



The International Journal for
Translation & Interpreting
Research
trans-int.org

Relating visual images to subtitle translation in *Finding Nemo*: A multi-semiotic interplay

Yuping Chen

The University of Sydney
yche1694@uni.sydney.edu.au

Wei Wang

The University of Sydney
wei.wang@sydney.edu.au

DOI: 10.12807/ti.108201.2016.a05

Abstract: Focusing on the role of semiotic interplay in subtitling, this paper draws upon Systemic Functional Linguistics (SFL) -informed multimodality and semiotic translation to address the subtitle translation in *Finding Nemo*. In so doing, the subtitle translation is analysed at Kress and van Leeuwen's three metafunctional levels with a view to relating dynamic visual images to subtitle translation, while previous studies only consider interrelations between static images and isolated filmic units. This paper examines the types and functions of semiotic interplay, and the concrete influence that such interplay brings to bear on subtitle translation. This study finds that there are six types of semiotic interplay, operating at three metafunctional levels, which perform different functions to facilitate the subtitling process and which exert different impacts on subtitle translation.

Keywords: visual modes, SFL-informed multimodality, semiotic translation, subtitle translation, animation film

1. Introduction

Subtitle translation has long been regarded as a type of translation activity concerned with verbal language. Though verbal language is one of the major vehicles conveying meanings in subtitle translation, the function of non-verbal modes, such as visual images and audio tracks, should not be ignored. Effectively, it appears that the audiovisual text – the carrier of subtitle translation – is predominantly a multimodal text, in which both verbal languages and visual images play crucial roles in the meaning-making process. Drawing on theories in SFL-informed multimodality analysis and semiotic translation, this paper examines how the semiotic interplay between the verbal and visual modes influences the subtitle translation in one animation film – *Finding Nemo*. Thus, this paper addresses the research questions as to: 1) how the semiotic interplay between the visual and verbal modes function to address the key issues in subtitle translation, and 2) what concrete influences can be seen in subtitle translation.

Although recognition of the function of visual images has led to abundant semiotic research on subtitle translation (e.g. Chuang, 2006; Díaz-Cintas & Remael, 2007; Perego, 2009), the dominant research perspective taken in these studies has been principally confined to the semiotic interdependence between verbal messages and static visual images. Considering the two sub-

modes of visual images in audiovisual texts, i.e. static images and dynamic images (Pérez-González, 2014), this paper intends to go beyond the intersemiotic relations between verbal messages and static visual images, in its exploration of the role of both static and dynamic images in the Chinese subtitle translation of the Disney animation film *Finding Nemo*. As one of the most popular English animation films worldwide in recent times, *Finding Nemo* has attracted a large number of viewers in mainland China. Since only a relative few among Chinese audiences can understand English, subtitling becomes indispensable for the film's circulation in China. Indeed, much of the credit for its success in there can be attributed its Chinese subtitle translation which, despite some isolated imperfections, is generally of outstanding quality. *Finding Nemo* has therefore been considered as a good example for analysis of semiotic interplay in subtitle translation.

With a view to analysing subtitle translation in terms of static and dynamic images, this paper begins with considerations of key issues and concerns in subtitle translation. Then, a theoretical model is proposed for the analysis of subtitle translation. Subsequently, a case study is presented with examples from *Finding Nemo*. Finally, this paper discusses the present research findings in relation to previous studies, and concludes by arguing that both static and dynamic images play vital roles in ensuring a high quality subtitle translation.

2. Key issues in subtitle translation

In light of the previous research in subtitle translation (e.g. Georgakopoulou, 2009; Giovanni, 2003; Pettit, 2009; Remael, 2003), three key issues in subtitle translation have been identified in this paper – namely, overcoming technical constraints, bridging cultural gaps, and maintaining the narrative flow in audiovisual texts.

Overcoming technical constraints, including temporal and spatial ones, is one of the key concerns in subtitle translation. As film dialogues are usually delivered at a faster pace than a translation rendered graphically on the screen (Delabastita, 1989), subtitlers usually have to compress or reduce the text in subtitle translation. This reduction is due to the dual constraints of time (how fast audiences can process the text) and space (subtitles cannot occupy any more than 20% of the screen). To overcome these constraints, positive feedback from the audiovisual context is useful for completing the translated version as a holistic entity (Gottlieb, 1992). It is recognised that the relations between verbal modes and visual modes are vital in clarifying the concrete function of visual text in subtitling process.

The next key issue in subtitle translation is that of bridging cultural gaps. Though a common issue in all translation activities, it carries different implications in subtitle translation where the role of visual modes is more conspicuous and frequently considered.

A cultural sign is a sign which contains culture-specific information, verbal or nonverbal, transmitted aurally or visually. Since signs occur in every instance of human action and interaction, it is clear that semiotics (including areas such as film studies) is deeply embedded in the dynamics of cultures and societies (Giovanni, 2003). An audiovisual text accordingly offers a cultural representation of the world through both language and visual images. Intending to go beyond the sole examination of the subtitling of verbally-expressed cultural messages, this paper investigates how to handle cultural information transferred through visual modes, rather than through verbal language alone. Because an audiovisual text is such an integrated and

inseparable entity, it is illogical for the subtitling process to address only those culture-specific artefacts that are expressed verbally but neglect the non-verbally stated ones.

Subtitle translation concerns not only rendering verbal dialogues, but also and more importantly, maintaining narrative flow in audiovisual texts. As Remael (2003, p. 233) argues, "...film dialogue is not just 'dialogue', it is also a narrative". Accordingly, subtitles must have their own sequential structure to reconstruct the narrative flow for viewers who cannot understand the source verbal dialogues. The assembly of this sequential structure relies on both the verbal and non-verbal dimensions, given that audiovisual texts feature multiple modes of communication. In view of this, this paper also explores how subtitles interact with visual images to realize the narrative flow of the film.

Together with the two theoretical foundations discussed below, these three key issues in subtitle translation will contribute to the construction of a theoretical model for this paper.

3. A theoretical model

In its construction of a theoretical model for its analysis, this paper draws upon SFL-informed multimodality and semiotic translation as twin theoretical foundations. In so doing, the selected subtitle translation is analysed at three metafunctional levels by integrating some key theories in semiotic translation. These theories in semiotic translation include the interrelations between Representamen, Object and Interpretant, and the theories of deduction and induction (see details below).

The three metafunctional levels for the multimodal analysis of subtitle translation are representational, interactive and compositional. Kress and van Leeuwen's (2006) terminologies of "representational", "interactive" and "compositional" are used to replace Halliday's "ideational", "interpersonal" and "textual" because this paper deals not only with language analysis, but also and more importantly, with the semiotic interplay between verbal language and visual images. Accordingly Kress and van Leeuwen's terms emerge as being better suited to the research focus of the present study. Additionally, different analytical units are used at these three metafunctional levels. At the representational level, the analytical unit is frame/shot, because frame/shot constitutes uncut or unedited camera action (Iedema, 2001), which focuses on the content of stills. At the interactive level, the analytical unit is scene, because scenes comprise more than one shot with time and space continuity. So the camera actions are edited or cut and these edited actions highlight the interplay. At the compositional level, the analytical unit is sequence, generic plot stage, and even the entire film, with a view to involving subtitle translation in the larger filmic context. A sequence is composed of "a group of scenes having dramatic unity" (Oumano, 1985, p. 160). Stages are those elements which tell us where we are in the overall filmic text: "this is where things are starting to go wrong", "this is where we enter another argument", "this is where things come to a kind of conclusion" (Iedema, 2001, p. 190).

This study carries out an SFL-informed multimodal analysis of subtitle translation at these six filmic levels. At the representational level, frame/shot-based semiotic interplay is the major concern for investigating the interdependence between static visual images and verbal messages. At the interactive level, the scene is the basic analysis unit because the filmic "cut" between shots strongly demonstrates the interaction between participants in

the visual world or that between viewers and the framed world. This is not to say that there is no interaction in a still, but rather that a much more vivid and strong interaction can be detected from the dynamic images. At the compositional level, the sequence/stage/whole text is the basic unit to examine how the holistic contextual visual information influences the subtitle translation in *Finding Nemo*.

The second theoretical foundation is semiotic translation, regarding which some key theories are invoked in this study. The first theory concerns the interrelations between Representamen, Object and Interpretant. Representamen is that which represents the sign (e.g. the word “book”), and that the Representamen of the word “book” stands for an actual existing Object (i.e. a physical book) to an Interpretant, i.e., to a specific interpretation of what a book is. The interrelation is used at the representational level to examine the interplay between the visual images, the source-language verbal messages and the subtitles. The second key theory of semiotic translation is deduction and induction. Deduction “entails construction and articulation of the most feasible possibilities through the enaction within physical world contexts” (Gorlée, 1994, p. 190). This concept is used at the interactive level to examine how the visual mode in preceding or following shots functions to influence subtitle translation. The other concept is induction, which “is the product of the enaction within physical world contexts” (Gorlée, 1994, p. 190). This concept is used at the compositional level to examine impacts of the visual mode in preceding or following scenes, sequences, or stages of subtitle translation.

The integration of SFL-informed multimodality and semiotic translation constitutes the basic structure of a theoretical model as shown in Figure 1.

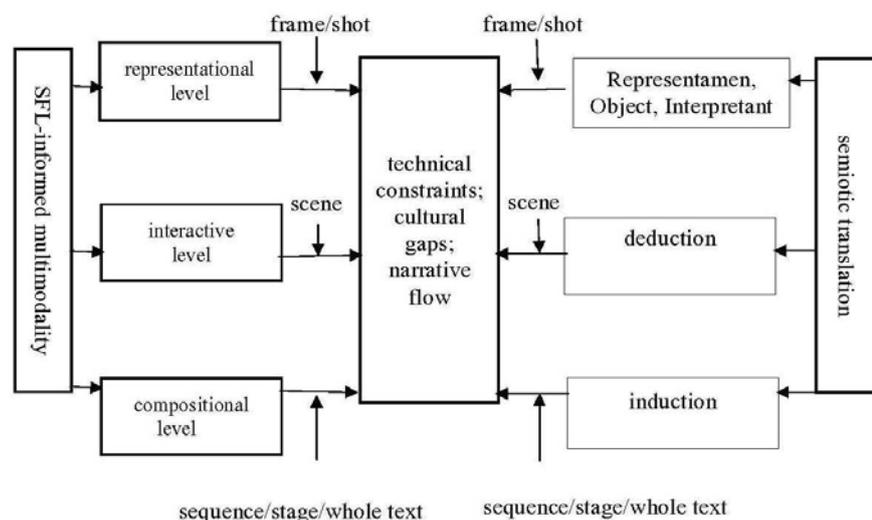


Figure 1. A theoretical model for subtitle translation

The foregoing figure illustrates that at the representational level, the semiotic interplay between verbal mode and static visual images is analysed at the unit of frame/shot by resorting to the interrelations between Representamen, Object and Interpretant. At the interactive level, the semiotic interplay is examined at the unit of scene by referring to the concept of deduction with a view to detecting the influencing power of the visual images presented in preceding or following shots. In this way, static images acquire motion, and the role of dynamic images can be revealed in subtitle translation. At the compositional level, the research unit is sequence/stage/whole text. The semiotic interplay is investigated in relation to the visual information in

preceding or following scene/sequence/stage. At each metafunctional level, the different functions of the semiotic interplay will also be investigated to explore how to address these issues in subtitle translation with the involvement of visual images.

In the following section we apply this theoretical model to a case study of illustrative examples taken from *Finding Nemo*.

4. Case study

Finding Nemo is an American animation film written and directed by Andrew Stanton, released by Walt Disney Pictures in 2003. It tells the story of an overprotective clownfish named Marlin who, accompanied by a blue tang named Dory, searches for his abducted son Nemo in a journey that takes him as far as Sydney. Along the way, Marlin learns to take risks of his own and allow Nemo take care of himself. This film was launched in mainland China on 28 July 2013 and garnered record box-office receipts, peaking at 35,530,000RMB Yuan¹. This paper analyses the subtitled version circulating in mainland China.

In this research we identified subtitles in 34 shots (i.e. to be analysed at the representational level), 35 scenes (i.e. to be analysed at the interactive level) and 10 sequences and stages (i.e. to be analysed at the compositional level) in the whole film. The key criteria for this data selection are as follows. Foremost, the subtitles to be analysed should demonstrate interplay between visual and verbal modes, while those subtitles without such semiotic interplay may be neglected for our purposes. Second, as for how to judge whether the subtitle translation indeed contains semiotic interplay between visual and verbal modes, the spoken source messages are compared with the written target subtitles to examine the similarities and differences in expressions at lexical and clausal levels. Any discrepancies between the spoken messages, the subtitles and the images are inspected for possible interrelations between the visual and verbal modes. If such interrelations between the visual images and the subtitles are confirmed for a given instance, it will be selected for further detailed analysis.

Based on the selection criteria mentioned above, subtitles in 34 shots, 35 scenes and 10 sequences and stages are identified for detailed analysis with the findings presented below.

4.1 At the representational level

The accentuated information at the representational level is content or participants, verbal or visual, in the frame/shot unit. Examination of the relationships between the verbal and visual modes allows two major categories of semiotic interplay to be identified: verbalized-visual image and non-verbalized-visual image. Literally, verbalized-visual images indicate that relevant visual images are verbalized in subtitles, while non-verbalized-visual images signify the opposite. Furthermore, based on the ideational features of identification, activity and circumstances discussed in Royce (2002, pp. 193–194), this paper examines how these three features are presented in subtitles, in order to demonstrate how different linguistic levels (i.e. lexical and clausal level) in subtitles are influenced.

- *Identification*: Who or what are the represented participants, or who or what is in the visual frame (animate or inanimate)?
- *Activity*: What processes are there, or what action is taking place between the actor(s) and the recipient(s) or object(s) of that action?

¹ Information retrieved 6 Dec. 2014 from <http://58921.com/film/854/boxoffice>.

- *Circumstances*: What are the elements that are locative (i.e. concerned with the setting), are of accompaniment (i.e. participants not involved with the action), or are of means (i.e. participants used by the actors)?

Meanwhile, the functions of the semiotic interplay in the subtitling process are also investigated. Table 1 below presents more detailed information in this regard.

Table 1. Semiotic interplay at the representational level in Finding Nemo.

Categories of semiotic interplay	Numbers of instances	Functions	Affected linguistic levels in subtitles					
			Lexical level			Clausal level		
			I*	A*	C*	I*	A*	C*
verbalized- visual image	17	to maintain the narrative flow	7	4	2	/	4	/
	2	to fill cultural gaps	1	/	/	/	/	1
non-verbalized-visual image	15	to overcome technical constraints	13	/	0	/	2	/

*Note: "I" stands for "identification"; "A" stands for "activity"; "C" stands for "circumstances"

Table 1 shows that there are 19 (i.e. 17+2) instances of verbalized-visual images and 15 non-verbalized-visual images. The former class serves to guarantee the narrative flow of the film and fill the cultural gaps between the source dialogues and the target subtitles, while the latter class substantively operates to overcome technical constraints. As for the influenced linguistic levels in subtitles, the lexical level is the most affected (i.e. with total $7+1+13+4+2=27$ instances), while only 7 (i.e. $4+2+1$) instances concern the clausal level. At lexical level, 21 instances relate to the identification of representational meaning, 4 to activity and 2 to circumstances. At clausal level, no instance relating to identification is found, while 6 concerning activities and 1 about circumstances are identified.

The following presents one instance of verbalized-image and one of non-verbalized-image.

Instance 1 (verbalized image) [Shot at 21:50]:



ST (source text): *On my honour, or may I be chopped up and made into soup.*
 TT (target text): *以我的名誉担保，如果撒谎就把我剁碎做成鱼翅汤。*

In Chinese, “鱼翅汤” (i.e. shark fin soup) specifically refers to the soup made from shark fins. Because of the visual image of these sharks, “鱼翅汤”

(i.e. shark fin soup) in the subtitle becomes a more easily acceptable version for Chinese viewers. That is to say, the expected Interpretant can be ensured through the Representamen of the Object (i.e. the visual image of the sharks). If “soup” is literally translated as “汤” (i.e. soup), the humour and ironic tone will be lost. So, the translation of “soup” into “鱼翅汤” (i.e. shark fin soup) takes Chinese viewers’ cultural background into consideration and facilitates the viewing process by integrating the visual and verbal messages in the subtitle. The visual message exerts impact on the lexical level in the subtitle, specifically on the identification meaning.

Instance 2 (non-verbalized image) [Shot commencing 23:10]:



(23:11):



ST: Come here. Group hug.

TT: 过来

In this instance, with a view to overcoming the temporal constraint of subtitle length, the translation of “Group hug” is deleted in the subtitle. There might be two reasons to justify this approach. On the one hand, “Come here” and “Group hug” are spoken by two different sharks (i.e. the former is the small blue shark and the latter is the small brown shark). The big shark with the mouth open wide is not speaking but crying. And no time is left for the translation of “Group hug” after it is spoken, because the next subtitle emerges almost at the same time as the brown shark finishes uttering “Group hug”. On the other hand, the non-translation of “Group hug” does not entail loss of any essential information due to the visual presence of the group hug of the three sharks (in the shot at 23:11). In other words, viewers can implicitly understand the expected Interpretant through the interrelations between the Representamen (i.e. “过来”) and the Object (i.e. the visually demonstrated group hug of the three sharks).

4.2 At the interactive level

The basic analytical unit at the interactive level is the scene, which includes multiple shots having continuity in time and space. The change of shots brings about edited camera movements. These edited camera movements justify the

consideration of the role of dynamic visual images in subtitle translation, in the sense that the visual depiction in preceding or following shots exerts impact on translation of the immediate subtitles. In view of naming these bi-directional influences, this paper employs Martin and Rose's (2007) concepts of anaphor (i.e. the reference that points back) and cataphor (i.e. the reference that points forward). Accordingly, two types of semiotic interplay are identified: cataphoric deduction and anaphoric deduction.

The aim of adopting these referential concepts (anaphor and cataphor) in the analysis of subtitle translation is two-fold. First, it is to demonstrate that the visual images in preceding or following shots/scenes/sequences/stages can function to influence the respective subtitles. In this way, subtitle translation is no longer an activity of transferring verbal meanings from one language to another in stills, but rather involves the visual mode across a much lengthier setting. Furthermore, it aims to provide subtitlers with a tool to cut subtitles in practice. It is known that the length of each utterance that the subtitlers hear and the duration of the accompanying shots are two decisive factors in tailoring subtitles to satisfy the unavoidable temporal and spatial constraints. However, precisely how to produce concise and faithful subtitles in a way that facilitates splicing them into the action is still an issue of predominant importance in subtitling process. In other words, the consideration of visual images can provide subtitlers with more choices when splicing subtitles in practice.

Meanwhile, the ideational features of identification, activity and circumstances of dynamic visual images will also be discussed at the interactive level, with a view to revealing how the subtitles are specifically influenced. The findings at the interactive level can be found in Table 2 below.

Table 2. Semiotic interplay at the interactive level in Finding Nemo.

Categories of semiotic interplay	Number of instances	Impacted linguistic level and semiotic function										
		Lexical level			Function	Clausal level			Function			
		I*	A*	C*		I*	A*	C*				
cataphoric deduction	15	10	6	2	2	to maintain the narrative flow	5	/	4	1	4	to maintain the narrative flow
											1	to overcome technical constraints
anaphoric deduction	20	18	16	1	1	to maintain the narrative flow	2	/	2	/	1	to maintain the narrative flow
											1	to overcome technical constraints

*Note: "I" stands for "identification"; "A" stands for "activity"; "C" stands for "circumstances"

Table 2 shows that there are 15 instances of cataphoric deduction, in which 10 relate to the lexical level in subtitles and 5 the clausal level; and 20 instances of anaphoric deduction, in which 18 concern the lexical level in subtitles and 2 the clausal level. In addition, the shared function of these semiotic interplay on lexical level is to create the strong link among shots so as to maintain the narrative flow of the film and provide an enjoyable viewing experience for viewers, while at clausal level, two kinds of functions are found, namely to maintain the narrative flow and to overcome the technical

constraints. This table also shows that the lexis with the feature of identification is still the most heavily influenced part in subtitles (i.e. 22 out of 35). The following part illustrates some instances.

Instance 3 is a visually-based cataphoric deduction. The explosion visually shown in the shot at 25:22 explains why these balloons should be “a bit dodgy”, which in turn justifies the deduction of “很容易破” (i.e. easy to explode) in the subtitle. If there was no such visual image (i.e. explosion of the naval mines), the translation of “a bit dodgy” into “很容易破” (i.e. easy to explode) would be far less justifiable. Though the image of mines shown in the shot at 20:32 might logically secure viewers’ understanding of “a bit dodgy” as “be probable to explode”, this direct expression in the subtitle “很容易破” (i.e. easy to explode) further reinforces the logical link between these two shots. This demonstrates the function of the visual image in this translation. The interrelation of these two visual images guarantees the strong link between these two shots. It means the two shots with almost 5 minutes time span are closely related in this way. Meanwhile, if “a bit dodgy” was literally translated into “要躲着点儿” (i.e. avoid being close to), the subtitle would be repetitive in expression because the former part of it is “可是要离远点儿” (i.e. keep distance). In this way, the translation of “a bit dodgy” into “很容易破” (i.e. easy to explode) becomes a better version.

Instance 3 (cataphoric deduction) [Scene at 20:32-25:22]:



©2003 Disney-Pixar

ST: *Mind your distance, though. Those balloons can be a bit dodgy.*

TT: 可是要离远点儿，这些气球很容易破。

(25:22):



©2003 Disney-Pixar

Instance 4 is an example of anaphoric deduction. In the shot at 05:47, the visual image shows that Nemo is stuck in a coral. In the shot at 05:56, Nemo’s father is concerned whether Nemo has been hurt and says “Sometimes you can’t tell ‘cause fluid rushes to the area.” He was afraid that Nemo might have been hurt already since he believes it is easy for flood to rush to Nemo’s head after his being stuck up-side-down in the coral and, thus, might make Nemo feel uncomfortable. So, thanks to the visual image present in the shot at 05:47 (i.e. Nemo got stuck in a coral up-side-down), “fluid” is justifiably

translated into “血” (i.e. blood) in the subtitle in the shot at 05:56 because the literal translation of “fluid” into “液体” (meaning fluid) does not make sense to Chinese viewers in this case. In this way, the two shots are closely linked visually and verbally.

Instance 4 (anaphoric deduction) [Scene at 05:47-05:56]:



©2003 Disney-Pixar

(5:56):



©2003 Disney-Pixar

ST: Sometimes you can't tell 'cause *fluid* rushes to the area.

TT: 有时血会冲到那儿, 有感觉吗?

Instance 5 is another example of cataphoric deduction. “Gill” is the name of the black fish. The translation of “Gill” is deleted in the subtitle in the shot at 50:38 because his visual image is present in the shot at 50:40. In doing so, no essential information will be lost. Viewers can still logically infer it is Gill that the starfish is talking to. This does not mean that express translation of “Gill” into “吉儿” (i.e. Gill) would necessarily make for ‘bad’ subtitling; rather, it illustrates how the visual and other aspects of the scene provide more strategy choices for subtitlers in practice. In this case, omission of Gill’s name does not affect the transfer of meaning. This instance differs from Instance 3 and Instance 4 in that Instance 5 involves a dialogue (i.e. between the starfish and Gill, the black fish). In this case, the camera movement is cut to shoot the interlocutors in different shots, while in Instances 3 and 4, the shot edit is driven by the logical flow of the plot.

4.3 At the compositional level

At this level, the basic analytical unit can be the sequence, the generic stage, or even the text as a whole. Compositional meaning stresses the relevance to the context. In films, the context refers to all the relevant information, verbal or visual, expressed over the whole span of the film, as opposed to the isolated information in single frames where subtitles are presented. This paper highlights the influencing power of visual messages, in the holistic film sense, on subtitle translation. At this level, the concept of induction is employed.

Instance 5 (cataphoric deduction) [Scene 50:38-50:40]



(50:40):



ST: *Gill*, this isn't a good idea.

TT: 这个主意不好。

Drawing on Martin and Rose's (2007) concepts of anaphor and cataphor, we examine two types of semiotic interplay – namely, anaphoric and cataphoric induction – with a view to detecting the influencing power of visual images in preceding or following scenes, sequences and stages. Meanwhile, the ideational features of identification, activity and circumstances of dynamic visual images will also be discussed at this level with a view to revealing how the subtitles are specifically influenced. Table 3 below provides more detailed information regarding the categorization and functions of semiotic interplay at the compositional level.

Table 3: Semiotic interplay at the compositional level in *Finding Nemo*.

Categories of semiotic interplay	Number of instances	Impacted linguistic level and semiotic function						Function
		Lexical level			Clausal level			
		I*	A*	C*	I*	A*	C*	
anaphoric induction	1	1	/	/	/	/	/	to overcome the constraints
	8	1	/	1	/	/	6	to maintain the narrative flow
cataphoric induction	1	1	/	/	/	/	/	to maintain the narrative flow

*Note: "I" stands for "identification"; "A" stands for "activity"; "C" stands for "circumstances"

As Table 3 shows, 9 instances of anaphoric induction and 1 instance of cataphoric induction were found. This tells us that the visual images in succeeding scenes, sequences and stages exert more obvious influencing power than those in preceding scenes, sequences and stages. Among the 9 instances of anaphoric induction, only 1 instance functions to overcome the

constraints, while all the remaining 8 instances function to maintain the narrative flow. Regarding the role of semiotic interplay in cataphoric induction, it is to maintain the narrative flow of the film. Table 3 also shows that the clauses in the ‘circumstances’ feature category are the most heavily-influenced linguistic part in subtitles (i.e. 6 out of 10). Three instances are presented below.

In instance 6, “Jellyman” is translated into “水母克星” (i.e. jellyfish-buster). This is a reasonable version because of the visual image demonstrated in the 48:07 shot, showing how Marlin (i.e. the clown fish) survived the jellyfish forest. This interrelation between the visual image in the shot at 48:07 explicates the translation of “Jellyman” into “水母克星” (i.e. jellyfish-buster) in the subtitle in the shot at 55:24.

Instance 6 (anaphoric induction) [sequence at 48:07-55:24]:



©2003 Disney-Pixar

(55:24):



©2003 Disney-Pixar

ST: *Oh, intro. Jellyman, offspring. Offspring, Jellyman.*

TT: 介绍一下, 水母克星, 孩子们, 水母克星。

Instance 7 relates to a sequence with a shared theme, i.e. Nemo’s first day at school. The above two shots take place in different places, i.e. one is in the school (i.e. the shot at 11:24) and the other outside the school (i.e. the shot at 13:06). The translation of “the drop off” into “悬崖” (i.e. the cliff) in the shot at 11:24 closely links to the visual image shown in the shot at 13:06. Though “drop off” can also refer to the daily ritual of dropping off kids in school, the content in several other shots in between these two shots shown above, in which Nemo’s father is shocked to death when he hears the words “the drop off” and terrified to know that his son is going to “the drop off”, narrows down the meaning of “drop off” to “cliff”. English-speaking viewers could easily adjust their mental imagery from a school “drop off” to the picture of a cliff by induction based on the visual information in the following scene. However, the literal translation of “the drop off” into “落下处” (i.e. a place to fall off) will not make any sense to Chinese viewers; by rendering it as “the cliff”, the interrelation between the verbal expression in the subtitle in the shot at 11:24 and the visual image of the cliff in the shot at 13:06 gives viewers a logical

progression to the plot and the narrative flow is maintained. Otherwise, the link between these two sequences would be severed.

Instance 7 (cataphoric induction) [Sequence at 11:24-13:06]:



(13:06):



ST: *I had a tough time with my oldest out at the drop off.*

TT: 我的老大去峭壁时, 我很担心。

Instance 8 involves two scenes: one concerning how Nemo's father Marlin searches the ocean for his son, and one about how Marlin finally finds Nemo. The subtitle in one shot of the latter scene (at 1:00:27) is produced by referring to the visual image shown in one of the shots in the former scene (at 52:23). So "with a bunch of sea turtles" is explicited into "骑着海龟" (i.e. riding the turtle) in the subtitle. If "with a bunch of sea turtles" is rendered literally as "和一群海龟一起" (i.e. with a group of turtles), then even though it still makes sense it is less specific than "骑着海龟" (i.e. riding the turtle), and for Chinese viewers at least, the narrative link between these two scenes will be not so strong.

Instance 8 (anaphoric induction) [sequence at 52:23-1:00:27]:



(1:00:27):



ST: *Now he's with a bunch of sea turtles on the E.A.C...*

TT: 现在他正在东澳流中骑着海龟

5. Discussion

This paper analyses the subtitle translation of *Finding Nemo* at three metafunctional levels. At the representational level, by focusing on the interrelations between static visual images and verbal languages in frames/shots, two types of semiotic interplay – namely, verbalized image and non-verbalized image – are identified. The study finds that instances of verbalized-visual image (i.e. 19 instances) are more than that of non-verbalized-visual image (i.e. 15 instances). These 34 shots in *Finding Nemo* are all instances of subtitle translation involving semiotic interplay between verbal and visual modes. This numerical difference (19 versus 15) supports the argument in this study that visual information tends more toward being verbalized rather than non-verbalized in the subtitles in *Finding Nemo*. This might be contradictory to the research findings of some previous studies (e.g. Gottlieb, 1992; Georgakopoulou, 2009), which argue that there is no need to transfer visual messages into verbal ones in subtitles with the aim of overcoming technical constraints in subtitles. What we have found here, however, verifies that the visual mode can not only function to compress or reduce subtitles (e.g. Instance 2), but can also be verbalized to bridge cultural gaps (e.g. Instance 1). It appears that when there is occasion to fill cultural gaps in subtitles, verbalized-visual image would be the choice, while for overcoming technical constraints, non-verbalized-visual images are employed. Furthermore, since the representational meaning focuses on the content or the participants (static images), lexis is the most influenced linguistic unit in the subtitles, with the most heavily impacted item being expressions concerning identification (21 of 34 instances). This, on the other hand, shows that the visual mode representing who or what should be mostly referred to might be of great utility in subtitle translation performed at the frame/shot unit level.

Representational meaning is the basic and the most fundamental implication that viewers can obtain from the pictorial elements presented in audiovisual texts. However, since audiovisual texts, such as films, comprise both static pictures and moving images, this relates to semiotic interplay at the interactive level.

At the interactive level, 15 instances of cataphoric deduction and 20 instances of anaphoric deduction were found. The fact that more anaphoric deduction was identified might be attributable to audiences more naturally recalling the visual images in preceding shots than anticipating those in succeeding shots. This finding also highlights the function of dynamic images in subtitle translation, whose role, however, has not been given its deserved attention in previous studies. Given that visual images in the same scenes enjoy temporal and spatial continuity, even when they are not presented in the

same shots, the intrinsic link between these images can still be used to benefit the subtitle translation. And the subtitles, produced with the consideration to dynamic images, can create a logical structure for film viewers and facilitate the viewing process.

In addition, lexis featuring identification is the more heavily influenced linguistic unit in subtitles (6 instances of cataphoric deduction and 16 of anaphoric deduction; please refer to Table 2). This finding sheds light on the argument that within the filmic unit of the scene, the represented participants in frames or shots are the most active elements that influence subtitle translation. In other words, who or what is the primary subject of visual frames is still the most valuable concern in the subtitling process, while the visual images relating to what takes place and what is the setting of the event do not exert great impact on the subtitling. This demonstrates that the interactions between the participants play a more obvious role in subtitling process than the interactions between the viewers and the visual world. The active involvement of who or what in subtitle translation points to the important function of the former type of interaction, while the less active involvement of activity and circumstances indicates that interactions between viewers and the visual mode are not so frequently considered in subtitling process.

Furthermore, at the interactive level, this study also found that the visual images are mostly used to explicitate the verbal source messages via cataphoric deduction (11 of 15 instances), while the visual images are employed to either replace the verbal source messages (11 of 20 instances) or explicitate the verbal messages (9 of 20 instances) via anaphoric deduction. This demonstrates that the visual images in preceding and succeeding shots exert different impacts on the verbal messages in subtitles. This finding further complements the relevant findings in previous studies on the interrelations between verbal and visual texts. One of the most influencing categories is offered by Barthes (1977). It is proposed that there are four types of relations between verbal words and visual information, namely, “anchorage” (verbal words further specify visual information), “relay” (verbal texts and visual images stand in a complementary relationship), “redundancy” (the visual and verbal communicate the same information) and “substitution” (the linguistic code has been replaced by a kinetic sign). The finding in this paper complements Barthes’ classification in the following three ways. First, Barthes’ (1977) categorisation is based on the analysis of print texts, so the interrelations between verbal and visual texts are also print-text based. Thus, Barthes did not address the function of dynamic images, such as those in preceding and succeeding film shots, and much less how such visual images function in subtitle translation. Second, Barthes’ (1977) proposed assortment only concerns the effects that verbal texts can have on visual information, but not vice versa. The finding in this paper fills this gap. Third, Barthes’ (1977) classification is entirely unrelated to translation studies, so it only contemplates a bilateral relation between visual and verbal texts, while the finding in this paper complements the interrelations between verbal and visual modes by expanding this relation to a trilateral level: source verbal messages, visual messages, and target verbal messages (i.e. subtitles).

Regarding how semiotic interplay functions to address the three key concerns in subtitle translation, this study has found that at the interactive level the influenced lexis in subtitles mainly functions to maintain the narrative flow of the film by strengthening the links between shots, while the impacted clauses in subtitles function not only to maintain the narration, but also to overcome spatial and temporal constraints. It is interesting to notice that no instances relating to the third key concern, i.e. to bridge cultural gaps,

were identified. This might be due to the fact that in *Finding Nemo*, there are not many culture specific information conveyed in the verbal and visual modes entailing cultural differences between Western and Chinese cultures. As for the visual mode, while previous studies (e.g. Kovačič, 1996; Georgakopoulou, 2009) have discussed how to use visual images – especially those in preceding or succeeding filmic units – its role in overcoming the technical constraints in subtitling is still an under-researched territory. To that end, this paper proposes that the visual images in both preceding and succeeding shots can influence the subtitles in the immediate text. Meanwhile, no relevant research has examined the function of maintaining cinematic narrative flow in terms of subtitle translation. This paper finds that subtitles can function to maintain the narrative flow of the film by relating visual information to verbal texts.

At the compositional level, less instances of semiotic interplay (10 in total) are identified than that at the representational (34 in total) and interactive levels (35 in total). This shows that the visual messages in preceding and succeeding scenes, sequences and stages exert less impact on subtitle translation than those in the immediate frames/shots (i.e. at the representational level) and in preceding and following shots (i.e. at the interactive level). In these 10 instances, only 1 instance concerns cataphoric induction. This demonstrates that visual images presented in advance provide viewers with more inferences or insights to refer to the subtitle translation in later sequences or stages. Additionally, these inferences are revealed more obviously at clausal level than at lexical level, as witnessed by the finding that more clauses (6 of 10) are influenced in subtitles at the compositional level than in subtitles at the representational level (7 of 34) and the interactive level (7 of 35). With regard to who or what is usually linguistically presented at the lexical level, we observe that although activity and circumstances are expressed more frequently at clausal level, at the compositional level there is more influence on clauses than on lexis .

Furthermore, it is also found that the visual mode is used almost equally to replace (4 of 10) or explicitate (6 of 10) verbal messages at the compositional level. This differs to the findings at the representational and interactive level, i.e. the visual mode is mostly found to explicitate visual messages. The decisive factor lies in the function of the semiotic interplay: when there is temporal or spatial constraint to overcome, the visual images are used to support the verbal messages; otherwise, the visual images are employed to explicitate the verbal messages.

6. Conclusion

This paper integrates SFL-informed multimodal analysis and the theory of semiotic translation in its examination of the semiotic interplay between verbal language and dynamic visual images. The paper explores the role of visual images in helping to shape subtitle translation, whilst obeying technical constraints, filling cultural gaps and forging narrative flow in films. It is found that subtitle translation is never confined to static frames or shots, but involves much larger filmic units, such as scenes, sequences, stages and even the whole text. The theoretical model constructed in this paper applies semiotic theories to the analysis of subtitle translation by highlighting the roles of both static and dynamic images in the subtitling process. In so doing, it affirms that subtitle translation involves the holistic analysis of audiovisual texts rather than the mechanical combination of isolated stills.

References

- Barthes, R. (1977). *Image, music, text*. London: Fontana Paperbacks.
- Chuang, Y. T. (2006). Studying subtitle translation from a multi-modal approach. *Babel*, 52(4), 372-383.
- Delabastita, D. (1989). Translation and mass-communication: Film and TV translation as evidence of cultural dynamics. *Babel*, 35(4), 193-218.
- Díaz-Cintas, J. & Remael, A. (2007). *Audiovisual translation: Subtitling*. Manchester and Kinderhook: St. Jerome Publishing.
- Georgakopoulou, P. (2009). Subtitling for the DVD industry. In J. Díaz-Cintas & G. Anderman (Eds.), *Audiovisual translation: Language transfer on screen* (pp. 21-35). Basingstoke and New York: Palgrave Macmillan.
- Giovanni, E. D. (2003). Cultural otherness and global communication in Walt Disney films at the turn of the century. *The Translator*, 9(2), 207-223.
- Gorlée, D. L. (1994). *Semiotics and the problem of translation: With special reference to the semiotics of Charles S. Peirce*. Amsterdam: Rodopi.
- Gottlieb, H. (1992). Subtitling. A new university discipline. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting. Training, talent, experience* (pp. 161-172). Amsterdam: John Benjamins.
- Iedema, R. (2001). Analysing film and television: a social semiotic account of hospital: an unhealthy business. In T. v. Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 183-206). London and New Delhi: Sage.
- Kovačič, I. (1996). Subtitling strategies: A flexible hierarchy of priorities. In C. Heiss & R. M. B. Bosinelli (Eds.), *Traduzione multimediale per il cinema, la televisione e la scena* (pp. 297-305). Bologna: Clueb.
- Kress, G., & van Leeuwen, T. (2006) *Reading images: The grammar of visual design* (2nd ed.). London: Routledge.
- Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause* (2nd ed). London: Continuum.
- Oumano, E. (1985). *Film forum: Thirty-five top filmmakers discuss their craft*. New York: St. Martin's Press.
- Perego, E. (2009). The codification of nonverbal information in subtitled texts. In J. Díaz-Cintas (Ed.), *New trends in audiovisual translation* (pp. 58-69). Bristol, Buffalo and Toronto: Multilingual Matters.
- Pérez-González, L. (2014). *Audiovisual translation: Theories, methods and issues*. London and New York: Routledge.
- Pettit, Z. (2009). Connecting cultures: Cultural transfer in subtitling and dubbing. In J. Díaz-Cintas (Ed.), *New trends in audiovisual translation* (pp. 44-57). Bristol, Buffalo and Toronto: Multilingual Matters.
- Remael, A. (2003). Mainstream narrative film dialogue and subtitling: A case study of Mike Leigh's 'Secrets & Lies'. *The Translator*, 9(2), 225-247.
- Royce, T. D. (2002). Multimodality in the TESOL classroom: Exploring visual-verbal synergy. *TESOL Quarterly*, 36(2), 191-205.