



VNU Journal of Foreign Studies

Journal homepage: <https://jfs.ulis.vnu.edu.vn/>

## INTERSEMIOTIC THEORIES: TOWARDS A FRAMEWORK FOR THE ANALYSIS OF THE INTERPLAY OF LANGUAGE AND VISUAL IMAGES IN MOVIE POSTERS

Nguyen Thi Thuy Linh\*

*Faculty of Linguistics and Cultures of English-Speaking Countries,*

*VNU University of Languages and International Studies, No.2 Pham Van Dong, Cau Giay, Hanoi, Vietnam*

Received 24 February 2024

Revised 11 August 2024; Accepted 28 August 2024

**Abstract:** The present paper aims to propose a preliminary adapted framework to analyze the relations between language and images in movie posters. After critically reviewing some major approaches to intersemiosis with both their strengths and weaknesses, the study decides to employ the Rhetorical Structure Theory (RST) discursal approach in general and the specific framework of Bateman's (2008) Genre and Multimodality (GeM) for the purposes of the current study. The review also suggests the need to include compositional features into the discussion of text-image relations. Putting all these issues into consideration, the paper finally develops an adapted analytical framework that employs Bateman's (2008) GeM as a platform and combines the layout and RST layers of GeM with compositional features for a more thorough examination of the roles of the visual and verbal elements. The application of the adapted model is then illustrated by the analysis of the theatrical release poster of *What women want* (2000).

**Keywords:** intersemiotic theories, movie posters, framework, Bateman's (2008) Genre and Multimodality

---

\* Corresponding author.

Email address: [linhnghuyen804@gmail.com](mailto:linhnghuyen804@gmail.com)

# CÁC LÝ THUYẾT LIÊN TÍN HIỆU: HƯỚNG TỚI KHUNG LÝ THUYẾT PHÂN TÍCH TƯƠNG TÁC HÌNH ẢNH - CHỮ VIẾT TRONG ÁP PHÍCH PHIM

Nguyễn Thị Thùy Linh

*Khoa Ngôn ngữ và Văn hóa các nước nói tiếng Anh, Trường Đại học Ngoại ngữ, Đại học Quốc gia Hà Nội,  
Số 2 Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam*

Nhận bài ngày 24 tháng 02 năm 2024

Chỉnh sửa ngày 11 tháng 08 năm 2024; Chấp nhận đăng ngày 28 tháng 08 năm 2024

**Tóm tắt:** Nghiên cứu này nhằm kiến tạo một khung lý thuyết phù hợp để phân tích mối quan hệ giữa ngôn ngữ và hình ảnh trong các áp phích phim. Sau khi xem xét các đường hướng cơ bản trong phân tích tương tác hình - chữ trong các diễn ngôn đa thức và chỉ ra ưu và nhược điểm của các đường hướng này, nghiên cứu đề xuất lựa chọn đường hướng Lý thuyết cấu trúc hùng biện tiếp cận từ góc độ diễn ngôn nói chung và khung lý thuyết Thể loại và đa phương thức (GeM) của Bateman (2008) nói riêng để đạt mục đích nghiên cứu cụ thể của đề tài. Tổng quan lý thuyết cũng chỉ ra sự cần thiết của việc xem xét các đặc điểm bố cục thiết kế trong phân tích mối quan hệ hình - chữ. Sau khi cân nhắc tất cả các yếu tố trên, nghiên cứu đề xuất một khung phân tích phù hợp, ở đó sử dụng khung lý thuyết Thể loại và đa phương thức (GeM) của Bateman (2008) như một mô hình nền tảng và kết hợp lớp phân tích cấu trúc hùng biện và lớp phân tích bố cục của mô hình này với đặc điểm thiết kế để đạt tới mô tả chi tiết và đầy đủ hơn về vai trò của các yếu tố hình và chữ trong poster. Việc kết hợp này được minh họa bằng việc phân tích một poster cụ thể là poster của phim *What women want* (2000).

*Từ khóa:* các lý thuyết liên tín hiệu, áp phích phim, khung lý thuyết phân tích, khung Thể loại và đa phương thức của Bateman (2008)

## 1. Introduction

Since the early days of multimodal discourse analysis (MDA), intersemiosis has always been a central area of multimodal research. Multimodality is viewed as “textual combinations of different modes and their integration in terms of structure, discourse semantics, and rhetorical function” (Stockl, 2019b, p. 50). Therefore, the very nature of multimodality not only implies that all modes contribute in their own ways to the overall textual meaning but also puts emphasis on the ways, or mechanisms, in which the semiotic resources work in combination to make meaning.

This interplay of semiotic resources has been expressed in a number of terms. Matthiessen (2009) calls it “semiotic harmony” with the observation that “one interesting - and critical - aspect of the division of semiotic labour among the denotative semiotic systems is the extent to which they operate in semiotic harmony with one another” (p. 11). Royce (1998) employs the term “intersemiotic complementarity” where “visual and verbal modes semantically complement each other to produce a single textual phenomenon” (p. 26). Other terms are also suggested like “mutual elaboration” of modes (Jewitt, Bezemer & O’Halloran, 2016, p. 91), “dialogicity” of signs (Jewitt et al., 2016, p. 111) or “meaning multiplication” (Bateman, 2014, pp. 5-7). Despite the differences, all these terms refer to the interplay, and varying levels of integration between systems with different kinds of semiosis. In this present study, intersemiosis is used to refer to the interplay of two specific modes of language and

visual images in a multimodal artifact.

The issue of text-image relations is even more burning in the current changing semiotic landscape. Since the 19<sup>th</sup> century, thanks to advancements in technological innovation and social changes in the values associated with images, the area has witnessed fundamental changes with a shift toward image-centricity (Stockl et al., 2019). In an image-centric discourse, images become dominant and central thanks to their “semantic centrality”, which means they lead the interpretation of the multimodal text, and/or their “perceptual dominance or salience”, that makes images the textual ‘entry point’ for reading paths and meaning construal (Stockl et al., 2019, p.7). This phenomenon has been observed in different contexts, for example print news stories (Caple, 2013), online newspaper homepages (Knox, 2007), Instagram (Caple, 2019), or Internet memes (Zappavigna, 2019). As it is opposite to the traditional dominance of linguistic elements, this shift seriously influences the status and functionality of language, text-image relations and the development of multimodal genres (Stockl, 2019a). Therefore, there is an urgent need for further studies in the field to explore the relative status and relations of images and language in different image-centric genres. The present paper is in line with these studies, aiming to examine the text-image relations in movie posters.

Movie posters are an interesting yet under-researched member of the image-centric genre family. At a glance, a movie poster may show the dominance of images with the layout space taken up by images. However, the actual relations between images and language in movie posters and how they are integrated to fulfill the general communicative purposes of movie posters has not been explored. Most studies on movie posters, including Chen and Gao (2013), Iftikhar, Shahnaz and Masroor (2019), Peng (2022), Linh (2021) and Rondon and Heberle (2022), only focus on the meaning making of images without the consideration of linguistic elements. Maiorani (2007, 2014) is the only exception that analyzes the slogans and images in three poster campaigns for The Matrix film trilogy to identify their advertising strategies. However, the verbal signs are only limited to the slogans, which in fact are slightly different from the taglines in ordinary movie posters. Other linguistic elements such as movie titles, actors and actresses, the cast, director and other information have not been studied.

Therefore, in order to seek the answer to the question: “What kinds of relations are established between language and images in the posters to promote the movies?”, this paper aims to critically review some approaches to text-image relations in order to develop a relevant framework of intersemiosis to provide insights into the interplay of visual images and language in this image-centric genre.

## **2. Major Approaches to Text-Image Relations**

Since the late 1990s, various attempts have been made to investigate the nature of intersemiotic relations to explain what features make multimodal text visually - verbally coherent, including Martinec and Salway (2005), Kong (2006), Royce (2007), Taboada and Habel (2013), to name but a few. Although different frameworks have been proposed, these studies can be grouped into four major approaches to text-image relations (Bateman, 2014). This section will critically review these four approaches and choose the most suitable one for the present study.

### ***2.1. Multimodal Cohesion and Text-Image Relations***

The first approach, proposed by Royce (1998, 2007) and then further developed by Martinec (1998), Baldry and Thibault (2006) and Liu and O’Halloran (2009), employs the

popular and ready applicable term of cohesion as a method of description in the context of text-image relations. ‘Cohesion’, as a technical term introduced in Halliday and Hasan (1976), is defined as a non-structural linguistic resource, which means cohesion is a way of relating linguistic entities across all kinds of syntactic and other structural boundaries. Therefore, this concept can be well applied to multimodality to investigate the web of cohesive ties across text and image as they are parts of a single ‘textual’ unit. Royce (2007) then proposes a system of intermodal cohesive relations including repetition, synonymy, antonymy, meronymy, hyponymy and collocation, which are equivalent to the theory of verbal lexical cohesion. Liu and O’Halloran (2009) further develop the theory by examining the integration of information from different modalities rather than simply documenting their ‘linkages’. Some further extensions have been made to Royce’s (2007) framework, including the addition of polysemy, intersemiotic logical relations and intersemiotic parallelism, to reach this purpose.

Although the analysis of intersemiotic cohesive devices is quite explicit and straightforward, there is concern about its effect in interpreting the texture and meaning of the multimodal text (Bateman, 2014). Just like in linguistic texts when cohesive devices cannot assure the coherence of the text, what the examination of intersemiotic cohesive devices can tell about how the text is working still remains unclear. Particularly in this research when the purpose of examining text-image relationships is to explore how they help to promote the movies, this approach seems to be ineffective and irrelevant.

### ***2.2. Using Metaphor for Text-Image Relations***

The second approach, pioneered by Forceville (1996), is appeal to the notion of metaphor, particularly conceptual metaphor. If cohesion is a non-structural concept, metaphor is actually a structural relation, which means metaphor deals with semantic organizations rather than just a lexical, or word-based operation. In fact, many cognitive linguists, Forceville included, consider metaphor as a conceptual phenomenon and state that metaphor is a cognitive operation rather than one situated in the linguistic system. This viewpoint leads to the occurrence of the new concept of ‘pictorial metaphor’, which refers to the phenomenon when either or both source domain and target domain are presented pictorially. Regarding the text-image relations, this concept provides a “well-established means of talking about relations of ‘similarity’ or ‘transfer’ of meanings across potentially very different domains” (Bateman, 2014, p.185). In other words, when the target domain or the source domain is image and the other is text, the transfer of properties from a source domain to a target domain can be suggested as a means of enabling information to flow across modalities (Forceville, 2002, 2009).

Although this approach provides a well-established means of examining text-image relations, its application in the present study is problematic. In movie posters, this approach can analyze the relation between image and taglines as they both express the content of the movie visually and verbally but it is not applicable to other textual elements like movie title or actors’ and actresses’ names. Therefore, this approach cannot help to reach the overview of text-image relations and more importantly the promotional strategies regarding text-image relations in the posters.

### ***2.3. Grammar Approach to Text-Image Relations***

The third approach starts from the linguistic ‘stratum’ of grammar with the assumption that when language and image occur together in a text to serve similar or conjoined communicative purposes, similarities in their organization will be expected. Studies in this approach, e.g., Martinec and Salway (2005), Unsworth (2007) or Kong (2006), tend to focus

on two aspects of grammar, which are clause transitivity and clause combining relations. For the first area, analysis aims at finding the correspondence between ‘grammatical transitivity’ of language and ‘visual transitivity’ of images, both of which follow the principles of systemic-functional grammar. The grammar of visual images can be analyzed using Kress and van Leeuwen’s (2006) framework. The second area describes the ways in which grammatical clauses can be combined into larger ‘clause complexes’ with two main types of relationships: (1) relations of projection and (2) relations of expansion with three sub-types of elaboration, extension, and enhancement. This concept is then expanded to multimodal texts and these relations are not only applied to linguistic clauses but also to text-image relations.

Several attempts have been made to construct a classification system for text-image relations that employ these two grammatical concepts. The first framework by Martinec and Salway (2005) proposed two dimensions for finer classification of the grammatical clause-combinations, which are status and logico-semantic relations. The former indicates the relative importance of texts and images in the relationship with two options of equal and unequal status, while the latter characterizes the specific logical relationships held between text and image in terms of expansion and projection. Other attempts that could be considered extensions of Martinec and Salway’s (2005) framework are Unsworth’s (2007) examination of school textbooks and other educational materials and Kong’s (2006) multilevel descriptive framework. These frameworks show some adaptations but still share similar principles with Martinec and Salway (2005).

Although this approach seems practical and in line with the analysis of visual images in the posters with the employment of Kress and van Leeuwen’s (2006) framework, the actual problems lie in the analysis of linguistic texts. As most texts in movie posters are proper nouns (e.g. movie title, actors’ and actresses’ names), the transitivity analysis and examination of clause combining relations are problematic. Moreover, this approach rules out the contribution of both layout and multimodal features of texts to their model, while these factors, in fact, can provide much useful information on the text-image relations.

#### ***2.4. Discourse Approach to Text–Image Relations***

The last approach looks at text-image relations from the discursive perspective, which generally considers discourse to be made up of ‘discourse moves’ to serve the communicative goals pursued by a speaker or writer. These ‘discourse moves’ traditionally are linguistic expressions and then extended to visual images. Two separate models of discourse have been employed to discuss text-image relations, which are Conjunctive Relations by Martin (1992) and Rhetorical Structure Theory (RST) by Mann and Thompson (1988).

The first branch of discursive approach uses Martin’s (1992) system of conjunctive relations with four main classes: additive, comparison, time and consequential/causal. This system is somehow related but not limited to grammatical clause combining relations in the third approach. Instead, Martin (1992) examines discourse semantic relations, which have realisations scattered over the entire grammatical system. The multimodal development of this system has ended up with a number of different models to describe linking relations between visual and verbal elements including van Leeuwen (2005), Liu and O’Halloran (2009) or Martin and Rose (2008). This diversity shows the complexity of discourse semantic relations. Just like coherence in linguistic discourses that can be created in various ways, the text-image relations in multimodal discourse can be explicitly and implicitly suggested by a variety of means. The implicit elements are hardly categorized in the analytical frameworks, thus raise the issue of reliability of the analysis. Moreover, according to Bateman (2013), the vacillation across these

analytical frameworks is partly due to the lack of empirical work with far more extensive collections of data in the field, which may help to target the issues of reliable analysis and accurate coding.

The second branch follows the tradition of studies of the rhetorical combination of visual and textual material with the application of Rhetorical Structure Theory (RST) by Mann and Thompson (1988), an approach to text developed within linguistics and discourse studies. In contrast to the conjunctive relations, which focuses on the concrete step-by-step interpretation of linguistic units, RST is more concerned with intentions and the knowledge and beliefs of speakers and hearers. RST provides a repertoire of 'rhetorical relations', each is defined by its realization and its effect when successfully applying the given relation on the speaker and hearer's beliefs and intentions. The RST analysis aims at constructing a single overarching hierarchical analysis that covers all the text and respects all the definitions. This hierarchical structure then shows precisely how the components of the text and their combinations contribute to achieving the goals of the text as a whole (Bateman, 2013). As the analysis mainly focuses on communicative intentions, the pictorial elements are well considered fragments of the RST hierarchical structure, which assures the multimodal application of RST.

So far, this section has reviewed four different approaches to text-image relations with the employment of different linguistic theories including cohesion, metaphor, grammar and discourse. After considering both advantages and drawbacks of each approach, it can be concluded that RST discourse approach seems to be the most relevant to the present study for a number of reasons.

Firstly, RST approach is particularly relevant to the purpose of the present study. RST was originally developed for the analysis of persuasion and communicative effectiveness; thus it is widely used for the analysis of advertisements, posters included. It is less concerned with the concrete step by step interpretation of forms and focuses more on intentions and rhetoric. This is suitable for the examination of how movie posters use verbal and visual resources to promote the movies. In other words, the study aims at identifying the promotional strategies in terms of multi-semiotic choices in these posters rather than simply describing the semiotic features of the posters. With the adoption of RST, a rhetorical structure of a poster may show the role of each element in the poster, being linguistic or visual element, and their relationship in forming the overall message of the poster.

The second reason for the choice of RST model is its relevance to the sample of the present study. In most other approaches to text-image relations, there seems to be a clear distinction between text and image. However, in movie posters, this separation is quite vague. The linguistic texts are often designed multimodally with color, typographic features and even mixed with images. For example, a company logo may include words or letters but its design is prone to picture or symbol. RST provides a good solution to this unclear distinction as it considers the rhetorical relations among different components of the poster, which can be text, picture or a combination of both. In other words, a clear division of language and visual image is not necessary in this approach. The second problem lies in the nature of linguistic texts in the posters. They are mainly noun phrases and their meanings are not similar to the picture. As mentioned above, this causes difficulties to the employment of other approaches like metaphor or grammar. However, noun phrases can be well treated as a rhetorical element of the poster, thus this is not a problem to the application of RST model.

However, the RST approach also has its own drawbacks. Firstly, RST analysis relies on

the linear nature of what is being described, which can be a challenge to visual materials with spatial organization. Although reading order can be a solution to this problem, it is claimed to be subjective and can be influenced by reading behaviour (Stockl, 2004). The second problem lies in the identification of RST segments. While grammatical clauses usually serve as the minimal unit in traditional RST, the decision of multimodal RST segments is more challenging with unclear grammatical units. Furthermore, each multimodal RST segment may serve different purposes, leading to multiple simultaneous discourse relations in the RST structure. Finally, similar to other approaches to text-image relations, there was no attention to issues of layout and composition in RST model, which in fact may significantly influence these relations.

These problems might be indicative of a need for an adaptation of the RST model. A more careful decision of the RST segments and inclusion of related compositional issues like layout, salience and reading path is required for a more reliable and thorough analysis.

### 3. An Adapted Framework to Analyze the Interplay of Language and Visual Images in Movie Posters

As discussed above, this study aims to develop a framework to analyze the interplay of language and visual images in movie posters that can combine RST analysis and compositional features like layout or salience. In search for a framework that can include all these issues, Bateman's (2008) Genre and Multimodality (GeM) has been chosen as it can provide a good platform for the examination of both layout and RST structure of the posters. This section, therefore, will discuss this framework in detail and show how it can be employed to form the paper adapted framework.

Bateman's (2008) GeM aims at exploring the interaction and combination of multiple modes within single artefacts. As GeM treats the multimodal page as a multilayered semiotic artefact, the model includes different layers, which are illustrated in Table 1. The starting point for analysis is the identification of base units in the base layer, which include all the units actually occurring on the page that can be called upon subsequently to take on particular roles or functions within other layers. At this stage, no elements could be omitted for any reasons, which assures the full coverage of what is being analyzed. After the base units are identified, they are combined according to their layout grouping and spatial proximity, their mutual spatial relationships and their particular layout and formatting properties to form the layout layer. This layer consists of three domains, which are layout structure, area model and realization information, to examine the layout structure of the multimodal document. Similar process is used to form the rhetorical layer, which describes the discourse relations that hold between the base units using an extension of RST. Lastly, the navigational layer analyzes the base units and layout units with navigation functions such as pointers, entries and indices.

**Table 1**

*The Layers of GeM Model (Cited in Hiippala, 2012, p. 108)*

Layer name	Descriptive function	Analytical unit and examples
Base layer	Provides a list of base units that may be analysed as a part of other layers.	<i>Base units</i> : sentences, headings, drawings, figures, photos, captions, list items, etc.

Layout layer	Groups the base units together based on similar properties in the three domains below.	<i>Layout units</i> : paragraphs, headings, drawings, figures, photos, captions, list items, etc.
<i>Structure</i>	The hierarchical structure between layout units.	
<i>Area model</i>	The placement of layout units in a layout.	
<i>Realisation</i>	Typographical or visual features of layout units.	
Rhetorical layer	Describes the rhetorical relations holding between the identified rhetorical segments.	<i>Rhetorical segments</i> : base units with rhetorical functions
Navigational layer	Describes the navigational structure by defining pointers, entries and indices.	<i>Pointers, entries and indices</i> : base units and layout units with navigation functions

As can be seen, the layout layer can bring a more thorough understanding to the compositional features of the posters. Compositional meaning is realized by three interrelated systems which are information values, framing and salience (Kress & van Leeuwen, 2021). While framing refers to grouping or ungrouping of elements, information values indicate the arrangement of these elements with three options: left-right, top-bottom and center-margin, which seem inadequate to explain the complex relations among various components scattered all over the poster. In this case, GeM can be a good solution. In GeM, with more specific information on both the overall hierarchical structure of the layout units and the position and features of each of these units, this layer can provide a sound empirical basis for comments on Information value and Framing of the posters. Therefore, in the present study, the analysis of compositional meaning will start with the layout layer and the two categories of Information value and Framing in Kress and van Leeuwen's (2021) will work as the perspective for judgements on the arrangement of these layout units. Similarly, the third category of salience can also be integrated to evaluate the prominence of the layout units. The order of elements in terms of salience creates the "reading path" (van Leeuwen, 2005). By reintroducing linearity in spatially structured texts, the principle of reading path helps to solve the problem of RST in dealing with the linear nature of visual materials.

The rhetorical layer is in line with the chosen RST approach to deal with the relationship between language and visual images. By means of a fine-grained theory of rhetorical structure, Bateman (2008) identifies a broad range of meaning relations that can hold between and within page elements, which are based on 25 classical rhetorical relations by Mann and Thompson (1986) and five additional relations drawn from Halliday and Matthiessen (2004). This helps to identify the functional contributions made by various semiotic elements to the intended communicative goals of the posters as a whole.

While the rhetorical layer carries in it the advantages of RST approach, it can overcome some RST problems. Firstly, the rhetorical segments are decided from the exhaustive list of base units in the base layer by combining base units with rhetorical functions. This may increase the reliability of the decomposition of units of analysis as all minimal elements of the posters



are involved. The second concern about RST is about the non-sequentiality of multimodal segments, which is different from the linear order of linguistic segments. To resolve this, the GeM model restricts rhetorical relations to “pairs (sets) of document parts (segments/spans) which are adjacent in any direction”, which can be expressed with respect to the area model (Bateman, 2008, p. 158). Most importantly, with the cross-checking with the Layout layer, the RST structure is considered together with layout and compositional features of the posters, which may provide more thorough understanding regarding the relationship between different visual and verbal components of the movie posters.

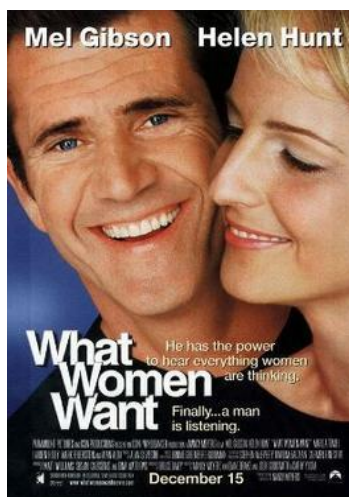
In fact, GeM projects are often conducted with corpus design with the potential of building multimodal corpora; however, this leads to the major problem of this approach: the application of GeM is too labor intensive. Both the meticulous analysis of each layer and the manual annotation to build the annotated corpora require a considerable amount of time and resources. Therefore, in this project, corpus analysis is not applied and I only analyze these layers manually to examine the design of the posters.

#### 4. An Application of the Adapted Framework to the Analysis of *What Women Want* (2000)

This section will present how the adapted framework is employed to analyze movie posters. The theatrical release poster of *What women want* (2000), a top-grossing movie with a North American domestic gross of \$182 million and a worldwide gross of \$374 million (Wikipedia), is analyzed as an example to illustrate the analysis using the model.

##### Figure 1

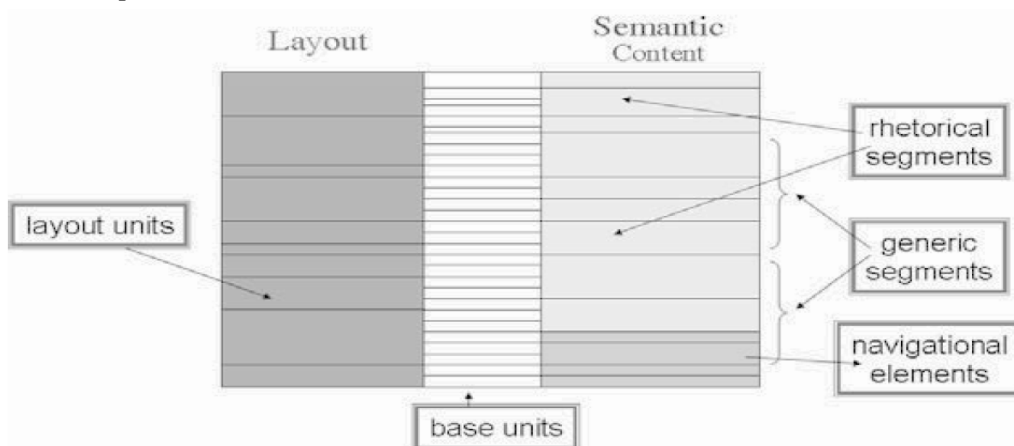
*Theatrical Release Poster of What Women Want (2000) (Source: Wikipedia)*



Using Bateman’s (2008) GeM, the analysis follows three main stages. After the base units are identified in stage 1, they are grouped into the layout units to examine the layout layer in stage 2 and into the rhetorical segments to explore the RST structure in stage 3. The base units include everything that can be seen in the poster and should not be larger than layout units and rhetorical segments. Notably, as can be seen from figure 3, these three units in three different layers are identified separately and do not necessarily overlap each other. Although the analysis goes through all three stages, my discussion will focus on the last two stages to explore the relations between visual images and language in the poster.

**Figure 2**

*The Distribution of Base Elements to Layout, Rhetorical, Generic and Navigational Elements (Bateman, 2008, p. 109)*



**4.1. The layout layer**

In stage 2, the researcher works visually from the page, decomposing all elements of the page in terms of their visual unity, their features and layout position. Table 2 lists all 15 layout units in the poster and their spatial information is illustrated in Figure 3. As the visual image covers the whole poster and expresses a single idea, it is treated as a single unit. Similarly, while L7 consists of many different base units, these base units form one block with similar format and information, thus they are combined into one layout unit of L7. L10 also has similar format as L7 but it is separated to a new line with a different layout position, thus it is considered a new layout unit. After the layout units are decided, the page decomposition is transformed into a hierarchical structure in Figure 4. In this hierarchical structure, the layout units are grouped according to their meaning, e.g. Movie title, Release date or Billing block (i.e. names of producers, key crew members, and others involved in the production of a film).

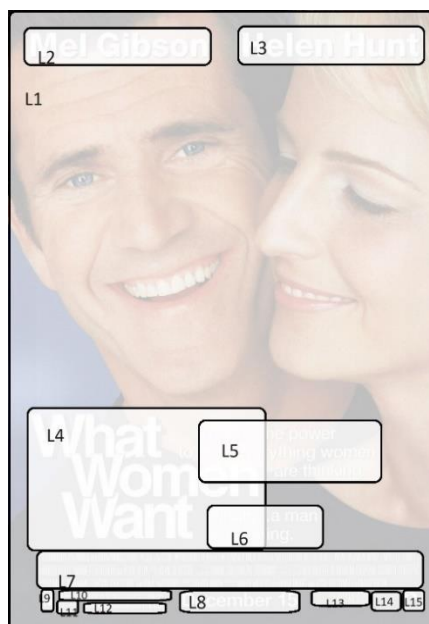
**Table 2**

*The List of Layout Units in the Poster*

L1	The image of the couple	L9	
L2	Mel Gibson	L10	SOUNDTRACK AVAILABLE ON COLUMBIA/SONY MUSIC SOUNDTRACK
L3	Helen Hunt	L11	DO SOMETHING GREAT
L4	What Women Want	L12	www.whatwomenwantmovie.com
L5	He has the power to hear everything women are thinking.	L13	PRODUCED BY NANCY MEYERS
L6	Finally...a man is listening.	L14	IT'S A GUYFRIEND'S FOLLOU BY PARAMOUNT PICTURES ALL RIGHTS RESERVED
L7		L15	
L8	December 15		

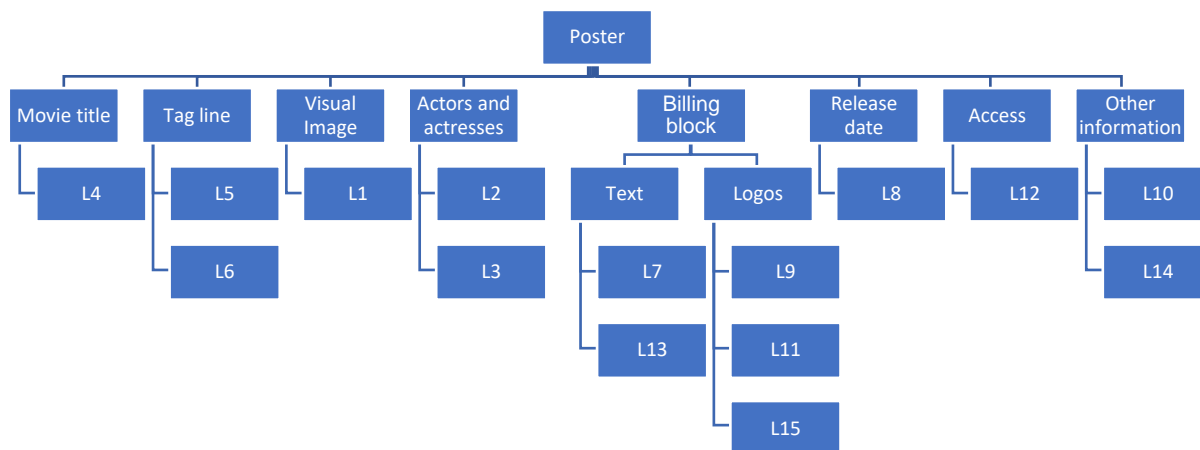
**Figure 3**

*The Area Model of Each Layout Unit*



**Figure 4**

*Layout Structure for The Poster*



As can be seen from Figure 5, except for the image of the couple (L1), which covers the whole poster, the position of the other layout units classifies them into three groups. The first group is at the top of the page with the names of the main actor and actress (L2 and L3). The second group consists of the movie title (L4) and the two taglines (L5 and L6) at the lower center of the page, leaving the rest (from L7 to L15) to the third group at the bottom. Notably, group 2 and group 3 are placed next to each other with similar color, and the only differences are the typographical features including font size and type. This layout separates group 1 from group 2 and 3 and creates the Top-Bottom relation in the poster. This choice of Information Value is popular in advertisements where the upper part shows the idealized or generalized essence of the information and the lower part presents more ‘down to earth’ or more practical

information (Kress & van Leeuwen, 2021). In this poster, the actor’s and actress’s names together with the couple’s faces, which is also the salience of the image, create the ‘ideal’ world in the movie while other information in group 3 presents the ‘real’ information of how the movie is created and how it could be accessed. The taglines are at the border of these two groups as they are related to both the plot line (Ideal) and the viewers’ background knowledge (Real) with the statement “Finally...a man is listening”. Therefore, their relevant position is at the lower center of the poster.

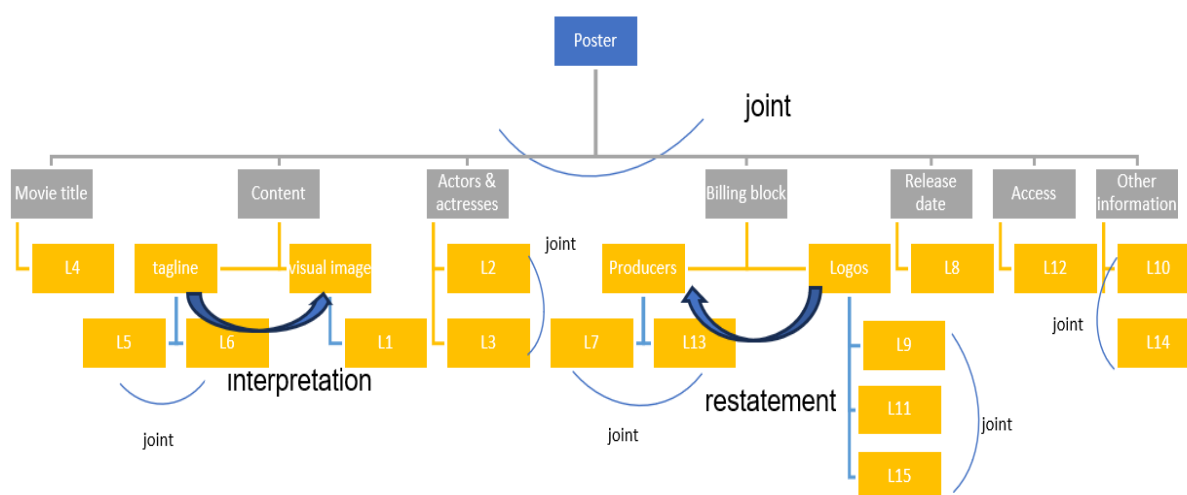
The spatial information and multimodal features of the layout units also decide their salience, and thus their roles in the poster. To this point, the concept of Reading path can be employed to grade the level of Saliency among the layout units. The visual image is the most salient element as it covers all the poster with the biggest size and colorful image. The next most prominent element is the movie title thanks to its big size, followed by the actor’s and actress’s names as they are clearly separated from the other linguistic elements, positioning at the top of the poster. The fourth most prominent unit are the tag lines, which are also isolated from other linguistic elements. The rest layout units have a similar level of salience as they are all put at the bottom of the page with small size, similar font type and color. Among these units, the release date receives a bit more salience with slightly bigger font size.

**4.2. The RST layer**

Stage 3 deals with the rhetorical layer and the result is presented in Figure 5. Notably, in this poster, the layout units coincide with the RST segments thanks to their self-containment and mobility. Therefore, for the sake of convenience, I still use the layout coding (from L1 to L15) to refer to these RST segments. They are then combined into seven RST groups according to their semantic meaning and rhetorical functions.

**Figure 5**

*RST Structure for the Poster*



As can be seen from Figure 5, the joint relations are frequently used at all levels of the RST structure. Each rhetorical group presents one aspect of the movie, thus they are rhetorically equal to each other. Similar relations can be found at the lower level, for example L5-L6, L2-L3, L7-L13, L9-L11-L15 and L10-L14. The popular employment of joint relations shows the loose connection among different components of the poster. In other words, the RST segments

are self-contained and loosely linked to each other with unclear cohesion.

The only exceptions are the cases of Content and Billing block. Firstly, the content of the movie is expressed in the both taglines and visual image. While the visual image sets out the context and the main vibe of the movie visually, the taglines interpret the meaning of the image. The image of the couple cannot tell the invisible ability of the man in hearing women's thinking, which is the key element of the movie. This is why the taglines are essential and they help to clarify the interpretations and support the right understanding of the photo. The second case is in the Billing block, where the production companies are mentioned both verbally and visually through their logos. As they simply repeat each other, the RST relation is restatement.

These two exceptions further contribute to the Reading path with the principle that repeated information tends to be more prominent. As the two taglines interpret the visual images, they reinforce the dominant role of the visual image or the content of the movie in the poster. Similarly, although producers are put in the billing block with less salience, their importance is slightly increased with the repetition of information with the appearance of several company logos.

### **4.3. Summary of findings**

The above analysis has figured out three main strategies that have been used to promote the movie regarding the interplay of visual and verbal elements. Firstly, both language and visual image are employed to provide information on different aspects of the movie, creating a joint rhetorical relation among them. With the popular use of the joint relations, the RST segments are loosely linked to each other and their main connection is actually through the layout and compositional features, particularly the salience. Specifically, in terms of composition, these units harmonize as their multimodal designs form an integrated whole of the poster but at the same time, they compete with each other for salience. More important information like the content, the movie title and the actors'/actresses' names are made prominent due to their position, size, color, font type or repetition. Less important information tends to be put at the bottom of the page with less salience. Finally, this movie poster employs the traditional information value of Top-Bottom or Ideal-Real, which is popularly used in advertisements. By following this convention, the poster brings the familiar feeling to viewers and makes it easier for viewers to access the important information simply at a quick glance.

Although this finding still supports the early assumption that movie posters are a member of the image-centric genre family, it clarifies the actual relations between language and visual images in the poster. Being the most prominent element in the poster, the visual image is the 'entry point' for reading paths and steer viewers' perception and evaluation of the poster. However, in terms of rhetorical and semantic relations, except for the taglines, other linguistic elements are actually of equal roles to visual images. Regarding the logos, a type of visual element, they simply restate the verbal elements with equal salience; therefore, their roles are equal both visually and rhetorically.

The finding also contributes to the theory of intersemiosis. It serves as strong evidence to support the need to include issues like layout and compositional features to the traditional approaches to intersemiosis for a more thorough examination of the text-image relations.

## **5. Conclusion**

The present study has critically reviewed some approaches in intersemiosis to indicate their strengths and weaknesses and then develop an adapted framework for a specific purpose

of analysis. The application of this adapted model is then illustrated by the analysis of the poster of *What women want* (2000). Although further research with more samples is required to testify its practicality, the present study can serve as the first step in forming a relevant framework to analyze the relations between language and visual images in the movie posters and how they are employed to promote the movies.

### References

- Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis*. Equinox.
- Bateman, J. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Palgrave Macmillan.
- Bateman, J. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge.
- Bateman, J. (2013). Using multimodal corpora for empirical research. In C. Jewitt, (ed.), *The Routledge handbook of multimodal analysis* (2nd edition) (pp. 238–252). Routledge.
- Caple, H. (2013). *Photojournalism: A social semiotic approach*. Palgrave Macmillan.
- Caple, H. (2019). “Lucy says today she is a Labordoodle”: How the dogs-of-Instagram reveal voter preferences. *Social Semiotics*, 29(4), 427–447.
- Chen, Y., & Gao, X. (2013). Interpretation of movie posters from the perspective of multimodal discourse analysis. *GSTF International journal on education (JEd)*, 1(1), 1-24.
- Forceville, C. J. (1996). *Pictorial metaphor in advertising*. Routledge.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- Halliday, M. A. K., & Matthiessen, C. M. (2004). *An Introduction to functional grammar* (3rd edition). Edward Arnold.
- Iftikhar, S., Shahnaz, A., & Masroor, F. (2019). Multimodal discourse analysis of the poster covers of Academy award winning animated feature movies. *PUTAJ – Humanities and Social sciences*, 26(2), 49-80.
- Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality* (1st ed.). Routledge. <https://doi.org/10.4324/9781315638027>
- Knox, J. (2007). Visual-verbal communication on online newspaper homepages. *Visual communication*, 6(1), 19-53.
- Kong, K. (2006). A taxonomy of the discourse relations between words and visual. *Information design journal*, 14(3), 207-230.
- Kress, G., & van Leeuwen, T. (2006). *Reading images: the grammar of visual design* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203619728>
- Kress, G., & van Leeuwen, T. (2021). *Reading Images: The Grammar of Visual Design* (3rd ed.). Routledge. <https://doi.org/10.4324/9781003099857>
- Linh, N. T. T. (2021). A multimodal discourse analysis of romantic comedy movie posters. *VNU Journal of Foreign Studies*, 37(3), 79-93. <https://doi.org/10.25073/2525-2445/vnufs.4647>
- Liu, Y., & O'Halloran, K. (2009). Intersemiotic texture: analyzing cohesive devices between language and images. *Social semiotics*, 19(4), 367-388.
- Maiorani, A. (2007). ‘Reloading’ movies into commercial reality: A multimodal analysis of The Matrix trilogy’s promotional posters. *Semiotica*, 166(1/4), 45-67.
- Maiorani, A. (2014). The Matrix phenomenon. In S. Norris & C. Maier (eds.), *Integrations, images and text: A reader in multimodality* (pp. 269-279). De Gruyter Mouton.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: toward a functional theory of text organization, *Text*, 8(3), 243-281.
- Martin, J. R. (1992). *English Text: systems and structure*. Benjamins.
- Martin, J. R., & Rose, D. (2008). *Genre relations: mapping culture*. Equinox.
- Martinec, R. (1998). Cohesion in action. *Semiotica*, 120(1-2), 161-180.
- Martinec, R., & Salway, A. (2005). A system for image\_text relations in new (and old) media. *Visual Communication*, 4(3), 337-371.

- Matthiessen, C.M.I.M. (2009). Multisemiosis and context-based register typology: Registerial variation in the complementarity of semiotic systems. In E. Ventola & J. Moya (eds.), *The world told and the world shown* (pp. 11- 38). Palgrave Macmillan.
- Peng, Z. (2022). A multimodal discourse analysis of movie posters from the perspective of visual grammar — A case study of "Hi, Mom". *Theory and Practice in Language Studies*, 12(3), 605-609. <https://doi.org/10.17507/tpls.1203.22>
- Rondon, T., & Heberle, V. (2022). Marvel's captain america civil war poster: an analysis based on the grammar of visual design. *Revista X*, 17(3), 726-751.
- Royce, T. (1998). Synergy on the page: exploring intersemiotic complementarity in page-based multimodal text. *Japan Association for Systemic Functional Linguistics Occasional Papers*, 1(1), 25–49.
- Royce, T. (2007). Inter-semiotic complementarity: A framework for multimodal discourse analysis. In T. Royce and W. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 63-109). Lawrence Erlbaum Associates.
- Stöckl, H. (2004). In between modes: language and image in printed media. In E. Ventola, C. Charles & M. Kaltenbacher (eds.), *Perspectives on multimodality* (pp. 9–30). John Benjamins.
- Stöckl, H. (2019a). Image-centricity – When visuals take center stage: analyses and interpretations of a current (news) media practice. In H. Stöckl, H. Caple & J. Pflaeging (eds.), *Shifts towards image-centricity in contemporary multimodal practices* (1st ed.) (pp. 19-41). Routledge. <https://doi.org/10.4324/9780429487965>
- Stöckl, H. (2019b). Linguistic multimodality – multimodal linguistics: a state-of-the-art sketch. In J. Wildfeuer, J. Pflaeging, J. A. Bateman, O. Seizov, & C.I. Tseng (Eds.), *Multimodality: Disciplinary thoughts and the challenge of diversity* (pp. 41–68). De Gruyter.
- Stöckl, H., Caple, H., & Pflaeging, J. (Eds.). (2019). *Shifts towards Image-centricity in contemporary multimodal practices* (1st ed.). Routledge. <https://doi.org/10.4324/9780429487965>
- Taboada, M., & Habel, K. (2013). Rhetorical relations in multimodal documents. *Discourse Studies*, 15(1), 65-89.
- Unsworth, L. (2007). Image/text relations and intersemiosis: towards multimodal text description for multiliteracies education. In L. Barbara, & T. B. Sardinha (eds.), *Proceedings of the 33rd International Systemic Functional Congress (33rd ISFC)* (pp. 1165-1205). Pontificia Universidade Catolica De Sao Paulo (PUCSP), Sao Paulo, Brazil. [http://www.pucsp.br/isfc/proceedings/Artigos%20pdf/59pl\\_unsworth\\_1165a1205.pdf](http://www.pucsp.br/isfc/proceedings/Artigos%20pdf/59pl_unsworth_1165a1205.pdf)
- van Leeuwen, T. (2005). *Introducing Social Semiotics*. Routledge.
- Zappavigna, M. (2019). “And then he said ...No one has more respect for women than I do” Intermodal relations and intersubjectivity in image macros. In H. Stöckl, H. Caple, & J. Pflaeging (eds.), *Shifts towards image-centricity in contemporary multimodal practices* (1st ed.) (pp. 204-225). Routledge. <https://doi.org/10.4324/9780429487965>