愛知淑徳大学論集－文学部篇－　第 47 号　2022.3　99-116

Interlingual Film Subtitles and Disruption of Narrative Coherence

Thomas Kabara

Key Words：audiovisual translation, cognitive science, narratology

1. Introduction

What makes viewing a film such a fertile experience is that meaning comes from a variety of directions. The words, images, sounds, performances of the actors, and our own knowledge and experiences come together to formulate narrative, characters, themes, and underlying messages in our minds.

Indeed, the process of constructing a narrative seems almost hopelessly convoluted. For any given text, sources of narrative information are numerous, and each one affects the audiences understanding to a wide variety of degrees. For a literary text, the text itself, i.e., the words on the page, provide explicit information as well as implicit clues that lead the reader to infer basic narrative meaning. Cinematic texts add sound and image to the meaning-making apparatus. Extratextual sources, including paratexts and intertextual references also contribute to narrative construction. Moreover, in today's transmedia storytelling landscape, extratextual sources do not just color audiovisual texts, they provide necessary information and fill in the blanks left in primary texts in order to build a coherent narrative. To add to all this, foreign-language audiovisual content depends on translations (via subtitling or dubbing) as a narrative input, making the process even more complex.

When we watch a non-subtitled film, we combine the various elements providing story information to create a coherent whole. But when we watch a subtitled film, we read the subtitles first. Then we attend to the image. Does this mean we understand the story through the images on the screen or through the mediation of translation? If the latter, this mediation likely affects our understanding of the explicit contents of the story, as well as the way we interpret meaning.

The goal of this article is to examine how audiences construct a narrative from a subtitled film as compared to a non-subtitled film. Taking a cue from cognitive psychology, I begin with a discussion of the task of formulating a narrative in the mind of the viewer

using all of the sources of narrative input in films both subtitled and non-subtitled. Then, utilizing the research of audiovisual translation scholars, I explore how new narrative inputs (namely, interlingual subtitles) interact with the other inputs. Finally, I examine the presence of a hierarchy of narrative inputs that affect how we understand a film.

Ultimately, this article argues that the role of film subtitles as conveyor of narrative information from the source text is a fluid one. In order to make a coherent narrative, filmmakers not only direct the viewer's attention toward necessary narrative information, they also direct the viewer's inferential reasoning processes toward convergent conclusions. In a subtitled film, the filmmakers cede partial control of the narrative to the subtitling apparatus and this disrupts the coherence of the source text. The breakdown in narrative coherence engenders a hierarchy of semantic channels: unlike in a non-subtitled film, the viewer of a subtitled film is confronted with options about where to look and what to believe when seeking the "narrative truth" of a film, and she must therefore prioritize one semantic channel over others. I call this **attentional prioritization** and **semantic prioritization** respectively. In the end, the subtitling apparatus decides attentional prioritization, but semantic prioritization is a complex and fluid process that combines input from the source text film, the target text interlingual subtitles, and the inferential processes of the viewer.

2. Narrative Coherence

In a pivotal scene in Paul Thomas Anderson's 2007 film *There Will Be Blood*, the main character Daniel Plainview (Daniel Day-Lewis) negotiates a business deal with a Paul Sunday (Paul Dano). Daniel is a cunning and ruthless oilman looking for prospects in California in the early twentieth century. Paul is a naïve-seeming farmer eager to sell Daniel information about where to find oil. The two men strike a deal where Paul reveals the location of the oil for cash. Afterwards, Daniel warns Paul: "If I travel all the way out there and I find that you've been lying to me, I'm going to find you, and I'm going to take more than my money back. Is that alright with you?" Daniel's threat is clear, but his final words are enigmatic. "Is that alright with you?" colors his threat with a hint of civility even friendliness. He is giving Paul a choice. He wants him to make an informed and rational decision. And yet his words intimate the violence of a gangster.

As Daniel delivers his final enigmatic line, he reaches out abruptly to shake Paul's hand. The gesture work in tandem with his words to create a complex character. He is cold-blooded but fair or, at least, he wants to present an image of fairness in order to gain leverage over Paul. Dialogue and image blend to create a coherent whole. Nothing is made explicit. Daniel's threat, his faux friendliness, his self-assurance must all be inferred from

image and dialogue.

The Japanese subtitles for Daniel's final line read:

金を取り返すだけじゃ　済まさん　わかってるな？

The translation is not entirely satisfying. The meaning is still clear if only implied just as in the original. But the paradoxical characterization of Daniel is missing. In the Japanese translation, he is only communicating his own interests; he is not presenting a façade of polite and rational interaction. The richness of the original is missing and that richness suggests a theme about the dehumanizing effects of capitalism.

Nevertheless, the image of Daniel's quasi-friendly gesture remains unaltered in the subtitled version. It is there for viewers to observe and interpret. But the image and subtitle do not square up as nicely as the image and dialogue do in the original. The subtitled version is more contradictory than complex. Is Daniel supposed to be a tough-talking gangster or a just a businessman? This raises questions about how we understand a character or story when watching a subtitled film: where do we look when we are looking for meaning in a subtitled film? And how do we decide what is meaningful?

2.2 Directing the Viewer's Attention and Inferences

In his book, *The Philosophy of Motion Pictures*, Noel Carroll argues that narrative films convey their stories by manipulating the viewer's attention. Carroll suggests the attention management hypothesis as a way to describe how films work in our minds:

> Motion picture makers communicate to spectators by controlling their attention. Through cinematic sequencing, the moviemakers select what the viewer sees as well as the order in which she sees it along with the duration of what she sees and the scale. The motion picture maker articulates her intentions to the audience by guiding our attention (Carroll 2007: 122).

Primarily, filmmakers achieve this control through editing and camera movement. But filmmakers have many means of manipulating attention, including lighting, blocking, sound (e.g., dialogue), etc. In fact, by following basic principles of visual perception, filmmakers can control our gaze even when there is the potential for more audience volition, as in a long shot in long take.

In an article on film scholar David Bordwell's blog, cognitive psychological researcher Tim Smith demonstrates how viewer attention is manipulated with an eye-tracking experiment conducted as part of the Dynamic Images and Eye Movement (DIEM) project. In his study, Smith recorded the eye-movements of eleven adults viewing the

aforementioned scene from *There Will Be Blood*. An eye-tracking device was used to record fixations, where the eyes are mostly still and visual processing is occurring, and saccades, where the eyes move quickly from one point to another and visual processing ceases. The results of the study reveal two related findings: a high degree of "attentional synchrony," where viewers' gazes tend to cluster around the same area of the frame, and a low degree of actual volitional gazing (Smith 2011). This confirms what Smith et al. found in previous research: viewers of moving images yield little control over where they look and when; in other words, their gaze is controlled exogenously (Smith and Henderson 2008).

Moreover, steering attention steers the processing of information. Through eye-tracking research we can measure the cognitive effort we concentrate on a task. This is known as the task-evoked pupillary response—a phenomenon in which the pupil dilates when confronted with a cognitively-challenging task (Beatty 1982). Thus, these studies do more than just measure when and where we are looking, they establish what we are concentrating our cognition on and how much how cognitive effort we are making. So, by establishing where we look and for how long and what our pupillary response is, studies like Smith's research above can show us that when we watch a film, the filmmakers direct our attention and, in doing so, direct our cognition to their communication goals.

This is significant in the meaning-making process in narrative film. Directing the viewer's attention leads to directing her cognition, which then leads to directing her inferential reasoning. Research in cognitive psychology shows us that inference is an essential process in the comprehension of any narrative text, literary and cinematic. No text makes explicit every element necessary to construct a coherent narrative. Inference must be deployed for the inevitably implicit elements in a text. This inferential reasoning wields a major influence on how we generate meaning from a narrative text.

When reading a story, we are almost constantly engaging our powers of inference. Researchers divide inference into two major types: automatic inference, where the process is fast and outside the reader's control, and strategic inference, where the process is slower and at least partially initiated by the reader. As can be seen from the definition of automatic inference, even elementary reading comprehension tasks can be categorized as inference. This includes connecting a pronoun to its referent (anaphoric inference). But strategic inference involves more advanced reading routines, such as predictive inference, where the reader forecasts future events. For example, when we read about a dog snarling at a stranger, we can predict the dog will attack. We then proceed with the story and our prediction is either confirmed or denied.

With simpler inferential reasoning, readers will draw the same or similar conclusions

from the hints provided; but with more complicated inferences, like predictive inferences, personal knowledge and experience play a greater role and individual differences will arise. Cognitive psychologists call this convergent and divergent thinking. Convergent thinking is used to connect ideas and find solutions to problems with a single well-defined answer. Conversely, divergent thinking is useful in solving problems with multiple ill-defined answers (Gerrig and Wenzel 2015: 379). For example, a reader exercises convergent thinking when encountering the sentence "The dog spotted the boy and snarled at him." At a simple, grammatical level, the reader must infer the referent for "him" in an anaphoric inference; but since there is no room for leeway about who "him" is, we can expect capable readers to all draw the same conclusion. On the other hand, the reader might exercise divergent thinking when making predictions about the events that will follow the above sentence. Some readers will guess the dog will attack, others may guess the dog will be pacified.

Any text offers a mixture of opportunities for convergent and divergent thinking for readers. Which type of thinking will prevail is often guided by the author of the text. This can be true of the director of a film, too. The aforementioned scene in *There Will Be Blood* prompts both convergent and divergent thinking. Daniel's line "I'm going to take more than my money back" provides a strong enough hint to lead viewers to draw the same conclusion about what Daniel means; but his unusual follow-up "Is that alright with you?" muddies the narrative. The viewer's thoughts can turn in several different directions. She may be tempted to try to ascertain narrative truth (Is Daniel really the type of person to follow through with this threat?). Or she may be tempted to search for and infer a major theme of the film (free market capitalism can make monstrous threats seem like rational free choices). The filmmakers have constructed a purposefully ambiguous narrative so as to support many inferential paths: some of which help the viewer understand the situation, others of which allow the viewer to interpret the situation.

Thus, inferencing processes affect not only how we understand a literary text and create a model of the narrative in our minds, they also affect how we interpret the text. The difference between understanding a text and interpreting is complex and is fraught with overlap; however, a general distinction can be made. Cognitive psychology researchers, Susan R. Goldman, Kathryn S. McCarthy, and Candice Burkett, have shown a distinct division in approaches to reading a text. They identify two approaches: a literal stance and an interpretive stance. "A literal stance orients readers to constructing what the text says based on the propositions and connections among them in the text, using prior knowledge to the extent necessary to create a coherent representation of the situation referenced by the text" (Goldman et al. 2015: 387). Meanwhile, an interpretive stance triggers the viewer to look

beyond the surface of the text and probe for implicit meaning. Goldman, et al. state that this stance "depends on integrating what the text says with prior knowledge of a variety of sorts, including knowledge of motivated human action, text genres and their characteristics, plot structures, character types, moral and philosophical systems,
and pragmatic aspects of the communicative event" (Ibid).

Moreover, readers can be externally prompted to adopt an interpretive stance. In a study where college students were divided into two groups and given different essay prompts—one designed to elicit an interpretive response, the other designed to elicit a literal response, the students who were prompted to adopt an interpretive stance did so (McCarthy et al. 2015). This is significant, since it shows that our approach to a text can be manipulated exogenously. And the above scene from *There Will Be Blood* illustrates this.

Thus, filmmakers communicate by controlling both our attention and our reasoning processes. They show us what they want us to see or guide our gaze toward what they want us to look at. In doing so, they steer not only our attention, but also our comprehension, as well as our inferential reasoning.

But there are many inputs in the meaning-making process for an audiovisual text. Watching and understanding a film is much more than a visual exercise, it, at the very least, also involves auditory input as well. While normally a filmmaker has as much control over the auditory input as the visual, the mere presence of multiple channels of expression complicates the notion of attentional control. What is more important: what we see or what we hear? Are they equally important? Adding extratextual sources of narrative information, such as interlingual subtitles, complicates the matter even further. And compounding that complication is the slipperiness of translation in audiovisual texts.

3. Disruption of Narrative Coherence

Carroll's attention management hypothesis differentiates between reading a written text and watching a film. He states explicitly that to view a film is not to "read" a film as the cognitive processes are different:

> We process the flow of information delivered to us by cinematic sequencing through an iterated series of hypotheses to the best explanation where our abiding concern is the search for coherence. This involves something quite different than the exercise of rudimentary reading skills (Carroll 2007: 121).

For Carroll, a film must be understood holistically whereas a written text is understood by building meaning from smaller units (words) in combinations governed by rules (grammar). Since images are not decomposable as words are and there are no hard rules to film (only rules of thumb), the construction of meaning for the two media follow different paths.

But a film narrative is pieced together from different sources. Sound and image work together to form a coherent story. But does the fact that there is more than one source of narrative input upset this coherence? What about the presence of foreign-language subtitles in a film? Does this draw viewers back into reading territory, and, therefore, affects the cognitive processes that go into understanding a film? To understand how subtitles affect our thinking about a film, it is helpful to look at the how the various sources of narrative information in a film interact with each other.

Danish translation studies scholar Henrik Gottlieb has attempted to classify translations according to their media characteristics. For example, according to Gottlieb, film subtitles are an example of what he calls a diasemiotic translation of a polysemiotic text. In his effort to formulate a taxonomy of audiovisual translation, Gottlieb has proposed dividing all source texts into two types: monosemiotic and polysemiotic. Monosemiotic texts are texts that use only one semiotic channel of expression (e.g., written language), whereas polysemiotic texts are texts that use two or more parallel semiotic channels of expression (Gottlieb 1997: 143). Thus, a book, a medium which relies entirely on a single semiotic channel (written language) to convey meaning, would be categorized as a monsemiotic text, while a film would qualify as a polysemiotic text, since it features at least three semiotic channels, including image (non-verbal), sound (non-verbal), and dialogue.

Gottlieb's taxonomy also offers several categories for translation types based on the characteristics of the media involved. Only two of these categories are relevant to the present article: isosemiotic and diasemiotic. Isosemiotic refers to translation in which the same semiotic channel of expression in the source text is used in the target text. An example would be a translation of a non-illustrated novel where the written language of the source text is conveyed as written language in the target text. Foreign-language film dubbing would be another example. Diasemiotic translation, on the other hand, refers to translation where a semiotic channel of expression that is different from the one used in the source text is used in the target text. Gottlieb's quintessential example of this is film subtitling, as spoken dialogue is converted into written language (Gottlieb 2005: 36). Thus, as stated above, interlingual film subtitling can be classified as a diasemiotic translation (shifting spoken language in the source text to written language in the target text) of a polysemiotic text (audiovisual content).

Gottlieb's classifications raise some immediate issues, however. First, the concept of a monosemiotic text has been roundly criticized by Klaus Kaindl, who questions whether there is such a thing as a "monosemiotic" text in the strictest sense. He points out that even in the quintessential "monosemiotic" text, an unillustrated book, "the colour of the book cover, the paper quality, the layout and the typography already have semiotic qualities" (Kaindl 2013: 260). In other words, Kaindl finds the presence of paratext makes any text polysemiotic.

So, does that make every text polysemiotic? This too is problematic. As mentioned above, Gottlieb defines audiovisual contents as polysemiotic. He borrows Frederic Chaume's definition of an audiovisual text to support this classification. Chaume states that "An audiovisual text is a semiotic construct comprising several signifying codes that operate simultaneously in the production of meaning" (Chaume 2004: 16). Ostensibly, Chaume is referring only to the signifying channels that are available to the viewer during the immediate process of viewing a film, in other words, the auditory, visual, and verbal output used to convey narrative information in a movie. This would include sound effects and music, cinematographic images and their arrangement, and character dialogue and narration, all of which combine to form coherent meaning. Thus, there are multiple channels of information as part of the text itself.

In a non-subtitled film, generally, the various semantic channels reinforce each other to form a coherent narrative. But, the addition of interlingual subtitles can disrupt this coherence. Subtitles add a new semantic channel to the mix; but one with characteristics that separate it from the other channels. First, although it appears concurrently with other channels, it is not experienced concurrently. Second, it alters the source text, generating new narrative information. Alteration of the source text implies conflict among the various channels of expression found in film. The words of a subtitle may not match the images, sounds, or dialogue of the source text. Even when they do match, viewers cannot read and explore the image and sound simultaneously. This must be done consecutively.

3.2 Conflict Among Semantic Channels

The above framework raises questions about the existence of a hierarchy of channels in film. This includes non-subtitled films. If we assume such a hierarchy exists, then what channel do viewers prioritize unconsciously or otherwise? Viewers prioritizing one channel over another could refer to one of two things: 1. focusing more attention on a certain channel over any other channel throughout a film or at a given moment, or 2. consulting a certain channel and weighing its contents more heavily than the contents of any other channel when attempting to grasp the film's overall content (e.g., the story in a narrative film) or interpret

that content. These two modes of prioritizing can be referred to as **attentional prioritization** and **semantic prioritization**. The presence of a hierarchy raises a further question: If a contradiction arises in the content between two or more channels, which channel do viewers refer to as the most reliable source of information when attempting to grasp the meaning of the film?

In discourse on how viewers piece together the narrative of a non-subtitled film from various pieces of information we perceive during the viewing process, it is often assumed that viewers take in information from all semiotic channels more-or-less simultaneously since these pieces of information generally combine to form a coherent whole. However, occasionally conflict may arise between the information provided in the different channels resulting in a narrative contradiction. Invoking the usefulness of the concept of a cinematic narrator, Seymour Chatman refers to this as "a conflict between two mutually contradictory components of cinematic narrator" and points out that this is unique to multi-channel media such as film, which has, what he calls, a visual track and an auditive track (1990: 136).

Examples of this narrative clash can be found in films in which a voice-over narrator's descriptions do not match the image information. This kind of conflict arises in a number of films, such as *Rashomon* (1950), *Badlands* (1973), *The Usual Suspects* (1995), *Fight Club* (1999) and *Memento* (2000). In these cases, what the viewer learns through narration clashes with what they see on screen; the viewer must then reconcile this difference by, in a sense, "selecting" which piece of information (the dialogue or the image) he or she will use to form a coherent narrative. Although, this "selecting" is almost never done consciously and is rather done automatically based on conventions of film narrative, genre narrative, film viewing, and our natural inclination to believe our sensory perception that is unmediated and unfettered by human intention.

Indeed, in the case of non-subtitled films, it is often assumed that image occupies the foremost position in film. In his account of Terrance Malick's *Badlands*, Chatman describes a contradiction between Holly's account of her escape with Kit and the visual information being displayed on screen. The visual information is patently more sordid. In short, he notes a conflict between what is told and what is shown. He argues that this sort of conflict must be reconciled through a cinematic narrator and that narrator delivers the "true" story content visually. As he states, "the partially unreliable narration of Badlands arises explicitly from a conflict between two mutually contradictory components of the cinematic narrator. Normally, as in Badlands, the visual representation is the acceptable one, on the convention that seeing is believing" (Chatman 1990: 136).

Peter Verstraten criticizes Chatman for this assumption about the necessity of a

cinematic narrator and the power of image over all auditive information. Verstraten views this as an artificial hierarchy. He argues that the viewer cannot assume that the visual information is more accurate when seeking the "truth" of the story, but rather it is more "correct to say that Malick's road movie concerns itself with the clash between the auditive and the visual tracks itself." The contradiction is the point of the story, and its narrative truth is deliberately slippery and difficult to define. He states his position succinctly: "if words can lie and images can comply with untruthful words, where does the true version of events reside?" (Verstraten 2009: 136-138).

Perhaps this is true in films that make a theme out of conflict between the auditive and visual tracks; however, not all films with conflicting information in the auditive and visual tracks necessarily aim to develop a theme about the nature of our perception or "truth" in storytelling. In some cases, contradictions between the voice-over narration and the visual information of a film are used to illustrate the unreliability of the character giving the narration. In *Memento*, for example, a conflict arises between what Leonard, who suffers from anterograde amnesia, tells the other characters, including an off-screen presence, who serves as a de facto stand-in for the audience, throughout the film, and information shown on the visual track. Since Leonard's mind is clearly not reliable, we are left wondering which story to believe: the story told in Leonard's quasi-voiceover narration or the visual information. On the one hand, we may not care. Unless we are invested in re-constructing the most accurate story of Leonard's fictional life, it matters little what the "truth" of the story is, as the film is constructing a theme that calls into question the reliability of memory and belief. On the other hand, what the flashback images do show is the narrative "truth" about Leonard's state-of-mind: Leonard is utterly confused and even his most cherished memories are not reliable even to himself. It is the visual track that provides this information.

The films unusual structure and how it is treated confirms this theme. Because its plot proceeds backwards, *Memento* can be called a "puzzle movie." It invites reformulating the narrative elements into a coherent fabula (chronological order of the story) either in one's own mind or even in physical form. In fact, some ambitious fans of the film have already reformulated it to play in chronological order and uploaded it onto the Internet; and the official region one limited edition DVD has a hidden feature, or "Easter egg," that allows viewers to watch the film in chronological order. Referring to *Memento* as a puzzle film implies that with a bit of work the viewer can reformulate the plot into a coherent story.

The point of puzzling out the film's fabula is to uncover the objective truth of the events of the fabula. And objective truth comes from the camera and nowhere else. Leonard has no

functioning long-term memory, and as we discover late in the film, he willingly distorts his own written records of history for his own sense of satisfaction. He lies. Therefore, we have no reliable verbal record to the story events—leaving us with only the visual information that feels objective to construct the "true" story.

Thus, hierarchy seems to emerge when conflict arises between the two tracks, and a viewer seeking narrative coherence must prioritize one track over another. or when one track supersedes another track, as when subtitles command the attention of viewers at the expense of image information.

3.3 Adding Subtitles to the Hierarchy

It is widely accepted that subtitles are a limited translation, if they can be called "translation" at all. Gottlieb states that due to limitations on the number of characters that can be displayed per second and thus the necessity for "condensation" of the source text, "language professionals tend to disagree as to whether subtitling is indeed translation, and even the subtitling industry is often reluctant to grant this type of language transfer the status of 'real' translation" (Gottlieb 2004: 219). Using the same reasoning, practitioners in Japan bolster this idea with the often repeated mantra that "subtitling is not translation." In fact, these subtitlers have produced a volume in 1992 dedicated to this idea entitled *Film Subtitling is not Translation* (aka *Eiga jimaku wa honyaku de wa nai*).

In Japan, at least, the difficulty of producing film subtitles was recognized early. In February of 1931, the Japan office of Paramount Pictures distributed one of its American films, *Morocco*, with subtitles in Japan for the first time. Tamura Yoshihiko was commissioned to create the subtitles for the film. In the February 1, 1931 issue of *Kinema Junpo* Tamura remarked on his first attempt at creating subtitles: "In translating everything, the audience would have their attention occupied by reading and would not be attentive to the screen, and at the very least, there is a chance they wouldn't understand the next scene" (Tamura 1931, my translation).

Tamura limited his subtitling to an average of thirty titles per reel in the 92-minute film. Thus, the question of where viewer attention is directed was pivotal from the very beginning of subtitling in Japan. In 1931, Tamura made the decision to limit how much of the audience's time he was going to occupy with his subtitles, recognizing that audiences would invariably turn their attention to the super-imposed words at the expense of the image. In a sense, he directed viewers to prioritize the image over the subtitles.

As noted earlier in this article, a filmmakers task is to direct the viewer's attention. The significance of this for subtitled films is that all of the control a director can wield over

where viewers look potentially disintegrates when subtitles appear, since viewers relying on subtitles will immediately and involuntarily look to them for information as soon as they appear. This has repercussions for who controls the viewer's gaze: the director or the subtitling apparatus. Note that we can only say that the subtitling apparatus controls viewer gaze and not the subtitlers themselves since they must follow strict rules about character or word count, and they can only translate the dialogue found in the source text. Although subtitlers make choices about how the original dialogue is translated, they cannot make decisions about length of the subtitles or the duration they remain on screen (other than severely limiting the number of subtitles as Tamura did; but audiences' demands for accurate translations preclude subtitlers from following Tamura's lead). Indeed, as a rule the appearance of a given subtitle does not extend beyond a single shot. As soon as the image cuts to a new camera angle, the subtitle from the previous shot will disappear. In the new shot there will either be a new subtitle or no subtitle. Thus, the first thing we look at when with every cut in a scene with much dialogue is the new subtitle: with every cut our attention shifts back to the subtitles. This would not necessarily be the case in scenes shot in a single take or with little cutting. In other words, the director can maintain control of the viewers' gazes when using long takes even in subtitled versions of the film.

Looking closely once again at the aforementioned scene from *There Will Be Blood* can test the above assumption about single takes. The scene in question lasts one minute and 45 seconds. As mentioned above, it is shot as a single take with a wide angle lens. All four characters in the scene have at least one line of dialogue, but some two-thirds of the thirty-six lines of dialogue is spoken by one character, Paul, as he explains to the others where on the map oil has be discovered. Each of the 36 lines of dialogue is translated and rendered as 36 subtitles, averaging one subtitle approximately every two and half seconds.

Although the content of the dialogue is necessarily condensed, as is usual for subtitles, not much is lost in the translation. The content of the dialogue is mostly banal, covering the location of the oil, family composition, and names. The language used in this scene is simple and the actors deliver these lines slowly. At the end of the scene, Daniel vaguely threatens Paul.

With a few exceptions the subtitles appear one right after another, giving viewers little time to scan the image. Although this could be measured empirically, it is difficult to imagine that with the quick succession of subtitles, however simple they may be, the viewer's attention is anything but occupied with reading. The key moment described at the beginning of this article illustrates this well. As scene in Figure 2, toward the end, Daniel, abruptly raises his hand in a grand gesture. It is both threatening and mysterious and it

provides a climactic moment in a tense discussion between the characters. As it turns out, Daniel is merely offering to shake hands with Paul as is customary when striking a deal. However, this is precisely when he delivers his vague threat to Paul (“I’m going to take more than my money back. Is that alright with you?”). This line summarizes Daniel’s character nicely, but then again, so does the simple hand-raising gesture. Thus, the source text dialogue and the image of the gesture dovetail into a coherent characterization of Daniel.

But in the target text, the line “Listen!” is uttered simultaneously with the gesture; thus, a Japanese subtitle is also displayed simultaneously with the gesture. Of course, where the viewer looks at this particular moment is dependent on a host of factors and an eye-tracking study would provide empirical evidence, but with the factor of attentional synchrony and viewer tendency to look to the subtitles first for information, it is probably safe to say that when seeing this moment for the first time, a viewer would have his or her attention directed at the subtitle rather than Daniel’s meaningful gesture.

Perhaps the viewer can take in the visual information as well as the subtitle all at once and combine them in her mind into the same coherent whole that is there in the original. Eye-tracking research on subtitled videos seem to suggest this is the case. Numerous studies have been done using a combination of eye-tracking with questionnaires. Participants are asked to view a short video clip with foreign-language subtitles with an eye-tracking device. The eye-tracking measures not only where the viewer looks and for how long, but also how much cognitive effort a given task requires via the task-evoked pupillary response. Afterwards, they are given questionnaires to test their ability comprehend and retain both image and language information. This research shows that viewers do attentionally prioritize subtitles, but they take in considerable image information as well. In fact, some eye-tracking studies have shown that viewers regularly do have time and take advantage of that time to scan the image as well (Perego, Del Missier, Porta, and Mosconi 2010), (Bisson, Van Heuven, Conklin, and Tunney 2012). Additionally, viewers grasp information from both channels easily. When viewing a video clip, viewers were able to pay attention to both subtitles and image content and were able to recall information from both afterward, suggesting at least the possibility of taking in content from both channels adequately (Perego, Del Missier, Porta, Mosconi 2010). These studies also measure if we spend more time processing text or image or if we sacrifice cognition of image information in favor of subtitles. The answer is no. There is no apparent trade-off between text and image processing.

Other studies have expanded on this research to include even more cognitive tasks. Specifically, they test viewers’ ability to follow a subtitled video with annotated surtitles

containing meta-linguistic information. According to one study, we do not sacrifice narrative comprehension or cognition of image information even when we are saddled with multiple textual inputs and cognitive load is increased. In these studies participants were shown foreign-language video clips either with standard subtitles or with annotated subtitles. Eye movement and fixation duration were measured with an eye tracking device. Afterwards, participants were tested on the ability to retain information from the text or from the images. The findings show that when viewers watched the annotated video clips, they spent less time looking at image content and expended more cognitive effort processing the content, but they were still able to retain both graphic and textual information just as well as when viewing content with standard subtitles only (Kunzli and Ehrensberger-Dow 2011). In summary, the eye-tracking research suggests that viewers can and often do manage to perceive information from a wide variety of channels even if they are presented all at once, and they combine this information to construct a narrative.

But none of this research contradicts the idea of a hierarchy of channels; in fact, it implicitly confirms an important underlying observation: viewers take in each channel separately. The viewer reads the subtitle as an isolated piece of information and then, time-permitting, seeks further information or confirmation from the image. If the viewer has no time to scan the image, which, despite what the above findings suggest, can be the case, then by virtue of providing the only information fully perceived, the subtitles not only are attentionally prioritized but also semantically prioritized (i.e., they are what the viewer refers to when piecing together the "truth" of the story). In addition, if a conflict arises between the meaning of the images/sounds and subtitle, the coherence-seeking viewer must prioritize one source of narrative truth over another. This is in direct contrast to the view that non-subtitled films are understood holistically rather than as pieces of information coming from separate channels in non-subtitled films.

In this sense, it is difficult to see anything but a hierarchy of semiotic channels in the subtitled film viewing experience. Rather than three semiotic channels (image, sound, dialogue) perceived more or less simultaneously and forming a coherent whole, only two channels (subtitles and sound) can be perceived simultaneously and one of those channels may conflict with the others in the same way the images and narration conflict in films like *Memento*. Thus, one (or perhaps two) channel(s) must be attentionally prioritized and semantically prioritized over the remaining other channels. To summarize, if we seek coherence, then we must prioritize.

Nevertheless, this does not guarantee that the subtitle channel is the one that is prioritized when grasping the meaning of a film. Some eye-tracking studies have found that

viewers sometimes skip reading subtitles when more salient features in the image, such as motion, appears simultaneously with the subtitle (Bisson, Van Heuven, Conklin, and Tunney: 2012). Thus, viewers do exercise some control over which channel they attentionally prioritize.

Furthermore, subtitles are delimited by the other semantic channels, and as a result, they are usually designed to offer coherent reinforcement to the viewer. As Delia Chiaro points out, subtitles are leashed to the image channel. She argues that the translator is bound by the visual information on the screen and therefore any translation of spoken speech would have to conform to that information, a requirement not shared by an isosemiotic translation of a monosemiotic text (e.g., a written translation of a written work). She elaborates: "For example, in a novel, no matter how such features are conveyed for the target reader, the idea of the objects in question will remain in the reader's mind and imagination; in contrast, with filmic products many references are in full view on screen, leaving the translator with little room to manoeuvre" (Chiaro 2009: 155). For example, if a Japanese film with English subtitles features a close-up of a bottle of green tea with the label in clear view and a character makes reference to the beverage in the source text dialogue, a subtitler would be bound by that visual information when translating. She could not use a substitution strategy to translate the reference from green tea to, say, Coke. A translator of a monosemiotic text would be free to use a substitute since she need only consider the internal coherence of her translation. But a subtitler must respect coherence between two channels at once: image and written word. If not, a conflict between image and subtitle would arise.

Moreover, in cases where viewers are proficient in the language of the source text dialogue, that viewer can compare the meaning of that dialogue with the meaning of the subtitle. Indeed, it is not rare for viewers to critique subtitles for straying too far from the meaning of the source text (see Tachibana Takeshi's article on Toda Natsuko's subtitles for *Apocalypse Now* or fans' vociferous complaints about the subtitles for *The Lord of the Rings*). Thus, even the presumably often ignored dialogue channel of a subtitled film can draw the attention of some viewers and even be used as a measuring stick for evaluating the accuracy of the translation. This is why Diaz-Cintas and Ramael refer to subtitles as a "vulnerable translation" (Diaz-Cintas and Ramael 57: 2007). With this vulnerability, the subtitles content must defer to the dialogue channel of the source text.

Thus, in the same way Tamura limited his subtitle count for *Morocco*, today's subtitlers often mute the authority of their work by redirecting the viewer to find meaning in image or, to some extent, source text dialogue. Or perhaps it is more accurate to say the subtitling

apparatus redirects the viewer. While it is true subtitlers do make translational decisions, often those decisions are made for them by the cinematic apparatus. Nevertheless, as complaints about translation inaccuracies by fans and critics illustrate, subtitlers are bound to rewrite much of the source text in their work due not to their own professional arrogance but to that same subtitling apparatus that limits their choices. The phenomenon of fan subtitling demonstrates just how far fans are willing to go to obtain subtitles that reflect the source text with minimal change, even if those subtitles leave many culturally specific references untranslated (Kabara 2019). Such subtitles are designed to reinforce the source text semantic channels, but if any conflicting information were to arise between the source text image information or dialogue and the subtitles, the viewer would need to semantically prioritize the various channels, and she may decide to prioritize the former two channels over the subtitles.

4. Conclusion

This article has argued that in non-subtitled films, filmmakers form coherent narratives by directing the viewer's attention and inferential reasoning. They do this by integrating all of the semantic channels (sound, image, dialogue) in a film into a coherent whole. However, in subtitled films, the filmmaker's power to direct attention and inference is disrupted by the subtitling apparatus, which adds a new semantic channel to the experience. Because this new channel comes from outside, it does not integrate with the core text fully. Unlike image and dialogue, subtitles are not taken in simultaneously with all the other channels. We read subtitles first and then engage with the image. Also, subtitles alter the source text dialogue. In doing so, they disrupt the coherence of dialogue and image formulated by the filmmaker, as in the example above from *There Will Be Blood*, where image, source text dialogue, and subtitle form a fuzzy match thanks to the subtitles lack of subtlety. This shifts the way the viewer makes inferences about the film. Nevertheless, this does not mean that subtitles necessarily distort the filmmaker's intentions. Rather, this new semantic channel builds onto the source text by adding a new source of narrative information. The images, dialogue, and sound of the source text remain available for viewers to consult when forming the narrative in their minds. As a result, a two-fold hierarchy of semantic channels forms. The subtitling apparatus decides where the viewer will look first for narrative information (attentional prioritization), but then the viewer must decide where she can find "narrative truth" or the true story of the film (semantic prioritization). Thus, the role subtitles play in conveying source text information is a fluid one.

References

Beatty, Jackson. "Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources." *Psychological Bulletin* 91 (1982): 276-292.

Bisson, Marie-Josee, Walter J.B. Van Heuven, Kathy Conklin, Richard J. Tunney. "Processing of Native and Foreign Language Subtitles in Films: An Eye Tracking Study." Applied Psy*cholinguistics* (2012): 1-20.

Carroll, Noël. *The Philosophy of Motion Pictures*. Malden, MA: Blackwell Pub., 2007.

Chatman, Seymour Benjamin. *Coming to Terms: The Rhetoric of Narrative in Fiction and Film*. Ithaca, NY: Cornell UP, 1990.

Chaume, Frederic. "Film Studies and Translation Studies: Two Disciplines at Stake in Audiovisual Translation." *Translators' Journal* 49.1 (2004): 12-24.

Chiaro, Delia. "Issues in Audiovisual Translation." *Routledge Companion to Translation Studies*. Ed. Jeremy Munday. New York: Routledge, 2009. 141-165.

Diaz-Cintas, Jorge and Aline Ramael. *Audiovisual Translation: Subtitling (Translation Practices Explained)*. Manchester: St. Jerome, 2007.

Gerrig, Richard J. and William G. Wenzel. "The Role of Inferences in Narrative Experiences." *Inferences During Reading*. Ed. Edward J. O'Brien, Anne E. Cook, Robert F. Lorch, Jr. Cambridge University Press, 2015. 362-385.

Goldman, Susan R., Kathryn S. McCarthy, and Candice Burkett. "Interpretive Inferences in Literature." *Inferences During Reading*. Ed. Edward J. O'Brien, Anne E. Cook, Robert F. Lorch, Jr. Cambridge University Press, 2015. 386-410.

Gottlieb, Henrik. *Subtitles, Translation and Idioms*. University of Copenhagen: Center for Translation Studies and Lexicography, 1997.

Gottlieb, Henrik. "Subtitles and International Anglification." *Nordic Journal of English Studies* 3.1 (2004): 219-230.

Gottlieb, Henrik. “Multidimensional Translation: Semantics turned Semiotics.” *Challenges of Multidimensional Translation: Conference Proceedings* (2005): 33-61.

Kabara, Thomas. “The Cultures of Professional Subtitling and Fansubbing: Tradition and Innovation in Audiovisual Translation in Japan.” *Diverse Voices in Translation Studies in East Asia*. Eds. Nana Sato-Rossberg and Akiko Uchiyama. Oxford, UK: Peter Lang Ltd., 2019. 123-145.

Kaindl, Klaus. “Multidmodality and Translation.” *Routledge Handbook of Translation Studies*. Eds. Carmen Millán and Francesca Bartrina. New York: Routledge, 2013. 257-270.

Kunzli, Alexander and Maureen Ehrensberger-Dow. “Innovative Subtitling: A reception Study.” *Methods and Strategies of Process Research: Integrative Approaches to Translation Studies*. Eds. Cecilia Alvstad, Adelina Hild, and Elisabet Tiselius. Amsterdam: Amsterdam: John Benjamins Publishing, 2011. 187-200.

McCarthy, Kathryn S. and Susan R. Goldman. “Comprehension of Short Stories: Effects of Task Instructions on Literary Interpretation.” *Discourse Processes* 52.7 (2015): 585-608.

Perego, Elisa, Fabio Del Missier, Marco Porta, and Mauro Mosconi. “The Cognitive Effectiveness of Subtitle Processing.” *Media Psychology* 13 (2010): 243-272.

Shimizu, Shunji. *Eiga jimaku wa honyaku de wa nai*. Eds. Natsuko Toda and Ueno Tamako. Tokyo: Hayakawa Shobo, 1992.

Smith, Tim. “Watching You Watch There Will Be Blood.” Weblog post. *Observations on Film Art*. Ed. David Bordwell and Kristen Thompson. N.p., 14 Feb. 2011. Web. 12 Feb. 2014.

Smith, Tim and John Henderson. “Attentional Synchrony in Static and Dynamic Scenes.” *Journal of Vision* 8.6 (2008): 773.

Tamura, Yoshihiko. *Kinema Junpo* Feb. 1, 1931

Verstraten, Peter. *Film Narratology*. Toronto, ON: University of Toronto Press, 2009.