

Nana Zhou*

Competition among visual, verbal, and auditory modalities: a socio-semiotic perspective

<https://doi.org/10.1515/sem-2022-0060>

Received June 3, 2022; accepted April 16, 2024; published online June 13, 2024

Abstract: This article presents a fresh perspective on the interplay among visual, verbal, and auditory modalities, positing that these modalities, as semogenic resources, compete to express dynamic meanings. The theoretical paradigm emphasizes that whether a modality or an element within a modality gets or loses semantic status, it will elicit an additional layer of social meaning to depict a comprehensive picture of a story together with an explicit semiotic meaning. The article adopts a qualitative method to analyze the data, which are drawn from *The Good Wife* and *My Roommate is a Gumiho* and annotated in ELAN 6.3. It was found that modal competition can shed light on the dynamic meaning-making processes in semiotic and societal orientations. Modal competition may distort space and time of different stories, and reconstruct a different discursive spatio-temporal dimension in the TV world. It can diversify the dynamic orientations from New to Given in visual, verbal, and auditory texts of multimodal discourses to tell stories. Modal competition provides a lens to understand the multidimensional reality and to appreciate the aesthetics of a modern TV series.

Keywords: modal competition; socio-semiotic perspective; modern TV series

1 Introduction

After multimodal discourse analysis was initially put forward (Kress and van Leeuwen 1996; O'Halloran 1999), a fundamental issue is to reconsider the statuses of different modalities as the semogenic resources in meaning-making processes. The default condition is that all modalities have an equal stance in making meanings to produce discourses in context (Kress 2010; Norris 2019). This viewpoint was coined “modal democracy” (Krug and Frenk 2006), which meant the “democratic stance that all modalities are equal” (Page 2010: 4). Many scholars have explored how different

*Corresponding author: Nana Zhou, Yangzhou University, Yangzhou, China,
E-mail: zhounazi@163.com. <https://orcid.org/0000-0001-9566-2768>

modalities interact with each other to construct multimodal discourses, and have developed terms such as intersemiotic complementarity (Royce 1998, 2007), intersemiosis (Lim-Fei 2004), and co-contextualization (Thibault 2004). It was argued that when different modalities synergize to produce multimodal discourses, they might play equal, superordinate or subordinate roles in making discursive meanings (Norris 2004). These studies are oriented towards discourse-semantics in different genres.

This article focuses on the uneven distributions of visual, verbal, and auditory modalities and their elements in the dynamic meaning-making processes in modern TV series. To respond to the phenomenon, I put forward modal competition to analyze the interplay among these modalities from a socio-semiotic perspective. The metaphorical description underlines the asymmetry of the modalities in making meanings in a modern TV world where visual, verbal, and auditory modalities do not just complement each other, but also compete to get the semantic statuses of instantiating their semogenic resources to tell a story. To be specific, a modality instantiates its semogenic resource to tell a story if it gains a semantic status in modal competition. Otherwise, it does not participate in the meaning-making process. No matter whether a modality gains or loses semantic status, it will elicit an additional layer of social meaning. It is motivated by the narrative, determined by the TV series author,¹ but oriented to the audience.

The article aims to address three issues: (1) how intermodal competition sheds light on the uneven distributions of visual, verbal, and auditory modalities in the dynamic meaning-making processes in the semiotic and societal orientations in a modern TV series; (2) how intramodal competition explicates the interaction among multiple elements within a modality to tell a story in different aspects in the semiotic and societal orientations in a modern TV series; (3) how modal competition distorts time and space of two stories, and reorganizes New-Given information in the TV world. The data are drawn from modern TV series (*The Good Wife* and *My Roommate is a Gumiho*) which are taken as “at least a first step to understanding semiosis in everyday life” (O’Halloran 2004: 4), and annotated in ELAN 6.3 as “a free, multimodal annotation tool” (Lausberg and Sloetjes 2009: 841). Modal competition responds to the tension of the dynamic meaning-making processes when different modalities instantiate their semogenic resources to tell stories in modern TV series.

¹ The TV series author is borrowed from the film author as a collective term referring to “the author of the script, the producer, actors and actresses, photographers, etc.” (Lothe 2000: 31) in the film-making. Here the term refers to those who determine how to produce TV series.

2 Paradigm of analyzing modal competition

The article proposes that visual, verbal, and auditory modalities compete to get the semantic statuses of instantiating their semogenic resources to make meanings in a story-telling activity in modern TV series. Some get the semantic statuses of instantiating the semogenic resources to tell the same story in different dimensions in the semiotic orientation while others lose in modal competition. In most cases, all gain the semantic statuses, and reach a temporary balance to tell a story. But either gaining and losing a semantic status will elicit an additional layer of social meaning that contributes to a comprehensive picture of a story together with an explicit semiotic meaning (see Figure 1). The competition among verbal, visual, and auditory modalities is oriented towards dynamic meaning-making processes instead of static and abstract inter-semiotic relations.

Apart from the competition among different modalities, the elements within a modality compete to get the semantic statuses to tell a story in different aspects. Intramodal competition is more prevailing and flexible than intermodal competition, since multiple elements within a modality have the potential to participate in dynamic meaning-making processes, either foregrounding or backgrounding some aspects of a story in the modern TV world. Intramodal competition may yield two results: one is that all the elements within a modality gain their semantic statuses to tell a story in different aspects; the other is that some lose their semantic statuses driven by the strong motivation of generating additional layers of social meanings. Even though some elements of a modality lose their semantic statuses, the modality does not lose its semantic status since other elements of the same modality contribute to telling the story in other aspects. As the plot moves on, other new elements enter intramodal competition and gain the semantic statuses of instantiating the semogenic resource to show how the story develops. It happens that some regain the semantic statuses to tell the story after they lose in the previous intramodal competition. The modern TV series, as a specific genre, can actualize the loss and the regain of the elements in the TV world filled with color, noise, and words by utilizing some video production techniques.

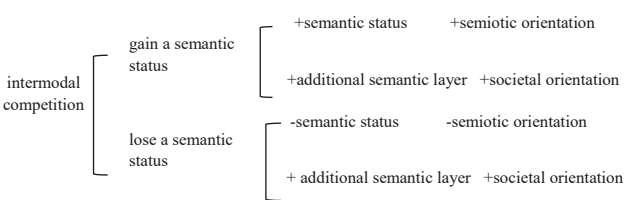


Figure 1: The paradigm of analyzing intermodal competition.

Different from the real world in a unified spatial-temporal dimension, the modern TV series can create an artistic world in which space and time may be warped to achieve a specific narrative purpose. Modal competition can shed light on such phenomenon: visual, verbal, and auditory modalities compete to get the semantic statuses to reconstruct different but relevant stories in receiving time and receiving place in the TV world, which diverges from the original spatio-temporal dimensions of the stories. The conflict between the occurrence of the stories in the material world and their reproduction in the TV world makes the narrative more attractive and aesthetic. Besides, modal competition can diversify New-Given structures when different modalities participate in the meaning-making processes with their unique principles of organizing information.

3 Competition among different modalities

As argued, visual, verbal, and auditory modalities compete to get the semantic statuses of instantiating the semogenic resources in the story-telling activities in the modern TV world under the heading of intermodal competition.

Take *The Good Wife*, for example. It tells the story of Jeff, an independent contractor in the US National Security Agency, who is getting a weekly lie-detector test at the beginning of Episode 18, Season 5. Visual, verbal, and auditory modalities participate in the meaning-making process to tell the story in different aspects: (1) the verbal expressions in the weekly lie-detector test (verbal text); (2) the visual image of Jeff on the screen after 8 s (visual text); (3) the voices of Jeff and the examiner, and the sounds of pressing a key on the keyboard after Jeff answers each question (auditory text). However, these three modalities do not get their semantic statuses simultaneously. Only verbal and auditory modalities initially gain the semantic statuses to indicate what is happening in the first 8 s in intermodal competition. No visual images depict the story except the black screen. Afterwards, the visual modality enters intermodal competition and gains the semantic status to show Jeff on the screen. Then, visual, verbal, and auditory modalities reach a temporary balance to tell the audience that Jeff is getting a polygraph test in different dimensions. The competitions among these three modalities will be illustrated in two phases as follows.

In Phase I (from 00:00:00.000 to 00:00:08.000), language gets the semantic status of verbalizing the lie-detector test in the semiotic orientation; for example, “Is your first name Jeff? Yes.” This direct speech between Jeff and the examiner gains the semantic status in intermodal competition, creating an authentic atmosphere to resonate with the audience in the societal orientation. Meanwhile, the auditory modality wins the semantic status of instantiating its semogenic resource to tell

the same story. The voices of Jeff and the examiner identify two speakers in the polygraph test in the semiotic orientation. The sequential order of their voices implies their unequal positions in the workplace in their societal orientation. The sounds of pressing a key on the keyboard indicate that a polygraph instrument is used in the test in the semiotic orientation. When the sound of pressing a key on the keyboard gains the semantic status, it signals the end of a question and the start of another one in the test in the societal orientation. As far as visual modality is concerned, no visual image shows what is happening except the black screen. The loss of visual images implicates the confidentiality of the lie-detector test in the experiential meaning, and is expected to arouse the audience's curiosity about the background of the story in the interpersonal meaning in the societal orientation. The relevant data are annotated in ELAN 6.3 to show the competitive results among verbal, visual, and auditory modalities in this story in Figure 2.

The loss of visual images is not common in a modern TV series that is produced to entertain the audience via dynamic colorful images. To fill the visual vacuum, the audience must make full use of the verbal expressions and the auditory elements to piece together a story with the additional layer of meaning that the loss of visual images elicits. And as the image of Jeff appears on the screen after 8 s, the audience will be clear about what Jeff looks like and where the story occurs.

In Phase II (from 00:00:08.000 to 00:00:10.300), verbal and auditory modalities maintain the semantic statuses of instantiating their semogenic resources to indicate that the weekly polygraph test continues. The visual image of Jeff enters the

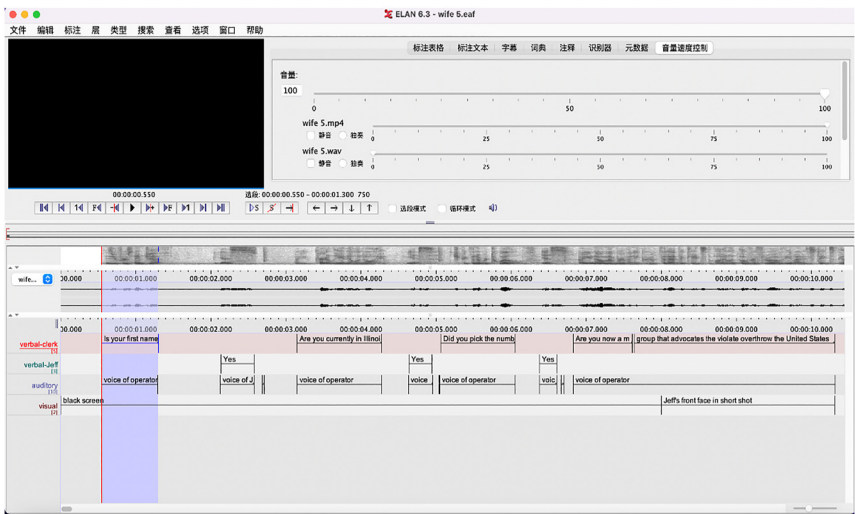


Figure 2: The competitive results among verbal, visual, and auditory modalities in Phase I.

intermodal competition and wins the semantic status of instantiating the visual resource to show what he looks like and how he behaves in the semiotic orientation. That the visual image gains the semantic status is strongly driven by the fact that Jeff is not adequately described by the verbal or auditory elements. When the visual image appears, it establishes the sociosemiotic relations with the word *Jeff* and his voice quality, which inform the audience who Jeff is. Besides, the information structure is prominent in the image: the image of Jeff, as new information, appears in the center of the screen, which does not comply with the principle that new visual information appears on the right, as discussed in an advertisement (Feng 2011). Visual, verbal, and auditory modalities reach a temporary balance in a clear image to show that Jeff is getting a polygraph test in Phase II (see Figure 3).

To sum up, intermodal competition is process-oriented, in which verbal, visual, and auditory modalities gain or lose the semantic statuses of instantiating their semogenic resources to tell a story in the semiotic orientation in a modern TV series. Gaining and losing semantic status can generate additional layers of meanings in the societal orientation that unfold in a more comprehensive picture of what is happening. The competitive result will change when a new modality enters the competition or when a modality withdraws from intermodal competition. The dynamic of intermodal competition manifests the aesthetics of visual, verbal, and auditory modalities in telling a story in the TV world.

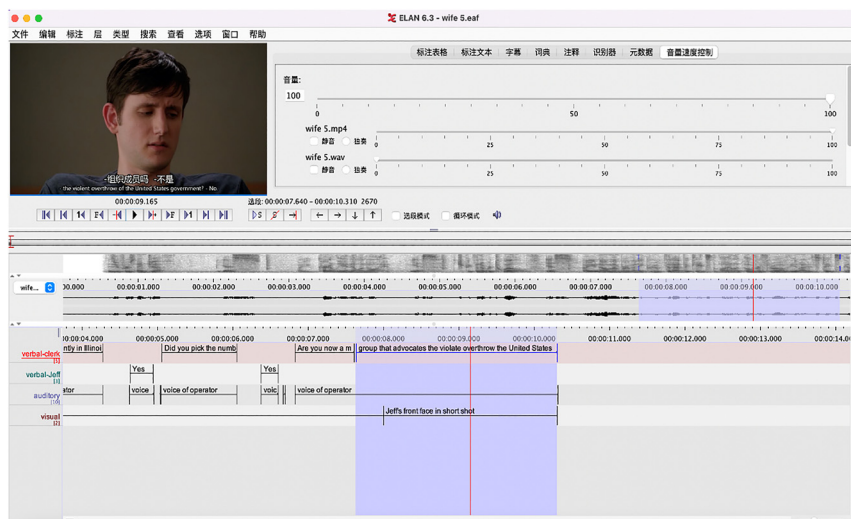


Figure 3: The temporary balance among visual, verbal, and auditory modalities in Phase II.

4 Competition among different elements within a modality

Multiple elements within a modality compete to get the semantic statuses to tell a story in different aspects in the modern TV world under the heading of intramodal competition.

Take *My Roommate is a Gumiho* (Episode 3) for example. The South Korean TV series tells of a romance between a 24-year-old female college student, Dam Lee, and a male *gumiho* (a nine-tailed fox in Korean mythology) named Wuyeo Shin. At the end of this episode, Dam runs back home after getting off the bus on a rainy night because she is not carrying an umbrella. Then she rushes into an umbrella accidentally, and finds that it is her roommate Wuyeo who is worried about her and goes out to look for her as it rains heavily.

More than one auditory element is involved in the scenario, including Dam's sigh, puffing, and clattering as she runs, the sounds of raindrops falling on the ground and the umbrella, the background song 우연이 아닌것만 같아서 ('It doesn't seem to be an accident') and the voices of Dam and Wuyeo. These auditory elements gradually participate in the auditory competition to describe the romantic encounter on the street on a rainy night. The auditory competition yields five results as follows: a temporary balance with no auditory loss, the loss of all the auditory elements, a triumph of the background song, another temporary balance with no auditory loss, and then the single auditory element (the continuance of the background song). They will be specified in five phases.

In the first phase, Dam decides to run back home and then rushes into an umbrella by accident on a rainy night. Before Dam runs, she sighs, which is introduced as a new auditory element. That the sigh gains the semantic status implies that she is depressed at the awful situation in the societal orientation. As Dam runs, the sounds of raindrops falling on the ground, puffing, and clattering enter the auditory competition. They cannot beat each other, but reach a temporary balance to indicate that Dam is running on the street in the rain in the semiotic orientation. These auditory elements create a depressing but hurried atmosphere in the societal orientation that stands in sharp contrast to Dam's happiness when she meets Wuyeo later. However, when she runs, the sigh withdraws from the auditory competition, which implies that Dam accepts the terrible situation in the societal orientation. Then the sounds of raindrops falling on an umbrella participate in the auditory competition, and gain the semantic status to indicate the weather condition in the semiotic orientation. That the sounds of raindrops on an umbrella gain semantic statuses implies that Wuyeo comes in the societal orientation. The dynamic of auditory competitions shows a noisy but real world that is filled with multiple sounds (see Figure 4).

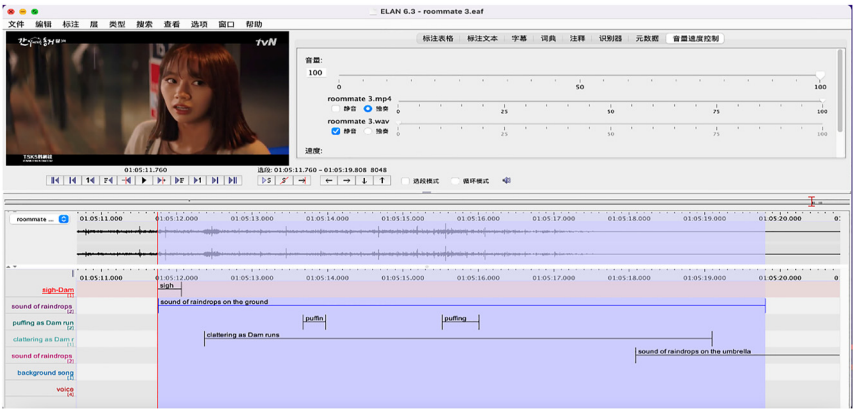


Figure 4: The competitive result of the auditory elements in Phase I.

In the second phase, Dam stops running and Wuyeo shows up under the umbrella. All the auditory elements are muted and lose the semantic statuses of instantiating the auditory semogenic resource to tell the romantic encounter between Dam and Wuyeo on the street on a rainy night (see Figure 5). The loss of all the auditory elements symbolizes Dam's surprise as Wuyeo suddenly appears in front of her with an umbrella in the societal orientation. It cannot happen that all the sounds from the external world are suddenly muted in reality. But it is

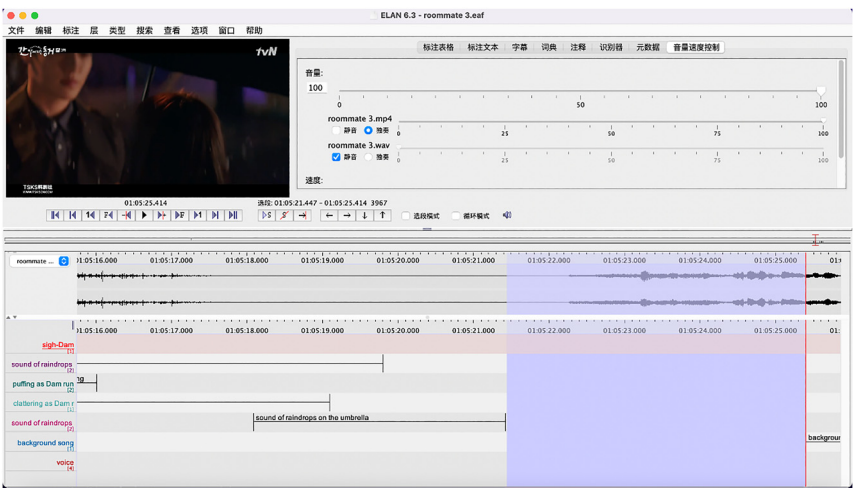


Figure 5: The auditory vacuum in Phase II.

possible to create an artistic world such as the modern TV series and the films in which a multitude of video production techniques are utilized to achieve this specific artistic effect. The loss of all the auditory elements does not last long since the TV series author expects to tell a love story in the real world. The generic feature of modern TV series does not allow all the sounds to be set on mute.

In the third phase, Dam is surprised and smiles at Wuyeo. In this phase, only the background song 우연이 아닌것만 같아서 (‘It doesn’t seem to be an accident’) fills the auditory vacuum, and gains the single semantic status to beautify the scenario in the semiotic orientation (see Figure 6). The song does not come from the external world, but is produced as nondiegetic music that cannot be heard by the characters in the TV series (Giannetti 2014). Its gain of the semantic status creates a romantic atmosphere with “a touch of emotive color” (Kress and van Leeuwen 2001: 20) in the societal orientation. The single auditory element is expected to resonate with the audience by means of the tender and romantic rhythm and lyrics. It can be recognized as the auralization of Dam’s surprise and happiness at the given moment when she meets Wuyeo on the exhausting rainy night.

In the fourth phase, Dam asks why Wuyeo is here and Wuyeo asks Dam to answer his call even though he is nearby. The sounds of raindrops falling on the ground and the umbrella regain the semantic statuses in the auditory competition so as to inform that it is still raining when Dam and Wuyeo meet in the semiotic orientation. Their regaining of sematic statuses indicate that Dam recovers from the surprise of meeting Wuyeo in the societal orientation. The background song

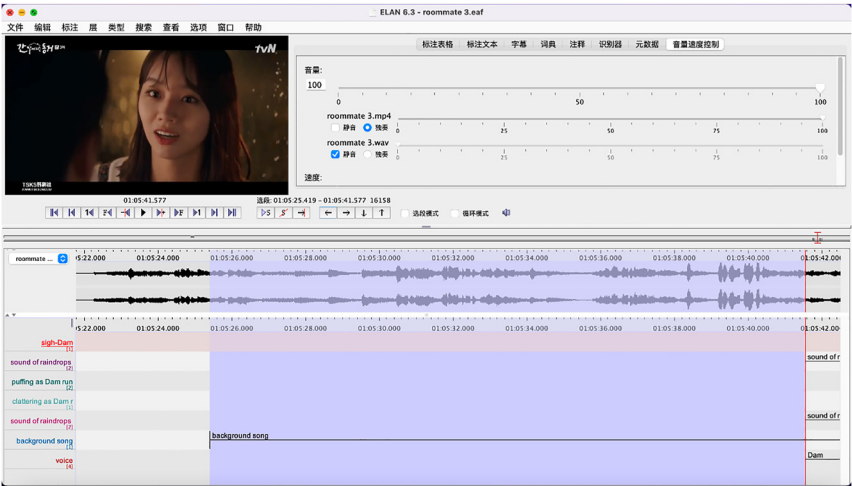


Figure 6: The auditory competitive result in Phase III.

maintains its semantic status to imply that her feeling of happiness does not go away in the societal orientation. These auditory elements reach a temporary balance in this intramodal competition, and create an auditory world filled with the noises from the external world and the background song of making a tender and romantic atmosphere (see Figure 7). The competition between the background song and these auditory elements from the external world implies the narrative conflict between the reality and the artistic world in the narrative TV series. Besides, when Dam talks with Wuyeo, their voices gain the semantic statuses to confirm their identities in the semiotic orientation. As their voices gain the semantic statuses, it signals the moment that Dam and Wuyeo realize that they have special feelings for each other in the societal orientation, which is expected to impart to the audience in direct speech.

In the final phase, Dam and Wuyeo look at each other. The background song maintains its semantic status to beautify the romantic encounter, but the other auditory elements lose in the auditory competition again (see Figure 8). The song gets the single semantic status for 25 s without any interference from the auditory elements in the external world, which implicates that Dam and Wuyeo cannot perceive anybody or anything but each other in the societal orientation. The loss of the auditory elements from the external world recreates the emotional space between Dam and Wuyeo to resonate with the audience.

In sum, the auditory modality plays the important role of telling what happens in the TV world, and implying how Dam and Wuyeo feel at the end of the episode. The auditory elements compete to gain the semantic statuses of instantiating the auditory

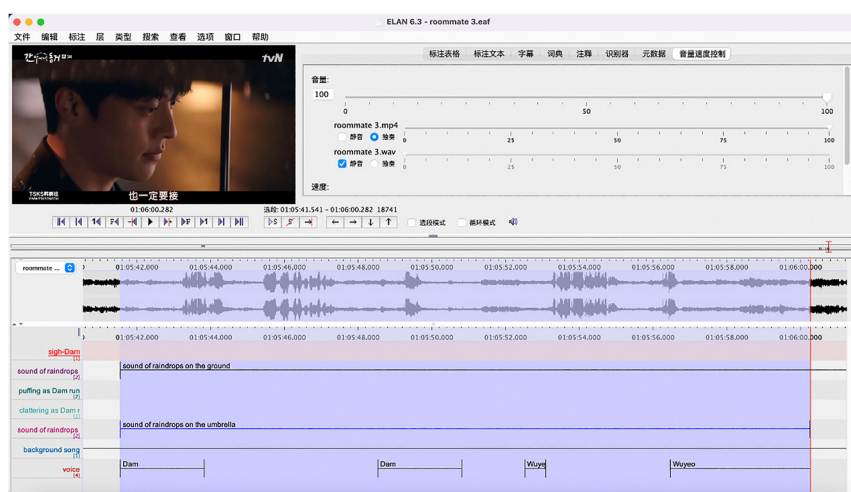


Figure 7: The competitive result of the auditory elements in Phase IV.

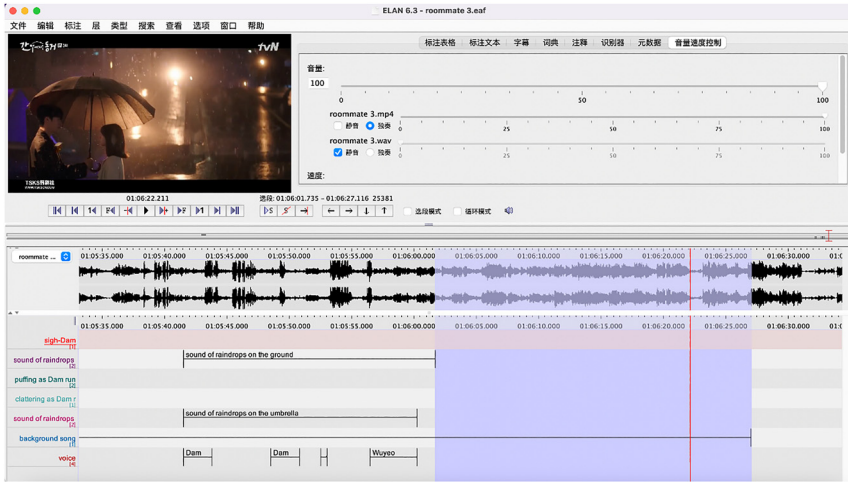


Figure 8: The competitive result of the auditory elements in Phase V.

semogenic resource to tell the story in different aspects in the semiotic orientation. Whether an auditory element wins or loses the semantic status in the auditory competition, it will render an additional layer of meaning in the societal orientation, which depicts a comprehensive picture of the story. In this example, the more sounds from the external world gain the semantic statuses, the more authentic an atmosphere will be created to arouse the audience's interests in the story; the more auditory elements from the external world are muted, the more likely the audience will be invited to the characters' internal worlds. When all the auditory elements lose the semantic statuses, it symbolizes the auralization of a character's emotional change. As the background music gets the single semantic status to beautify the scenario, it creates an emotional space where the TV series author invites the audience to go. And the integration of the background music and the sounds from the external world creates an illusion of living in between dreams and reality. The diversified organizations of these auditory elements are genre-oriented, since modern TV series allows the artistic reproduction of auditory elements to tell a story by means of the video production techniques.

In addition, these auditory elements as a whole compete with the visual images to get the semantic statuses to tell the audience what happens. They do not beat each other, but reach an intermodal balance to tell this story in different dimensions. Take Phase III for example. The background song *우연이 아닌것만 같아서* ('It doesn't seem to be an accident') gets the semantic status to beautify the encounter between Dam and Wuyeo. And Dam's facial expression changes from

surprise to smile when she finds that it is Wuyeo who holds the umbrella for her on the rainy night. The TV series author invites the audience to see what happens to Dam and Wuyeo in visual images and to perceive the protagonists' emotional change in the background song and their facial expressions. As for the verbal modality, it does not get the semantic status in Phase III. It is not until Phase IV that language gets the semantic status to verbalize Dam's doubt about why Wuyeo is here and Wuyeo's demand that Dam should answer his call even though he is nearby. The dialogue implies that Wuyeo has a strong affection for Dam and cares about her. In the story, visual, auditory, and verbal modalities compete to get the semantic statuses to tell the audience the romantic encounter between Dam and Wuyeo, and eventually arouse their interest in this TV series. However, these three modalities do not get the equal chance of telling the story. And some lose the semantic statuses; for example, the loss of auditory modality in Phase II, and the loss of verbal modality in Phases I, II, III and V, to achieve specific purposes. The competitive results in the five phases are listed in Table 1.

Table 1: The competitive results of visual, auditory, and verbal modalities in this scenario.

Modality	Phase I	Phase II	Phase III	Phase IV	Phase V
Visual	✓	✓	✓	✓	✓
Auditory	✓		✓	✓	✓
Verbal				✓	

5 Modal competition for spatio-temporal reconstruction and diverse information structures

Modal competition can shed light on the spatio-temporal differences of the stories and the diversification of New-Given information structures in the TV world in which the TV series author reproduces the stories in an artistic manner.

5.1 Modal competition for semantic status

Take *The Good Wife* (Episode 4, Season 2) for example. It simultaneously tells two stories at the beginning of this episode: one is that Alicia, a lawyer at the Lockhart

Table 2: The division of labor of visual, auditory, and verbal modalities.

	Visual	Auditory	Verbal
Story one – deposition	– Semantic status	+ Semantic status	+ Semantic status
Story two – delivery of an envelope	+ Semantic status	+ Semantic status	– Semantic status

& Gardner law firm, makes a deposition to State Attorney Childs about the Northbrook killings; the other is that a yellow envelope containing the tape of this deposition is secretly delivered to a woman journalist at the *Chicago Sun-Times*. Though visual, verbal, and auditory modalities all gain the semantic statuses, their labor is divided to tell two stories (see Table 2).

In the story of the deposition, the verbal expressions together with some auditory elements gain the semantic statuses to tell how the deposition of States of Attorney Childs is made by the lawyer Alicia from Lockhart & Gardner. Direct speech verbalizes how the deposition is conducted in the semiotic orientation. The voices of Alicia and Childs confirm their identities in the deposition in the semiotic orientation. Verbal and auditory modalities reach a temporary balance to create an authentic atmosphere for the audience with high fidelity in the societal orientation. In contrast, the visual modality is forced to withdraw from the intermodal competition, which, to some extent, arouses the audience’s curiosity about the setting of the deposition, and leaves the visual room for the delivery of the envelope in story two.

In the story of delivering an envelope, the visual images gain the semantic statuses to show how a yellow envelope containing the tape of the deposition is delivered to a woman journalist at the *Chicago Sun-Times* by a woman and a young man in the semiotic orientation: a woman walks with a yellow envelope in her left hand on the street; she puts it on the outdoor dining-table; a young man picks it up and goes skateboarding along the street; then he enters the office of *Chicago Sun-Times* and puts it on the office desk where a woman is sitting; afterwards, the woman opens the envelope and takes a tape out; she puts the tape in a cassette and plays it; she turns to the cassette and then is lost in thought. At the same time, some auditory elements get the semantic statuses to indicate the delivery of the envelope in the semiotic orientation, for instance, the clack of high heels on the street, the rustle of the envelope put on the outdoor dining-table, the friction noise of going skateboarding on the street, the rustles of the envelope put on the office desk, fetched, and opened, the rustle of the empty envelope put aside, and the clicks when the woman puts the tape in the cassette and plays it. These auditory elements inform the audience of what is happening in the dynamic and noisy world. The visual

images and the auditory elements show how an envelope is delivered in the semiotic orientation and imply how the news of the deposition is leaked in the societal one. But language withdraws from this story-telling process and leaves room for the verbalization of the deposition in story one.

There is a remarkable discursive feature in this example. Different from so-called synergy or intersemiotic complementarity (Royce 2007), there is no corresponding relationship between the visual images and the verbal expressions. The non-synergy stems from the intermodal competition in which visual and verbal modalities get the separate semantic statuses to tell two different stories. The verbal expressions verbalize the deposition of Childs by Alicia while the images visualize the delivery of an envelope by different persons. Though these two stories occur in different times and places in the material world, they are synchronized at the beginning of this episode, which will be further explicated in Section 5.2. The auditory elements are split into two parts, which gain the semantic statuses to tell these two stories simultaneously. The division of auditory labor creates an effect of intertwining two worlds where these two stories occur. But the auditory modality as a whole smooths out the narrative conflict between visual images and verbal expressions by participating in two story-telling activities. The dual narrative implies that these two stories are related to each other. Besides, the background music incorporates the rhythm to the movement of the plot (Norris 2004), which creates a nervous atmosphere in the scenario. The relevant data are annotated in ELAN 6.3 in Figure 9.

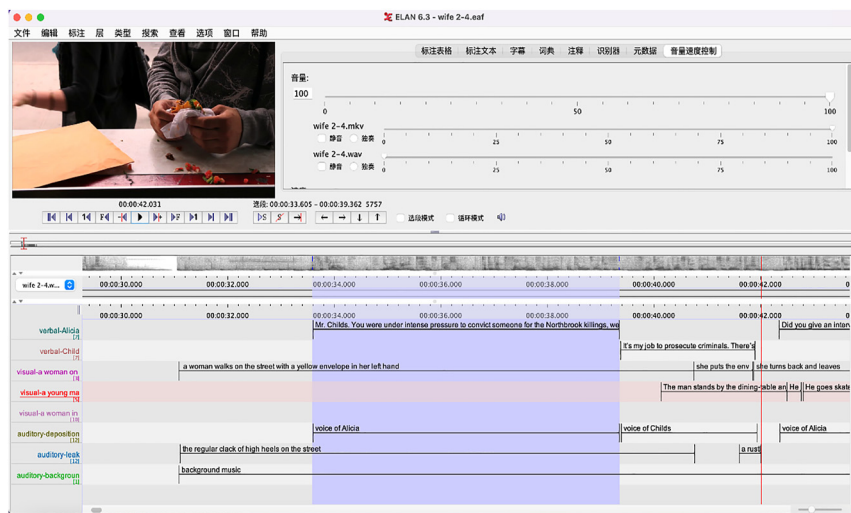


Figure 9: Visual, verbal, and auditory elements in the dual narrative annotated in ELAN 6.3.

5.2 Modal competition for the reconstruction of the discursive spatio-temporal dimension

Modal competition may warp space and time of different stories. Here I borrow two temporal concepts from pragmatics, receiving time (RT) and coding time (CT; Fillmore 1971), and reinterpret them in multimodal discourse analysis: RT refers to the time when a story in multimodal discourse can be received or perceived, which is audience-oriented, whereas CT denotes the time when a story occurs in multimodal discourse, which is participant-oriented. Sometimes, RT overlaps with CT, for instance, when the third party who is present on the spot is also the audience.

Like RT and CT, I propose receiving place (RP) as the place where a story in multimodal discourse can be received, which is audience-oriented, and coding place (CP) as the place where a story happens in multimodal discourse, which is participant-oriented. RP coincides with CP when a participant is also the audience. RP/CP is similar to RT/CT in reference, but different in the semantic orientation. These two pairs of concepts can be applied to explicate the relation of different stories in different spatial and temporal dimensions.

At the beginning of *The Good Wife* (Episode 4, Season 2), it seems that the deposition synchronizes with the delivery of the yellow envelope as they are presented simultaneously. But it turns out that the deposition (CT₁) precedes the delivery of the envelope (CT₂) along the unidirectional temporal dimension (Huang 2014) in light of their cause-effect relationship. As regards the places where these two stories occur, the verbal deposition is implied to be conducted in an office at Lockhart & Gardner (CP₁) while the yellow envelope is delivered across the street and then to a woman journalist at the Chicago Sun-Times (CP₂) presented in visual images. The auditory elements also imply coding time and coding places of

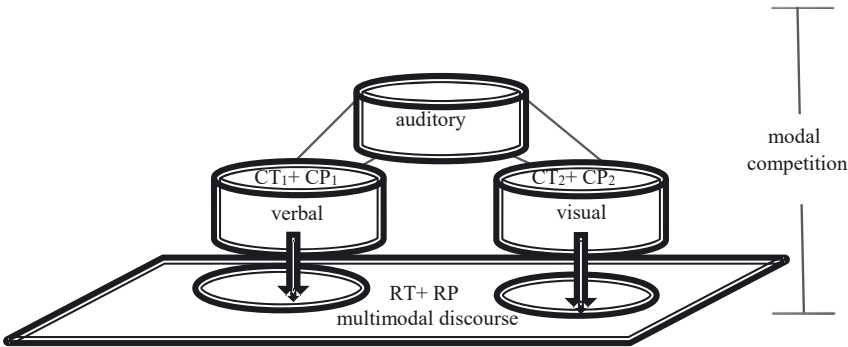


Figure 10: Modal competition for the reorganization of CT and CP as RT and RP.

these two events though they are presented in simultaneity. Modal competition breaks off the spatial and temporal relations between two stories (CT_1/CT_2 and CP_1/CP_2), and reconstructs a different discursive spatio-temporal dimension (RT/RP) in the TV world (see Figure 10). Modal competition sheds light on the innovation of the spatial and temporal patterns in the meaning-making processes of multimodal discourses, and builds a bridge between the external reality and the discursive world.

Besides, the reason for the delivery of the yellow envelope is to leak the deposition of Childs. Logically speaking, the deposition functions as the context of the leak. A conventional narrative strategy is to place the deposition before the delivery to realize the contextualizing relation instead of synchronizing them. But intermodal competition smashes their logical connection by instantiating separate semogenic resources to tell these two stories in simultaneity, which distorts the spatio-temporal relationship of two stories in the material world and blurs their contextualizing relation.

It is unusual to have a dual CT and CP simultaneously in the whole episode since it does cause confusion about the spatio-temporal logics of the stories. It is not commonly practiced in a commercial TV series because it aims to entertain the audience instead of confusing them. Therefore, after the journalist takes a tape out from the envelope and plays it, the voices of Alicia and Childs with high fidelity become a tape-recording when Alicia says “You stand by the prosecution of Wyatt Stevens” (see Figure 11). At the given moment, the deposition becomes a tape-recording deposition, which explicates that the deposition has been made before the delivery of the envelope containing this tape. The deposition plays the role of contextualizing the delivery of the yellow envelope by which the deposition is leaked. Afterwards, visual, auditory and verbal modalities enter a new intermodal competition, and all gain the semantic statuses to tell the same story that the journalist at the Chicago Sun-Times becomes interested in the tape of recording the deposition of State Attorney Childs about the Northbrook killings. These three modalities together with their elements reach a temporary balance and instantiate their semogenic resources to tell this story in different aspects in an identical spatio-temporal dimension in the TV world.

5.3 Modal competition for the diversification of information structures

Different from monomodal discourses, more than one modality gets the semantic status to make meanings in multimodal discourses with more than one information structure generated by different semogenic resources. Modal competition diversifies

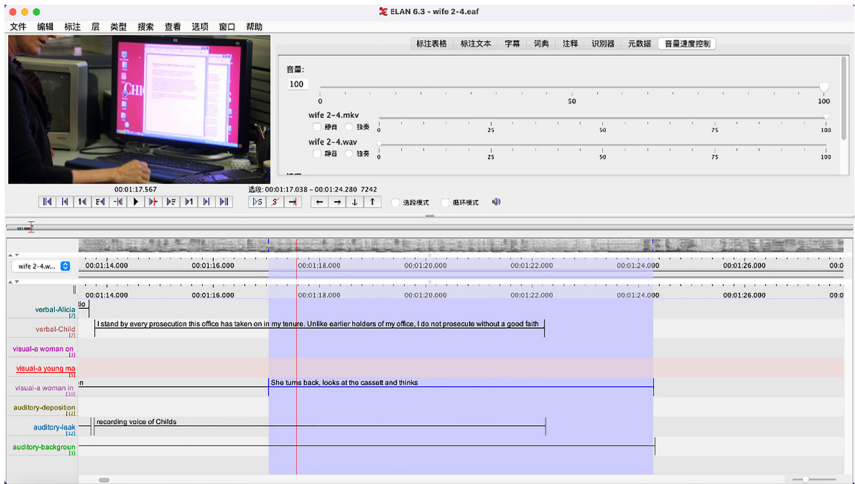


Figure 11: The contextualizing relation between deposition and delivery of an envelope in intermodal competition.

the dynamic orientations of information structures in which “new becomes given” (Matthiessen 1992: 43) in the TV world.

As different modalities and their elements compete to tell stories, they will generate intermodal and intramodal information structures of Given and New. Take the beginning of *The Good Wife* (Episode 4, Season 2) for example. Intermodal competition yields a parallel information structure of macroNew (Martin and Rose 2007) and macroGiven in multimodal discourse, while intramodal competitions evoke different dynamic orientations from New to Given within monomodal texts in the scenario (see Figure 12).

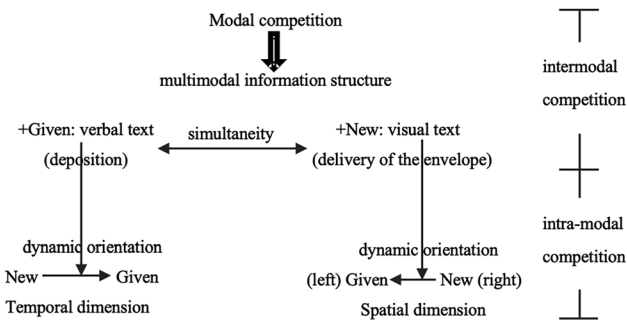


Figure 12: Modal competition for information structure in dynamic orientation from New to Given.

As argued, the deposition functions as the context of the delivery of the yellow envelope, which means that verbal text is contextualized for visual text in this scenario. It is a marked organization in which verbal text as context synchronizes with visual text for which verbal text is contextualized. Though the simultaneity enhances the difficulty of understanding the contextualizing relation between verbal and visual texts, it is justified to organize them in a parallel structure since the deposition (as macroGiven) has been presented in the previous episode.

Since verbal and visual modalities play the separate roles of telling two different stories, the multimodal information structure is bipartite. Generally speaking, each story introduces the new information, which then becomes Given as another piece of new information is introduced. But the dynamic orientation of information structure is different in verbal and visual texts: in verbal text, it is the temporal dimension that determines the dynamic orientation of information structure from New to Given as “a verbal text unfolds over time in a dynamic, sequential way” (Painter et al. 2012: 133); in visual text, it is the spatial dimension that indicates the flux from New on the right “to which viewers must pay special attention” (Kress and van Leeuwen 2020: 217) to Given on the left.

Different from a visual element as either Given or New in the static images, a visual element functions as New and then Given since New turns into Given in the dynamic visual materials. For instance, the image that a woman holds a yellow envelope in her left hand initially appears on the right of the screen as New information in visual text (see Figure 13). Then she walks to the left of the screen

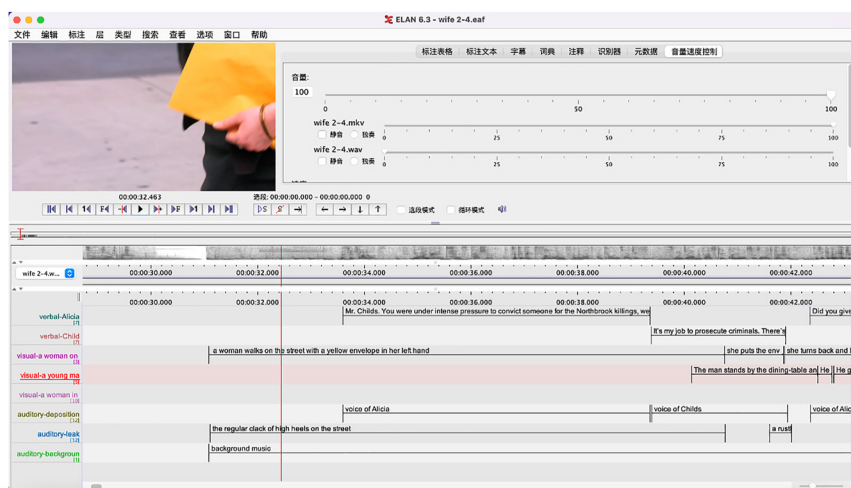


Figure 13: New information on the right of screen in visual text.

as Given information. Like the dynamic orientation of information structure in visual text, a new verbal expression becomes Given as the dialogue in the deposition goes on. As far as the auditory information structure is concerned, an auditory element that initially occurs is recognized as New. As the plot moves on, it becomes Given in auditory text. From the multimodal perspective, a new auditory element which gets the semantic status to tell the same story together with the corresponding new visual element, forms the multimodal New information, and then becomes a part of the multimodal Given information.

6 Conclusion

Multimodality is a mechanism for us to understand the world. Modal competition provides a perspective to explore the interaction among visual, verbal, and auditory modalities together with their elements. The metaphorical concept posits that visual, verbal, and auditory modalities compete to gain the semantic status to tell a story or different but relevant stories in simultaneity in a modern TV series.

Modal competition responds to the socio-semiotic nature of modalities as the semogenic resources on the premise of their equal stance of making discursive meanings. Some gain the semantic statuses of instantiating their semogenic resources to tell a story in a modern TV series but others do not. Gaining and losing the semantic status are strongly driven to elicit an additional layer of meaning (experiential and/or interpersonal) in light of the societal nature of a modality. When all the modalities get the semantic status, they reach a temporary balance to make meanings. The temporary balance is easily broken when a new modality is introduced in the competition or when the existing one withdraws from the competition. Both semiotic and societal meanings contribute to telling a story to the audience on the opposite side of the screen. When a modality outweighs others in time duration and/or frequency in modal competition, it implies that the TV series author attempts to manipulate the audience's emotion. Modal competition can warp space and time of the stories and reconstruct a different spatio-temporal dimension in multimodal discourses to achieve a specific artistic effect in the TV world. It can diversify the dynamic orientations of Given/New and macroGiven/macroNew in multimodal discourses, which provides the audience with a wide range of perspectives to perceive the information.

The analyses of how visual, verbal, and auditory modalities together with their elements compete to gain the semantic statuses present an intertwining world of art and reality for the audience to show the “humans exchange narrative in contemporary society” (Cumming et al. 2017: 1). The new perspective on the interplay among visual, verbal, and auditory modalities provides a lens for further

studies on semiotic issues, for instance, the removal of redundancy or contradiction arising from the multiplicity of the semiosis (Hodge 2017), on narrative discourse analysis in light of the actantial model of narrative grammar (Greimas 1988) from the multimodal and dynamic perspectives, and on the shifts of the external and internal focalizations (Genette 1993) in multimodal discourses.

References

- Cumming, Samuel, Gabriel Greenberg & Rory Kelly. 2017. Conventions of viewpoint coherence in film. *Philosophers' Imprint* 17. 1–29.
- Feng, Dezheng. 2011. Visual space and ideology: A critical cognitive analysis of spatial orientations in advertising. In Kay O'Halloran & Bradley Smith (eds.), *Multimodal studies: Exploring*, 55–75. New York & London: Routledge.
- Fillmore, Charles. 1971. Toward a theory of deixis. *University of Hawaii Working Papers on Linguistics* 3(4). 219–241.
- Genette, Gérard. 1993. *Fiction and diction*, Catherine Porter (trans.). Ithaca & London: Cornell University Press.
- Giannetti, Louis. 2014. *Understanding movies*, 13th edn. Boston: Pearson Education.
- Greimas, Algirdas. 1988. *Maupassant: The semiotics of text*, Paul Perron (trans.). Amsterdam & Philadelphia: John Benjamins.
- Hodge, Bob. 2017. *Social semiotics for a complex world: Analyzing language and social meaning*. Cambridge: Polity Press.
- Huang, Yan. 2014. *Pragmatics*, 2nd edn. Oxford: Oxford University Press.
- Kress, Gunther. 2010. *Multimodality: A social semiotic approach to contemporary communication*. London: Routledge.
- Kress, Gunther & Theo van Leeuwen. 1996. *Reading images: The grammar of visual design*. Abingdon: Routledge.
- Kress, Gunther & Theo van Leeuwen. 2001. *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.
- Kress, Gunther & Theo van Leeuwen. 2020. *Reading images: The grammar of visual design*, 3rd edn. London: Routledge.
- Krug, Christian & Joachim Frenk. 2006. Enter the matrix: Interactivity and the logic of digital capitalism. In Myriam Diocaretz & Stefan Herbrechter (eds.), *The matrix in theory*, 73–94. Amsterdam: Rodopi.
- Lausberg, Hedda & Han Sloetjes. 2009. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments & Computers* 41(3). 841–849.
- Lim-Fei, Victor. 2004. Developing an integrative multi-semiotic model. In Kay O'Halloran (ed.), *Multimodal discourse analysis: Systemic-functional perspective*, 220–246. London: Continuum.
- Lothe, Jakob. 2000. *Narrative in fiction and film: An introduction*. Oxford: Oxford University Press.
- Martin, James & David Rose. 2007. *Working with discourse: Meaning beyond the clause*. London: Continuum.
- Matthiessen, Christian M. I. M. 1992. Interpreting the textual metafunction. In Martin Davies & Louise Ravelli (eds.), *Advances in systemic linguistics: Recent theory and practice*, 37–82. London: Pinter.
- Norris, Sigrid. 2004. *Analyzing multimodal interaction: A methodological framework*. New York: Routledge.
- Norris, Sigrid. 2019. *Systematically working with multimodal data research methods in multimodal discourse analysis*. Hoboken: Wiley Blackwell.

- O'Halloran, Kay. 1999. Interdependence, interaction and metaphor in multisemiotic texts. *Social Semiotics* 9(3). 317–354.
- O'Halloran, Kay. 2004. Introduction. In Kay O'Halloran (ed.), *Multimodal discourse analysis: Systemic-functional perspective*, 1–7. London: Continuum.
- Page, Ruth. 2010. Introduction. In Ruth Page (ed.), *New perspectives on narrative and multimodality*, 1–13. New York: Routledge.
- Painter, Clare, James Martin & Len Unsworth. 2012. *Reading visual narratives: Image analysis in children's picture books*. Sheffield: Equinox.
- Royce, Terry. 1998. Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal text. *JASFL Occasional Papers* 1(1). 25–49.
- Royce, Terry. 2007. Intersemiotic complementarity: A framework for multimodal discourse analysis. In Terry Royce & Wendy Bowcher (eds.), *New directions in the analysis of multimodal discourse*, 63–110. Mahwah, NJ: Lawrence Erlbaum.
- Thibault, Paul. 2004. *Brain, mind, and the signifying body*. London: Continuum.