning hinged on its closed nature; hiding it away from the hearing public is an approach in sync "with ideologies of disability and difference as deficiencies to be excluded from public view" (Ellcessor, 2015).

Public awareness and appreciation of the importance of captioning appears to have increased in the era of streaming, with research suggesting that a majority of young hearing people prefer to watch content with captions when possible (Youngs, 2021). That advocacy groups have found success in expanding accessibility services using lawsuits is indicative of changing attitudes from the general public, as broadcasters and streaming platforms are keen to avoid bad publicity and public criticism (Neves, 2018; Podlas, 2018). For example, a second lawsuit brought against Netflix and other major studios in 2015 over the non-captioning of song lyrics was decided in favour of the studios (Choi, 2017; Podlas, 2018). In the wake of that lawsuit, however, Netflix updated its TTSG with the directive to "subtitle all audible song lyrics that do not interfere with dialogue" (English TTSG, I.15). In the summer of 2018, public outcry over captions that had been erroneously censored for season 2 of *Queer Eye,* as well as criticism surrounding the captioning of Jamaican characters in season 2 of *Luke Cage* (Ausiello et al.,

2018; Ratcliff, 2018), saw swift resolution from Netflix, who publicly affirmed its commitment to “delivering a great experience to our deaf and hard of hearing members” (Netflix CS, 2018).

Ellcessor (2012) notes that although the rights of people with disabilities are protected by the state, the enactment and enforcement of access measures through telecommunication laws and corporate actions are indicative of “a neoliberal framework within which people with disabilities must act as ‘consumer citizens’” (p.344). She cautions that it is imperative to keep deaf audiences at the center of discussions of captioning to ensure that civil rights are not usurped by market interests and that industry standards reflect the access needs of deaf audiences rather than the business needs of the digital media industry (2012).

Captioning for the deaf and audio description for the blind are the two most well-known instruments of media accessibility (Greco, 2016). Within a human rights framework, accessibility “describes the degree to which an environment, service or product allows access by as many people as possible, in particular people with disabilities” (WHO & World Bank, 2011). Greco (2016) defines media accessibility from the perspective of AVT as “a set of theories, practices, services, technologies, and instruments providing access to audiovisual media content for people that cannot, or cannot properly, access that content in its original form” (p. 33). Greco emphasizes the universalist nature of accessibility, highlighting that, within the field of AVT, disability and language difference are both types of barriers to access. When captioning is situated within the field of AVT, as it is in Europe, it is grouped together with subtitling, and the specific needs of deaf audiences may thus be overlooked.

Because many other audiences also benefit from closed captioning, including English language learners, individuals with audio processing or attention deficit disorders, and people in noisy environments, closed captioning is often cited as the “electronic poster-child of universal

design" (Udo & Fels, 2010). This "rhetorical widening" of the captioning audience to include hearing audiences is a major component of captioning awareness campaigns today (Downey, 2008; Zdenek, 2015a). This approach reflects "interest convergence," a concept Dolmage describes as "the idea that conditions change for minorities only when the changes can be seen (and promoted) as positive for the majority group as well" (2005). While this tactic has historically been successful in attracting institutional buy-in for captioning from studios and broadcasters, this framing risks further sidelining the deaf audiences who are the primary intended recipients of closed captioning. The deaf audience on its own has – and continues to be – judged as "too small to be of commercial importance" (Ellcessor, 2012, p. 330).

Hamraie critiques the concept of universal design, which is typically framed as "a common sense strategy for making built environments more usable for all people" (2016, p. 286). They argue that the focus on "common sense" and "all people" distances the approach from the notion of disability (Hamraie, 2016). Disability is "an experience of discrimination resulting from inaccessible built environments, rather than an inherent pathology or impairment in the body" (Hamraie, 2016, p. 287). Disability-neutral approaches to universal design frame disability solely as a functional limitation divorced from the context of disability culture and politics (Hamraie 2016). When the framing of universal design is used to advocate for closed captioning, it is another practice that risks sidelining deaf people as the primary audience of closed captioning.

According to the DCMP, the goal of captioning is to provide deaf audiences equal access to content by ensuring that "the meaning and intention of the material is completely preserved" (DCMP, 2022). While captions are intended to provide "equal access," transforming sound and speech into text creates a different, unequal experience, similar to the "equivalence in difference"

dilemma described by Jakobson (2004). Occasionally, written text can provide superior access to audio, clarifying difficult to understand dialogue or terminology, while at other times, it omits information due to spatial and temporal constraints (Zdenek, 2015a). A key principle of universal design as described by Connell, et al. (1997) is the notion of equitable use: "All users should be given the same means to use a product or service. If this is not possible, users should be given an equivalent opportunity for use" (Udo & Fels, 2010, p. 210). In the case of captioning, Udo & Fels note that entertainment and enjoyment need to be evaluated as important factors in how successful captioning is at providing an equitable experience. A critique of captioning must interrogate whether and how captions achieve the goal of providing equal access: do the choices captioners make contribute to an equitable viewing experience for deaf audiences? Are captions preserving the meaning and intention of the material? The choices captioners make reflect their conception of what constitutes equal access for deaf people. It is thus pertinent to explore the conditions under which captions are produced.

Chapter 5: How Captions are Created

The captioning industry today is built for scale and efficiency, a result of the captioning mandates of the Telecommunications Act of 1996, the CVAA, as well as the boom in content production from streaming platforms (Downey, 2008; Zdenek, 2015a). As discussed in Chapter 4, deaf individuals and activists developed captioning technologies and they captioned film and television before the content boom. This is no longer the case in the captioning industry today. In this chapter, I ask how do the constraints of production affect the content of captions and captioners' approach to equal access? How does the scale of Netflix's operation affect captioning processes?

I begin by describing the contemporary film and television industry's reliance on vendors to produce captions, how vendors recruit and train captioners and the steps in captioning workflows. I argue that the industry's reliance on freelance labour and outsourcing is a result of cost-cutting, reflecting studios and broadcasters' continued reluctance to make space in their budgets for captioning services. I note that the low pay rates complicate formal training initiatives, and they impact how much time captioners can dedicate to their work. I explain how captioning style guides are employed as training manuals and quality assurance documents. I outline the different sections in Netflix's TTSG and describe how captions are evaluated based on how well they adhere to the TTSG's instructions.

The Vendor Model

To accommodate the volume of content that needs to be captioned, major studios, networks and streaming platforms contract the majority of subtitling and captioning work to localization vendors, who in turn subcontract tasks to a global network of freelance and in-house editors and translators (Udo & Fels, 2010; Zdenek, 2015a). The work of captioning has been

historically undervalued and underpaid, often framed as an afterthought and an unwelcome extra post-production expense (Neves, 2018; Udo & Fels, 2010). Much of the work is outsourced to countries like India and the Philippines, which have significant English-speaking populations (Pond, 2010), and it is often rushed, with short turnaround times expected (Udo & Fels, 2010). Largescale localization vendors typically advertise captioning positions as "English editor" roles because their work usually includes English subtitling as well as captioning. English editors create and edit captions, as well as other English subtitle files including subtitles that have been translated into English, English templates used as the basis for translation, forced narrative subtitles and dialogue lists[1] (Deluxe, 2022; Pixelogic, 2022).

In the early days of captioning, many deaf and hard-of-hearing individuals were involved in the development and production of captioning and captioning technology (Downey, 2008). Captioning today is performed by hearing individuals, who may have limited connections or experiences with deaf people or Deaf culture. This circumstance is unique in the field of AVT: while subtitlers typically share the language and possibly the culture of the intended recipients of their translation, captioners often know little about deafness and Deaf culture, leaving them "to try and guess the needs, preferences, expectations and requests" of their audience (Morettini, 2012, p. 325).

Vendors recruit individuals with undergraduate degrees in English, communication, journalism, or other humanities and social studies fields to fill English editor positions (Deluxe, 2022; Iyuno, 2022; Pixelogic, 2022; Zdenek, 2015a). Job postings emphasize attention to detail, thorough knowledge of spelling and grammar, an ability to work well under pressure, computer

[1] Dialogue lists are transcriptions of all dialogue and speech sounds used as a reference for dubbing studios. Forced narrative subtitles supplement dubbed versions of video. They contain subtitle translations for plot-relevant elements that are not dubbed, such as onscreen text, song lyrics or archival footage.

literacy, good audio sensitivity and a high degree of film or cultural literacy (Deluxe, 2022; Iyuno SDI, 2022; Pixelogic, 2022; Zdenek, 2015a). Although, as of the early 2000s, many European universities with AVT programs began offering specialized captioning training, formal professional training or a background in disability, caption, or accessibility studies are not often requirements for editor roles (Zarate, 2021).

Clark (2013) rails against this approach to staffing captioning positions, opining that this type of recruitment attracts "a workforce of young people, mostly female, with commercially useless liberal arts degrees" who are "actively harmful to captioning" due to their carelessness and lack of life experience. He takes issue with what he perceives to be the attitude of captioners – that it is an unserious job, done "until something better comes along" rather than a career (Clark 2013). I certainly agree that captioning needs to be taken more seriously within the post-production industry, but I would argue that the poor compensation for the work is at the heart of the issue. Vendor rates can be as low as $1 to $2.25 USD per minute of runtime, which means captioners need to work quickly to make a living wage (Klein, 2022; Pixelogic, 2022).[2] If some of the workforce it attracts is young and uncommitted, it is because it is not sustainable for long-term employment and opportunities for advancement are limited. Visibility for the work that English editors do is also poor. While translator credits are increasingly *de rigueur*, appearing at the end of subtitle tracks for most original content on streaming platforms, captioners are rarely ever credited (Blair, 2021).

I also strongly disagree with Clark's assertion that a liberal arts degree is of no benefit to captioning. Captioners need to be inquisitive, creative problem solvers, and good writers. Cultural literacy is of paramount importance, as effective captioning involves recognizing and

[2] The rates for captioning in India, the Philippines, and countries where work is outsourced are lower, relative to cost-of-living estimates.

describing references and allusions (Zdenek, 2015a). Zdenek describes cultural literacy for captioners as “the ability to be an ideal reader of the text, to recognize the intertextual associations that every text creates with prior texts” (2015a, p. 221). Tools like search engines, online encyclopedias, fan sites, and music identifying apps aid captioners in accurately transcribing dialogue and describing important sonic allusions, but captioners need to be able to recognize when an allusion is being made in the first place (Zdenek, 2015a). In my experience, as a person who has herself entered the captioning industry with a liberal arts degree and who has trained editors on captioning, an education in the liberal arts provides a solid foundation towards this type of cultural literacy.

As for Clark’s misogynistic remark that women make up the majority of the captioning workforce, there is a historical basis for his comment. The Telecommunications Act precipitated the rapid expansion of the captioning industry, and thus in the 1990s through to the early 2000s, the profession did skew more female: many ex-courtroom stenographers, an occupation that skewed largely female, were recruited for captioning positions on the basis of their skills as transcriptionists and fast typists (Downey, 2008). However, in my experience, the field today is more gender balanced. When I worked in-house for a captioning vendor, my department skewed more male at a ratio of 3:1, but the global network of English editors I work with is diverse in age, background, and gender. Rather than gender, I have observed that interest in the content is a better predictor of caption quality. If a captioner is a fan or familiar with the subject matter of the content they are captioning, it reduces the amount of time they need to spend on external research and increases their focus on the captions. Familiarity with names and terminology aids in the accurate transcription of dialogue and makes it easier for the captioner to spot a line that has been misheard or mistranscribed.

Since a formal education in captioning is not a requirement for English editor positions, training mostly happens on the job, relying on style guides, instructional videos, and whatever degree of training the vendor offers. A major issue with this approach is the pay structure: while some coordinators, managers, and editors have full-time salaried or hourly positions, freelancers are paid either by video runtime or wordcount. This means additional administrative tasks, like training or responding to feedback, are frequently unpaid, and thus there is less incentive for freelancers to make themselves available for more than the minimum training they need. The lack of emphasis on training combined with the low pay rates is reflective of industry attitudes that captioning is straightforward transcription and not a specialized accessibility service requiring skilled labour (Udo & Fels, 2010).

Caption Workflows

Captioning is part of the post-production process and is seldom overseen by members of the creative team (Udo & Fels, 2010). In my experience with Netflix Originals and other pre-release content (especially if an auteur director is involved), content producers are sometimes hands-on, sharing extra resources or specific requests, notes, or feedback from the director. This is often because in addition to being captioned, the content is being simultaneously translated and/or dubbed into many languages, so the assistance is often more oriented towards translation than it is captioning. Nevertheless, it is beneficial when captioners receive special instructions about the content from the creative team or are able to reference a director's annotated version of the shooting script, for example, to ensure the captions respect the creative intent of the producer.

When a company has been contracted to provide captioning, the work is managed by project managers or coordinators, who assign origination and proofing tasks to freelance captioners. Most captioning for pre-recorded programming is done entirely within specialized

software, many of which are cloud-based (DCMP, 2022). Origination, or the creation of the caption file, is done by playing a video that has an imbedded timecode and either spotting a script/dialogue list while adding nonspeech information or by manually timing and transcribing all of the content (Milinkiewicz, 2020; Netflix, n.d.-e)[3]. Spotting refers to the process of timing text events, setting precise in and out times for each caption (Zarate, 2021). In recent years, automated speech recognition (ASR) software has been used to aid in the origination process, with origination editors correcting the ASR-generated file (Apptek, 2020). In some cases, the origination process involves a complex conform, where an existing English subtitle file or template file is adapted into captions by an editor, who adds nonspeech information and brings the file closer to verbatim[4] (Netflix, n.d.-g).

Next, the caption file is reviewed by at least one other editor before it is sent through internal and/or external quality control checks. The number of proofs and checks done depends on the type of content and how much the client is willing to pay for captioning services. Vendors will typically prioritize higher-profile projects, such as new theatrical releases, broadcast television and streaming services' original content, over lower-profile projects, such as reruns or back catalogues of older shows. Most, if not all of Netflix Originals' captions go through a manual quality control check (Netflix, n.d.-e).

Freelance editors often have little visibility beyond the immediate task at hand. If they have questions or reservations about how to approach something in their project, they may be able to alert their coordinator, but in most cases, due to the short timelines and fast turnarounds typical in captioning, issues are resolved by someone else. Freelancers also have little control

[3] This process is followed for pre-recorded content. Live content, on the other hand, is often captioned via respeaking, and thus typically involves different workflow, preparation, and software. (Romero-Fresco, 2009)
[4] The recommended reading speed for Netflix's English language templates is 20% lower than for its captions at 17 characters per second and thus they are typically heavily condensed (Template TTSG, I.23).

over what work they are assigned. Some vendors may make efforts to coordinate work according to their editors' interests or strengths and mimic the workflow of translation, where the same editors work on the entirety of a season or series to improve consistency. This practice facilitates an important element of captioning that Zdenek calls "series awareness," which is "a working knowledge of how previous episodes have been captioned" (2015a, p. 247). He states that captioners "need to look beyond the individual episode, treating each episode as a 'fragment' in a larger whole that unfolds over time, to see how recurring sounds across episodes have been captioned" (2015a, p. 247). While this is often studios' and vendors' goal, it is not always possible. Unfortunately, because captioners are not credited, it is virtually impossible to gauge how widespread this practice is.

Style Guides

If the "first wave" of captioning advocacy was the movement to raise public awareness of the importance of media accessibility for the deaf and to make captions accessible to television and film viewers, the second wave is preoccupied with ensuring the accessibility of new media and pushing for better quality captions. Defining and articulating what quality captioning entails is currently a pressing concern within the captioning industry (Romero-Fresco, 2021). Although captioning was initially developed to make film and television accessible to deaf audiences, it is now also used on newer media, including streaming platforms, social media videos, video games, and virtual reality technology (Kritzer & Smith, 2020). Captioners working in these new fields risk carrying over practices that are antiquated or have not been sufficiently interrogated or questioned in the context of twenty-first century media's capacities and contemporary deaf audiences' desires (Neves, 2008).

Evaluating caption quality is a challenging task due its subjective nature and the different perceptions, needs, and expectations held by caption producers and consumers (Neves, 2018; Romero-Fresco, 2021). Devising a quality assessment model to suit all stakeholders involved in captioning is especially difficult considering the marked separation between the researchers who devise models and guidelines, the broadcasters and networks that commission captions, the political and social bodies that legislate and advocate for captioning provisions, the companies that provide captioning services, and the deaf and hard-of-hearing users, who are often cut out of the process altogether (Neves, 2008; Romero-Fresco 2021). Despite how challenging it is to create a model to measure quality that satisfies all stakeholders involved, quality is nevertheless judged daily by captioners, quality assurance editors, broadcasters, and viewers (Pedersen, 2017).

Within the captioning industry, the current most popular method for evaluating caption quality relies on style guides: captions are assessed based on how well they adhere to a studio or vendor's particular style guide (Kuo, 2020; Zarate, 2021; Zdenek, 2015a). This method, which purports to be somewhat objective, focuses on whether captions follow the orthographical and reading speed rules and recommendations that are laid out in the guides. As captioning firms, networks, and streaming platforms seek to streamline captioning processes to accommodate ever-increasing volumes of content, guidelines are a useful tool to improve the efficiency of quality assurance processes (Zdenek, 2015a). However, although the last two decades have marked significant advancements in video technology and major changes in distribution models, captioning conventions and guidelines have remained virtually unchanged since their inception (Udo & Fels, 2010; Zdenek, 2015a).

Style guides are comprehensive instruction manuals captioners follow to caption content. Captioning style guides may differ on stylistic recommendations, but they generally cover the same topics, including "accuracy, reading speed, screen placement, line breaking, timing, grammar, punctuation and nonspeech information" (Zdenek, 2015a, p. 54). As Zdenek observes, style guides are "light on theory," and their instructions are "offered up as truths in no need of justification" (2015a, p. 52), making it difficult to decipher the rationale behind some of their pronouncements. Their focus is usually on technical parameters and style, rather than offering practical instructions for content-related issues (Kuo, 2020). Most captioning vendors have their own proprietary in-house style guides that captioners are trained on and instructed to follow (Zdenek, 2015a; Zarate, 2021). These style guides are generally not publicly available, however some publicly funded broadcasters and non-profits, most notably the DCMP and BBC, have made their guidelines freely available (BBC, 2022; DCMP, 2022). If the networks or studios commissioning captions have their own stylistic preferences or style guides, the captioning vendor shares those instructions with the captioner.

The Netflix Timed Text Style Guide

Captioners working on Netflix Original content must follow Netflix's Timed Text Style Guide (TTSG) rather than any in-house style guides. Netflix makes its 47-language specific TTSGs available to the public online on its Partner Help Center (Netflix, n.d.-j). Captioners primarily need to refer to three guides: the English Timed Text Style Guide (aka English TTSG), the Timed Text General Requirements (aka General TTSG), and the Subtitle Timing Guidelines (aka Timing TTSG). The English TTSG is divided into two sections, Subtitles for the Deaf and Hard of Hearing (SDH) and English Subtitles, which is for non-English language content (English TTSG I & II).

The English SDH TTSG contains 18 subsections that dictate:

- instructions regarding content accuracy and editing
- technical restrictions (e.g. line length, reading speed, required font specifications)
- orthographical instructions (e.g. when to employ italics; how to use punctuation)
- formatting instructions (e.g. how to format nonspeech information; how to indicate multiple speakers; line breaks)
- stylistic preferences (e.g. how to transcribe numbers, dates, and song lyrics; how to approach foreign dialogue)
- some genre-specific and content-specific instructions (e.g. a modified italics rule for unscripted content and sensitive language)

Generally speaking, captioning style guides are hard to study because of access challenges and a lack of transparency; they typically credit organizations or companies rather than individual authors, they do not cite when changes were introduced, nor do they offer references to the specific research or sources that have informed them (Zdenek, 2015a; Zarate, 2021). Netflix's TTSG rectifies the problem of access, since it is available online, and it does have a vague update history as of December 2015. Netflix's information management does leave something to be desired, as additional important information about captioning is scattered through a few other points in the Partner Help Center, including the FAQ and QC sections (Netflix, n.d.-a; Netflix, n.d.-e). The English TTSG, like most other caption style guides, provides specific directives for captions, but it does not explicitly articulate qualitative standards for captioning, nor does it cite the research that informs it. Regardless, the availability of the TTSG presents a valuable opportunity to study the conventions and expectations of the current largest subscription-based streaming platform (Maglio, 2022).

The notion of style guides as iterations of what quality captions should look like is reinforced by Netflix's quality control process which focuses exclusively on how well captions adhere to the TTSG. Caption files for Netflix Originals are reviewed by a localization quality control editor, commonly known as a "QCer," who may be a Netflix representative or a third-party contractor.[5] The QCer watches the captioned video in Netflix's proprietary subtitling software Originator, looking for typographical mistakes and violations of the TTSG (Netflix, n.d.-e). The QCer is able to make corrections they deem necessary directly on the caption file. QCers generate a report that lists the errors, which are categorized according to Netflix's list of "Manual QC Error Messages," as well as the changes made (Netflix, n.d.-i). The report is shared with the vendor, but in my experience, this report rarely makes its way back to the captioner/s who originally worked on the file. If the QCer detects a serious number of issues, it may trigger a redelivery request from Netflix who will share the report with the vendor and ask that the caption file be corrected and resubmitted (Netflix, n.d.-b). Netflix employs this QC practice because it is relatively fast, its standardization accommodates a high volume of content, and it allows Netflix to keep quantifiable metrics around how well the vendors perform (Netflix, n.d.-e). However, one significant drawback to this approach is that it focuses on the negative – the QCer is focused on identifying mistakes and the reports are only shared if they contain mistakes and corrections. This means negative feedback is shared with captioners far more frequently than positive feedback, as within this model, no news is good news. Moreover, this approach does not consider quality in relation to the content. The QC process operates on a right/wrong binary, relative to the TTSG and not the specific content. This effectively makes captioners' first loyalty

[5] Only Netflix Originals and "higher profile" content are subject to a full QC review. Titles that Netflix expects will be "less popular" are subjected to a spot QC review, wherein the QCer performs spot checks on the beginning, middle and end of the caption file (Netflix, n.d.-e).

to the TTSG, rather than to the content, complicating their ability to provide equal access to the material.

This examination of the caption production process and the current structure of the captioning industry reveals a few reasons why captioners may struggle in their pursuit to provide equal access to audiovisual content. Firstly, captioners work in isolation – they are siloed off from both the content producers, the deaf audiences they serve, and in many cases, from receiving feedback about their work. The industry's reluctance to improve compensation for captioning work means captioners have to work quickly to earn a living, and thus they may not be able to dedicate as much time as they might need for research to produce quality captions. Without formal training or personal experiences with deaf people or Deaf culture, captioners predominantly rely on style guides for training and to guide them as they seek to make content accessible to deaf audiences. Reliance on style guides can lead to issues in caption quality, as I will explore in Chapter 7. Because style guides are written in general terms, some of their directives may not be suitable for all the unique contexts and situations captioners may encounter in their work.

Style guides are designed to provide captioners with instructions on how to make content accessible in line with industry expectations of captioning and the studio's vision of media accessibility. A study of Netflix Originals' closed captions thus needs to evaluate how the TTSG reflects and articulates standards for quality captioning in its directives. To do so, in Chapter 6, I present the DCMP and FCC's eight principles for quality captioning and examine how the principles of accuracy, completeness, readability and synchrony come into conflict due to the TTSG's timing instructions and its target reading speed.

Chapter 6: Principles of Quality Captioning

Because style guides embody the norms and regulations of the industry (Pedersen, 2020), a useful starting point to critique them and uncover the rationale behind them is looking at regulating bodies and advocacy groups' recommendations. In the case of American captioning, as I establish in Chapter 4, the foremost of these would be the FCC and DCMP. In this chapter, I ask, how does the Netflix TTSG address the needs of deaf audiences? How are industry standards articulated in style guides?

I introduce the eight criteria of quality captioning identified by the FCC and DCMP, which will inform my analysis of Netflix Originals' captions in Chapter 7. My focus here is mostly on accuracy, completeness, readability, and synchrony, as I explain the TTSG's timing and reading speed requirements and provide a brief review of some of the literature that informs and challenges these metrics and practices in the hopes of demystifying the TTSG's "truths" that are presented without explicit justification (Zdenek, 2015a).

Principles of Quality Captioning

The FCC oversees captioning for American broadcast television, and as such, those captions are required to follow the FCC's Rules for Caption Quality as presented in Title 47 of the Code of Federal Regulations (FCC, 2021; Telecommunication, 2022). Although these specific rules do not apply to captions on streaming platforms, they influence the captioning industry broadly. As described in Chapter 4, the DCMP, a federally funded non-profit, is another influential organization within the captioning industry, and it has been supplying captioning guidelines since the 1950s (DCMP, n.d.). Its style guide, the Captioning Key is a highly influential guide that includes a statement on elements of quality captioning (DCMP, 2022). Between them, the FCC and DCMP present eight criteria of quality captioning: accuracy,

completeness, synchrony, readability, placement, consistency, clarity, and equal access (DCMP, 2022; FCC, 2021).

According to the FCC, accurate captions "match the spoken words in the dialogue and convey background noises and other sounds to the fullest extent possible," (FCC, 2021) "without substituting words for proper names and places, and without paraphrasing, except to the extent that paraphrasing is necessary to resolve any time constraints" (47 CFR 79.1). Captions must also appear in synchrony "with their corresponding spoken words and sounds to greatest extent possible" and they must be complete, running from the start to the end of the video (FCC, 2021). Lastly, the FCC requires that captions are correctly placed, not running off the edge of the screen or blocking important on-screen text or content (FCC, 2021). According to the DCMP (2022), readability refers to the pace at which captions are displayed: audiences need to be able to read each caption completely. Consistency, which is "crucial to viewer understanding" involves "uniformity in style and presentation of all captioning features" (DCMP, 2022). Clarity is achieved by including speaker identification and non-speech information as part of the "textual representation of the audio" (DCMP, 2022). Lastly, equal access "requires that the meaning and intention of the material is completely preserved" (DCMP, 2022).

These elements of quality captioning offer a point of entry to analyze the Netflix TTSG to understand the goals and reasoning behind some of its many "truths" offered up without justification (Zdenek, 2015a). An examination of the TTSG through this lens reveals how these elements work in concert with one another and how they come into conflict. There is a breadth of creative decision-making on the part of the captioner as they negotiate *how* to achieve accuracy and synchrony to the "fullest/greatest extent possible" (FCC, 2021). The goal of the TTSG is to facilitate "an effortless experience" for the caption user, who should "be able to read…closed

captions without being reminded they are reading" (Netflix, n.d.-k) – in other words, the media accessibility ideal of providing equal access to the material.

Accuracy Versus Timing

The ideal of accuracy is reflected in the first section of the English TTSG, which instructs captioners to "include as much of the original content as possible" without "simplify[ing] or water[ing] down the original dialogue" (English TTSG, I.1). It states that "truncating the original dialogue should be limited to instances where reading speed and synchronicity to the audio are an issue" and "when editing for reading speed, favour text reduction, deletion and condensing, but do not paraphrase" (English TTSG, I.1). Netflix captions are a maximum of two lines, which can each hold a maximum of 42 characters (English TTSG, I.2, I.10). To achieve the goal of "including as much content as possible," captioners have to navigate the TTSG's exacting timing rules and its restrictions on reading speed, line length, and on the formatting of dual speaker captions.

Netflix identifies 20 characters per second (CPS) as the ideal reading speed for adult programming (English TTSG, I.14). Most captioning software, including Netflix Originator, display the CPS calculation for each individual caption, so when creating and editing caption files, the captioner can see how the addition or removal of each character affects the reading speed (Bond, 2019). Netflix measures the reading speed of individual captions, as well as the average reading speed of the entire caption file (Netflix, n.d.-c). To accommodate dialogue-heavy content, Netflix offers a degree of flexibility and does allow "certain" captions "to go up to 30% over the reading speed limit" (Netflix, n.d.-f). Netflix does caution, however, that if the average reading speed of the entire file exceeds 20 CPS by more than 30%, it will be rejected by their QC team (Netflix, n.d.-f).

Netflix expounds the concept of synchrony in the Subtitle Timing Guidelines, which provides instructions for subtitle and caption timing that all timed-text providers must follow (Timing TTSG). Netflix's timing rules are some of the strictest and most precise in the industry – while most vendors and style guides do counsel captioners to time their captions to the corresponding audio and to avoid crossing shot changes where possible, Netflix's TTSG provides frame-specific timing guidelines to achieve these goals (Zarate, 2021). Per the TTSG, captions must be timed to the audio, starting "on the first frame of audio or as close to it as possible, using the waveform as reference" and they may end up to a half second past the end of the audio (Timing TTSG, 1). In a continuous sequence of captions, the captioner is to create "an even run of [captions]" by extended the end time of the previous caption to two frames before the start time of the subsequent caption (Timing TTSG, 1). Adjacent captions must have a minimum two-frame gap between them, but if a bigger gap is needed between two captions, the gap should be at least 12 frames wide (Timing TTSG, 4).

Regarding shot changes, the TTSG recommends that captions start on the first frame of a shot change if the audio starts within half a second after the shot change (Timing TTSG, 2). Similarly, captions can be timed to end two-frames before the shot change if the audio ends within half a second of the shot change. Captions "may cross shot changes when the dialogue they represent also crosses the shot change," but they must do so by at least 12 frames (Timing TTSG, 3). Netflix acknowledges that following these prescriptions can be a challenge and suggests the strategy of "borrowing time" to help captioners make content fit the timing rules (Timing TTSG, 5). Borrowing time entails merging adjacent captions or re-spotting and re-segmenting dialogue in a scene, as sometimes changing caption division improves reading speed metrics (Timing TTSG, 5). However, this strategy is hampered by a separate, newer rule added

in July 2020 regarding captions that contain two speakers: “Text in each line in a dual speaker [caption] must be a contained sentence and should not carry into the preceding or subsequent [caption]. Creating shorter sentences and timing appropriately helps to accommodate this” (English TTSG, I.6).

Netflix states that the plethora of timing rules they expect captioners to follow are designed to “allow the [captions] to fit neatly within the edit of the content” (Timing TTSG, 2). The reality, however, is the soundscape of a film or the pace at which characters speak seldom perfectly conforms to Netflix’s timing and reading speed prescriptions. This means captioners have to conform the content to fit the timing restrictions, complicating captioners’ commitment to accurately transcribing the dialogue and their pursuit of providing deaf audiences with an equivalent representation of the audio. Netflix acknowledges that following these timing rules while also representing the audio accurately will involve compromise and asks captioners to “apply good judgement as many timing decisions may feel subjective” (Timing TTSG, 3). However, because of the rigidity of the timing rules and Netflix’s QC process, which evaluates how well captions adhere to the TTSG, captioners are more likely to edit and condense dialogue to avoid breaking any timing rules or exceeding the readability allowance.

Reading Speed

In Chapter 7, I will explore in greater depth how the practice of editing affects the way the meaning and intention of content is preserved in the captions. To provide some context as to why Netflix set 20 CPS as its target reading speed for deaf audiences, I will first review the scholarship on caption reading speed. I devote special attention to this topic because the debate surrounding reading speed and the extent to which captions should be edited is among the most

sensitive and contentious issues within caption studies and the captioning industry (Ofcom, 2005; Romero-Fresco, 2009; Zdenek, 2015a).

The consensus among researchers who study the reading skills and speed of deaf and hard-of-hearing students is that deaf children and young adults lag behind their hearing counterparts in learning literacy skills (Bélanger & Rayner, 2015; Bélanger et al, 2018; Burnham et al., 2008; Conrad, 1977; De Linde & Kay, 2006; DiFrancesca, 1971; Kritzer & Smith, 2020; Neves, 2008; Romero-Fresco, 2009; Shroyer & Birch, 1980). Deaf children are at a disadvantage relative to hearing children who are immersed in oral and written language learning from a young age, hearing and reproducing speech and speech sounds as they learn to read (Bélanger, et al 2018). According to Torres Monreal & Santana Hernández (2005), improving deaf students' reading comprehension and reading speed is further complicated by a lack of "the encyclopedic language knowledge necessary to understand texts" (p. 380). While literacy research indicates prelingually and profoundly deaf students struggle to learn how to read, it does not indicate that hard-of-hearing students or individuals who acquire hearing loss post-lingually face any difficulty with reading skill or speed (DiFrancesca, 1971; Conrad, 1977). This research on deaf literacy has shaped how broadcasters, vendors and regulating bodies approach caption speed.

Calculating a target reading speed is challenging because of the number of variables involved, including the quantity and complexity of the dialogue, the amount of visual information displayed on screen, the viewer's level of interest in the content, the amount of hearing the viewer possesses, the viewer's age, the viewer's level of education, and the viewer's level of exposure to captioning (Jensema, 1998; De Linde & Kay, 2006; Szarkowska & Gerber-Moròn, 2018; Zdenek, 2015a; Zarate, 2021). The experience of reading captions is markedly different from reading a static text, as one may read such a text at any pace they choose, whereas

the reading speed of captions is fixed, with viewers forced to follow the captions at the pace they are displayed (Neves, 2018). Romero-Fresco (2009) describes caption speed as a technical, economic, political, and ideological issue: technical and economic because of the calculations involved and the amount of labour required to meet required reading speeds, and political and ideological because reading speeds reflect how broadcasters and vendors perceive the reading skills and needs of their deaf and hard-of-hearing audiences.

In the 1950s and ‘60s, the NAD heavily edited dialogue in film captions by simplifying vocabulary to a lower reading level and truncating dialogue to lower the reading speed (Downey, 2008). In the 1970s, as WGBH and PBS developed closed captioning programs, the two institutions were divided on how to proceed: Doris Caldwell at PBS believed in creating captions as close to verbatim as possible, while Phil Collyer at WGBH advocated editing captions to broaden the service’s appeal to as wide a population as possible (Downey, 2008). Caldwell cautioned that “rewriting content for purposes of simplification may be both unwise (posing the danger of distorted information) and unwarranted (serving well only a minority of our target audience)” (in Downey, 2008, p. 91). Caldwell’s belief was rooted in part in PBS’ preliminary research which indicated deaf viewers found verbatim captions to be acceptable (Downey, 2008, p. 91).

This debate continues today amongst captioning providers, researchers, and viewers. Romero-Fresco (2009) surmises that while some broadcasters tend to push for verbatim captioning to keep costs down, many researchers within caption studies cite reading speed research as a reason to edit captions (Burnham et al., 2008). The majority of deaf organizations and viewers, however, advocate for verbatim captioning, as they contend that editing is a form of censorship that denies deaf audiences equal access to content (Romero-Fresco, 2009; Neves,

2008). Deaf audiences consistently request verbatim captioning and when surveyed about caption quality, they often cite edited dialogue as a significant source of frustration (Morettini, 2020; Romero-Fresco, 2009; Szarkowska et al., 2011; Szarkowska & Moròn, 2018; Zarate, 2021; Zdenek, 2015a). Some more current studies suggest that deaf people are more exposed to captions and texted media through the use of social media, video games, internet streaming and video-on-demand services, which may result in faster average reading speeds (Kritzer & Smith, 2020; Lång et al, 2021). This would reinforce Jensema's (1998) observation that increased exposure to captions improves deaf audiences' comfort with faster reading speeds. More research is needed in this area, as most research about deaf literacy has focused on the group's weaknesses rather than strengths (Bélanger & Rayner, 2015).

Despite deaf audiences stating a preference for verbatim captioning, some researchers are reluctant to accept that verbatim is a suitable approach. For example, Sancho-Aldridge calls for the need to "disentangle the politically sensitive issue of 'access' from the practical issue of which style, in real terms, provide[s] deaf viewers with the most information" (1996, p. 24). Romero-Fresco cautions that verbatim captions, if displayed too quickly, may not actually provide access to audiences if they cannot be read (2009). As a result, most professional captions are created with a target reading speed that is set by a regulating body, the broadcaster commissioning them, or the vendor.

While the TTSG does not specify how Netflix decided upon 20 CPS as the target reading speed for its captions, its decision was likely informed by the research reviewed above, as well as its own research. In Chapter 7, I will analyze how Netflix captions are edited to accommodate reading speed and discuss how this practice affects the way the meaning and intention of the original content is conveyed. I also examine how the TTSG employs orthographical conventions

to enforce clarity in captions and assess how consistency and clarity come into conflict in its directives surrounding speaker identifiers. Lastly, I analyze the TTSG's directives regarding the captioning of foreign language dialogue, which challenges Netflix's commitment to providing equal access to audiovisual content through its captions.

Chapter 7: Netflix Captioning Practices in Action

In this chapter, I examine three captioning practices in action, analyzing instructions articulated in the TTSG and engaging in close readings of excerpts of captions from a variety of Netflix Original series and film to examine how the instructions are interpreted and acted on by captioners. I ask, how do Netflix closed captions serve the needs of deaf audiences? How does the TTSG affect the content of captions and captioners' approach to equal access? How do captioners perceive the needs of deaf audiences?

I begin by looking at the process of editing captions to meet timing and reading speed requirements, and I engage in a close reading of captions excerpted from *Gilmore Girls: A Year in the Life* (Sherman-Palladino, 2016) to see how the edits impact the representation of the meaning and intent of the content. Next, I look at how the TTSG's instructions surrounding orthographical conventions ensure captions are clear and consistent – and examine how the directives regarding speaker identifiers put clarity and consistency into conflict in a variety of contexts. Lastly, I evaluate the TTSG's instructions related to the captioning of foreign dialogue, illuminating the assumptions Netflix makes about its audience and illustrating how the directives to translate, cue, and transcribe the dialogue affect the meaning conveyed to the audience.

Editing Captions

While captioners do strive to create verbatim captions for Netflix, its caption timing rules, line length restrictions and 20 CPS limit (English TTSG) complicate this endeavour, as does the speed at which dialogue is delivered. While Netflix simultaneously instructs captioners not to "simplify or water down the original dialogue," the TTSG acknowledges that reducing the character count by editing dialogue may be the only way to meet the target reading speed (English TTSG, I.1). English editors who create templates for translation are especially familiar

with the practice of editing, as condensing content is an industry standard in subtitle translation (Ferriol, 2013; Kovačič, 1994). Netflix's recommendations as to how the captioner is to edit the text are minimal: the captioner is asked to "favor text reduction, deletion and condensing" but is told not to paraphrase (English TTSG, I.1). When the target reading speed cannot be met, the captioner is forced to choose which words and phrases can be omitted without corrupting the message or deviating from the creative intent of the content producer. However, edited dialogue is an intralingually translated text, presenting a fundamentally different text from the dialogue as it is spoken. It is thus important to assess caption editing strategies to judge how they affect meaning, how they change the viewing experience for deaf audiences and how they affect captioners' pursuit of providing equal access to audio.

The TTSG's lack of concrete directives about how to reduce and condense are part of an industry-wide assumption that captioners should be able to intuit which dialogic elements can be removed when adapting dialogue to improve reading speed (Kovačič, 1994; McIntyre & Lugea, 2015). In-house style guides typically provide more comprehensive instructions, especially in cases where editors work on both English templates and captions. The target reading speed of English translation templates for Netflix is an even more restrictive 17 CPS (Template TTSG, 23) and thus most English editors are well-versed in editing text to accommodate reading speed. Reduction and condensing involve removing smaller elements of a phrase, whereas deletion is the total removal of a phrase (Kovačič, 1994). The subtitle editing process Kovačič describes can also be applied to captioning:

> When the [captioner] is short for space, he/she evaluates the relative relevance of individual segments of a given message. Relying on the viewers' ability to apply adequate cognitive schemata or frames and to draw on either previous information in the

> story or their general knowledge of the world, the [captioner] leaves out the part of the message he/she considers the least relevant for understanding the message in question, for perceiving the atmosphere of a situation or the relationship amongst the participants involved, and eventually for the general understanding and reception of the story (1994, p. 250 in De Linde & Kay, 2006[6]).

The notion of "relevance" is context-based and hinges on the judgment of the captioner. The captioner decides which elements of dialogue can be omitted while still ensuring the viewer has access to the intended message (Kovačič, 1994). This process is subjective and interpretive, and different captioners will edit dialogue differently based on their own evaluations of relevance (De Linde & Kay, 2006). Captioners will consider relevance firstly as it relates to the individual utterance, considering how the meaning of the sentence is affected by removing parts of the dialogue. Then captioners consider relevance to plot, assessing based on the context of the scene and/or the preceding scenes. Plot is typically an immediate and relatively straightforward basis on which to judge relevance, especially in cases where the captioner is working on an individual episode or segment of a film. Plot-relevance is also top of mind for captioners because it is emphasized throughout the English TTSG in other areas of creative decision making, such as when to include sound cues and what lines of dialogue to prioritize (i.e. foreground vs. background) (English TTSG, I.17). Relevance to characterization and atmosphere, however, are not as immediately discernable because they develop more gradually over time (McIntyre & Lugea, 2015). Thus, important dialectal features that contribute to these aspects may by excised because they are considered less relevant or are less accessible to the captioner. As Kovačič

[6] Excerpt cited in De Linde & Kay, 2006, but the particular excerpt does not appear to be part of the published work.

notes, "[the captioner's] own context may be insufficient or inadequate for him to make the right decision" (1994, p. 251).

The TTSG lists two practices that captioners are *not* to follow. Firstly, paradoxically, the TTSG mandates that "transcription of the source language should follow the word choice and sentence order of the spoken dialect. Slang and other dialectal features should not be changed" (English TTSG, I.1). While captioners are counselled not to "change" dialectal features or slang, words that fit into these categories are commonly deleted, as will be discussed below. Secondly, captions are not to be paraphrased (English TTSG, I.1). When reducing text, the captioner is supposed to remove content, but not add or change the dialogue that remains. Thus, captioners favor edits that allow for words or phrases to be removed without compromising the sentence's grammatical integrity, leaving what remains a functional, fluent sentence. Though the TTSG provides no rationale for this instruction, paraphrasing is often cited by audiences in caption reception studies as a disruptive practice, particularly by hard-of-hearing or lip-reading individuals (Morettini 2020; Szarkowska et al., 2016). Szarkowska et al (2016) suggest that when captions do not match the exact words spoken, they cause "perceptual confusion" because the intersemiotic cohesion between the visual and auditory channels is disrupted (p. 198).

Romero-Fresco (2009) and Zarate (2021) identify several categories of words and phrases that are most frequently omitted when dialogue is edited: discourse markers, connectors, intensifiers, repetitions, hesitations, and fillers. These types of words and phrases are characteristic of oral speech, which is less formal and structured than written English (Zarate, 2021). To illustrate how removing these elements affects content, I have excerpted several captions from *Gilmore Girls: A Year in the Life* (Sherman-Palladino, 2016), a Netflix-produced sequel to the WB show *Gilmore Girls* (Palladino et al., 2000-2007). The dialogue in *Gilmore*

Girls is well-known for its speed and its density, packed full of pop culture references and witty repartee; scripts for the series were often 20 pages longer than that of the average hourlong comedy (Austerlitz, 2020). The series' creator Amy Sherman-Palladino required word-perfect delivery from the actors, who were not permitted to deviate from the script (Grady, 2016). This poses a unique challenge for the captions, which have to be edited significantly to be displayed at a readable pace.

Gilmore Girls centers on the relationship between Lorelai and Rory, a mother-daughter duo who have an especially close relationship, and their relationship with Lorelai's parents, Emily and Richard. Lorelai left her wealthy, upper-class parents' home in Connecticut when she was sixteen years old and pregnant with Rory. In the original series, Rory was a teenager attending high school, while in the Netflix Original sequel, Rory is a freelance writer in her early thirties, floundering in her career. In the below scene from episode one, "Winter," Lorelai and Rory are discussing Rory's Brooklyn apartment she has recently vacated, leaving her without a place to live.

Table 2

English Closed Captions from Gilmore Girls: A Year in the Life, Winter, 00:10:24-00:10:52

In Time and Out Time	Speaker	Netflix Closed Caption	Dialogue Spoken (if different)
00:10:24.03 00:10:25.22	Rory	It was a perfectly decent apartment.	
00:10:26.00 00:10:29.01	Lorelai	I liked how you had to avoid the walls or you'd stick to them.	I liked how you had to avoid **touching** the walls or **your clothes would** stick to them.
00:10:29.03 00:10:31.02	-Rory -Lorelai	-One wall. -See, in some apartments,[7]	

[7] *Gilmore Girls: A Year in the Life* predates the rule about dual speaker timing. If this episode was captioned today, this line would need to be radically altered to form a complete sentence. "-See, in some apartments, you don't have to avoid any walls at all," a 67-character long sentence, would need to be edited down to 42 characters.

00:10:31.04 00:10:32.17	Lorelai	you don't have to avoid any walls.	you don't have to avoid any walls **at all.**
00:10:32.19 00:10:34.23	Rory	It was a great place. I just was never there.	
00:10:35.01 00:10:37.04	Rory	Do you know how many nights I spent there this year?	Do you know how many nights I **actually** spent there this year?
00:10:37.06 00:10:38.19	-Lorelai -Rory	-Three? Eighteen? -No. No.	-Three? -No. -Eighteen? -No.
00:10:38.21 00:10:40.05	-Lorelai -Rory	-Forty-six? -No.	
00:10:40.07 00:10:43.12	Lorelai	Is this *The Price Is Right?* If I'm a little under I win, over, I lose?	Is this **like** *The Price Is Right* **where** if I'm a little under, I win, over, I lose?
00:10:43.14 00:10:45.11	Rory	I don't know exactly how many nights, okay?	I don't **actually** know exactly how many nights, okay?
00:10:45.13 00:10:47.19	Rory	It was not a lot. Now the building's going condo.	**But** it was not a lot **and** now the building's going condo,
00:10:47.20 00:10:50.23	Rory	I'm not ready to buy a condo, so it seemed like the time to say...	**and** I'm not ready to buy a condo, so it **just** seemed like the right time to say…
00:10:51.02 00:10:52.21	-Rory -Lorelai	-Bye-bye, Brooklyn. -Bye-bye, Brooklyn.	

Note. From "Sherman-Palladino, A. (Writer & Director). (2016, November 25). Winter (Season 1, Episode 1) [TV episode]. In D. Palladino & A. Sherman-Palladino (Executive Producers), *Gilmore girls: A year in the life.* Warner Brothers Television; Netflix. https://www.netflix.com"

Some of the faster lines in this scene are spoken at 25 and 26 CPS and include more text than can comfortably fit on two lines of 42 characters. The back and forth between Lorelai and Rory when Lorelai is guessing "Three? Eighteen?" cannot be displayed synchronously with the audio because of the speed. The caption is only 37 frames long, so if it were to be split into two, the individual captions would be below the 20-frame minimum length that the TTSG allows (General TTSG, 1). As a result, the captions present a compromised version of the exchange, linearizing the dialogue so that is read one speaker at a time, rather than altering the content of

the exchange. The rest of the dialogue in the scene, however, required a significant degree of editing to bring the reading speed within the allowed 20 CPS threshold.

Discourse markers – words such as “so,” “well,” “I mean,” “you know,” “just,” “okay,” “now,” etc. – are typically the first to be excised when editing captions, as they can be easily removed without compromising the grammatical integrity of the rest of the utterance, since they are typically spoken at the start or end of sentences. Discourse markers function to connect, organize, and manage speech (Cambridge Dictionary, n.d.). They provide interpersonal information, indicating relationships and intersubjective positioning (McIntyre & Lugea, 2015). They may mark a change in topic, indicate items in a series, signify shared knowledge, serve as response tokens to acknowledge someone is paying attention in a conversation, demonstrate attitude or to soften a statement by making it sound less direct (Cambridge Dictionary, n.d.). Removing some discourse markers makes minimal impact on the overall message of each line above, but it has the effect of formalizing the speech. The captions thus present a “quantitatively impoverished” and “ennobled” version of the source, producing sentences that are more “elegant” and “readable” than the original (Berman, 2000). The dialogue loses some of the tone and attitude that the discourse markers like “actually” and “at all” bring to the exchanges, which changes the rhythm of the dialogue. If the same strategy is used on all of the faster dialogue in *Gilmore Girls,* the effect is homogenizing and potentially damaging to the characterization process. Formalizing the dialogue between Rory and Lorelai could lessen the contrast the show establishes in their relationship versus the relationship of Lorelai with her upper-class parents who speak more formally than Rory and Lorelai.

Connectors are similarly commonly targeted for removal. When words like “and,” “but,” and “though” that begin clauses are deleted, the remaining dialogue is restructured into more,

shorter sentences (Romero-Fresco, 2009). This reconfiguration is popular in Netflix captions because it helps captioners follow the rule that prohibits sentences in dual-speaker captions from continuing into the preceding or subsequent caption (TTSG, I.6). In the line about Rory's building "going condo," this alteration changes the way the rhythm of dialogue is represented. Long run-on sentences in exchanges between Lorelai and Rory are a hallmark of the series and of their dynamic, especially when they are anxious or worked up about events. When chopped up into discrete sentences, Rory's rant about her apartment going condo appears more rational and less like a panicked defense of her decision to give up her apartment. Berman cautions that the rhythm of theatre – in this case, a script – is vulnerable to destruction through translation (2001, p. 292). By re-segmenting dialogue into shorter sentences, the rhythms of the characters' speech are changed. In an analysis of the dialogue in a dinner scene in the original *Gilmore Girls* series, Blackman (2020) points out that Emily speaks in shorter, frequently interrogative sentences, that when combined with her delivery, convey passive aggression. Blackman observes that when the four main characters are together, the speed of their dialogue remains the same, but all of their sentences – except Richard's – are shorter, resembling a series of clipped punch lines. The significance of the shorter sentences in the group dynamic is diluted when longer sentences elsewhere are split and abbreviated.

Assessing whether repetitive dialogue can be removed can be difficult. In situations where a single word or short phrase is repeated in rapid succession many times, captioners will often elect to include two or three to conserve space. In many instances, deliberate repetition is played for comedic effect. In the below exchange between Lorelai and her partner Luke about whether Luke ever wanted to have children, they repeat the phrase "that's the kid" an absurd seven times within 15 seconds.

Table 3

English Closed Captions from Gilmore Girls: A Year in the Life, Winter, 00:55:25-00:55:57

In Time and Out Time	Speaker	Netflix Closed Caption	Dialogue Spoken (if different)
00:55:25.03 00:55:29.03	Luke	I figured if you wanted another kid you would say something.	**Well,** I figured if you wanted another kid, you would say something.
00:55:29.05 00:55:31.12	Lorelai	I figured if you wanted another kid, you'd say something.	
00:55:31.14 00:55:32.21	-Luke -Lorelai	-I brought it up once. -When?	
00:55:32.23 00:55:34.13	-Luke -Lorelai	-Five, six years ago. -I don't remember.	
00:55:34.15 00:55:35.16	Luke	At the little league game.	**We were** at the little league game.
00:55:35.18 00:55:39.06	Luke	The kid playing right field walked off mid-inning to use the bathroom.	**There was that** kid playing right field **and he** walked off in **the middle of an** inning to use the bathroom,
00:55:39.08 00:55:41.13	Luke	I turned to you and I said, "That's the kid."	**and** I turned to you and I said, "That's the kid."
00:55:42.09 00:55:44.09	-Lorelai -Luke	-"That's the kid"? -Yes.	
00:55:44.11 00:55:46.15	Lorelai	That's the kid is how you say you want to have a kid?	"That's the kid" is **your way of saying** you want to have a kid?
00:55:46.17 00:55:47.15	Luke	In so many words.	
00:55:47.17 00:55:49.22	Lorelai	No, that is not you saying you want to have a kid.	No, **"that's the kid"** is not you saying you want to have a kid.
00:55:50.00 00:55:53.00	Lorelai	That's the kid is acknowledging the kid about whom you said,	"That's the kid" is you acknowledging that **that's the kid who's standing there** about whom you **just** said,
00:55:53.02 00:55:55.12	-Lorelai -Luke	-"That's the kid." -I thought you knew what I meant.	-"That's the kid." -I thought you knew what I **was getting at.**
00:55:55.14 00:55:57.15	Lorelai	Impossible, because I don't speak "huh"?	

Note. From "Sherman-Palladino, A. (Writer & Director). (2016, November 25). Winter (Season 1, Episode 1) [TV episode]. In D. Palladino & A. Sherman-Palladino (Executive Producers), *Gilmore girls: A year in the life.* Warner Brothers Television; Netflix. https://www.netflix.com"

The deeper into the conversation they get, the faster Lorelai speaks. If the full line at 00:55:50.00 was captioned, it would reach 34.3 CPS. As a result of the speed of delivery, two instances of "that's the kid" were omitted and the content had to be paraphrased slightly to compensate. However, although the joke surrounding "that's the kid" is conveyed in the captions, the climax of the exchange is different, and perhaps less satisfying, as it does not preserve the extent of the verbal gag. The trade-off is indicative of the compromises captioners make, trying to balance reading speed and preserving creative intent of the original message.

Non-verbal hesitations, such as "um" and "uh" are also classified as discourse markers (Cambridge Dictionary, n.d.). Ums and uhs, as well as other hesitations, like stuttering, stammering and sentences that change track midway through are more commonly found in unscripted material, where individuals are speaking *ex tempore* rather than performing from a script. Filler words – of which "like" is the most infamous – are also characteristic of unscripted material. Captioners have to exercise discretion when deciding how to represent this type of speech and how much of these types of words to omit. On the one hand, when faithfully transcribed, this type of speech can be difficult to follow, but captioners risk misrepresenting the articulacy, confidence, or dialect of the speaker if these features are neglected (Zdenek, 2015a). De Linde and Kay refer to this conflict as the "dilemma of accuracy," noting that "a balance has to be struck between the clarity of a particular utterance and its stylistic function" (2006, p. 4). In the interest of clarity, a manner of speaking cue, such as [stuttering], [stammering] or [hesitating] may be employed, or the captioner may selectively represent the hesitations. Sometimes intensifying adjectives and adverbs, such as "really," "much," and "very," are classified as filler and removed. These types of words are often in service of emphasis, hyperbole, or exaggeration and thus their removal from a line may lessen the impact of the statement. Captioners may elect

to remove some profane intensifiers when editing profanity-heavy content, but these decisions need to be handled with care lest they inspire criticisms of censorship.

Audience Attitudes Towards Editing

The practice of editing captions is ostensibly done to accommodate deaf audiences, and thus it is important to consider how audiences receive edited captions. While the practice of editing captions is done with care, there is considerable antagonism from deaf audiences about the practice in principle, as I outlined in Chapter 6, but also in practice. I believe this is rooted in two issues: outdated research on caption-speed preferences that dictate Netflix's target reading speed and a lack of transparency and consistency in editing practices. Netflix's target reading speed is applied across all adult content, regardless of genre, and irrespective of how dynamic the visuals are or the speed at which dialogue is delivered. This means that some slower-paced content is captioned verbatim, while other, faster-paced content may be heavily edited. The viewer, however, is unaware of the TTSG rules and the target reading speed dictating the captioner's choices and is thus liable to misinterpret edited captions as failures of accuracy. When questions are raised in forums inquiring why captions are sometimes different from dialogue, one popular suggested reason is laziness – that either Netflix or the captioner was careless or was working off an inaccurate script (see forum posts McGraw-Herdeg, 2012; rBles, 2018; underbuster, 2016). The reality, however, is that editing content is more labour intensive and is done deliberately. This discrepancy between what audiences seem to want versus the accessibility service that Netflix is offering needs to be examined and researched more closely, as perhaps the practice needs to be changed. Neves (2008) believes that editing techniques are acceptable as long as the effect of "easy and enjoyable reading is achieved" (2008, p. 137). Evaluations of caption reading speed need to expand their focus beyond reading comprehension

to consider audience enjoyment, as well (Udo & Fels, 2010). If deaf audiences are comfortable reading captions at a pace faster than Netflix's 20 CPS rate, Netflix is not delivering an equitable viewing experience with its captions. Alternatively, Netflix could offer multiple caption tracks, providing one that is as close to verbatim as the restrictions of timed text allow, and a second at 20 CPS for audiences that read more slowly. This tension between accuracy, synchrony and readability demonstrates how difficult it is to accommodate the diverse needs of deaf audiences with a single caption file.

Caption Consistency and Clarity

Consistency in captioning is framed as a concern of formatting and style, rather than a terminological issue. This is markedly different from subtitling, where consistency is of paramount importance. Netflix is diligent about using the same translations for titles, character names, places and recurring phrases, ensuring that the same translations for these terms are used in subtitles, dubbing scripts, the Netflix user interface, trailers, marketing material, and intellectual property (Netflix, n.d.-h). To manage consistency, Netflix employs a spreadsheet-based term management tool called the KNP, which stands for Key Names and Phrases (Netflix, n.d.-h). In the first column, the English editor inputs plot-pertinent recurring terms, which are then categorized as a person, phrase, organization, or location (Netflix, n.d.-h). The second column is reserved for a brief explanation of the term and its significance to the material, providing enough context for translators to create an appropriate translation. Each subsequent column in the KNP is reserved for a different language, and translators input their translations into their designated columns. Although a captioner may consult the KNP to familiarize themselves with important terms, there is no space on the KNP to track captioning-related consistency issues. Netflix does not offer any consistency tracking documents for captioning. If

the vendor or captioner creates a document to track captioning consistency, it is strictly used and shared internally and not shared with Netflix. This neglect of captioning echoes the industry's tendency to prioritize subtitling and translation over accessibility services. Translations are more highly valued by Netflix and other studios and broadcasters because each language is associated with a market; each language into which content is translated represents millions more people Netflix stands to gain as subscribers.

Consistency

Despite Netflix's lack of resources for tracking caption consistency, consistency is an important concern within captioning. The same specific sound or pieces of music may appear across multiple episodes, serving as leitmotifs, references, or gags (Zdenek, 2015a). For example, in the comedy series *Wet Hot American Summer: First Day of Camp* (Showalter & Wain, 2015), a recurring gag involves characters throwing things offscreen that, when they land, inexplicably sound like a piece of pottery being smashed. As a result, the cue [pottery smashing] appears when a can of TAB is tossed on the kitchen floor[8], a framed gold record is smashed[9], and a can of vegetables is thrown aside[10], making the joke explicit to audiences who do not have access to the soundtrack.

It is difficult to document sonic elements for consistency, considering sound cannot be easily recorded into a spreadsheet for reference the way translations are. Nonetheless, it is a worthwhile endeavour to try to do so with descriptions, to at least indicate which sounds the captioner should keep an ear out for, especially considering that different captioners or vendors may be working on different segments or episodes of the same series. It is rarely possible for

[8]Season 1, Episode 5 (Lee et al., 2015)
[9] Season 1, Episode 7 (Showalter et al., 2015a)
[10] Season 1, Episode 8 (Showalter et al., 2015b)

editors to consult previous episodes to check how things have been captioned in the past (Zdenek, 2015a).

Netflix's emphasis on consistency is focused on orthographical and stylistic conventions to convey speech and sound information in an intuitive manner. Much like a play or script relies on conventional, easy-to-follow formatting to differentiate dialogue, stage direction and who is speaking, captioning style guides establish conventions to ensure captions are clear and consistent in how they telegraph speech, identify speakers, describe sounds, format song lyrics, and indicate pauses, interruptions, accents, and emphasis (Zarate, 2021). The way these elements are represented is not standardized across the industry. Different guidelines suggest different approaches – for example, the Netflix TTSG stipulates that sound cues and speaker identifiers should be formatted in square brackets and in mixed case, while other guidelines suggest uppercase letters and round brackets (English TTSG I.16; Zarate, 2021). Despite differences in approach, style guides' instructions are designed to provide a clear visual separation between dialogue and nonspeech information, establishing internal consistency. Through consistency and repeated exposure to captioning conventions, audiences are able to process captions more efficiently (Zarate, 2021).

Speaker Identifiers: Clarity and Consistency in Conflict.

Employing punctuation and captioning conventions consistently reinforces clarity via repetition – if ellipses, for example, are consistently used to indicate pauses, the audience recognizes the meaning of the ellipsis by virtue of its repetition in similar contexts (Zdenek, 2015a). Adherence to these conventions ensures cohesion within the individual text and within the broader category of captioned media. However, sometimes captions prioritize consistency to the detriment of clarity when the TTSG's rules are not well-suited to particular genres or

contexts. Because captioners are bound to follow the TTSG, they may be reluctant to question whether the application of some directive may hinder clarity and lead to a worse viewing experience for the audience. One example of this type of misguided loyalty to the TTSG is the convention surrounding speaker identification, commonly referred to as "speaker IDs."

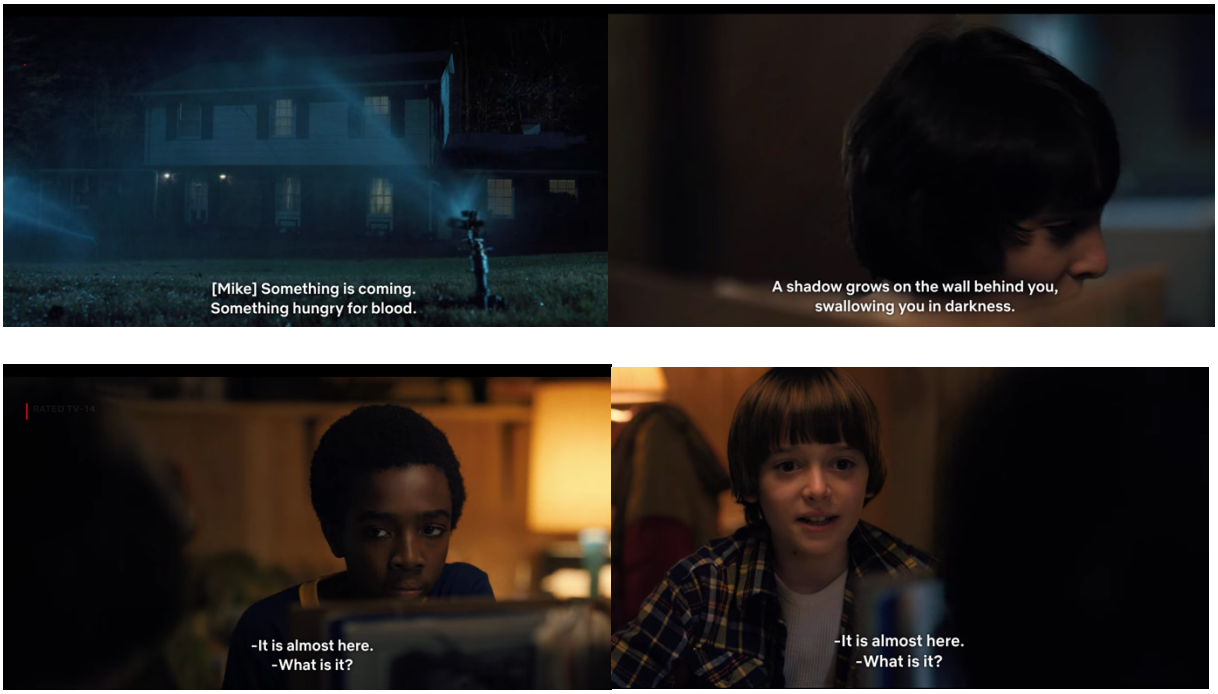
A speaker ID labels the start of dialogue with the name of the person who is speaking when the speaker cannot be visually identified (English TTSG, I.16). Speaker IDs are needed for voiceover narration, internal monologues, when a speaker's mouth is not facing the camera or if there are many people in the frame making it unclear who is speaking. Most style guides, the TTSG included, only explain when they are needed in general terms and assume that from context, it will be clear how they function.

From the point at which a speaker ID is introduced, unless a change in speaker is visually indicated (i.e., cutting to another person speaking) or the captions indicate a new speaker, the subsequent lines are delivered by the same person. The speaker ID only appears once, at the start of their dialogue, rather than reappearing in each subsequent caption if their dialogue continues. This principle can be observed in the first episode of *Stranger Things* (The Duffer Brothers, 2016). Four of the main characters, Mike, Lucas, Dustin and Will, are introduced for the first time, gathered in Mike's basement, playing Dungeons and Dragons.

Figure 2

English Closed Captions from The Disappearance of Will Byers, S1E1 of Stranger Things

00:01:44- 00:01:54



Note. From "The Duffer Brothers (Writers & Directors). (2016, July 15). Chapter one: The vanishing of Will Byers (Season 1, Episode 1) [TV series episode]. In D. Cohen, Duffer Brothers, K. Gajdusek, C. Holland, S. Levy, I. Paterson, M. Thunell, B. Wright (Executive Producers), *Stranger things.* Twenty-One Laps Entertainment; Netflix. https://www.netflix.com/"

The first shot in this scene is an exterior of Mike's house, and we hear him leading the game. The speaker ID establishes that it is his voice. The next caption accompanies a close-up of Mike, his face partially covered by the dungeon master's guide. The third caption spans two shots, showing Lucas' reaction, before cutting to Will asking, "What is it?" Thus, in the example below, all the lines are spoken by Mike until we see Will say "What is it?" The change in

speaker is indicated visually, by showing Will speaking, and textually in the captions by a hyphen.

The TTSG's only other instruction regarding speaker IDs complicates the captioner's mission to provide clear captions:

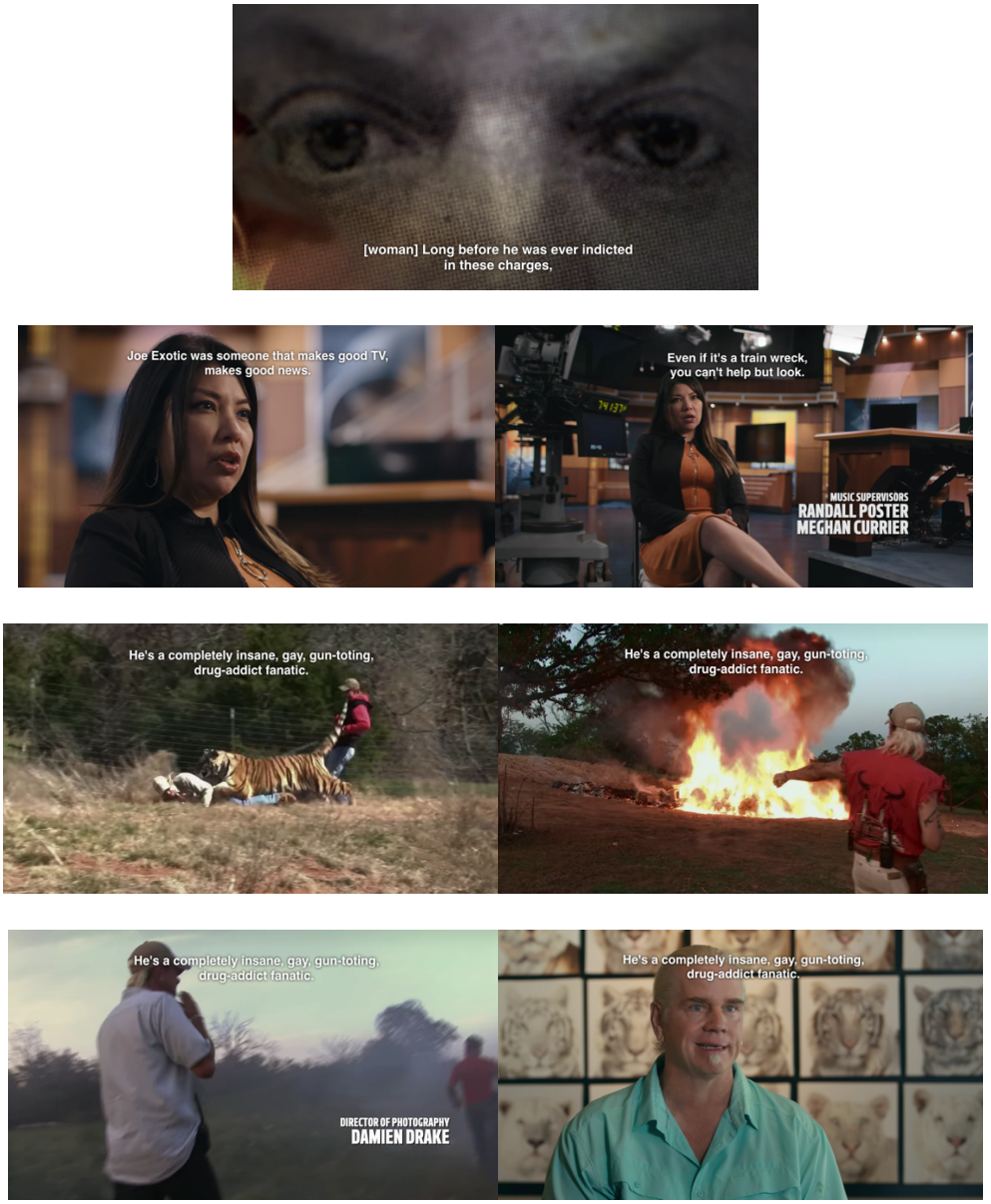
> When a speaker ID is required for a character who has yet to be identified by name, use [man] or [woman], or [male voice] or [female voice], so as not to provide information that is not yet present in the narrative. If the same identifier is used multiple times in one scene, number should be added to distinguish them, for example [man 1]" (English TTSG, I.16).

The rationale behind this guideline is to avoid giving the audience watching captions more information than the hearing audience has. In some situations, explicitly identifying a character can be a spoiler for the audience, particularly if a character's identity is supposed to remain unknown to the audience, if they have a secret identity or are about to introduce themselves. However, this directive does not consider other situations, where revealing the name of a character or speaker may be beneficial or essential for viewer comprehension. The speaker ID conventions outlined in the TTSG often work well for scripted, narrative content, because the visuals are typically easy to follow; the logic of speaker IDs is often facilitated by editing, as directors typically structure shots and cuts around the delivery of dialogue. However, challenges abound once these rules are applied to unscripted and documentary content, which involve more voiceover dialogue, shot changes, and B-roll footage.

In "Not Your Average Joe," episode one of Netflix's documentary series *Tiger King* (2020), the opening montage consists of many interviews, news reports and B-roll footage edited together at a rapid pace. In the first three and a half minutes, nineteen different people speak.

Figure 3

English Closed Captions from Not Your Average Joe, S1E1 of Tiger King, 00:02:42-00:02:58



Note. From "Chaiklin, R. & Goode, E. (Directors). (2020, March 20). Not your average Joe (Season 1, Episode 1) [TV episode] In R. Chaiklin, E. Goode, C. Smith, F. Stevens (Executive Producers), *Tiger king.* Royal Goode Productions; Netflix. https://www.netflix.com"

In the first shot of the sequence in Figure 3, [woman] refers to Sylvia Corkill, the reporter seated in the second and third stills. She says, "Long before he was ever indicted in these

charges, Joe Exotic was someone that makes good TV, makes good news. Even if it's a train wreck, you can't help but look" (Chaiklin et al., 2020). The subsequent line, "He's a completely insane, gay, gun-toting, drug-addict fanatic," (Chaiklin et al., 2020) is actually said by Doc Antle, the man who appears in the seventh still. No speaker ID was used at the start of Doc's line because he appears on screen while it is being said, which is consistent with the TTSG instruction to only use speaker IDs when the speaker cannot be visually identified (English TTSG, I.16). However, Doc Antle only appears on screen as he is saying the final word, "fanatic." Up until that final half-second, someone watching who does not have access to the audio track is liable to assume that statement is a continuation of Sylvia's dialogue. This presumption is especially problematic because she is being interviewed in her workplace, her tone is professional, and she is discussing the appeal of Joe Exotic as the generator of news stories. Doc Antle's statement, on the other hand, is rather inflammatory and it establishes that he is not fond of Joe Exotic. In this case, the captioner's effort to be consistent in their application of the speaker ID rule generates unnecessary confusion rather than clarity.

The practice of using [man] and [woman] rather than names also needs to be interrogated. Because captions are supposed to provide equal access to deaf audiences, the captioner needs to acknowledge that while the hearing audience may not know someone's name, they do have access to the sound of their voice. While the speaker's name may be "information that is not yet present in the narrative" (English TTSG, I.16), their voice is part of the narrative. Hearing audiences can often track who is speaking based off of the voice alone (Zarate, 2021). When combined with context clues, hearing audiences can conclude the identity of a speaker before they are explicitly introduced. Some actors have easily recognizable voices, and thus, for example, if the film *Forrest Gump* (Zemeckis, 1994) were to open with off-screen narration by

Tom Hanks, a large portion of the hearing audience would be able to put together the sound of Tom Hanks' voice with the knowledge that Tom Hanks plays the main character to conclude that it is the character Forrest Gump delivering the dialogue. To identify the main character as [man] is not helpful to someone who does not have access to the audio, and it makes the captioner appear ignorant to something that is obvious to the hearing audience.

Similarly, when a viewer selects an episode of the docuseries *Chef's Table* on Netflix, they can see the episode title, which is the name of the chef being profiled in the episode, and when accessing Netflix on a browser, the chef's image and name appear as the episode loads. The first episode of *Chef's Table* season 4 (Fried et al., 2018) opens with voiceover accompanying B-roll footage of the episode's subject, Christina Tosi, baking cookies. The first nine captions read:

Table 4

English Closed Captions from Christina Tosi, S4E1 of Chef's Table, 00:00:10-00:00:34

In Time & Out Time	Speaker	Netflix Closed Caption	Shot Description
00:00:10.09 00:00:11.21	Christina	[woman] *Let's be honest.*	Voiceover dialogue. Stacks of cookies on a sheet pan
00:00:11.23 00:00:14.16	Christina	*You're not going for a cookie* *for sustenance.*	
00:00:19.04 00:00:23.01	Christina	*You're going for a cookie* *for the beauty of indulgence,*	Christina adding ingredients to a stand mixer
00:00:23.03 00:00:25.17	Christina	*for the spirit of just letting free* *and being like,*	Shots of the stand mixer mixing batter
00:00:25.19 00:00:28.22	Christina	*"I know maybe this cookie isn't the thing*	Close up of Christina scooping cookie batter
00:00:28.23 00:00:31.06	Christina	*that's gonna round out my diet* *for the day,*	
00:00:32.10 00:00:34.14	Christina	*but it's what's gonna bring me joy...*	Shots of the cookie dough on the sheet pan
00:00:32.10 00:00:34.14	Christina	and remind me that life's too short	Cut halfway through line to Christina talking head
00:00:38.06 00:00:40.05	Christina	to worry about how many cookies I ate today."	

Note. From "Fried, A. (Director). (2018, April 13). Christina Tosi (Season 4, Episode 1) [TV episode]. In A. Del Deo, A. Fried, M. Hilliard, D. Lillegard, B. McGinn, L. Nishimura, D. O'Malley & M. Weaver. (Executive Producers), *Chef's table.* Chef Pictures One: Netflix. https://www.netflix.com"

Despite seeing the B-roll footage of Tosi baking cookies and knowing this episode is about her, the captioner followed the TTSG directive and identified her as [woman] because she has not been explicitly introduced within the show. Referring to the subject of a documentary as [woman] misrepresents the importance of the person speaking and what they are saying. Later in the same episode, [man] is used to identify David Chang, Tosi's mentor, three seconds before his name appears onscreen introducing him. In a subsequent scene, [man] is also used to indicate the off-screen voice of a vendor selling Tosi ice cream. Using generic speaker IDs like [man] and [woman] in this situation comes across as obtuse, as it makes the captioner appear willfully ignorant to the subject of the documentary and in practice, it erases the relative importance of the character or person speaking. It is unlikely that any viewer – hearing or deaf – could select the episode from the Netflix interface, watch the opening scene and not infer that it is Tosi speaking before her name appears onscreen.

In the case of *Stranger Things,* in Figure 2, the captioner broke the speaker identification rule[11] and used their names rather than [boy 1], [boy 2], et cetera. This improves the viewing experience for the audience, who is thus able to easily follow which boy says which lines. The rest of their first scene is important in establishing characterization, as the boys are squabbling over the best approach to fighting the Demogorgon in their game of Dungeons and Dragons. The

[11] It is likely that the speaker ID rule was added sometime post-2016, as the Speaker Identifier section has been revised multiple times since 2018. The TTSG's update history, unfortunately, does not clearly indicate which directives in that section were changed.

speaker IDs help make the audience aware that Mike leads the game, that Lucas wants to attack the Demogorgon, that Dustin panics in the game and when the dice roll off the table, and that Will is honest about the wayward dice roll. These are all important details to each of the characters as the story progresses and would not come across as clearly without the use of their names in the speaker IDs, as the audience would have to remember which boy corresponded with which number.

When there are many speakers who have not been formally introduced, the numbering system can become overwhelming very quickly. For example, in the opening sequence of "Not Your Average Joe," the following eighteen speaker IDs are present in the first three minutes, in order of appearance: [man 1], [interviewer], [man 2], [man 3], [handler], [reporter 1], [reporter 2], [operator], [man 4], [Joe][12], [reporter 3], [man 5], [man 6], [man 7], [reporter 4], [reporter 5], [reporter 6] and [woman] (Chaiklin et al., 2020). It is incredibly difficult to keep track of which speaker ID corresponds to which person. In a situation like this, even if the people speaking have not been formally introduced, using their names is the best way to ensure the audience can follow who is speaking. In their study of deaf and hard-of-hearing audiences' preferences regarding nonspeech information, Harkins et al. (1995) found that audiences expressed a preference for explicit identification of off-screen narrators and speakers, most especially in situations where "circumstances could confuse the viewer" (p. 11).

Another captioning quirk observable in the Tiger King scene in Figure 3 is that many captioning rules are applied on a sentence-by-sentence basis, rather than a contextual basis. Captioners will interpret the TTSG based on how it applies to individual sentences, rather than consider the context in which those sentences appear. For example, in the case of Doc Antle's

[12] Joe is [man 4] until he says his own name, "introducing" himself in the narrative.

dialogue in Figure 3, the captioner interpreted that Doc was visually identifiable as the speaker of his line because he appeared onscreen while the sentence was spoken – regardless of the fact that he was only visible for the final half second of the time that caption was displayed. If Doc had said two sentences and appeared during the second sentence, the captioner would have used a speaker ID. The captioner is primed to interpret the guidelines this way because the TTSG presents other rules in terms of sentences rather than context, such as the rules about quotation marks (English TTSG, I.13).

These practices involving speaker IDs represent how the captioner prioritizes their loyalty to the TTSG over their duty to provide an equitable and clear viewing experience for the deaf audience. While I would argue that the directive should be rewritten, I wonder if the captioner or reviewer of "Not Your Average Joe" considered how confusing the use of speaker IDs would be to the deaf audience, or if they were aware that by following the directive, they were not providing an accessible experience for their audience.

Quotation Marks

Regarding quotation marks, the TTSG stipulates, "If the quote extends beyond more than one subtitle, use an open quote at the beginning of the first subtitle, at the start and end of sentences within the quote and an end quote at the end of the last subtitle" (English TTSG, I.13). Before July 2020, the TTSG rule instructed captioners to use a single set of quotation marks to encompass a quote, regardless of whether it extended into multiple captions.

The quotation marks only serve to indicate that speech is quoted and do nothing to help the viewer understand when quotes start and end. When each sentence is enclosed in quotation marks, sequences containing multiple quotes become more difficult to decipher. This occurs often in stand-up comedy routines when comedians re-enact conversations between multiple

people. In a sequence from his stand-up special *Aziz Ansari Live at Madison Square Garden* (2015), Ansari re-enacts a hypothetical conversation between an emigrant and his cynical friend:

Table 5

English Closed Captions from Aziz Ansari: Live at Madison Square Garden, 00:02:58-00:03:17

In Time & Out Time	Captions as they appear on Netflix (following old quotation rules)	Captions per English TTSG 2022 (following current quotation rules)
00:02:58.22 00:03:01.11	‘Cause you know they had friends that were dicks	‘Cause you know they had friends that were dicks
00:03:01.13 00:03:03.14	that were like, “What? You can’t move to America.	that were like, “What? You can’t move to America.”
00:03:03.16 00:03:05.17	You don’t know anyone there. You don’t speak the language.	“You don’t know anyone there. You don’t speak the language.”
00:03:05.19 00:03:07.05	What are you going to do for work?”	“What are you going to do for work?”
00:03:07.07 00:03:08.16	“I don’t know, man. We’ll figure it out.	“I don’t know, man. We’ll figure it out.”
00:03:08.19 00:03:12.08	Worst-case scenario, we’ll cook food and we’ll sell it to white people, okay?”	“Worst-case scenario, we’ll cook food and we’ll sell it to white people, okay?”
00:03:14.11 00:03:17.11	“You got a Master’s degree in Chemistry. You don’t know how to cook.”	“You got a Master’s degree in Chemistry. You don’t know how to cook.”

Note. From “Ansari, A. (Director). (2015). *Aziz Ansari: Live at Madison Square Garden.* [Comedy Special]. Oh Brudder Productions; Netflix. https://www.netflix.com”

This idiosyncratic rule change exemplifies the opaque nature of captioning guidelines. No explanation was provided to justify this change, which leaves the captioner guessing as to its rationale. My best guess is that it reflects a captioning convention the author is used to, as it does not improve clarity or readability, nor does it reduce the total character count per caption.

When captions are clear and consistent, they facilitate intersemiotic translation, presenting an equivalent message in a different code (Jakobson, 2004) for deaf audiences to access and enjoy audiovisual content. Thus, it would be beneficial for Netflix to better prioritize tracking consistency with a tool similar to the KNP to help captioners identify speakers and represent important sounds and pieces of music consistently. The speaker ID rule, in particular,

should be revisited by Netflix. It is flawed by design because it does not take into consideration that hearing people's access to the sounds of voices makes them able to recognize and track speakers. Captioners' adherence to the speaker ID rule demonstrates a tendency to follow the TTSG at the expense of providing a clearer, more accessible viewing experience for deaf audiences. It also demonstrates a reluctance by Netflix and by captioners to offer superior access to deaf users – as Berman (2000) and Zdenek (2015a) both observe, occasionally translation provides superior access to content, clarifying and rendering explicit something that may have been more subtle in the original text. While introducing a character or speaker's name in advance of the narrative gives the deaf audience access to more information than the hearing audience has, it reflects and balances that hearing audiences do have access to speakers' voices. Furthermore, it is much easier for captioners to identify speakers by name, rather than methodically count and track the number of voices per scene, ensuring the numbering system is employed accurately. The tendency to err on the side of inferior over superior access is a captioning trend I will continue to explore in my discussion of foreign dialogue captioning.

Foreign Languages, Accents, and Dialects

The TTSG instructs captioners to "include as much of the original content as possible" (English TTSG, I.1). Though not explicitly stated, this directive applies specifically to captions for English-language content, since this approach cannot be followed for content that is translated into English. Section I.8 of the TTSG provides captioners with instructions on how to handle non-English dialogue:

- In instances of foreign dialogue being spoken:
    - If foreign dialogue is translated, use [in language], for example [in Spanish]
    - If foreign dialogue is not meant to be understood, use [speaking language], for example [speaking Spanish]
    - Always research the language being spoken – [speaking foreign language] should never be used

    - Accents or dialects require the same treatment, for example [in Spanish accent]
- Foreign words that are used in a mostly English line of dialogue do not require identifiers but should be italicized. Always verify spelling, accents, and punctuation, if applicable.
- Familiar foreign words and phrases which are listed in *Webster's* dictionary should not be italicized and should be spelled as in *Webster's* dictionary (e.g. bon appétit, rendezvous, doppelgänger, zeitgeist, etc.) (English TTSG, I.8)

The TTSG reduces all scenarios that involve foreign dialogue into three categories: dialogue that is to be "understood" via translation, dialogue that is not intended to be understood and is thus replaced by a cue indicating the language spoken, and foreign words that are to be transcribed if they appear in a mostly English sentence. These points provide concrete instructions based on a conception of the audience as a monolingual body and they are primarily concerned with whether or not something is intended to be understood. However, to the deaf viewer who sees these rules applied in the captions, it is not obvious why some segments of dialogue may be translated while others are cued or transcribed. Furthermore, they do not facilitate an equally accessible viewing experience for deaf audiences by selectively representing foreign dialogue.

Translation

Netflix elaborates on when to include translations in Section II.11, noting that "Foreign dialogue should only be translated if the viewer was meant to understand it (i.e. if it was subtitled in the original version" (English TTSG; II.11) The content producer decides which lines of foreign dialogue they would like translated for their audience and those translations are supplied by Netflix: either a liaison shares a timecoded document that lists translations approved by the showrunner/director, or Netflix supplies the vendor with a texted reference video. A texted reference is an "original" version of the video that contains subtitles that are to be copied into the captions. Captioners are not allowed to change the Netflix-supplied translations, but relevant

cues, non-verbal utterances, speaker identifiers, etc. can be added if needed. At most, a line may be edited if adding nonspeech information compromises the readability too much or if the line length requirements cannot be met. However, captioners are generally reticent to make any changes to client-provided translations out of fear that the client may object or find the creative intent of the producer compromised. When overseeing captioners' work, I have observed that in most cases when nonspeech information compromises the readability too severely, the captioner will sooner elect to omit them than change the translation. The director will often review the final caption and subtitle files to provide final approval.

Captioners are able to preface translations with [in language] cues because Netflix, like many streaming platforms, requires content producers to supply semi-texted video masters for their streaming service (Amazon, 2022; Netflix, 2021b). Video masters that are created for distribution can be textless, texted, or semi-texted. A textless video contains no onscreen text. Conversely, a texted video contains hardcoded subtitles, in addition to stylized creative textual elements, such as main titles, intertitles, location-establishing text, and credits. Finally, a semi-texted video does contain stylized textual elements, but it does not contain hardcoded subtitles. Semi-texted videos[13] are becoming the industry standard for streaming platforms because they are easier to localize and allow for the more seamless integration of subtitles into captions (Amazon, 2022; Netflix, 2021b).

On texted videos, when subtitles appear on the bottom of the screen, [speaking language] cues appear simultaneously in the caption track at the top of the screen. Zdenek evaluates how poorly this approach functions, noting that it is demanding on the viewer who is forced to engage in "visual gymnastics," continually reading two visually separated text streams at once (Zdenek,

[13] Netflix also refers to semi-texted videos as "non-subtitled."

2015a, p. 270). To illustrate his point, he critiques the captions on the DVD release of *Inglourious Basterds* (Tarantino, 2009), finding that in one 23-second segment, seven language identifying cues appear in the closed captions. Though this approach is clumsy, it is usually visually apparent that the caption track and subtitles were not created by the same person. In the case of *Inglourious Basterds* below, the hardcoded subtitles are in mixed case and a yellow font while the caption track is in white and uppercase. To the viewer who is unfamiliar with the inner workings of video production and captioning, it may appear as though the captioner is adding the language information the only way that they can in the face of a textually restrictive situation.

Figure 4

Still from Inglourious Basterds Demonstrating Captions on a Texted Video

Note. From "Zdenek, S. (2015b, October 1). Chapter 8: In a manner of speaking. [Blog post]. Retrieved from https://readingsounds.net/chapter8/"

On a semi-texted video, however, the translations for foreign dialogue are seamlessly integrated into the caption track, visually indistinguishable from the rest of the captions as they are all displayed in the same font and style. On Netflix, if an English series or movie contains foreign-language subtitles, the subtitles display by default if the user is based in an English-speaking country. If the user selects English CC, the subtitles are integrated into the caption track with the [in language] cue.

Prefacing a caption with an [in language] cue indicates to the audience that they are reading a translation rather than a transcription of the dialogue, and it conveys the specific

language the person is speaking. This language-identifying information is not immediately available to the hearing viewer, as the subtitles made for them contain solely English translations. Context and setting might establish what language a person is speaking, but regardless, the hearing audience has access to the sound of the language. Many hearing adults are able to identify or discriminate between foreign languages based on audio alone (Lorch & Meara, 1989; Bond, Stockmal & Muljani, 1998). In their language identification experiments, Muthusamy et al. (1994) determined that anglophone listeners were better able to identify languages they were more familiar with (i.e. French, German and Spanish) and that they made judgements based on whether they could recognize familiar words within the dialogue, the speaker's manner of articulation, and the prosody and tone of the spoken language (in Bond, Stockmal & Muljani, 1998). Thus, though captions are not supposed to serve as annotations that introduce more or different information than hearing people have access to (Zdenek, 2015a), the use of the [in language] identifier is an effective compromise to provide the deaf person with vital audio information. It is one of the ways in which the caption track clarifies content, providing superior access by presenting information textually rather than sonically (Zdenek, 2015a).

Netflix's [in language] instruction leaves it to the captioner to decide how frequently the cue should be included. In situations where speakers consistently speak the same language, captioners will typically place the [in language] cue in the first caption and assume that the descriptor applies for all subsequent captions covering that person's speech until the speaker or the scene changes.

Speaking Language Cues

While the transition from texted to semi-texted videos vastly improved the presentation of translated subtitles within the caption track, another major problem Zdenek (2015a) identified in DVD captions persists: the use of cues to represent foreign dialogue rather than transcriptions. Because the captioner must follow the TTSG and respect the producer's authority and creative intention, when they encounter a sentence or series of sentences in a foreign language for which no translation has been provided, the captioner will often default to using a [speaking language] cue. To ensure the correct language appears in the cue, the captioner may use the script to verify, or at larger localization companies, they may be able to request confirmation from a linguist via their coordinator.

The question of whether something is "meant to be understood" presumes the audience is monolingual, and it ignores the reality that despite the creator's intention, audiences may understand the foreign dialogue if they happen to be versed in multiple languages. The TTSG assumes that the audience using captions only needs access to the English dialogue and to the dialogue that has been translated into English. This approach leads to a significant discrepancy in the level of accessibility between the hearing viewer and the deaf viewer. While a hearing audience may not understand some lines of Italian spoken in a film, they nonetheless have access to the specific words spoken, to the sounds of the words and to the cadence of the language (Zdenek, 2015a). Replacing dialogue with a cue turns speech into two data points: how the dialogue is being delivered – though the TTSG example cue uses "speaking," the cue may indicate how the dialogue is delivered, for example shouting, whispering, singing etc. – and the language spoken. Zdenek is unequivocal in his belief that untranslated foreign-language words must be transcribed, "even if they read like nonsense to readers who may only know English" (2015a, p. 269). He argues that captioners have a duty to make their best effort to convey the

meaning of the text and believes that principal stands even if the language spoken is not English. Captioners "should make every effort to convey the *meaning* of the text regardless of the language in which that meaning is expressed, even if readers can't be assumed to understand the foreign-language speech sounds" (2015a, p. 269). The replacement of foreign language dialogue with a cue exemplifies an extreme version of what Berman calls "qualitative impoverishment" (2000). By substituting "the original terms with those that lack the 'sonorous' or 'iconic' richness of the original" the translation suffers in quality (Berman, 2000, p. 291). Furthermore, the use of a cue can occasionally be uneconomical, dwarfing the length of the original line. For example, replacing the Italian farewell *"arrivederci*" with [speaking Italian] or the more descriptive [saying goodbye in Italian] replaces an eleven-letter long word with phrases that are eighteen and twenty-seven characters long, respectively (Zdenek, 2015a). This adds bulk to the caption file while also patronizing the audience, implying they would not understand meaning of the word, and alienating the audience from the original phrase.

To facilitate more coherent captioning of foreign dialogue, content producers should be more aware of how their decisions on which lines of dialogue to translate affect closed captions. It is not always clear to audience why some lines are translated while others are not, and this confusion is often brought into starker relief when audiences watch content with closed captions. For example, in Martin Scorsese's 2019 film *The Irishman*, the titular Irishman, Frank Sheeran surprises Italian mobster, Russell Bufalino when he speaks to him in Sicilian Italian as they break bread together. The lines are translated for all audiences:

Table 6

English Closed Captions from The Irishman, 00:17:04-00:17:15

In Time and Out Time	Speaker	Netflix Closed Caption

00:17:06.15 00:17:09.21	Frank	[Frank in Italian] This bread is very good.
00:17:11.05 00:17:12.23	Russell	[Russell in Italian] Good, eh?
00:17:13.23 00:17:15.03	Frank	Where did an Irisher like you learn to speak Italian?

Note. From "Scorsese, M. (Director). (2019). *The Irishman* [Film]. Netflix. https://www.netflix.com"

Their conversation reveals that Frank was stationed in Russell's hometown of Catania during the Second World War and Russell approves of Frank's Sicilian accent. Later in the film, after decades of friendship, Frank and Russell are incarcerated together. In the prison cafeteria, they again break bread, but this time, no translation is provided. The caption track contains:

Table 7

English Closed Captions from The Irishman, 03:01:53-03:01:55

In Time and Out Time	Speaker	Netflix Closed Caption
03:01:53.05 03:01:54.22	-Russell -Frank	-*E buono, no?* -*Buono.*
03:01:55.00 03:01:57.01	-Russell -both	-[Russell] *Buono.* -[both chuckle]

Note. From "Scorsese, M. (Director). (2019). *The Irishman* [Film]. Netflix. https://www.netflix.com"

Despite the lack of translation, the hearing audience understands the callback to the previous scene, having heard the word *"buono"* when the men were complimenting the bread. The situation in the captions is awkward, as they cannot include a translation and replacing the dialogue with [both speaking Italian] would zap the scene of meaning. Scorsese likely did not include a translation to avoid patronizing or distracting the audience, considering *"buono"* should be familiar to the audience. Here the captions bend the rules in the TTSG to ensure the

deaf audience has access to the important dialogue, though it creates an imbalance relative to the scene in the restaurant.

Another complication regarding foreign dialogue presents itself in how captioners interpret the word "line" in the directive "foreign words that are used in a mostly English line of dialogue do not require identifiers, but should be italicized" (English TTSG, I.8). Many captioners interpret "line" to mean "sentence." If foreign words appear within an English sentence, captioners will transcribe it, whereas if foreign words constitute an entire sentence, they are more likely to cue it. This practice can drastically impoverish the quality of captions and the level of access they provide to the dialogue. This dynamic is at play in "This Is It," the pilot episode of the sitcom *One Day at a Time* (Calderón Kellett et al., 2017b). The show is about the Alvarez family, an intergenerational Cuban American family living together in the same house. The show derives a lot of humour from the dynamics between theatrical grandmother Lydia, who speaks English interspersed with a significant amount of Spanish, her daughter, Penelope, and her grandchildren, Elena and Alex. In the scene below, Lydia is exasperated with her granddaughter who does not want her family to throw her a *quinceañera* party.

Table 8

English Closed Captions from This Is It, S1E1 of One Day at a Time, 00:04:52-00:04:56

In Time and Out Time	Speaker	Netflix Closed Caption	Dialogue Spoken (if different)
00:04:52.23 00:04:56.02	Lydia	You need to do something about this little *sinvergüenza.*	
00:04:56.16 00:04:58.01	Elena	What does that even mean?	
00:04:58.03 00:04:59.06	Lydia	It means...	
00:04:59.18 00:05:02.18	Lydia	that you do not know enough Spanish to know that I'm insulting you.	

00:05:04.14 00:05:06.16	Lydia	*Oye...* [speaking Spanish]	*Oye, tu hermana no sabe* *la palabra "sinvergüenza."* (Hey, your sister doesn't know the word "scoundrel.")
00:05:07.15 00:05:09.11	-Lydia -Alex	-Ha! -[speaking Spanish]	*-¡Ha!* *-Oye, ¡que tonta!* (Ha, what an idiot!)

Note. From "Calderón Kellett, G. (Writer) & Royce, M. (Writer), & Fryman, P. (Director). (2017b, January 6). This is it (Season 1, Episode 1) [TV episode]. In G. Calderón Kellett, P. Fryman, M. Garcia, N. Lear, M. Royce (Executive Producers), *One day at a time*. Sony Pictures Television; Netflix. https://www.netflix.com"

Lydia refers to Elena as a *sinvergüenza,* a Spanish word meaning "scoundrel," but the insult is lost on Elena because she does not understand the word. When Lydia complains to Alex about Elena failing to comprehend her insult, the entire conversation, except an "oye" exclamation is replaced with cues. Alex jokes in Spanish, "¡que tonta!" which means, "What an idiot!" but this joke is not accessible to the deaf audience because the words are replaced by a cue. Even when played to a monolingual, anglophone audience the hearing viewer is generally able to understand the gist of the jokes that are made in Spanish, by virtue of the tone of delivery or the repetition of key phrases. *One Day at a Time* is especially good at bridging the gap between Spanish and English jokes. For many of the Spanish jokes and asides in the series, an English set up or follow up is integrated into the exchanges to aid audience comprehension. This structure can have a pedagogical effect, teaching the audience new words and phrases. For example, in the above scene, Lydia clarifies in English that "*sinvergüenza"* is an insult. Berman (2000) cautions that the destruction of rhythms in translation is a danger more to poetry and theatre than to prose. This threat is on display in the way the cueing of the Spanish dialogue ruins the comedic rhythms in *One Day at a Time.* Mike Royce, an executive producer on the show, observes that the jokes transcend language barriers – that, in fact, the characters are funnier when they speak both English and Spanish regardless of whether the audience understands Spanish

because of the comedic rhythm the two languages create. When the lines are not included in the captions, the deaf audience is deprived access to the show's comedic rhythm (Calderón Kellett et al., 2017a).

To remedy the inaccessibility generated by Netflix's foreign dialogue instructions, I would recommend that Netflix refine and revise the TTSG's instructions and embrace the transcription of foreign words and phrases, particularly with languages that are more familiar to English-speakers, like Spanish, German, and Italian (Bond, Stockmal & Muljani, 1998). When captioners interpret the directive to use a cue when dialogue is "not meant to be understood" (English TTSG, I.1), they ensure that dialogue absolutely *cannot* be understood. This does not provide deaf audiences with equal access to the content, but rather grossly inferior access. Changing these directives would also lead to greater consistency, providing only two options for clearly audible foreign dialogue. Perhaps captions could employ a hybrid method combining cueing and transcription to help bridge the accessibility gap and improve readability. The caption in *One Day at a Time* could thus look like this: [Spanish] *Oye, tu hermana no sabe la palabra "sinvergüenza"* (Calderon et al., 2017b). A cue plus an italicized transcription could improve the readability of foreign language dialogue, priming the audience that they are about to read dialogue in a different language.

Chapter 8: Conclusion

My study of Netflix Originals' closed captioning explored how captioners endeavor to make audiovisual content accessible to deaf audiences. Although captioners pursue the platonic ideal of providing equal access to the audio of TV shows and films, this goal is out of reach due to the inherent differences between sound and text, as well as the limitations of the form of closed captions and the conditions under which they are produced. Captions do provide access, but they must contend with achieving "equivalence in difference" (Jakobson, 2004). While I do argue that captioning needs to be considered distinctly and separately from translation subtitling, I assert that captioning evokes intralingual and intersemiotic translation. To describe how these translations affect the content and meaning of captions, I borrow from Berman's (2000) theories of the deforming tendencies of translation – a useful framework despite intending to be applied to interlingual translation.

To structure my study, I used D'Acci's circuit of media studies (2004). Her model suggests that cultural artifacts are influenced by the interconnected relations of the socio-historical contexts from which they emerge, the conditions of their production and the way they are received. In the context of Netflix Originals' closed captions, this meant reviewing the history and development of captioning technology. Important details emerged from this recapitulation of caption history – most notably, that deaf individuals and educators were instrumental in developing captioning techniques and thus, before the content boom precipitated by captioning mandates in the 1990s and 2000s, were more hands-on in caption production. The industry's rapid expansion alienated the deaf audience from the production process, leaving the task of articulating the captioning preferences and needs of the deaf to regulating bodies, like the FCC and the DCMP.

From there, I assessed the process of caption production, looking at how the industry is structured to accommodate the massive volumes of content that needs to be captioned and the conditions under which captioners are hired and work. This portion of my thesis was strongly informed by my experience as a professional captioner, as I was able to harness my insights to critique and contextualize some captioning practices. I establish the central role of style guides, explaining how Netflix uses the TTSG as a training manual, set of instructions, and quality control tool.

I chose to focus on three particular captioning practices in my analysis of Netflix captioning: editing captions to meet target reading speeds and to follow timing restrictions; the way orthographical conventions and speaker identifiers are employed with the intention of making captions clear and consistent; and how captions use a mixture of translations, cues, and transcriptions to represent foreign dialogue. To conduct my analysis, I engaged in a close reading of the TTSG and a selection of Netflix Originals' closed captions, calling attention to how the directives in the TTSG are interpreted and applied to the content. My analysis revealed that captioners often adhere to the TTSG to the detriment of caption quality, sacrificing caption quality in the name of loyalty to the guidelines. This is not surprising, given the way Netflix's quality control process is based around evaluating how well captions follow the TTSG's directives and it is difficult for style guides to anticipate how some contexts and content may require different approaches.

I prefaced my discussion of the practice of editing captions with a summary of the research surrounding deaf literacy and reading speeds. This context is important, as the TTSG distills how well deaf audiences can read and their preference for caption speed to a single value – the target reading speed of 20 characters per second. Reviewing the literature on reading speed

revealed how controversial the practice of editing is, with many studies presenting arguments against editing (Jensema, 1998; Morettini, 2020; Romero-Fresco, 2009; Szarkowska et al, 2011). In analyzing the way captions are edited for *Gilmore Girls: A Year in the Life* (Sherman-Palladino, 2016), I discuss how the truncation of dialogue typically targets discourse markers and dialogue that reflects characterization, formalizing spoken English. When content is consistently edited the same way, it has a homogenizing effect, potentially erasing how relationship dynamics are reflected in dialogue. I theorize that part of the deaf audience's vocal dislike of editing is a result of a lack of transparency – audiences may be unaware of the reading speed restrictions set by Netflix and thus misconstrue edits as failures of accuracy.

My focus on how orthographical conventions and speaker identifiers are employed in service of caption clarity and consistency demonstrated how these two values come into conflict. In my discussion surrounding speaker identifiers, I draw from *Stranger Things* (The Duffer Brothers, 2016) to illustrate the importance of identifying characters by name, even early on in a series – as long as there is no explicit reason not to, or if doing so would constitute a spoiler. I use *Tiger King* (Chaiklin et al., 2020) and *Chef's Table* (Fried et al., 2018) to explore a situation where adherence to the TTSG blatantly sacrifices clarity in the name of consistency.

Lastly, I assess how the TTSG instructs captioners to handle foreign dialogue. My close reading of the guidelines aimed to contextualize some of the instructions – for example, as a captioner and caption trainer, I am well aware of how skittish captioners are when it comes to adapting client-provided subtitle translations into captions. I explain how streaming platforms have embraced semi-texted videos, which allows for subtitles to be seamlessly integrated into the caption track. I explain how Netflix presents foreign dialogue on a meant to be understood/not meant to be understood binary, which is problematic because it means all dialogue not "meant"

to be understood and which does not appear in the midst of an English sentence is replaced with a cue rather than a transcription. I argue in favour of transcribing untranslated foreign dialogue to ensure deaf audiences have access to the words, even if they do not understand them. It is patronizing to replace foreign dialogue with a cue and outright denies deaf audiences access.

Recommendations for Netflix and the Captioning Industry

My choice to structure my study around D'Acci's circuit model of media studies (2004) reflects my desire to bring the key stakeholders in captioning together; to bring the audience of captions into direct contract with the producers of captions and the studios commissioning them and evaluating their quality. As Romero-Fresco (2021) notes, communication between these parties is restricted by nondisclosure agreements and industry competition – and as a result, efforts to improve and evolve captioning practices are stifled. The studios and broadcasters that commission captions and create style guides would benefit from revisiting current practices that may not necessarily provide deaf and hard-of-hearing audiences with the most accessible experience. I hope my study might instigate more discourse about captioning practices from North American captioning practitioners and scholars, as we are under-represented in the Euro-centric field of caption studies and audiovisual translation.

As a result of my study, I have a few recommendations for Netflix and for the captioning industry in general. To correct the over-reliance on style guides – particularly as quality control tools – I would urge Netflix to institute a "Deaf Quality Control" step. Rather than having hearing individuals review captions to evaluate how well they follow the TTSG, deaf and hard-of-hearing people could review captions to ensure they provide an accessible experience. Deaf people are an underemployed demographic (Garberoglio et al, 2019), so not only would creating a Deaf QC step provide valuable employment opportunities for deaf individuals, but it would

also address the current problem captioning faces of hearing individuals guessing at the needs of deaf audiences based on style guides. Here, the disability activism adage "nothing about us without us" could be reimagined as "nothing created for us without us." As my historical overview of captions emphasized, deaf people were once very involved in captioning, and I believe there is a need to return to that practice. Directly involving deaf people in the captioning process will only improve how well captions serve their needs.

Concomitantly, broadcasters and studios need to invest more money in captioning services. With the rise of ASR technology, the captioning industry is at a turning point. While on the one hand, there is an ever-increasing volume of audiovisual content available that needs to be captioned, ASR technology is not sufficiently developed to be tasked with captioning content on its own. If the media industry is unwilling to pay for captioners' labour, the work may be outsourced to artificial intelligence and lead to a lowering of caption standards to accommodate the financial savings. It is my hope that, instead, the renewed public interest in the creative work of captioning – as evidenced by the popularity of the captions for season 4 of *Stranger Things* – can be harnessed to pressure studios to invest and value captions like they do subtitle translations. Deaf audiences deserve the best possible access to audiovisual content and achieving this means centering the needs of deaf people, involving them in the production process and budgeting more resources for captioning.

References

Ansari, A. (Director). (2015). *Aziz Ansari: Live at Madison Square Garden.* [Comedy Special]. Oh Brudder Productions; Netflix. https://www.netflix.com

Amazon (2022). Packaging requirements – All other territories (global). Retrieved from https://videocentral.amazon.com/home/help?topicId=GQSN8LYGRF5BK4Q4

Apptek, (2020, June 1). ASR in captioning: Accessibility series article 7. Retrieved from https://www.apptek.com/post/asr-in-captions-accessibility-series-article-7

Ascheid, A. (1997). Speaking tongues: voice dubbing in the cinema as cultural ventriloquism. *The Velvet Light Trap, 40*(2), 32-41. https://www.proquest.com/scholarly-journals/speaking-tongues-voice-dubbing-cinema-as-cultural/docview/1306638946/se-2?accountid=14701

Ausiello, M., Gelman, S., Gelman, V., Mason, C., Mitovich, M.W., Roots, K., Schwartz, R., & Swift, A. (2018, June 29). Is Big Brother robot ridic? Why does Luke Cage parse patois? Is Salvation Veep a leap? And more TV Qs! [Commentary]. TVLine. Retrieved from https://tvline.com/2018/06/29/luke-cage-season-2-closed-captioning-jamaican-patois-tv-questions/

Austerlitz, S. (2020, November 23). Why 'Gilmore Girls' endures. *New York Times.* https://www.nytimes.com/2020/11/23/arts/television/gilmore-girls-cw.html

BBC. (2022, July). Subtitle guidelines, version 1.2.1. Retrieved from https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/

Bélanger, N., Lee, M. & Schotter, E.R. (2018). Young skilled deaf readers have an enhanced perceptual span in reading. *Quarterly Journal of Experimental Psychology, 71*(1), 291-301. https://doi.org/10.1080/17470218.2017.1324498

Bélanger, N. & Rayner, K. (2015). What eye movements reveal about deaf readers. *Current Directions in Psychological Science, 24*(3), 220-226. https://doi.org/10.1177/0963721414567527

Berman, A. (2000). Translation and the trials of the foreign. In L. Venuti (Ed.), *The translation studies reader* (pp. 284-297). Routledge.

Blackman, J. (2020, January 16). I studied the fast-talking style of Gilmore Girls way too closely & here's what I found…[Blog post]. Retrieved from https://prettyflycopy.medium.com/i-studied-the-fast-talking-of-gilmore-girls-way-too-closely-heres-what-i-found-a800c833422c

Blair, G. J. (2021, November 5). *Why are Netflix's subtitles so bad?* Hollywood Reporter. https://www.hollywoodreporter.com/business/business-news/why-are-netflix-subtitles-so-bad-1235043051/

Boatner, E.B. (1981). Captioned films for the deaf. *American annals of the deaf, 126*(5), 520-525. doi:10.1353/aad.2012.1192

Bochner, A.P. & Ellis, C.S. (2006). Communication as autoethnography. In G.J. Shepherd, J. St. John, & T. Striphas (Eds.), Communication as… Perspectives on theory (pp. 110-122). SAGE.

Bond, E. (2019, July 16). *How Netflix does subtitling for the world (ex China)* Slator. https://slator.com/how-netflix-does-subtitling-for-the-world-ex-china/

Bond, Z.S., Stockmal, V. & Muljani, D. (1998). Learning to identify a foreign language. *Language Sciences, 20*(4), 353-367. doi:10.1016/S0388-0001(98)00009-6

Burnham, D., Grebennikov, L., Jones, C., Leigh, G., Noble, W., Tyler, M., Carley, 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. http://www.jstor.com/stable/42658948

Butler, J. (2019). Perspectives of deaf and hard of hearing viewers of captions. *American Annals of the Deaf, 163*(5), 534-554. https://doi.10.1353/aad.2019.0002

Calderón-Kellett, G., Gomez, I., Lear, N., Machado, J., Miller, B., Moreno, R., Royce, M., Ruiz, M., (2017a, May 4). *An evening with One Day at a Time* [Panel discussion]. The Paley Center for Media, Los Angeles. Retrieved from https://www.youtube.com/watch?v=R3go4WFMjP0

Calderón Kellett, G. (Writer) & Royce, M. (Writer), & Fryman, P. (Director). (2017b, January 6). This is it (Season 1, Episode 1) [TV episode]. In G. Calderón Kellett, P. Fryman, M. Garcia, N. Lear, M. Royce (Executive Producers), *One day at a time*. Sony Pictures Television; Netflix. https://www.netflix.com

Cambridge Dictionary. (n.d.) Discourse markers. In *Cambridge Dictionary grammar*. Retrieved August 29, 2022, from https://dictionary.cambridge.org/grammar/british-grammar/discourse-markers-so-right-okay

Cawthon, S., Garberoglio, C.L., Palmer, J.L. & Sales, A. (2019). *Deaf people and employment in the United States: 2019.* Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Deaf Center on Postsecondary Outcomes.

Chaiklin, R. & Goode, E. (Directors). (2020, March 20). Not your average Joe (Season 1, Episode 1) [TV episode] In R. Chaiklin, E. Goode, C. Smith, F. Stevens (Executive Producers), Tiger king. Royal Goode Productions; Netflix. https://www.netflix.com

Choi, F. (2017). Equal access requires full captioning of music and song lyrics for the deaf and hard of hearing. [Comment] *Loyola of Los Angeles Entertainment Law Review, 37*(3), 237-270. https://digitalcommons.lmu.edu/elr/vol37/iss3/1

Clark, J. (2006). Comments on Josélia Neves's PhD thesis on captioning. Retrieved from https://joeclark.org/access/captioning/Neves/

Clark, J. (2013, January 19). If you want to caption competently, you cannot be: Someone doing captioning until something better comes along or because it's [Comment on blog post "The main factor that drives captioning quality is what clients are willing to pay for it."] https://seanzdenek.com/2012/05/09/the-main-factor-that-drives-captioning-quality-is-what-clients-are-willing-to-pay-for-it/#comment-13696

Connell, B.R., Jones, M., Mace, R., Mueller, J., Mullick, A., Ostroff, E., Sanford, J., Steinfeld, E., Story, M. & Vanderheiden, G. (1997). The principles of universal design. Version 2.0. Retrieved from https://projects.ncsu.edu/ncsu/design/cud/about_ud/udprinciplestext.htm#:~:text=The%20design%20of%20products%20and,for%20adaptation%20or%20specialized%20design.

Conrad, R. (1977). The reading ability of deaf school-leavers. *British Journal of Educational Psychology, 47*(2), 138–148. https://doi.org/10.1111/j.2044-8279.1977.tb02339.x

Conway, K. (2008). A cultural studies approach to semantic instability: the case of news translation. *Linguistica Antverpiensia 7*(2008), 29-43. DOI:10.52034/lanstts.v7i.207

D'Acci, J. (2004). Cultural studies, television studies, and the crisis in the humanities. In J. Olsson & L. Spigel (Eds.), *Television after TV: essays on a medium in transition* (pp. 418-445). Duke University Press.

Degeorges, I., Franck, N. & Jaubert, M. (Producers). (2021-present). *Lupin* [TV series]. Gaumont; Netflix. https://www.netflix.com/

DeLinde, Z. & Kay, N. (2006). *Semiotics of subtitling* (6th ed). Taylor & Francis.

Deluxe. (2022). English subtitler. [Job posting]. Retrieved from https://bydeluxe.wd5.myworkdayjobs.com/en-US/Deluxe_External/details/English-Subtitler_R0012788

Described and Captioned Media Program [DCMP]. (n.d.). Captioning timeline highlights. Retrieved from https://dcmp.org/learn/25-captioning-timeline-highlights

Described and Captioned Media Program [DCMP]. (2022) Captioning key. Retrieved from https://dcmp.org/captioningkey/

Díaz-Cintas, J. (2020). The name and nature of subtitling. In L. Boducki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation* (pp.149-171). Palgrave Macmillan.

Díaz-Cintas, J., Orero, P., & Remael, A. (2007). Media for all: a global challenge. In J. Díaz-Cintas, P. Orero, & A. Remael (Eds.) *Media for all.* Rodopi.

Dietz, M. (2021, October 6). *The difference between captions and subtitles (and why it matters for Netflix's 'Squid Game').* Life Hacker. https://lifehacker.com/the-difference-between-captions-and-subtitles-and-why-1847810948

DiFrancesca, S. (1971). Academic achievement test results of a national testing program for hearing impaired students, United States: Spring 1971. [Report]. Gallaudet College, Office of Demographic Studies.

Dolmage, J. (2005). Disability studies pedagogy, usability, and universal design. *Disability studies quarterly, 25*(4). https://dsq-sds.org/article/view/627/804

Downey, G. (2008). *Closed captioning*. Johns Hopkins University Press.

The Duffer Brothers (Writers & Directors). (2016, July 15). Chapter one: The vanishing of Will Byers (Season 1, Episode 1) [TV series episode]. In D. Cohen, Duffer Brothers, K. Gajdusek, C. Holland, S. Levy, I. Paterson, M. Thunell, B. Wright (Executive Producers), *Stranger things.* Twenty-One Laps Entertainment; Netflix. https://www.netflix.com/

Eco, U. (2001). *Experiences in translation.* University of Toronto Press.

Ellcessor, E. (2012). Captions on, off, on TV, online: accessibility and search engine optimization in online closed captioning. *Television and New Media, 13*(4), 329-352. https://doi.org/10.1177/1527476411425251

Ellcessor, E. (2015). Is there a sign for that? Media, American Sign Language interpretation, and the paradox of visibility. *Perspectives: Studies in Translatology, 23*(4), 586-598. https://doi.org/10.1080/0907676X.2015.1056814

Ellis, K. & Kent, M. (2015). Accessible television: The new frontier in disability media studies brings together industry innovation, government legislation and online activism. *First Monday, 20*(9), https://doi.org/10.5210/fm.v20i9.6170

Federal Communications Commission [FCC] (2019). Twenty-First Century Communications and Video Accessibility Act (CVAA) [Consumer Guide]. Retrieved from https://www.fcc.gov/sites/default/files/21st_century_communications_and_video_accessibility_act_cvaa.pdf

Federal Communications Commission [FCC] (2021). Closed Captioning on Television [Consumer Guide]. Retrieved from https://www.fcc.gov/consumers/guides/closed-captioning-television

Ferriol, J.L.M. (2013). Subtitle reading speed: A new tool for its estimation. *Babel, 59*(4). 406-420. doi:10.1075/babel.59.4.02mar

Fitzgerald, T. (2021, May 27). The number of cord cutters and cord nevers has tripled since 2014. Forbes. https://www.forbes.com/sites/tonifitzgerald/2021/05/27/the-number-of-cord-cutters-and-cord-nevers-has-tripled-since-2014/?sh=1c22f1ce1f12

Fried, A. (Director). (2018, April 13). Christina Tosi (Season 4, Episode 1) [TV episode]. In A. Del Deo, A. Fried, M. Hilliard, D. Lillegard, B. McGinn, L. Nishimura, D. O'Malley & M. Weaver. (Executive Producers), *Chef's table.* Chef Pictures One: Netflix. https://www.netflix.com

Gernsbacher, M. A. (2015). Video captions benefit everyone. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 195-202. doi.10.1177/2372732215602130

Grady, C. (2016, November 22). Why everyone on Gilmore Girls talks a mile a minute. *Vox.* https://www.vox.com/culture/2016/11/22/13554566/gilmore-girls-fast-talking-explained

Greco, G. (2016). On accessibility as a human right, with an application to media accessibility. In A. Matamala & P. Orero (Eds.), *Researching audio description, new approaches* (pp. 11-33). Palgrave.

Hall, S. (1980). Encoding/decoding. In S. Hall, D. Hobson, A. Lowe & P. Willis (Eds.), *Culture, media, language* (pp. 117-127). Routledge.

Hamraie, A. (2016). Universal design and the problem of "post-disability" ideology. *The Journal of the Design Studies Forum, 8*(3), 285-309. doi.org/10.1080/17547075.2016.1218714

Harkins, J.E., Korres, E., Singer, B. & Virvan, B.M. (1995). *Non-speech information in captioned video: A consumer opinion study with guidelines for the captioning industry.* Gallaudet University. Retrieved from https://dcmp.org/learn/static-assets/nadh126.pdf

Hoynes, W. (1994). *Public television for sale: Media, the market, and the public sphere.* Routledge.

Ivarsson, J. (1992). *Subtitling for the media: a handbook of an art.* Transedit.

Iyuno SDI (2022, July 15). Editor, subtitling. [Job posting]. Retrieved from https://iyuno-sdi.com/careers/editor-subtitling

Jakobson, R. (2004). On linguistic aspects of translation. In L. Venuti (Ed.), *The translation studies reader* (2nd ed.). (pp.138-143). Routledge.

Janz, M. (2022, July 7). *These memes about the gross captions in Stranger Things are sending me.* Elite Daily. https://www.elitedaily.com/entertainment/tweets-memes-strangerthings-captions-subtitles/amp

Jensema, C. (1998). Viewer reaction to different television captioning speeds. *American Annuals of the Deaf, 143*(4), 318-324. doi:10.1353/aad.2012.0073

Klein, R. (2022, July 25). *What's the true price of closed captioning services?* 3PlayMedia. Retrieved from https://www.3playmedia.com/blog/how-much-does-closed-captioning-service-cost/

Kourdis, E. (2015). Semiotics of translation: an interdisciplinary approach to translation. In Trifonas (Ed.), *International handbook of semiotics* (pp. 303-320). Springer.

Kovačič, I. (1994). Relevance as a factor in subtitling reductions. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2* (pp. 244-251). John Benajmins.

Kritzer, K.L. & Smith, C.E. (2020). Changing perspectives for the 21st century: Digital literacy and computational thinking for deaf/hard-of-hearing learners. In M. Marschark & H. Knoors (Eds.), *The Oxford handbook of deaf studies in learning and cognition*. Oxford University Press. doi:10.1093/oxfordhb/9780190054045.013.22

Kuo, A.S-Y. (2020). The tangled strings of parameters and assessment in subtitling quality: An overview. In L. Boducki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation* (pp. 417-436). Palgrave Macmillan.

Lång, J., Mehtätlo, L. & Vrzakova, H. (2021). Modelling gaze behaviour in subtitle processing: The effect of structural and lexical properties. *Journal of Audiovisual Translation, 4*(1), 71-95. doi:10.47476/jat.v4i1.2021.104

Lee, C. (Writer), Showalter, M. (Writer) & Wain, D. (Director). (2015, July 31). Dinner (Season 1, Episode 5) [TV episode]. In H. Bernstein, P. Principato, M. Showalter, J. Stern, & D. Wain (Executive Producers), *Wet hot American summer: First day of camp.* Netflix. https://www.netflix.com

Leigh, I. (2009). *A lens on deaf identities.* Oxford University Press.

Lorch, M. & Meara, P. (1989). How people listen to languages they don't know. *Language Sciences, 11*(4), 343-353. doi:10.1016/0388-0001(89)90025-9

Maglio, T. (2022, March 16). *Who is winning the streaming wars? Subscribers by the numbers.* IndieWire. https://www.indiewire.com/2022/03/how-many-subscribers-netflix-disneyplus-peacock-amazon-prime-video-1234705515/

Martínez Martínez, S. (2022). Subtitling for the deaf and hard-of-hearing. In *Encyclopedia of translation and interpreting (ENTI).* AIETI. https://doi.org/10.5281/zenodo.6370763

McIntyre, D. & Lugea, J. (2015). The effects of deaf and hard-of-hearing subtitles on the characterisation process: A cognitive stylistic study of The Wire. *Perpsectives, 23*(1), 62-88. doi.org/10.1080/0907676X.2014.919008

McGraw-Herdeg, M. (2012, July 11). *Why do subtitles sometimes not match the dialogue in the Netflix Instant Streaming version of the TV show "Bob's Burgers"?* [Online forum post].

Retrieved from https://www.quora.com/Why-do-the-subtitles-sometimes-not-match-the-dialogue-in-the-Netflix-Instant-Streaming-version-of-the-TV-show-Bobs-Burgers

Milinkiewicz, F. (2020, April 23). Watching the words go by – transcribing, spotting, captions or subtitles? *Supertext Blog*. https://blog.supertext.ch/en/2020/04/watching-the-words-go-by-transcribing-spotting-captions-or-subtitles/

Morettini, A. (2012). Profiling deaf and hard-of-hearing users of subtitles for the deaf and hard-of-hearing in Italy: A questionnaire-based study. *MonTI Monografías de Traducción e Interpretación, 4*(Special Issue), 321-348. doi:10.6035/MonTI.2012.4.14

Muthusamy, Y.K., Barnard, E. & Cole, R.A. (1994). Automatic language identification: a review/tutorial. *IEEE Signal Processing Magazine, 11*(4). doi:10.1109/79.317925

National Association of the Deaf [NAD]. (n.d.). Community and culture – frequently asked questions. Retrieved from https://www.nad.org/resources/american-sign-language/community-and-culture-frequently-asked-questions/

National Association for the Deaf (NAD), Western Massachusetts Association of the Deaf and Hearing Impaired and Lee Nettles v. Netflix, 11-30168-MAP. (2012). Retrieved from https://dredf.org/captioning/netflix-consent-decree-10-10-12.pdf

Netflix. (n.d.-a). Frequently asked questions. Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/sections/203516077-Frequently-Asked-Questions

Netflix, (n.d.-b). How do I know if an asset has passed Manual QC? Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/215652527-How-do-I-know-if-an-asset-has-passed-Manual-QC-

Netflix. (n.d.-c). How is reading speed measured? Do punctuation and spaces count? Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/115001352212-How-is-reading-speed-measured-Do-punctuation-and-spaces-count-

Netflix. (n.d.-d). Introduction + frequently asked questions. Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/categories/360005967673-Quality-Control

Netflix. (n.d.-e). Introduction to Netflix quality control (QC). Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/115000353211-Introduction-to-Netflix-Quality-Control-QC-

Netflix. (n.d.-f). Is there flexibility on the reading speed threshold? Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/115001352352-Is-there-flexibility-on-the-reading-speed-threshold-

Netflix. (n.d.-g). Missing content – identifiers. Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/216371818-Missing-Content-Identifiers

Netflix. (n.d.-h). Terminology tool overview: Key names and phrases (KNP). Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/5110291703443-Terminology-Tool-Overview-Key-Names-and-Phrases-KNP-

Netflix. (n.d.-i). Timed text – Manual QC error messages. Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/sections/360000285407-Timed-Text-Manual-QC-Error-Messages

Netflix. (n.d.-j). Timed text resources. Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/categories/1500000000781-Timed-Text-Resources

Netflix. (n.d.-k). Why are Netflix's standards for subtitles and closed captions so high? [FAQ] Retrieved September 10, 2020, from https://partnerhelp.netflixstudios.com/hc/en-us/articles/214969868-Why-are-Netflix-s-standards-for-Subtitles-and-Closed-Captions-so-high

Netflix. (2021a, October 22). English timed text style guide (English TTSG). [Style guide]. Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/217350977-English-Timed-Text-Style-Guide

Netflix. (2021b, October). Netflix branded delivery specifications (version OC-4-1). Retrieved from https://drive.google.com/file/d/1_bwbUs4NaF7Y-07_NHwQqsL5L6zPVMPN/view

Netflix. (2021c, July 9). Subtitle template timed text style guide (Template TTSG, I.23). Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/219375728-Timed-Text-Style-Guide-Subtitle-Templates

Netflix. (2021d, July 9). Timed text style guide: General requirements (General TTSG). [Style guide]. Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/215758617-Timed-Text-Style-Guide-General-Requirements

Netflix. (2021e, July 9). Timed text style guide: Subtitle timing guidelines (Timing TTSG) Retrieved from https://partnerhelp.netflixstudios.com/hc/en-us/articles/360051554394-Timed-Text-Style-Guide-Subtitle-Timing-Guidelines

Netflix. (2022). About Netflix. Retrieved from https://about.netflix.com/en

Netflix CS [@Netflixhelps]. (2018, June 28). *Delivering a great experience to our deaf and hard of hearing members is very important to us. We've also heard from fans about a similar concern in Marvel's Luke Cage season 2 — we're looking into this now.* [Tweet]. https://twitter.com/Netflixhelps/status/1012440356082757632

Neves, J. (2008). 10 fallacies about subtitling for the d/Deaf and the hard of hearing. *The Journal of Specialised Translation,* 10, 128-143. https://www.jostrans.org/issue10/art_neves.pdf

Neves, J. (2018). Subtitling for deaf and hard of hearing audiences: Moving forward. In L. Pérez-González (Ed.), *Routledge handbook of audiovisual translation* (pp. 82-95). Routledge.

Norwood, M. (1988) Captioning for deaf people: An historical overview. In J.E. Harkins & B.M. Virvan (Eds.), *Speech to text: Today and tomorrow.* Gallaudet Research Institute. Retrieved from https://dcmp.org/learn/80-captioning-for-deaf-people-an-historical-overview

Nović, S. (2020, January 16). *The frustrating fight for closed captions on in-flight entertainment.* CN Traveler. https://www.cntraveler.com/story/the-frustrating-fight-for-closed-captions-on-in-flight-entertainment

Office of Communication (Ofcom). (2005, January 6). Subtitling – an issue of speed? Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0018/16119/subt.pdf

O'Halloran, K., Tan, S. & Wignell, P. (2016). Intersemiotic translation as resemiotisation: a multimodal perspective. *Signata*, *7*(2016), https://doi.org/10.4000/signata.1223

Padden, C. & Humphries, T. (1988). *Deaf in America: Voices from a culture.* Harvard University Press.

Palladino, D., Polone, G., Sherman-Palladino, A. & Rosenthal, D.S. (Executive Producers). (2000-2007). *Gilmore girls* [TV series]. Warner Brothers Television.

Pedersen, J. (2017). The FAR model: Assessing quality in interlingual subtitling. *The Journal of Specialised Translation,* 28, 210-229. http://www.diva-portal.org/smash/get/diva2:1134906/FULLTEXT01.pdf

Pedersen, J. (2020). Audiovisual translation norms and guidelines. In L. Boducki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation* (pp. 417-436). Palgrave Macmillan.

Pixelogic. (2022). Pixelogic English Editor [Job posting]. Retrieved from https://apply.workable.com/pixelogicmedia/j/1E59797CD4/

Podlas, K. (2018). Viewer disability and television accessibility: Closed-captioning and video description requirements in today's television environment. *Cardozo Journal of Equal Rights and Social Justice*, 24(2), 233-270. https://heinonline.org/HOL/P?h=hein.journals/cardw24&i=253

Pond, S. (2010, April 27). *Hollywood outsources DVD captions to… India?* The Wrap. https://www.thewrap.com/hollywood-outsources-dvd-captions-india-16727/

Public Law 85-905, H.R. 13678, §§ 1-4 (1958) Retrieved from https://dcmp.org/learn/static-assets/nadh212.pdf

Ratcliff, A. (2018, July 10). *I rely on closed captions to enjoy a show and I don't appreciate Netflix's way of censoring them.* Self. https://www.self.com/story/queer-eye-netflix-closed-captions

[rBles]. (2018, April 12). *Why don't the subtitles always match?* [Online forum post]. Reddit. https://www.reddit.com/r/netflix/comments/8brc4z/why_dont_the_subtitles_always_match/

Remael, A. (2012). Media accessibility. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies: Volume 3* (pp. 95-101). John Benjamins Publishing Company.

Romero-Fresco, P. (2009). More haste less speed: Edited versus verbatim respoken subtitles. *Vigo International Journal of Applied Linguistics, 6*(1), 109-133.

https://www.researchgate.net/publication/228502664_More_haste_less_speed_Edited_versus_verbatim_respoken_subtitles

Romero-Fresco, P. (2021). Negotiating quality assessment in media accessibility: the case of live subtitling. *Universal access in the information society, 20*(2), 741-751. https://doi.org/10.1007/s10209-020-00735-6

Ruggiero, R.M. (1986, November 13-14). Impact of television captioning on hearing audiences [Conference presentation]. Annual Meeting of the California Educational Research Association, Marina del Rey, CA, USA.

Sancho-Aldridge, J. (1996). *Good news for deaf people: Subtitling of national news programmes.* Independent Television Commission.

Schlanger, T. (Host). (2020, Oct. 9). How a deaf, silent film actor pioneered closed captioning [Audio podcast episode]. In *Q.* CBC. https://www.cbc.ca/radio/q/friday-oct-9-2020-bahamas-hrishikesh-hirway-and-more-1.5755895/how-a-deaf-silent-film-actor-pioneered-closed-captioning-1.5755964

Scorsese, M. (Director). (2019). *The Irishman* [Film]. Netflix. https://www.netflix.com

Sherman-Palladino, A. (Writer & Director). (2016, November 25). Winter (Season 1, Episode 1) [TV episode]. In D. Palladino & A. Sherman-Palladino (Executive Producers), *Gilmore girls: A year in the life*. Warner Brothers Television; Netflix. https://www.netflix.com

Showalter, M. (Writer) & Wain, D. (Director). (2015a, July 31). Staff party (Season 1, Episode 7) [TV episode]. In H. Bernstein, P. Principato, M. Showalter, J. Stern, & D. Wain (Executive Producers), *Wet hot American summer: First day of camp.* Netflix. https://www.netflix.com

Showalter, M. (Writer) & Wain, D. (Writer & Director). (2015b, July 31). Day is done (Season 1, Episode 8) [TV episode]. In H. Bernstein, P. Principato, M. Showalter, J. Stern, & D. Wain (Executive Producers), *Wet hot American summer: First day of camp.* Netflix. https://www.netflix.com

Shroyer, E.H. & Birch, J. (1980). Captions and reading rates of hearing-impaired students. *American Annals of the Deaf, 125*(7), 916-922. doi:10.1353/aad.2012.1240

Spencer, S. (2021, December 27). *The 25 most-watched TV shows on Netflix in 2021.* Newsweek. https://www.newsweek.com/netflix-2021-tv-shows-most-watched-1662152

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. https://doi.10.1353/aad.2011.0039

Szarkowska, A., Krejtz, I., Pilipczuk, O., Dutka, L. & Kruger, J. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures, 17*(2), 183-204. http://dx.doi:10.1556/084.2016.17.2.3

Szarkowska, A. & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PLoS ONE, 13*(6). https://doi.org/10.1371/journal.pone.0199331

Szarkowska, A. (2020). Subtitling for the deaf and hard of hearing. In L. Boducki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation* (pp. 249-268). Palgrave Macmillan.

Tarantino, Q. (Director). (2009). *Inglourious basterds.* [Film]. The Weinstein Company.

Taylor, C. (2020). Multimodality and intersemiotic translation. In In L. Boducki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation* (pp. 83-99). Palgrave Macmillan.

Telecommunication, 47 C.F.R (2022). https://www.ecfr.gov/current/title-47

Telecommunications Act of 1996, Pub. LA. No. 104-104, 110 Stat. 56 (1996).

Torres Monreal, S. & Santana Hernández, R. (2005). Reading levels of Spanish deaf students. *American Annals of the Deaf, 150*(4), 379-387. doi:10.1353/aad.2005.0043

Udo, J.P. & Fels, D.I. (2010). The rogue poster-children of universal-design: closed captioning and audio description. *Journal of Engineering Design, 21*(2-3), 207-221. https://doi.org/10.1080/09544820903310691

[underbuster]. (2016, November 5). *ELI5: Why some subtitle words do not match the actual dialog?* [Online forum post]. Reddit. https://www.reddit.com/r/explainlikeimfive/comments/5bb069/eli5_why_some_subtitle_words_do_not_match_the/

VanArendonk, K. (2022, July 1). *Stranger Things 4's best descriptive captions, ranked.* Vulture. https://www.vulture.com/article/stranger-things-4-subtitles-captions-best-descriptions.html

World Health Organization (WHO) and World Bank. (2011). World report on disability. Retrieved from https://www.who.int/publications/i/item/9789241564182

Youngs, I. (2021, November 15). *Young viewers prefer TV subtitles, research suggests.* BBC News. https://www.bbc.com/news/entertainment-arts-59259964

Zarate, S. (2021). *Captioning and subtitling for d/Deaf and hard of hearing audiences.* UCL Press.

Zdenek, S. (2015a). *Reading sounds: Closed captioned media and popular culture*. University of Chicago Press.

Zdenek, S. (2015b, October 1). Chapter 8: In a manner of speaking. [Blog post]. Retrieved from https://readingsounds.net/chapter8/

Zemeckis, R. (Director). (1994). *Forrest Gump.* [Film]. The Tisch Company; Paramount Pictures.