

A Multimodal Analysis of Image-text Relations in Picture Books

Shuxuan Wu

English Department, Qingdao University of Science and Technology, Qingdao, China

Abstract—Image-text relations are expounded based on Systemic Functional Multimodal Analysis framework in this article. A thorough analysis of the logical relations between the visual and verbal elements in picture books is also given in terms of elaboration, extension, enhancement and projection. Furthermore, Divergence, a fundamental type of image-text relation in picture books, is thoroughly investigated. Generic framework of picture book is proposed for the exploration of image-text relations.

Index Terms—intersemiosis, multimodal discourse analysis, systemic functional linguistics (SFL)

I. PICTURE BOOKS

Picture book is a highly aesthetic genre, which weaves words and pictures together to tell a story. Meanings in picture books are inextricably constructed by art and text. Picture books are distinguished from the traditional storybooks that use pictures to supplement a text. Story books could be understood without reference to them. Illustrations undoubtedly explain the story to the readers, but some storybooks can be understood without them.

In picture books, words and pictures invite reader to create literal realities in the world to correspond with them. Picture books define the range of canonical characters, which reflect our modes of behaving, and thus they provide a map of possible roles and possible worlds. Picture books contribute to the formation and re-formation in our children of the cultural imagination, a network of patterns and templates through which we articulate and illustrate our experience. Picture books give children the shape of their identity as individuals and as members of a socially symbolic reality. In short, picture books could be analyzed from social semiotic perspective with relation to culture context and context of situation.

Lawrence Sipe (1998) states that “visual texts are on an equal footing with verbal texts” (P107). It seems necessary, in the logocentric society that we live in today, to make this point”. Schwarcz (1982) classifies 2 kinds of visual and verbal interactions: congruency (elaboration, alternation, amplification, and extension) and deviation (opposition and counterpoint). Maria Nikolajeva and Carole Scott (2000) points out that existing studies “do not recognize or explore ... the wide array and diversity of the dynamics (i.e. image-text relations) that picture books embody” (P225). The relationships between picture and text arose increasing academic concerns in children’s literature, but there is still lack of theoretic study of image-text interactions in picture books for children based on Systemic Functional Multimodal Analysis.

II. SYSTEMIC FUNCTIONAL MULTIMODAL ANALYSIS FRAMEWORK

We tend to take this for granted that we live in a multimodal society that makes its meanings through the combined use of semiotic resources (e.g. spoken and written language, visual imagery, mathematical symbolism, sculpture, architecture, gesture and other physiological modes). Multimodality refers to the fact that all texts combine and integrate diverse semiotic modalities. This multimodal perspective could be traced back to the pioneering work of scholars such as Gregory Bateson, Ray Birdwhistell, E. T. Hall, A. E. Scheflen and others in the 1950's and 1960's. However, until the publication of Kress and van Leeuwen’s (2006) *Reading Images: The Grammar of Visual Design*, the early work of these scholars has been further investigated based on Systemic Functional linguistics and social semiotics. A multimodal text, as a unit of meaning, combines semiotic resources to produce meaning. With development of computer technology in recent years, a large number of multimodal texts with more complex integrations of resources in printed pages come into view. meaning-making in such multimodal texts is dependent on an interplay of all semiotic resources of which language is only one component. The Meaning in a multimodal text is realized through the selection of semiotic choices from meaning potentials of all semiotic modes in the text. Social semiotic analysis of multimodal texts is based on M.A.K. Halliday’s (1978) theory of ‘metafunctions’ to identify three distinct but coexisting kinds of meanings that interplay within any text. Metafunctional analysis can be extended beyond the linguistic field thereby providing the descriptive basis for all meaning-making practices. Systemic Functional theory provides a social semiotic framework for the analysis of multimodal texts with context-dependent meaning studies and a comprehensive approach to language.

Table 1 shows a Systemic Functional Multimodal Analysis framework for investigating intersemiosis in printed multimodal texts that involve linguistic and visual forms of semiosis. Intersemiosis takes place within and across

Mini-Genres, Items and Components and on the expression stratum with regards to the materialization of the text.

TABLE 1
SF-MDA FRAMEWORK FOR PRINTED TEXTS: LANGUAGE AND VISUAL IMAGERY

IDEOLOGY		
GENERIC MIX		
REGISTERIAL MIX		
CONTENT Stratum	INTERSEMIOSIS Mini-Genres, Items and Components (Linguistic, Visual and/or Other)	
	LANGUAGE	VISUAL IMAGES
	INTERSEMIOSIS Discourse Semantics	
	Discourse	Intervisual Relations Work
	INTERSEMIOSIS Grammar	
	Clause complex Scene Clause Episode Word Group/Phrase Figure Word	Scene Clause Episode Word Group/Phrase Figure Word Part
DISPLAY Stratum	INTERSEMIOSIS Stratum Materiality	
	Typography/Graphology and Graphics	

III. IMAGE-TEXT RELATIONS

Image-text Relations of a multimodal text may be studied from their relative contribution of each mode to the construction of meaning in the text or the distribution of semiotic modes. For example, McCloud (1994) classifies image-text relations in terms of their equal/unequal contributions to meaning:

- word specific, where pictures illustrate but do not significantly add to a largely complete text;
- picture specific, where the picture dominates and words do not add significantly to the meaning of the image;
- duo specific, where words and pictures send essentially the same message;
- additive – words amplify or elaborate on an image or vice versa;
- parallel – words/image follow different courses without intersecting;
- montage – words are treated as integral parts of the picture;
- interdependent – image/words together convey an idea that neither could convey alone.

It gives a generalized view of visual-verbal balance of meaning.

In recent years, Systemic Functional Linguistics, particularly with reference to cohesion and inter-clause relationships, is widely applied in the theoretic study of image-texts relation. Generally, visual and verbal relations in multimodal texts are mainly explored from 2 perspectives: (1) inter-semiotic relations, examining how the pictorial and verbal elements complement each other to form a single and cohesive text by identifying cohesive links across structural-unit boundaries of cohesive chains. For example, Royce (2007) examines visual and verbal cohesive connection with categories of lexical cohesion developed by Halliday and Hasan (1985). (2) logico-semantic relations, derived from the inter-clause relations (Halliday 1994, 2004) to interpret the logical relation between image and text across structural syntagmatic units. Logical relations are not confined to clause relations in language, and they could account systematically for the relations across different semiotic modes. Halliday (1994) classifies 2 types of fundamental inter-clause relationships: *expansion* and *projection*. In *expansion*, the secondary clause expands the primary clause by: *elaboration*, *extension*, and *enhancement*. In *projection*, the secondary clause is projected through the primary clause, which instates it as: a locution or an idea. In *elaboration*, one clause elaborates on the meaning of another by further specifying or describing it. The secondary clause does not introduce a new element into the picture but rather provides a further characterization of one that is already there, restating it, clarify it, refining it, or adding a descriptive attribute or comment. In *extension*, one clause extends the meaning of another by adding something new to it. What is added may be just an addition, or replacement, or an alternative. In *enhancement* one clause enhances the meaning of another by qualifying it in one of a number of possible ways: by reference to time, place, manner, cause or condition.

IV. IMAGE-TEXT RELATIONS IN PICTURE BOOKS

The picture book has proved to be a fruitful field of study for inquiries into the image-text relations. Over the last 20 years, picture book is not regarded as a basically verbal artifact supported by pictures. Pictorial dimension, as an independent semiotic system in its own right, attracts increasingly academic concerns. Terms, such as Iconotext (Hallberg, 1983), imagetext (Mitchell, 1994), or synergy (Sipe, 1998) helps to convey the essence of this interaction. But there are a few Systemic Functional Multimodal Analyses of image-text relations. Image-text relation is expounded based on Systemic Functional theory across semiotic modes in this article. A though description of the logical relations

between the visual and verbal elements in picture books is also given in terms of elaboration, extension, and enhancement. Furthermore, Divergence, a fundamental kind of image-text relation in picture books, is thoroughly investigated.

A. Elaboration

Elaboration may be described as a relationship of similarity across semiotic modes while no new ideational element is introduced by the text or image. While the meaning is not simply repeated or duplicated, different sets of semiotic resources are employed. According to Perry Nodelman (1988), words and pictures can never simply repeat or parallel each other, because of the inherent differences between verbal and visual modes of communication. Images cannot explicitly assert attitudes towards the phenomena they display. They can, however, visually demonstrate attitudes, while words are incapable of directly expressing emotion through shape and color.

Elaborative relations in picture books can be identified from 3 perspectives: exposition, exemplification, and homospatiality. (1) Exposition is a sub-type of elaborating relationship where image and text reinforce each other by restating or reformulating meaning in some way. Visual structure is not same as the grammar of language in interpretation of experience and forms of social interaction. Image demonstrates its meaning in terms of color, shape, line, pattern, angle and etc. Visual and verbal elements in picture books express the same topic in different ways. For example, in *My Dad* by Anthony Browne, the text on page 8 “*My Dad can eat like a horse.*” provides a vivid description of the image displayed in the illustration. In the image, a man in pajamas with a horse head is sitting by the table with a fork and a knife in his hands. He is ready to dump a large plate of food including sausages, fried eggs, chips, tomatoes and etc. Image seems to be more specific in demonstration the similar meaning in picture books.

(2) In exemplification, image exemplifies text or text exemplifies image; text and image represent different levels of generality (Martinec & Salway 2005), and this is realized by a class-member relationship, where specified members are not an exhaustive set but rather, represent examples from that class set. An example could be found in *Something From Nothing* by Gilman, the text on page 8 only mentions “*my grandpa*”, while all the family members living in the same building are portrayed in the illustration to exemplify the text.

(3) Homospatiality is another type of concurrence between image and text, where different semiotic modes co-occur in one spatially bonded homogenous entity (Lim 2004). One instance is found where the words of the classical picture book *The Little House* by Burton on page 36 are arranged in the shape of a curving road, which is coherent with part of the text: *They rolled along the big roads, they rolled along the little roads, until they were way out of in the country.*

Generally speaking, there is no exact equivalence in the relationship between image and text, except for maps, diagrams, labels and etc. In the traditional illustrated books and picture books introducing regulations and rules, scientific information, and etc, verbal narratives are predominant. Pictures are usually subordinated to the words. A number of very significant children’s stories have been illustrated by various artists who may interpret the text by means of exposition, exemplification and homospatiality of visual elements. The many illustrated editions of Bible stories, folktales, Perrault’s, Grimm’s or Andersen’s tales fall into this category. The text is not dependent on the illustration to convey its essential message.

B. Extension

Extension is a kind of complementary relations of modalities where image and text complement each other to extend the meaning represented in the other mode. It includes three sub-types of extension in picture books: augmentation, distribution and divergence.

(1) Augmentation involves an image extending or adding new meanings to the text or the text extending the image by providing (an) additional ideational element(s). According to Systemic Functional Multimodal Analysis, the new ideational element in augmentation is represented by participants or circumstances. Table 2 demonstrates the classification of circumstances with examples (Halliday: 1994). Circumstance is usually expressed in language system but in pictures.

TABLE 2
CIRCUMSTANCE OF EXTENT AND LOCATION

	Spatial	Temporal
Extent	Distance walk for seven miles stop every ten yard	Duration stay for 2 hours pause every ten minutes
Location	Place work in the kitchen	Time get up at six o’clock

For example, on the 8th page of *Where the Wild Things Are*, the text reads “*and in and out of weeks and almost over a year to the wild things are*”, while the correspond meaning could not be represented in the illustration.

(2) Distribution refers to the construction of activity sequences with juxtaposed images and text jointly. According to Gill (2002), there are two kinds of distribution. Intra-process distribution refers to the portrayal by images and text of different aspects of a shared process. Inter-process distribution occurs when images fill a gap in the meaning in the text; image and text complement each other in that activities or processes are distributed across the two modes. A good example is found in *The Runaway Bunny* by Brown. There is no a single word on page 20-21. But a bigger colored

picture demonstrates that the runaway bunny changes into a flying rabbit with wings and his mother a tree in shape of a rabbit. This picture complements the previous text on page 19 and illustrates the transformation of the two rabbits.

(3) Divergence is used to describe the third type of extending relation, where the ideational meaning of the text is opposed or at variance to that of the image, or vice versa. This term is also applied to instances where the meanings in the text and image contradicted each other. According to Perry Nodelman, every picture book tells at least three stories: a verbal one, a visual one, and a third one that “tends to emerge from ironies created by differences between the first two” (1988). This kind of irony Nodelman has in mind here is in fact the divergence. For example, the text in Rosie’s Walk is only one sentence of 25th. The visual narrative is more complicated and exciting than the verbal one. The divergence between words and pictures shape the book’s impact and the reader’s involvement in decoding it. Babette Cole’s *Princess Smartypants* provides another fascinating example with ironically humorous counterpoint. The tasks that the princess gives her many admirers sound quite simple when described in words: “*stop the slugs from eating her garden,*” “*feed her pets,*” “*challenged [...] to a roller-disco marathon,*” “*invited [...] for a cross-country ride on her motorbike,*” “*rescue her from her tower,*” “*chop some firewood in the royal forest,*” “*put her pony through its paces,*” “*take her Mother, the Queen, shopping,*” and “*retrieve her magic ring from the goldfish pond.*” But pictures reveal the truly daunting nature of the tasks and the reasons why the princes fail: the slug is as large as a dinosaur, the pets are ferocious dragons, the tower is made of glass, the forest is enchanted, the goldfish pond is inhabited by an enormous shark, and so on. In each case, the unexpected realities featured in the pictures functions as an ironic counterpoint.

C. Enhancement and Projection

Language and image complement each other through enhancement when one mode provides meanings which expand another spatially, temporally or causally. While the test items did not target these relations, instances did occur in the stimulus material. In *Olivia* by Falconer, the text on the page 2 “*she even wears herself out*” is enhanced by the five portraits of Olivia in one illustration. Those portraits are arranged in a liner sequence to demonstrate how Olivia even wears herself out. Image and text are also found to complement each other through projection, the most congruent instance of this being illustrated texts with speech bubbles. In such text, the human participants (as sayers and/or sensors) are represented pictorially while their projected ideas and locutions are represented linguistically.

V. GENERIC FRAMEWORK OF PICTURE BOOKS

In one page of illustration of a picture book, image and text stylistically cohere into a single communicative act. The linguistic features of speech acts combine with other, non-linguistic and contextual features to create multimodal communicative acts in picture books. Visual and verbal features can combine into a single communicative act. The framework mainly consists of multimodal discourse systems at five levels: (1) context of culture, (2) context of situation and the function of the communicative activity, (3) genre mix, (4) content Stratum, and (5) display Stratum.

TABLE 3
GENERIC FRAMEWORK OF PICTURE BOOK

Culture Context (Ideology, Register)		
Context of Situation(field, tenor and mode) and the Function of the Communicative Activity		
Genre Mix		
CONTENT Stratum	INTERSEMIOSIS	
	LANGUAGE VISUAL IMAGES	
	INTERSEMIOSIS	
	Discourse Semantics(ideational meaning, interpersonal meaning , textual meaning)	
	Discourse Intervisual Relations	
	Work	
DISPLAY Stratum	INTERSEMIOSIS	
	Grammar	
	Clause complex Scene Scene	
	Clause Episode Clause Episode	
	Word Group/Phrase Figure Figure	
	Word Part	
	INTERSEMIOSIS	
	Stratum Materiality	
	Typography/Graphology and Graphics	

In this generic framework of picture books, genre refers to types of social activity including language activity, belonging to the social activity system. Picture book is a mixed genre that is born with visual and verbal semiotic modalities. Context of culture determines the social activities of the whole society, and genre is related to context of situation, which is a part of context of culture. Genre is determined by the three features of context of situation: the field of discourse, the tenor of discourse and the mode of discourse. The tenor of discourse is related to the choice of role relationship between the actual speaker and hearer. The actual speakers of a picture book are the writer and the illustrator. According to the framework, writers and the illustrators of picture books could freely make choices in both verbal and visual semiotic potentials and creatively weave word and picture together to tell a story. Therefore, the image plays a dominate role in the construction of meaning in picture books. Therefore, the interactions between image and

text are complicated.

It should be noticed that a lot of gifted artists create and illustrate picture books by themselves with keen insights in the nature of children and life, such as *The Tale of Peter Rabbit* by Beatrix Potter's, *Madeline* by Ludwig Bemelmans's, *Make Ways for Ducklings* and *Blueberries for Sal* by Robert McCloskey's, *The Little House* by Virginia Lee Burton's, *In the Forest*, Robert by Marie Hall Ets, *Harold and the Purple Crayon* by Crockett Johnson, *Little Blue and Little Yellow*, *Inch by Inch*, *On My Beach There Are Many Pebbles*, *Swimmy*, and *Frederick* by Leo Lionni, and etc. In other cases, artists produce picture books by adoption or rewrite fairytales, legends, folk stories and other traditional storied by themselves. Those artists are inspired by the classical storied and interpret the story in the version of picture books, such as *Millions of Cats* by Wanda Gag's, *The Three Robbers* by Tomi Ungerer, and tec. As brilliant painters, those artists skillfully employ visual semiotic means to develop the theme in picture books. It seems image in those works is more important than text in the interpretation of the meaning.

Language and image complement each other to create meaning in a multimodal text. Meaning is realized through verbal and visual choices made by the designer or illustrator of the text. The preference and selection of the semiotic modes definitely affect the image-text relations, which is not fully explored.

VI. CONCLUSION

Studies of picture books approach its pictorial dimension as an independent semiotic system in its own right, which does not necessarily concur with the verbal component, rather than as a mere prop to the verbal story. Both words and images make their own relatively autonomous contribution to the overall semantic, aesthetic and emotional effect of the picture book. Therefore, it has often been observed that the picture book is closer to other mixed narrative forms such as drama or film than to verbal fiction. From logico-semantic perspective, image-text relations of picture books are analyzed in terms of elaboration, extension, enhancement and projection. Context of situation, in particular, tenor of discourse is also related the image-text relations. It is hoped to shed some light on the theoretical frameworks for the understanding of multimodal semiosis.

REFERENCES

- [1] Gill, T. (2002). Visual and verbal playmates: An exploration of visual and verbal modalities in children's picture books. Unpublished BA (Hons) Thesis, University of Sydney.
- [2] Hallberg, Kristin. (1982). "Litteraturvetenskapen och bilderboksforskningen," *Tidskrift för litteraturvetenskap*, 3-4, 163-168.
- [3] Halliday, M.A.K. (1994). *An Introduction to functional grammar* (2nd edn). London: Edward Arnold.
- [4] Halliday, M.A.K. (2004). *An Introduction to functional grammar* (3rd edn). Revised by C.M.I.M. Matthiessen. London: Edward Arnold.
- [5] Halliday, M.A.K. & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Geelong: Deakin University Press.
- [6] Heffernan, J. & Blackwood, F. (2003). *Two summers*. Lindfeld: Scholastic.
- [7] Kress, G. & van Leeuwen, T. (2006). *Reading images: A grammar of visual design* (2nd edn). London: Routledge.
- [8] Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J.R. Martin & R. Veel (Eds), *Reading science: Critical and functional perspectives on discourses of science* (pp. 87-113). London: Routledge.
- [9] Lemke, J. (2006). Towards critical multimedia literacy: Technology, research, and politics. In M. McKenna, D. Reinking, L. Labbo & R. Kieffer (Eds), *International Handbook of Literacy & Technology*, volume 2.0. (pp. 3-14). Mahwah, NJ: Lawrence Erlbaum Associates.
- [10] Lim, V.F. (2004). Developing an integrative multi-semiotic model. In K. O'Halloran (Ed.), *Multimodal discourse analysis: Systemic functional perspectives* (pp. 220-246). London and New York: Continuum.
- [11] Martinec, R. & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication*, 4(3), 337-371.
- [12] Matthiessen, C.M.I.M. (2007). The multimodal page: A systemic functional exploration. In T.D. Royce & W.L. Bowcher (Eds), *New directions in the analysis of multimodal discourse*. Mahwah, NJ: Lawrence Erlbaum Associates.
- [13] McCloud, S. (1994). *Understanding comics: The invisible art*. New York: Harper Collins.
- [14] Mitchell, W. J. T. (1994). *Picture Theory. Essays on Verbal and Visual Representation*. Chicago: University of Chicago Press.
- [15] Nikolajeva, Maria, and Carole Scott. (2000). "The Dynamics of Picturebook Communication." *Children's Literature in Education* 31.4: 225-39.
- [16] Nodelman, Perry. (1988). *Words About Pictures: The Narrative Art of Children's Picture Books*. Athens and London: The University of Georgia Press.
- [17] O'Halloran, K.L. (Ed.). (2004). *Multimodal discourse analysis: Systemic-functional perspectives*. London and New York: Continuum.
- [18] Schwarcz, Joseph. (1982). *H. Ways of the Illustrator: Visual Communication in Children's Literature*. Chicago: American Library Association.
- [19] Sipe, Lawrence R. (1998). "How Picture Books Work: A Semiotically Framed Theory of Text-Picture Relationships." *Children's Literature in Education* 29.2:97-108.
- [20] Royce, T. (2007). Intersemiotic complementarity: A framework for multimodal discourse analysis. In T. Royce & W. Bowcher (Eds), *New directions in the analysis of multimodal discourse* (pp. 63-109). Mahwah, NJ and London: Lawrence Erlbaum Associates.
- [21] Royce, T.D. & Bowcher, W.L. (Eds). (2007). *New directions in the analysis of multimodal discourse*. Mahwah, NJ: Lawrence Erlbaum Associates.

Erlbaum Associates.

- [21] Unsworth, L. (2006). Towards a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique*, 5(1), 55-76.
- [22] Unsworth, L. (2008). Multiliteracies and metalanguage: Describing image/text relations as a resource for negotiating multimodal texts. In D. Leu, J. Corio, M. Knobel & C. Lankshear (Eds), *Handbook of research on new literacies*. NJ: Lawrence Erlbaum Associates.
- [23] Unsworth, L., Barnes, G. & O'Donnell, K. (2006-2008). New dimensions of group literacy tests for schools: Multimodal reading comprehension in conventional and computer-based formats. A linkage project funded by the Australian Research Council conducted by the University of New England and the New South Wales Department of Education and Training.
- [24] Unsworth, L. & Chan, E. (2008). Assessing integrative reading of images and text in group reading comprehension tests. *Curriculum Perspectives*, 28(3), 71-76.
- [25] Unsworth, L. & Cláirigh, C. (2009). Multimodality and reading: The Construction of meaning through image-text interaction. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 151-164). London: Routledge.
- [26] Unsworth, L., Thomas, A. & Bush, R. (2004). The role of images and image-text relations in group "basic skills tests" of literacy for children in the primary school years. *Australian Journal of Language and Literacy*, 27(1), 46-65.
- [27] van Leeuwen, T. (2005). *Introducing social semiotics*. Abingdon and New York: Routledge.

Shuxuan Wu was born in Inner Mongolia, China in 1978. She received her M.A in linguistics from Ocean University of China in 2004.

She is currently a lecturer in the School of Foreign Languages, Qingdao University of Science and Technology. Her research interests include SFL and children's literature.