Lawrence R. Sipe

How Picture Books Work: A Semiotically Framed Theory of Text-Picture Relationships

In this article, I provide an overview of the various ways in which the relationship between the words and pictures in a picture book has been conceptualized; I draw on several theoretical traditions and disciplines to describe this relationship. Although theorists have advanced many different conceptualizations of the text-picture relationship, they have not adequately explained what goes on in our heads as we relate words and pictures. Therefore, in order to describe more precisely what happens internally as we relate verbal and visual signs, I present a theory of text-picture relations that is based on the semiotic concept of "transmediation." By way of making the theory more clear, it is applied to the text-picture relationship in one double-page spread of Where the Wild Things Are.

Conceptualizing the Text-Picture Relationship

According to many writers, the essence of the picture book is the way the text and the illustrations relate to each other; this relationship between the two kinds of text—the verbal and the visual texts—is complicated and subtle. A variety of metaphors have been used to describe it. In several cases, the metaphor is drawn from music. Cech writes of the "duet" between text and pictures. Pullman utilizes the term "counterpoint," while Ward and Fox refer to the "contrapunctual" relationship. Alhberg talks about the "antiphonal" effect of words and pictures. On the other hand, using scientific metaphor in writing about illustration, Miller uses the idea of "interference" from wave theory, describing how two different wave patterns may com-
J. Ward and M. Fox, "A look at some outstanding illustrated books for children," p. 21

A. Ahlb erg in E. Moss, "A certain particularity: An interview with Janet and Allan Ahl berg," p. 21

J. H. Miller, Illustration, p. 95

W. Moebius, "Introduction to picturebook codes," p. 143

D. Lewis, "Going along with Mr. Gumpy: Polysystem and play in the modern picture book," p. 105

P. Nodelman, Words About Pictures: The Narrative Art of Children's Picturebooks, p. 223

P. Nodelman, Words About Pictures: The Narrative Art of Children's Picturebooks, p. 221

J. Schwar c, Ways of the Illustrator: Visual Communication in Children's Literature, p. 15

J. Golden, The Narrative Symbol in Children's Literature

P. Nodelman, Words About Pictures: The Narrative Art of Children's Picturebooks, p. 220

bine to form a complex new pattern. Moebius utilizes geological imagery to speak of the "plate tectonics" of text and illustrations.

Various alternative words and concepts seem to be pivotal in other theorists' description of the text-picture relationship. Lewis uses the idea of "polysystem": "the piecing together of text out of different kinds of signifying systems." Perry Nodelman describes how the text and pictures "limit" each other, and uses literary terminology to describe the relationship of text and pictures as one of "irony": The words and pictures comment ironically on each other, as it were. Nodelman also discusses Barthes's metaphor of "relaying," writing that, "By limiting each other, words and pictures take on a meaning that neither possesses without the other—perform the completion of each other that Barthes calls 'relaying.'" Schwarz conceives of two general categories of relationships between text and pictures: what he calls "congruency" and "deviation." In the category of congruency, the text and pictures are in a harmonious relationship. Sometimes the illustrations complement the text by "running ahead of the text and pushing the action forward." The text and the illustrations may also take turns in telling the story; this is what Schwarz calls "alternate progress." In deviation, Schwarz's second category of relationships, the illustrations "veer away" from the text by opposing it in some way. Another example of deviation is what Schwarz calls "counterpoint"—the musical metaphor, again—when the illustrations tell a different story from the text. Part of the enjoyment of this story lies in the reader-viewer's perception of both stories at once.

Golden describes five types of relationships between text and pictures: The categories seem to be differentiated by how much "work" the illustrations or the text do in conveying meaning. One limitation of this scheme is that, since it depends mainly on the relative amounts of power the text and the illustrations have, there is less attention to the dynamic way in which, as Nodelman puts it, "the words change the pictures and the pictures change the words." The text-picture relationship is not so much a matter of a balance of power as it is the way in which the text and pictures transact with each other, and transform each other.

The Synergy of Words and Pictures

My own descriptive term is "synergy," defined by the Shorter Oxford English Dictionary as "the production of two or more agents, substances, etc., of a combined effect greater than the sum of their separate effects." In a picture book, both the text and the illustration sequence would be incomplete without the other. They have a synergistic relationship in which the total effect depends not only on the
union of the text and illustrations but also on the perceived interactions or transactions between these two parts.

In adding to what these theorists have written about the relationship of words and pictures in picture books, I will focus on the phenomenological dynamics of the synergistic relationship itself. In other words, I want to try to explain more fully what happens in our heads: the process we engage in when we relate the verbal and visual texts of the picture book to each other. In order to do this, I will draw from literary theories of the type commonly referred to as “reader response” criticism (Wolfgang Iser), aesthetic criticism (G. E. Lessing and Wendy Steiner), linguistics and semiotics (Roman Jakobsen and Charles Peirce), and theories of literacy (Marjorie Siegel, Mark Sadolski, and Alan Paivio).

W. Iser, *The Act of Reading: A Theory of Aesthetic Response*

What Wolfgang Iser sees as he examines the reader-text relationship is a reader participating in the production of textual meaning. The actual literary work is realized through a convergence of reader and text. For Iser, the reader acts as co-creator of the work by supplying that portion of it that is not written but only implied. Each reader fills in the unwritten work or the “gaps” in his or her own way, thereby acknowledging the inexhaustibility of the text. Iser’s concept of gap-filling suggests that we can think of readers filling in some of the gaps in the verbal text of a picture book with information from the illustrations and of readers using information from the verbal text to fill in some of the gaps in the illustrations.

P. Nodelman, *Words About Pictures: The Narrative Art of Children’s Picturebooks*

Nodelman observes that words have a greater potential for conveying temporal information, whereas pictures have a greater potential for conveying spatial information. Does this mean that we might process verbal and visual information in different ways? There is a long tradition for this idea, beginning with the classic distinction, made by the German Romantic philosopher G. E. Lessing, between arts that are based on simultaneity of perception (painting, sculpture) and arts that are based on time sequence or successivity of perception (music, literature). Lessing argued that, since our experience of the world has two components or modes (space and time), all of the arts could be classified on the basis of which of these two modes were experienced while the viewer or listener was engaged in a work of art. We see a painting all at once; but in order to experience literature or music, we have to read or listen in a linear succession of moments through time. This raises the question of whether there could be arts that are based on both time and space—on simultaneity as well as successivity. Even in Lessing’s time, the opera was an example of such an art form; and it was because opera combined spatial arts and temporal arts that Wagner called his operas Gesamtkunstwerke (“assembled” or “put-to-
W. Steiner, *The Colors of Rhetoric*, p. 144

W. Steiner, *Pictures of Romance*


W. Steiner, *The Colors of Rhetoric*, p. 37

gether" works of art). Drama and film are also examples of art forms that we experience simultaneously in time and space. The picture book, of course, is another of these hybrid art forms. Wendy Steiner, a literary and aesthetic theorist, argues that “the illustrated text is a Gesamtkunstwerk, a mixture of artistic media epitomized in Wagnerian opera,” and, like opera, is a “gesture toward semiotic repleteness, combining several kinds of sign types and having them comment on each other.”

Wendy Steiner follows Lessing’s basic distinction by observing that the visual arts present us with a diffuse network, whereas temporal art leads us along a line or chain. The linguist and semiotician Roman Jakobsen makes a similar point: “Every complex visual sign, for example every painting, presents a simultaneity of various components, whereas the time sequence appears to be the fundamental axis of speech.” Jakobsen goes on to say that speech is not purely temporally linear, but that the “predominantly sequential character of speech is beyond doubt, and this primacy of successivity must be analyzed.” Written language is not purely linear, either; when we read, we are processing ever-greater sequences of letters, words, and sentences as meaningful wholes, until, finally, we may arrive at an understanding of the total structure of a poem or story. Nevertheless, to paraphrase Jakobsen, the *predominantly* sequential character of written language is beyond doubt.

Just as written language is not purely linear, painting and the visual arts are not purely spatial, either. When we look at a painting (or an illustration in a picture book), we look in a series of temporal moments at various parts of it. But, as Steiner comments, visual art is designed "to enlarge our ability to turn sequence into simultaneity, to allow us to form ever larger temporal flows into unified, atemporal structures." So, it's not the case that our perception of the image is static. We do look at the image in a series of temporal moments, but as we look our minds are forming "ever larger temporal flows into unified atemporal structures"; thus, to paraphrase Jakobsen once again, the *predominantly* simultaneous character of visual art is beyond doubt.

The different ways in which we experience written language and visual art have important implications for the ways in which we try to relate the words and the pictures in a picture book. Because of the primarily spatial nature of the pictures and our drive to form "unified atemporal structures," our tendency is to gaze on, dwell upon, or contemplate them. In contrast, the primarily temporal nature of the verbal narrative creates in us a tendency to keep on reading, to keep
going ahead in what C. S. Lewis termed “narrative lust.” There is thus
a tension between our impulse to gaze at the pictures—to forget
about time in creating an “atemporal structure”—and to not interrupt
the temporal narrative flow. The verbal text drives us to read on in a
linear way, where the illustrations seduce us into stopping to look.

This tension results in the impulse to be recursive and reflexive in our
reading of a picture book: to go backward and forward in order to
relate an illustration to the one before or after it, and to relate the text
on one page to an illustration on a previous or successive page; or to
understand new ways in which the combination of the text and pic-
ture on one page relate to preceding or succeeding pages. Picture
books have the effect of “loosening the tyranny of the one-way flow”
of the purely verbal text. Therefore, picture books seem to demand
rereading; we can never quite perceive all the possible meanings of
the text, or all the possible meanings of the pictures, or all the possi-
ble meanings of the text-picture relationships. Lemke observes that,
in illustrated informational texts and hypertext, there are “many possible
pathways through the textworld”; this is equally true of the picture
book. Margaret Meek argues that, “A picture book invites all kinds of
reading and allows the invention of a set of stories rather than a single
story.”

The reader/viewer’s construction of the relationship of text and pic-
tures is one of the “affordances” of picture books. But how does this
construction proceed? It is an intriguing idea that the interrela-

ship of words and pictures mirrors the thought process itself. Irving
Massey’s assertion that, “Thinking consists of a constant alternation
between image-making and word-making” is given support in reading
theory by the “dual coding” hypothesis of Mark Sadoski and Alan
Paivio, which suggests that cognition has two separate (though rel-
ated) structures: one for processing verbal information (either in
speech or written language) and one for processing nonverbal infor-
mation (such as visual stimuli). As we alternate our attention between
words and pictures in a picture book, then, we may be representing
the verbal and nonverbal information in separate cognitive structures;
following this, through the complex referential connections between
these two cognitive structures, we construct an integrated meaning.

A Theory of Transmediation

There is another theory from the field of literacy that may also prove
helpful in understanding how we construct the conversation between
words and pictures. Charles Suohor uses the concept of “transmedia-
tion” or “the translation of content from one sign system into an-
other.” This concept of transmediation may be heuristic for under-
standing the process we use in relating words to pictures in picture books. For in picture books, we must oscillate, as it were, from the sign system of the verbal text to the sign system of the illustrations; and also in the opposite direction from the illustration sign system to the verbal sign system. Whenever we move across sign systems, "new meanings are produced," because we interpret the text in terms of the pictures and the pictures in terms of the texts in a potentially never-ending sequence. Siegel draws from semiotic theory (and particularly the work of Charles Peirce) to explain the way in which we move across sign systems. For Peirce, a process of using signs consists of three parts: the sign itself, or representamen; the equivalent sign (called the interpretant) in the receiver of the sign; and the object for which the sign stands. The relationship among these three parts is usually represented visually by a triangle, called a semiotic triad:

![Semiotic Triad Diagram]

Thus, the sign or representamen "does not simply stand for an object, but tells something about the meaning of that relationship and this requires a third component, which he called an interpretant. The interpretant is another sign that represents the same object as the representamen, as its position in the semiotic triad indicates." According to Siegel, what happens in transmediation, when we move from one sign system to another, is that "an entire semiotic triad serves as the object of another triad and the interpretant for this new triad must be represented in the new sign system." In terms of what happens in picture books, we have two sign systems, and therefore two sets of triads:

![Picture Books Triads Diagram]

When we interpret the words in terms of the pictures, or move from the sign system of the words to the sign system of the pictures, the semiotic triad with the words as representamen becomes the object of a new triad, and the interpretant for this new triad changes accordingly:
Conversely, when we interpret the pictures in terms of the words, or move from the sign system of the pictures to the sign system of the words, the semiotic triad with the pictures as representamen becomes the object of a new triad, and the interpretant for this new triad changes as well:

The resulting process is a type of oscillation, as we adjust our interpretation of the pictures in terms of the words, and our interpretation of the words in terms of the pictures. And, because the meanings of the signs are always shifting (due to their adjustment and readjustment as each triad becomes the object in a new triad), this oscillation is never-ending. The possibilities of meaning in the word-picture relationship are inexhaustible.

**Words and Pictures in *Where the Wild Things Are***

To **concretize** all these abstractions and to elucidate the theory more completely, I will **analyze** a few of the possible semiotic triads in the text and pictures of the ninth opening of *Where the Wild Things Are*. The analysis is necessarily distorting and artificial, because it renders in laborious slow motion a process that in practice happens very quickly. It does, however, allow us to tease out an explanation of the steps in our meaning making. The text of this double-page spread reads,
And when he came to the place where the wild things are/they roared their terrible roars and gnashed their terrible teeth/and rolled their terrible eyes and showed their terrible claws.

Let us first consider the text alone, without reference to the illustration.

As we come to this point in the text, Sendak has carefully built our anticipation during the eight previous double-page spreads. Before we even opened the book, the title and the cover incited us to predict some encounters with “wild things,” and the vagueness of that phrase may have triggered all sorts of speculation about what these “things” will look like and how they will behave. The growth of the magical forest, which “grew/and grew—and grew” has been stretched out over three openings, an ocean has “tumbled by” with a “private boat,” and Max has sailed “in and out of weeks and almost over a year.” Now the text tells us that Max has finally “come to the place.” The four phrases

—roared their terrible roars
—and gnashed their terrible teeth
—and rolled their terrible eyes
—and showed their terrible claws

are representamens of objects for which our overall interpretant, if put into words, might be “horrorfic, savage monsters,” with razor-sharp teeth and claws, and wickedly cruel eyes. The representamens (Sendak’s words) allow us to construct this overall interpretant through their communication of sound (roars and gnashing), sight (teeth, eyes, and claws), and motion (eye-rolling and gesture), as well as our interpretation of the word “terrible.” The fourfold repetition of this word accentuates the monsters’ frightfulness. Our interpretant of these representamens may also include the inference that Max must be very frightened.

Now we uncover the illustration and attempt to disregard the text. The illustration contains many representamens of many objects. What is our interpretant of some of them? Max and his boat are on the extreme left, and we would normally look at him first, since our propensity is to “read” illustrations from left to right. In this case, however, it is probably the four wild things that first catch our eye. They are constructed primarily of curves and rounded shapes. Their claws (which resemble those on Max’s outfit), their horns, and even their teeth are slightly curved. Their bodies are the round shapes of stuffed animals. One of the wild things wears a child’s striped T-shirt and has a tail like the one on Max’s wolf suit. Another has chicken-like legs and feet. The pads on the other wild things’ feet and paws are puffy
and soft brown. The colors of their fur and faces are pastel and muted. The strongest color is the yellow of their eyes, which is only a bit more intense than the color of the rest of their faces and bodies. The wild things' hair looks soft and strokable, and they stand on soft green grass. The two wild things on the right-hand page can only be described as chubby. Their faces are humanoid (with broad, pudgy noses) and they stand upright. The wild thing nearest Max is perhaps the most formidable of the four: It is the most animal-like in its stance, and its face is more beastlike. It has an open mouth, red and lined with teeth. It has the longest claws, three horns, and a lion-like mane. Yet if it is so fierce, why has it allowed the smaller, goat-like wild thing to sit on its back? Why is its tail dragging on the ground, instead of quivering upright? If we turn our attention to Max, a glance assures us that he is not at all cowed by these beasts, even by the one closest to him. The expression on his face is one of disdain or exasperation, not fear. His hand is rebelliously on his hip, reminding us of his attitude in the third opening, where he was angry at his mother. The overall interpretant for this visual image, then, might be "mild menace, but nothing Max can't handle."

Having constructed interpretants for the two semiotic triads, we are now ready to move between the visual and the verbal sign systems. According to the theory of transmediation, in the movement from one sign system to another, "an entire semiotic triad serves as the object of another triad and the interpretant for this new triad must be represented in the new sign system." Let us take the picture triad as the object of the new triad we will build:

```
new interpretation of text

our interpretation of pictures

object pictures text
```

In this new triad, the new object to which the text refers is the entire picture triad we have previously constructed. The text is the representamen for this new object (or set of objects). We must therefore construct a new interpretant of the representamen (the text) since the representamen is now the referent for a new object. A "think-aloud" for this process of constructing a new interpretant for the text in light of the pictures might go something like this:
The text reads, "they roared their terrible roars"; but the mouths of two of the wild things are tightly closed, and the other two don't have their mouths open wide enough to make a terribly loud roar. The wild things are supposed to gnash their terrible teeth and show their terrible claws; but their teeth, being curved, are not so terrible. Neither are their claws, and I don't see any teeth being gnashed together. The wild things' eyes are almost circular, more like the glass orbs of stuffed toys than the eyes of fierce beasts. So I'll have to modify my interpretation of the words in the light of the pictures. There's menace here, especially in the creature closest to Max, but Max doesn't seem to be frightened out of his wits.

To complete this analysis, let us consider going in the other direction: constructing a new interpretant for the pictures when the object they represent is the textual triad:

A think-aloud in this case might be as follows:

Max doesn't seem frightened and the wild things (for many reasons) don't seem all that threatening. But the text says that they roared and gnashed their teeth and showed their claws. Those claws and teeth might not be needle-sharp, but they could probably still inflict some pretty terrible damage. And just because the beasts' mouths are either closed or only slightly open doesn't mean that they can't open them wide and let out a terrible roar. I wouldn't want to be caught in a dark alley with any of them, especially with the one who's closest to Max; that guy's paws are bigger than Max's whole head, and that rhinoceros horn looks dangerously sharp. Those horns look like devil's horns, so maybe they are terrible after all, and maybe Max is just too dumb to know the trouble he's in.

So this transmediation goes back and forth—oscillates—in a potentially endless process. Each new page opening presents us with a new set of words and new illustrations to factor into our construction of meaning. Reviewing and rereading will produce ever-new insights as we construct new connections and make modifications of our previous interpretations, in a Piagetian process of assimilation and accommodation. In other words, we assimilate new information and in the process we change our cognitive structures, accommodating them to the new information.
In this article, I have attempted to use the semiotic theory of transmediation to unfold the text-picture relationship; and I have used *Where the Wild Things Are* as an example to explicate the theory. I have argued that transmediation is a more precise way of looking at the text-picture relationship because it allows us to analyze the phenomenological process of relating words and pictures. Transmediation also makes it clear that visual texts are on an equal footing with verbal texts. It seems necessary, in the logocentric society that we live in today, to make this point and emphasize the significance of picture books in children's learning. Picture books, through transmediation, give children the opportunity to engage in an unending process of meaning making as every rereading brings about new ways of looking at words and pictures. In other words, picture books allow children to have multiple experiences as they engage in creating new meanings and constructing new worlds.

References


Copyright of Children's Literature in Education is the property of Kluwer Academic Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.