

[Kathy A. Mills & Len Unsworth](#)

[DOI: 10.1093/acrefore/9780190264093.013.232](#)

Multimodal Literacy

Summary

Multimodal literacy is a term that originates in social semiotics, and refers to the study of language that combines two or more modes of meaning. The related term, *multimodality*, refers to the constitution of multiple modes in semiosis or meaning making. Modes are defined differently across schools of thought, and the classification of modes is somewhat contested. However, from a social semiotic approach, modes are the socially and culturally shaped resources or semiotic structure for making meaning. Specific examples of modes from a social semiotic perspective include speech, gesture, written language, music, mathematical notation, drawings, photographic images, or moving digital images.

Language and literacy practices have always been multimodal, because communication requires attending to diverse kinds of meanings, whether of spoken or written words, visual images, gestures, posture, movement, sound, or silence. Yet, undeniably, the affordances of people-driven digital media and textual production have given rise to an exponential increase in the circulation of multimodal texts in networked digital environments. Multimodal text production has become a central part of everyday life for many people throughout the life course, and across cultures and societies. This has been enabled by the ease of producing and sharing digital images, music, video games, apps, and other digital media via the Internet and mobile technologies.

The increasing significance of multimodal literacy for communication has led to a growing body of research and theory to address the differing potentials of modes and their intermodality for making meaning. The study of multimodal literacy learning in schools and society is an emergent field of research, which begins with the important recognition that reading and writing are rarely practiced as discrete skills, but are intimately connected to the use of multimodal texts, often in digital contexts of use. The implications of multimodal literacy for pedagogy, curriculum, and assessment in education is an expanding field of multimodal research. In addition, there is a growing attention to multimodal literacy practices that are practiced in informal social contexts, from early childhood to adolescence and adulthood, such as in homes, recreational sites, communities, and workplaces.

Keywords:

multimodal literacy, language, modes, media, communication

Introduction

This article provides a systematic overview of multimodal semiotic theory, including the interaction between multimodal literacy theory and educational practice. It includes a brief history and outlines the foundational concepts of multimodality, modes, materiality, and semiotic resources. It explains the metafunctional theory that organizes three systems of meaning within language and multimodal texts. The central principle of intermodality is analyzed to account for the intersemiotic and semantic relationships between images and language in multimodal texts. The article concludes by synthesizing international research on multimodality and technology, writing, and reading in contemporary curriculum and assessment, with future directions for educational practice.

Multimodal literacy describes communication practices that use two or more modes of meaning (Mills, 2016). Multimodality has become a significant area of research given the broadened range of available designs and media forms in digitally networked and globalized textual ecologies. While there are varying definitions of multimodality, this article takes a social semiotic perspective following Halliday's systemic functional linguistics (Halliday, 1978)—that is, that language is fundamentally social and cultural. Language is dynamic and able to be modified by users, rather than being a static code (Jewitt, 2006). In other words, the meanings of texts, objects, and events are influenced by the contexts of situations within the culture or community (O'Halloran, 2009).

Language and literacy practices are inherently multimodal, because communication requires attending to diverse kinds of meanings, whether of spoken or written words, visual images, gestures, posture, movement, sound, or silence (Mills, 2011b). Yet, clearly, the affordances of people-driven digital media and textual production have given rise to an exponential increase in the circulation of multimodal texts in digitally-mediated environments. Multimodal text production has become a central part of everyday life for people across the whole of the life course—from young children to the elderly—and across cultures and societies. This has been enabled by the affordability, availability, and ease of production and sharing of digital content via the Internet and mobile technologies.

The increasing significance of multimodal literacy for communication has led to a growing body of research and theory to address the diverse potentials of modes and their intermodality for making meaning. The study of multimodal literacy learning in schools and society begins with the important recognition that reading and writing are rarely practiced as discrete skills, but are intimately connected to the use of multimodal texts, often in digital contexts of use. In addition, there is a growing attention to multimodal literacy practices that are practiced in informal social contexts, such as in homes and recreational sites. The implications of multimodal literacy for transformed pedagogy, curriculum, and assessment in education are far-reaching. Education must keep pace with the multimodal conventions and practices of communities, professions, and society. This involves embracing new potentials for reconfiguring student agency by providing opportunities for students to reshape semiotic resources in ways that are aligned with needs and interests of the signmakers (Kress, 2004).

Background: Multimodal Semiotics

The idea that communication extends beyond linguistic forms of semiotics is not new or unique to theories of multimodality. As Scollon and Scollon (2011) observed, while multimodality is a relatively new term, "language has been at the center of such interest for millennia." Prior to the social semiotic theory of multimodality, the first wave of Prague School semiotics from the 1930s to 1970s attended to non-linguistic modes of communication in semiotics.

For example, in the 1950s, the central theories of Birdwhistell (1952), Goffman (1959), Hall (1959), Ruesch and Bateson (1951), and Pike (1954), associated with the Natural History of the Interview, and the Palo Alto school of work on mammalian communication and systems theory, were key to early theorizations of kinesics and nonverbal communication. The Natural History of the Interview was a monumental project, begun at the Stanford University Center for Advanced Study in the Behavioral Sciences, that was essentially about enabling professionals, such as psychiatrists, to be able to explain the intuitive elements of their successful communication with patients (Leeds-Hurwitz, 1987). A significant dimension of the natural history project was the spoken interview, analyzed as filmed social interaction data, which was connected to the study of animal behavior and communication in situ, particularly mammalian communication (Scollon & Scollon, 2011).

Ruesch and Kee's book, *Nonverbal Communication: Notes on the Visual Perception of Human Relations* (1954), reflected the growing pragmatic interest in nonverbal and verbal forms of communication (Scollon & Scollon, 2011). In the 1970s, semiotics attended to art (Mukarovsky, 1976), costume (Bogatyrev, 1976), and theatre (Honzl, 1976), among other non-linguistic forms of communication. A significant development after this period was growing critique among linguistics and language scholars of the underlying models of language used to drive studies of nonverbal communication (Scollon & Scollon, 2011).

Multimodal semiotics has many trajectories, historical influences, and intersections with other approaches, particularly in contexts of education, such as the “multiliteracies” pedagogy of the New London Group (1996). However, multimodal semiotics can be traced to at least three main schools of thought—the social semiotic theory of multimodality (Hodge & Kress, 1988), multimodal discourse analysis (O'Halloran, 2004), and multimodal interactional analysis (Jewitt, 2014). Each of these three schools of thought has unique features, but they share a common foundation in Halliday's (1978) social semiotic theory of communication—also known as systemic functional linguistics. Systemic functional linguistics primarily is concerned with linguistic grammars, but it has been extended to elaborate other modes through the strands of multimodal theorization (Jewitt, 2006).

Beyond approaches to multimodal semiotics from systemic functional linguistic roots, there are many other historical contributions from different schools of thought. These include, but are not limited to, the work of Barthes (1967) and French semiotics, discourse analysis of Bernstein (1971), Marxian and Soviet psychology of Bakhtin (1986), and discernible influences of art history and iconography (Jewitt, 2014). Earlier historical influences include the Peircian-Sebeok semiotics of Charles Peirce (1934) and later, Sebeok (1976), which have been taken up by some of the educational literature (e.g., Roth, 2005). However, this article outlines the main theoretical and contemporary research on the social semiotic theory of multimodality and its application to classroom research on multimodal literacy.

Modes

Modes are defined differently across schools of thought and the classification of modes is somewhat contested (Mills, 2016). However, from a social semiotic approach, modes are the socially and culturally shaped resources or organized semiotic structures for making meaning. Specific examples of modes from a social semiotic perspective include written language, speech, gesture, movement, music, mathematical and scientific notation, attire, images, and the design of objects and environments. Modes are often used in concert. For example, speech is one of multiple modes that humans use to communicate, typically complemented by gestures, meaningful spatial arrangements among the speakers and listeners, and references to external objects that may be within the visual field

of the speaker (Mills, Chandra, & Park, 2013). Preferences in the use of modes of presentation, such as linguistic, auditory, gestural, and so on, differ according to the uses defined by culture and social context (Mills, 2011b). The regular pattern of use of modes is called a modal grammar, and these grammars have shared meanings within communities or cultures (Jewitt, 2006).

Each mode shares similarities with other modes, but also unique organizational principles, involving elements and conventions that do not have precisely equivalent meanings (Mills, 2011a; Semali & Fueyo, 2001). It is the lack of equivalence between modes that becomes a catalyst for transmediation—shifting or transferring semiotic content from one mode or sign-system. Transmediation is more than the straightforward reproduction of knowledge, because it involves a process of incremental knowledge transformation as users continually adapt their intentions for representing knowledge. This is required by designers in response to the possibilities and limitations of sign sign-making systems, including the affordances of digital platforms (Mills, 2011a). This is vitally important in the contemporary age, in which digital affordances of media engage users in particular ways that vary across different cultural tools or technologies.

Because social action often involves more than one focused interaction, due to simultaneous engagements (e.g., carrying out a text message exchange while carrying out an oral conversation with someone else), modal density is another important and related subconcept of modality. The modal density denotes the level of complexity and intensity of a focused communicative action (Norris, 2011). For example, writing an email while talking to a colleague involves high-level actions characterized by high modal intensity. This is necessary to simultaneously attend to writing, including the rapid movement of fingers on the keyboard, while maintaining a conversation with the colleague. This involves more subtle meanings of medium modal density to continue one's gestures, movement of the head, shifting gaze between the screen to the colleague, and so on.

Another related key concept is modal configurations, which pertain to the hierarchical order among the modes at operation in a meaningful action. For example, in a dinnertime conversation, multiple modes of meaning operate simultaneously, but with superordinate, equal, or subordinate levels of meaning, such as superordinate spoken words combined with subordinate gestures, gaze, facial expressions, movements, tastes, spatial arrangements of the seating, food, and so on (Norris, 2011). Among listeners in a dinner conversation, the modal configurations take on differing meanings and the modes take on different levels of importance than they do for the speaker.

It is important to note that modes and a user's choice of modes for any social purpose do not exist outside instantiations of ideologies. All modes of communication are ideological, whether an image, a song, or written text. While written language and ideologies have received much attention in critical discourse analysis, the analysis of ideologies that are realized in other modes of language have arguably received less attention. For example, in relation to images, Kress and van Leeuwen (2006, p. 4) observed that "neither power nor its use has disappeared,"; rather, the functioning of power potentially becomes more difficult to trace.

All modes of meaning can be used to convey the power or status of the created text, the viewer or reader, the represented subjects, and the relations between them. Thus, the reading or viewing of any multimodal text requires understanding of the historical, political, commercial, or ideological position of the text producers in relation to the reader or consumer. Similarly, readers must critically discern the interests that are served through the use of images, written words, gestures, sounds, and other modes. It is also important to consider that power is never definitively owned with certainty by any individual or social group, because power is contested, lost, and gained through social struggles (Fairclough, 1989).

Materiality

All multimodal texts possess materiality—a tangible, physical quality. For example, the mode of visual images can be rendered on different materials, such as paper, digital screens, or walls, or other objects, like billboards, cups, pillowcases, towels, or food packaging (Mills, 2016). Materiality matters to meaning-making. The semiotic function of the materiality of the sign is important, as is its intersection with the modes. The signifying resources of a medium—for example, the principles of narrative representations in painting—can often be used with other media, such as in digital design. Like modes, the medium can realize most of the choices from the ideational, interpersonal, and textual networks, but also have differing cultural and historic uses and constraints (Kress & van Leeuwen, 2006). Separating meaning-making from the materials—pencils, paper, screens, and so on—semiosis becomes produced as metaphors without material substance (Mills & Comber, 2015).

The selection of materials for communication, such as the voice (as occurs in speech), the limbs (as occurs in gestures), paper, paint, canvas, glass, or plastic (e.g., computer screens), is historical, cultural, and social (Kress & van Leeuwen, 2006). The user's choice of materials is influenced by factors that include the beliefs and practices established in the community or wider society (Mills, Davis-Warra, Sewell, & Anderson, 2016), the value afforded to those materials, and the ease of production and accessibility of those materials among users. The selection of materials is also open to modification as the materials are constantly transformed. Some theorists argue that there is an active and dynamic role of tangible materials, whether of toys, books, or tablets, in multimodal literacy learning (Mills, 2016).

Semiotic Resources

Semiotic resources are the systems of meaning available to the user. Central to this concept is the relationship between the representational resources, whether of words, images, and so on, and the meaning-making resources at hand (O'Halloran, 2005). In the context of theories of multimodality, semiotic resources include the many actions, materials, or artefacts that users employ to communicate meaning. This includes the human body—vocal chords, hands, facial expressions, gestures and so on, and the inanimate materials, technologies or media of communication, such as computer hardware and software, musical instruments, clay, paint, or pencils (van Leeuwen, 2005). The meaning potential of these resources is connected to their past uses and the norms or power relations that define their appropriation in social contexts of use. Multimodality has been used as a central organizing principle to build inventories of semiotic resources across modes that are available to cultural groups and social contexts (Jewitt, 2014).

Metafunctions

The concept of *metafunction* was developed within systemic functional linguistics (Halliday & Matthiessen, 2004) to refer to the three simultaneous meaning-making functions necessarily entailed in all instances of oral and written language. These are the ideational, interpersonal, and textual functions. Metafunctional theory has also been significant in understanding multimodal meaning-making, since researchers investigating the semiotics of images, displayed art, gesture, and

music from a systemic functional perspective have shown how the three metafunctions apply to the meaning-making resources of these modes (Kress & van Leeuwen, 2006; Martinec, 2004; O'Toole, 1994; van Leeuwen, 1999). The discussion here outlines metafunctions in relation to language and images only.

Ideational Meaning

The ideational metafunction has two aspects: the experiential and the logical. The experiential function refers to linguistic resources that enable meanings to be made about happenings and relations in the material world or in mental worlds, including the living, non-living, natural, artificial, and abstract participants in these happenings and the circumstances of time, place, and so on, in which they occur. In language, following systemic functional linguistic accounts, experiential meanings are realized grammatically by participants, processes, and circumstances of various kinds. Circumstances correspond to what are traditionally known as adverbs and adverbial phrases. Processes correspond to the verbal group, but they are semantically differentiated. There are four types of processes: material, mental, verbal, and relational. Material processes (action verbs), for example, express actions or events (like, “walk,” “sit,” or “travel”), while mental processes (sensing verbs) deal with thinking, feeling, and perceiving (e.g., “understand,” “detest,” or “see”), and verbal processes deal with saying of various kinds (e.g., “demand,” “shout,” or “plead”).

Relational processes link a participant with an attribute (e.g., She “is” tall, He “appears” tired), or identify or equate two participants (e.g., She “is” the School Captain, Cu “symbolizes” copper). The different kinds of processes have their own specific categories of participants. Material processes, for example, entail actors (those carrying out the process) and goals (those to whom the action is directed), while mental processes entail a sensor and a phenomenon (Mary *liked* what he liked; sensor–mental process–phenomenon). Relational processes that are attributive link a carrier to its attribute and, if they are identifying, they equate a token and a value (e.g., The captain [token] is the platoon commander [value]).

In their meta-functional account of images, Kress and van Leeuwen (2006) referred to experiential meaning as *representational meaning*. They distinguished narrative processes from conceptual processes. Narrative processes include vectors or action lines, which may be realized, for example, by the angle of a represented participant’s leg to suggest striding forward, an outstretched arm with a ball positioned in front of it to suggest throwing, or curved lines on each side of a pole to indicate vibration. Such vectors realize actional processes, but images also include reactional processes, where the vector is the sight line of the participant towards something being observed.

These processes can be transactional, where a participant acts on another participant, such as throwing a ball or looking at a picture, or they can be nontransactional, such as a participant running, or clearly looking at something, but the sight line does not lead to anything within the image. In images, verbal processes are realized by speech bubbles and mental processes by thought clouds. Conceptual images include those dealing with classification, such as various forms of visual taxonomic representations, and analytic images, which depict part/whole relationships, as well as symbolic images, where the meaning is different from that ostensibly portrayed. Circumstances of time, place, and manner can also be realized in images.

The logical metafunction refers to relations between happenings that are additive, temporal, cause, condition, purpose, manner, location, and comparative. These can be realized explicitly by conjunctions, such as *and*, *then*, *because*, *so*, *as if*, and so on, as well as by nouns, such as *the result*, *the effect*, or verbs, such as *cause*, *produce*, *enable*, and so on. Images do not have the same potential

to explicitly realize this range of logical relations. It is necessary to compare adjacent or successive images to infer logical relations from the similarities and differences between the images. Temporal succession, for example, can be inferred from differences in the depiction of daylight to night in successive images. Causal relations can be inferred from the depiction of an action in one image and the culturally understood result of the action in the subsequent image. However, while temporal relations of succession and simultaneity can usually be reliably inferred, causal relations cannot always be unambiguously distinguished from temporal succession (Painter, Martin, & Unsworth, 2013).

Interpersonal Meaning

The interpersonal metafunction is concerned with the construction of social relationships through the way people use language to interact with each other, as well as the way people take their own position through expressions of feelings and attitudes. According to Martin and White (2005a), interpersonal meaning deals with:

- Negotiation—focusing on the interactive roles of giving or asking for information and providing or requesting goods and services;
- Involvement—focusing on relative power, solidarity and affiliation of interactants through resources, such as familiar or formal address, colloquialisms, taboo expressions, and other resources for communicating inclusion or exclusion;
- Appraisal—focusing on attitude, including emotions we feel, judgments we make about human behaviors, appreciation we express about natural or human-created phenomena, the graduation or relative intensity of these evaluations, and engagement indicating how other voices or sources contribute to the evaluations.

Negotiation is realized grammatically by the mood and modality systems. Statements are realized by declarative mood, where the ordering within the clause is Subject followed by the Finite Verb (Grandad *does enjoy* his soup). Questions are realized by the interrogative mood, where the ordering of Subject and Finite Verb is reversed and the verb phrase is separated (*Does Grandad enjoy* his soup?). The realization of feelings and attitudes of the speakers/writers explicitly or implicitly in language was originally addressed through Halliday's account of modal verbs, and modal and comment adverbs (Halliday & Matthiessen, 2004). The modal verbs (*can, should, must, etc.*) indicate degrees of obligation or inclination, and modal adverbs, like *frequently, probably, and certainly*, enable negotiation of the semantic space between "yes" and "no," while comment adverbs, such as *happily, inexcusably, or exquisitely*, communicate affect, judgment, or appreciation. The personal stance aspect of interpersonal meaning was developed further in the appraisal framework for conceptualizing and analyzing attitude, graduation, and engagement (Martin & White, 2005b).

In multimodal texts, the roles that are realized in images through the interpersonal metafunction include those between the represented participants in the images and also between those represented participants and the viewer. The depiction of interactive relations between the viewer and image participants has been described by Kress and van Leeuwen (2006). Their account of interactive meanings in images includes options within the systems of contact, social distance, subjectivity, and objectivity.

The system of contact distinguishes between images where a character gazes directly at the viewer and images where there is no such gaze. Kress and van Leeuwen (2006) refer to images where the character gazes directly at the viewer as demand images, and where there is no such gaze, as an

offer. They interpret the former as overtly requiring participation by the viewer and the latter as indirectly, subtly, or incidentally involving the viewer.

The system of social distance (Kress & van Leeuwen, 2006) is realized by the size of frame, so that the represented participants are close or further away. If the entire body of the participant is visible, the character may appear more distant or remote. The extremes are commonly referred to as a close-up or a long shot, with mid-shot indicating commonly accepted interactive social distance.

The system of subjectivity includes options of involvement or detachment in viewer relations with the depicted characters. Involvement is concerned with positioning viewers to feel to a greater or lesser degree involved with the depicted participants. This is influenced by the horizontal angle (Kress & van Leeuwen, 2006). If the depicted participants are presented facing the viewer front on, there is a maximum involvement with them. On the other hand, if the participants are depicted at an oblique angle, it positions the viewer to be somewhat more detached from depicted participants. The greater the oblique angle, the more detached the viewer.

Also included in the system of subjectivity are options relating to the relative power accorded to the viewer or the depicted participants, and these are realized by the vertical angle of the image. If the depicted participants are seen from a high angle, with the viewer looking down on them, then they are depicted as if the viewer has power over them. If the depicted participants are seen at eye level, there is a sense of equality between the depicted participants and the viewer. When the viewer is positioned below the depicted participants, they are represented as having power over the viewer.

The system of objectivity refers to whether the view of an image is frontal or top-down. Most images in narratives are frontal images, and top-down views are rare, but when top-down images do occur, their narrative effect is very pronounced. One such image occurs in *The Lost Thing* (Tan, 2000), when the boy and the lost thing are outside the Federal Department of Odds and Ends, “a tall grey building with no windows.” The top-down view of the two characters on the pavement outside the very tall building emphasizes the de-personalized experience of their encounter with this institution.

Images convey attitude, including affect and judgement. Affect can be conveyed through facial expressions and body language, showing security and insecurity, happiness or unhappiness, and satisfaction or dissatisfaction. Images can also convey judgment of behavior. However, while affect is mostly inscribed through facial expression, judgments are more frequently invoked from the portrayal of experiential meaning in images. For example, in *The Tunnel* (Browne, 1989), the image of Jack, wearing a wolf mask and creeping into his sister’s room while she is asleep, may invoke a judgement of inappropriate behaviour. In some cases, images may also be considered to inscribe judgment. This seems to be the case when widely recognized symbols or visual metaphors are included in images (Unsworth, 2015; White, 2014).

Textual Meaning

The textual metafunction orchestrates the simultaneous presentation of ideational and interpersonal meaning. It provides the speaker or writer with strategies for guiding the listener/reader in the interpretation of text and correspondingly it provides the listener/reader with strategies for tracking and consolidating the integrated presentation of ideational and interpersonal meaning. Textual meanings are realized at the sentence level within each clause by the theme-rheme system. The theme is the point of departure (Halliday & Matthiessen, 2004), orienting the clause in its context, and is the element that is in first position in the clause. This is usually the subject (e.g., *The weather is*

mild in Australia), but another element, such as a circumstance (adverbial element) could be placed in theme position (e.g., *In Australia* the weather is mild). When some experiential element other than the subject is placed in first position in a clause, it is referred to as a marked theme, in that it draws more attention due to its relatively less frequent use.

The theme is commonly conflated with information that is the *given*, at the beginning of the clause. At the end of the clause is the less familiar information referred to as the *new*. These waves of information prominence, with predictive information in the theme position at the beginning of the clause and the accumulation of meaning as *new* toward the end of the clause, also occur across larger segments of text. What has been called the *topic sentence* of a paragraph, or the *hyper-theme* in SFL, predicts the pattern of clausal themes in a paragraph, and the conclusion of a paragraph (*hyper-new* in SFL) accumulates and distills the meanings presented in the *new* within the paragraph clauses. These waves of information flow also occur over expanses of text beyond the paragraph level, so that the introductory paragraph within, say a chapter or an essay, is the *macro-theme* predicting the *hyper-themes* of the subsequent paragraphs, and the concluding paragraph is the *macro-new*, accumulating and distilling the *hyper-new* of the preceding paragraphs (Martin, 1992; Martin & Rose, 2007).

In images, Kress and van Leeuwen (2006) have described the textual meta-function as composition, which relates representational and interactive meanings in images through three interacting systems—information value, salience, and framing.

Information value is the way in which image elements are endowed with specific informational values by their location in the various zones of the image: left/right, top/bottom, center/margin. Information values are culturally specific. The examples provided here are found in many Western texts, and can be applied to magazines, books, diagrams, and media interviews (Kress & van Leeuwen, 2006). The left and right zones realize the information values of *given* and *new* corresponding to the usual conflation in language of the *given* with theme at the beginning of clause and the usual location of the *new* toward the end of the clause. As well as in the examples provided by Kress and van Leeuwen, this has been shown to be the case in information books for children where, for example, in a book about floating and sinking, the everyday observational information and the image of boats floating is shown on the left, while the more technical visual representations concerning the principles of buoyancy are shown on the right (Unsworth, 2001).

The top and bottom zones endow information values that Kress and van Leeuwen refer to as the *ideal* and the *real*. The *ideal* at the top is the idealized or generalized essence of the information, while the *real* at the bottom is the more specific, concrete, practical information. In many advertisements, the top part indicates the promise of the product—its imagined or ideal effects, while the bottom part of the layout indicates more concrete information about the product itself. In textbooks, the top part of the layout typically deals with the more generalized, abstract conceptual information, while the bottom part deals with the specific, concrete observable information. For example, in a science book for primary school children, the concrete, every day, observable exemplars of scientific phenomena are often provided on the bottom part of the page, while the top part typically provides the technical images dealing with more generalized theoretical understandings (Unsworth, 2001).

Some layouts make more use of the center, placing one element in the middle and the other elements around it in a circular structure. This seems to be relatively uncommon in contemporary Western visualization (Kress & van Leeuwen, 2006). It does occur, albeit infrequently, in information books for primary or elementary school students. For example, a book for primary school children about ancient Rome shows the Roman gods in the center and smaller images depicting various Roman practices and beliefs associated with the gods around the margin (Unsworth, 2001). The central element, called the *center*, is the nucleus of the information, and the surrounding elements, the *margins*, are subordinate to, or dependent on, or sometimes peripheral to, the center. In these images,

the margins tend to be similar in terms of information value, although their relative marginality may relate to distance from the center, but there is no sense of *given* and *new* along the horizontal dimension, and no sense of ideal and real along the vertical dimension.

Salience occurs when the design of image elements attracts the viewer's attention to a greater or lesser extent due to factors such as being placed in the foreground or background, the relative size of the element within the image, contrasts in color, differences in sharpness of focus, and so on. Human, human-like, and animal participants also tend to be viewed as salient.

Framing can refer to the presence or absence of framing around the image as a whole or to the presence or absence of framing devices that create dividing lines within images. Framing devices connect or disconnect images and elements within images, indicating that they do or do not belong together in some sense. Where elements are completely disconnected and marked off, they are strongly framed. Where the elements are more integrated, they are weakly framed. Framing can be achieved by the use of frame lines or borders around elements, by discontinuities of colour or shape, or by white space. Connectedness can be achieved by vectors from within images and by devices like overlapping or superimposition of images. The more strongly framed an element is, the more it is emphasized as a separate piece of information.

For example, the form of framing of images, or lack thereof, is a very active feature of layout in constructing the nature of the narrative in literary picture books. In the picture book *Zoo*, Anthony Browne (1994) shows a boorish Dad, a quiet, unimposing but astute Mum, and two rough and tumble, food-oriented pre-teen sons, all visiting the London zoo. The framing around the images of the family is neither prominent nor straight, as if the images had been cut out from somewhere else. By contrast, once the family arrive at the zoo, the images of the zoo animals are all on the right-side pages and they are all heavily framed with thick, black, straight border lines. Different patterns of framing in a variety of picture books have been shown to relate interactive and representational meanings to the thematic interpretive possibilities of the literary narratives (Martin, 2008b; Painter et al., 2013; Unsworth, 2001). Framing is also significant in non-fiction and school textbooks (see, e.g., Unsworth, 2001).

Intermodality

Intermodality refers to the shared roles of different modes of meaning-making, such as image and language in constructing meaning in multimodal texts. Of course, other modes, such as typography, music, and sound, can also be involved. Music and sound, for example, are used significantly in digital versions of literary picture books, such as *The Heart and the Bottle* (Jeffers, 2009) and *Rules of Summer* (Tan, 2013); however, research on how music interacts with language and images in such texts has been limited (Barton & Unsworth, 2014). While there is some research on how other modes, such as typography, interact with language and images (Unsworth, Meneses, Ow, & Castillo, 2015; van Leeuwen, 2006), the discussion of intermodality here focuses on image-language relations, which have received much more attention in the literature.

While the joint contribution of language and images to the interpretive possibilities of picture books has long been accepted, in recent decades the inclusion of an increasing number of different kinds of images in an ever-widening variety of paper and digital media texts has established intermodality as a key dimension of 21st century literacy (Bateman, 2014; Bezemer & Kress, 2009; Luke, 2003). As Andrews (2004) noted: , “it is the visual-verbal interface that is at the heart of literacy learning and development for both computer-users and those without access to computers” (p. 63).

What is needed to explicate the central role of intermodality for literacy is an account of the intersemiotic semantic relationships between images and language to show how the visual and verbal modes interact to construct the integrated meanings of multimodal texts. While theorization of these intersemiotic semantic relations remains in its infancy, this has been the goal of a number of researchers working from the perspective of systemic functional semiotics derived from Halliday and his colleagues (Halliday & Matthiessen, 2004; Martin, 1992).

Most work from other theoretical perspectives has not had this focus. For example, Cohn's (2013a, 2013b) work on comics has been more concerned with categorizing formats of image-language relations, such as the use of speech balloons, captions, or "independent" image and text segments. In the study of Brazilian science textbooks by Roth and colleagues (Roth, Pozzer-Ardhenghi, & Han, 2005) the image-text relations were ambiguous and/or referred to relations among caption text and main text rather than between image and caption or image and main text (Unsworth, 2013). On the other hand, McCloud's (1994) descriptive work on comics did draw attention to the nature of image-language interaction in constructing meaning. McCloud identified four categories of image-text relations: (a) both words and pictures send essentially the same message, (b) words amplify or elaborate on an image or vice versa, (c) words and pictures follow different courses without interacting, and (d) the pictures and words go hand in hand to convey an idea that neither could convey alone (Bateman, 2014). It is this concern with how the resources of images and language interact to construct meaning that the systemic functional semiotic studies have sought to theorize.

Martin (2008b) and Painter and Martin (2011) argued that treating inter-modal relations as if they were intra-modal, cohesive, or logico-semantic ones, problematically assumes a simple equation of meanings in the different modalities, and that, at least in relation to children's picture books, "visual and verbal meanings are not realizations of an underlying meaning; rather they cooperate, bimodally, in the instantiation of a genre" (Martin, 2008b, p. 136). Painter and colleagues (Painter & Martin, 2011; Painter et al., 2013) emphasize that bimodal texts, like picture books, instantiate meaning choices from two meaning systems. They lay out the corresponding systems for the construction of ideational, interpersonal, and textual meaning in image and language and analyze the bi-modal texts to show the contributions made by each semiotic.

Differences in the meaning-making affordances of language and image are well documented (Kress, 1997, 2000, 2003a, 2004; Lemke, 1998; O'Halloran, 2008). In general, the resources of language are most apposite to the representation of sequential relations and the making of categorical distinctions, while the resources of images are most apposite to the representation of non-sequential, spatial, and comparative relationships. This suggests a logical basis for the distribution of meaning-making between images and text. But in some areas of meaning-making, both images and language have equally suited affordances. In such cases, the relevant meanings may be made by either or both semiotic systems, with each drawing on its own distinct range of options for realizing the particular meanings at stake. But since each semiotic system has its own affordances, there is not always a tidy complementarity between the two with respect to particular domains.

Some areas of meaning potential in language relate to two or more areas of meaning potential in images. For example, attitudinal meanings in language, such as un/happiness, dis/satisfaction, and in/security, relate to the visual depiction of affect and also to the visual depiction of ambience, while visual intercircumstance relations in images (shifts, contrasts, or continuities in depicted locations) do not have any specific correspondence in language. From this perspective, the questions of how choices combine across modalities, and how they complement one another, need to give priority to the distinctiveness of the two modes of meaning. To address this, Painter and colleagues (2013) draw on the notions of commitment and coupling (Martin, 2008a, 2008c, 2010).

Commitment refers to the amount of meaning potential that is taken up from any particular meaning system in the process of instantiating or creating a particular text (Hood, 2008). This can be seen in the depictions of Grandpa in different picture books. In *Dan's Grandpa* (Morgan & Bancroft, 1996), the image of the old Aboriginal man, and indeed all characters, is a black silhouette, with the Grandpa silhouette having white hair, so there is no commitment of meaning in relation to facial features. In John Burningham's (1984) famous picture book, the image of Grandpa is a simple line drawing, so while there is some commitment to facial features, it is very minimalist, with dots for eyes, a line for the nose, and vertical lines for the moustache. Then, in Norman's *Grandpa* (Norman & Young, 1998), the detailed realistic colour drawing by Young, shows much more commitment to the details of facial features. Similarly, a verbal description of a character as "an intelligent young English lad with blonde hair and blue eyes" commits more meaning than one describing him simply as "a lad."

Coupling refers to the repeated co-patterning within a text of realizations from two or more systems. Intermodally, in Browne's picture book *Gorilla* (1983), for example, the verbal text over several pages depicting the young girl Hannah's excursion with the gorilla, expresses her affect as happy, and this is consistently coupled with the visual depiction of solidarity between her and gorilla, realized by her arm around the gorilla's waist or her holding his hand. But, as Painter et al. (2013) pointed out, meanings are not always amplified through this kind of convergent coupling. Focalization in picture books, for example, may demonstrate divergent coupling, where focalization in the verbal text may be as a first-person narrator, but visual focalization may be in the third person, realized through observed images (e.g., which Kress and van Leeuwen refer to as *offers*), where the reader or viewer does not make eye contact with the character and is never positioned to see the image from, or along with, the character's point of view.

Converging ideational couplings create concurrence, while converging interpersonal meanings create resonance, and converging textual meanings create synchrony, but sometimes meaning committed in one mode is not committed in the other. For example, in the first two pages of McKee's (1987) picture book, *Not Now Bernard*, the first image shows a young boy talking to his father, who has his back to the boy and is about to hammer a nail into the wall. The second image shows the father crying out in pain at the moment he has hit his finger with the hammer and the little boy now with his back to his father's back and walking away. There is no verbal commitment to what is depicted visually.

Painter and colleagues (2013) consider the different affordances for meaning of the visual and verbal semiotics and the degree to which each commits meaning in a particular instance, as well as any convergent or divergent coupling relationships across the visual and verbal semiotics. Hence, rather than applying a limited taxonomy of image-language relations to a text, they provide a more open-ended but principled approach to exploring the ways instantiation of meaning from the two semiotic systems can construct new meanings at the intersection of image and language. While the focus here is on images and language in multimodal texts, the approach to theorizing about intersemiotic relations seems to have the capacity to embrace the differences in the semiotic workings and meaning-making potential of a range of different modes of meaning-making, such as images, language, intonation, and music.

Multimodal Literacy and Technology

The continual transformation of technologies for communication is associated with new potentials for configuring meaning differently (Jewitt, 2006). The exact nature of these meaning

potentials is the current focus of much research in relation to multimodality and new technologies, since the proliferation of digital technologies sets up transformed conditions and semiotic resources for multimodal interaction, from virtual reality to 3D printing. How these semiotic resources are used across different spaces (Mills & Comber, 2013, 2015), cultures (Mills, 2014), and communities of practices (Mills & Exley, 2014) is also a significant focus of multimodal literacy research.

Kress and van Leeuwen (2006) distinguish between three classes of production technologies: (a) technologies of the hand, such as pens, brushes, chisels, (b) recording technologies, such as audio-recording, photography, and film, and (c) technologies for synthesizing representations digitally, such as a keyboard and computer. There are slippages between these categories, and media that requires all three, such as a painting created with a brush that is later photographed and digitally synthesized and displayed in a digital presentation. There are also distribution technologies or media, which concerns the communication channel and the reproducibility of texts. Examples include the distribution of images via the Internet, which has made the dissemination of extraordinarily large numbers of texts and the connections between them possible. Some multimodal theorists, such as O'Halloran (2009), see that technologies also encompass organizational technologies—the agencies, organizations, and institutional bodies that regulate social action—such as governments, school systems, and corporations, and through which power relations are implicated in meaning-making.

New technologies do not always provide more potentials for meaning-making; rather, the potentials are different, and sometimes even constraining in certain ways (Mills, 2011a). For example, using pens or pencils, children can draw characters engaged in wider range of gravity-defying, imagined actions than are possible to create in certain comic creation software programs (Mills, 2011b). Similarly, a digital drawing application within a mobile device may have a more limited array of “paint tools” and functions than a professional artist has available, yet the technology affords a greater ability to “delete” virtual items than is possible on a physical canvas. The tools or crafts of communication have long- existed, and extend beyond digital or scientific technologies, such as computer software and hardware.

Multimodal Literacy and Writing

The very nature and constitution of writing is transformed in multimodal and digital communication environments. In the new textual ecologies of the 21st century, there is a greater multiplicity and circulation of multimodal and digital texts on the Internet and other media than ever before. With the ease of textual production on the web, and the increased accessibility of mobile and tablet devices, students are producers of material that is shared on the web beginning at much younger ages and with greater frequency than in previous generations. For example, social media platforms continue to make available new features that have seen text messaging and microblogging posts (short posts with a word count) expand from predominantly written text and hyperlinks to include combinations of emoticons, still images, gifs, memes, videos, and other media. The changes to the semiotic landscape, and our human predisposition toward multimodal meaning-making, requires that writing theorists attend to more than a single mode in writing curriculum and pedagogy. For writing educators, these changes call for in-depth knowledge of the semiotic relationships between the modes, the “division of labor” between them in representations, and the enabling and constraining potentials of the modes for different social purposes (Iedema, 2003).

It has been well established that children combine multimodal symbolic systems, such as talking, drawing, singing, and role-playing, long before their communicative interests can be served by the written linguistic forms of their culture (Kress & Bezemer, 2008; Siegel, 2006). Research on

children's composing processes within social semiotic frameworks has begun to focus on digital media, extending semiotic principles established in studies of print-based writing to the incorporation of multiple media in compositions (Ranker, 2009). These have included exploring sign-making in video- interaction (Adami, 2009), young writers' incorporation of multimedia into their writing as compositional elements (Dyson, 2001; Ranker, 2009), and young filmmakers' deployments of semiotic tools (Gilje, 2010). Others have researched the semiotic potentials of combining modes in digital storytelling (Hull & Nelson, 2005). These studies have contributed to understanding how children combine, shift, or transform meanings in multimodal contexts of digital composition (Mills, 2011a).

A sample of research on the changed dimensions of pedagogy in multimodal composition classrooms focuses on important social dimensions, such as power relations (Mills, 2007), classroom discourses (Mills, 2006b), situated practice (Mills, 2006c), transformed practice (Mills, 2008), time, space, and text (Mills & Exley, 2014), and architectonic meanings of the social and built space (Mills, 2010b). This body of work points to historically deep-seated issues of equity and marginalization for certain groups that are not remedied by introducing powerful forms of multimodal design in the curriculum. Socio-economically disadvantaged groups have vastly different levels of access to the semiotic resources, modes, and materials of written composition than economically privileged groups.

Another important body of research analyzes the multimodal textual products across a range of text types, such as children's animations (Burn & Parker, 2003), multimodal mapmaking in the classroom (Clark, 2011), hip hop as media literacies (Turner, 2012), complex claymation projects (Hepple, Sockhill, Tan, & Jennifer, 2014; Mills, 2011b), and critical multimodal literacies (Hamston, 2006; Mills, 2006a), to name a few. Future directions for multimodal composition research will include new understandings of the nature of multimodal composition across emergent text types, platforms, and social spaces for writing in an evolving digital communication environments, from social media to video gaming.

Researchers are asking new questions about digital ethics, politics, and online surveillance (Luke, Sefton-Green, Graham, Kellner, & Ladwig, 2018), epistemologies and multimodal practices (Mills, Davis-Warra, Sewell & Anderson, 2016), the assessment of multimodal writing (Unsworth & Chan, 2009), the changing materiality and embodiment in multimodal writing (Haas & McGrath, 2018), and the role of the senses in multimodal composition (Mills, Unsworth, & Exley, 2018). There is a multisensorial revolution in multimodal and digital practices that has renewed material features and affordances, and which requires continual sensorial adaptations in our cultural and social practices of communication (Mills, 2016).

Multimodality and Reading

Reading in the 21st century increasingly involves the integrative construction of meaning from print and images, with the inclusion of a growing variety of image types in an ever-widening range of paper and digital media texts (Bezemer & Kress, 2009; Kress, 1997; Kress & vanLeeuwen, 1995). The images are rarely gratuitous decoration or inconsequential additions, and are frequently integral to the interpretation of the text, whether in electronic or paper media (Rowell, Kress, Pahl, & Street, 2013). Although the fundamental principles of reading and writing have not changed, the process has shifted from the serial cognitive processing of linear print text to parallel processing of multimodal text-image information (Luke, 2003).

The need to reconceptualize reading comprehension to take account of the ways in which images and image-language interaction contribute to the meanings that can be made from texts is now

reflected in the national language curricula of many countries and is addressed in international reading tests, such as PISA (OECD, 2013), but has not yet achieved recognition in national literacy tests in Australia, the United States, or England. Notwithstanding, research indicates that different kinds of image-text relations vary in comprehension difficulty. Strategies for negotiating the comprehension of image-language relations distinguish among proficient and non-proficient readers (Unsworth & Chan, 2008, 2009). Since multimodality in relation to reading has, to date, principally concerned the integration of images and language, this section principally addresses the issues mentioned above in contexts of reading paper media texts. It concludes by noting additional factors concerning the integrative reading of images and language in digital texts.

The multimodal nature of reading is now recognized in government-mandated English or other national language curricula in a number of countries, including the United States, Canada, Australia, Singapore, and Sweden (ACARA, 2015; BCMOE, 2009; NYSED, 2012; Singapore, 2008; Sweden, 2009). For example, in the United States, documents such as the New York State Common Core Learning Standards for English Language Arts and Literacy (NYSED, 2012) address students' capacities to negotiate meaning-making through image-language interaction in their text comprehension, requiring students to "explain how specific aspects of a text's illustrations contribute to what is conveyed by the words in a story" (NYSED, 2012, p.18).

In the Australian Curriculum: English, the multimodal nature of the English curriculum is clearly established, since it "aims to ensure that students. . . learn to listen to, read, view, speak, write, create and reflect on increasingly complex and sophisticated spoken, written and multimodal texts across a growing range of contexts with accuracy, fluency and purpose" (ACARA, 2015, p. 3). Throughout all grade levels in this national curriculum, the content descriptions include explicit attention to the image-language interface in text interpretation (see Exley & Mills, 2012; Unsworth, 2014).

In the light of such explicit, mandated curriculum requirements for developing students' multimodal reading, one might expect that the national testing of reading would assess students' comprehension of meanings constructed in texts through image-language interaction (Unsworth, 2014; Unsworth & Chan, 2009); however, this would not appear to be the case in the United States, England, or Australia (Unsworth, 2017). For example, in England, past papers for the Standard Achievement Tests for the assessment of the national curriculum were available from the Assessment and Qualifications Authority (AQA) on its "Testbase" website [<http://www.testbase.co.uk/sec/pastpapers.php>]. An analysis of the reading tests for Key Stage 2 (grades 4–6) in 2013 and 2015 showed that only one test item in the entire 2013 Key Stage 2 test, and no items at all in the 2015 test, necessitated the reading of images. Similarly, in Australia, there has been minimal and decreasing attention to the role of images in reading comprehension in the National Assessment Program in Literacy and Numeracy (Unsworth, 2013, 2017; Unsworth & Chan, 2008, 2009).

The multimodal nature of the reading process in the 21st century has been well recognized by literacy researchers (Alvermann, Unrau, & Ruddell, 2013; Coiro, Knobel, Lankshear, & Leu, 2014), has been well documented in government-mandated curriculum documents in a number of countries, including Australia and the United States, and has been extensively taken up in classroom teaching (Callow, 2013; Mills, 2010a; Mills & Unsworth, 2016). Similarly, in the PISA international literacy tests, approximately 30% of test items involved images in what is referred to as "non-continuous text" (Unsworth, 2017).

However, in view of the undisputed pervasive impact of national literacy testing on curriculum implementation and pedagogic practice, this large-scale, high-stakes assessment of reading will need significant research-based development if schooling is to effectively embrace the reality of multimodal reading in today's world. Furthermore, the evolving nature of multimodal reading as a

social process entails many more issues than it is possible to canvass here. These include the distinctive ways in which images and language collaborate in online texts, including the reading of interpolated animated images (Chan & Unsworth, 2011), the multimodal nature of typography and its meaning-making interaction with language and image (Unsworth et al., 2015; van Leeuwen, 2006), and the recognition of sensory literacies and bodily engagement in multimodal reading in digital contexts (Mills, 2016).

Multimodal Literacy Curriculum and Pedagogy

Research on multimodal literacy curriculum and pedagogy is a fast-growing field. Researchers, educators, and curriculum makers see the need for transformed learning conditions to meet the multifarious cultural and technological contexts of students' current and future societal participation (Mills, 2015a). Meaningful and sophisticated communication for different social contexts can be achieved through knowledge of multimodal grammars or metalanguages that extend beyond linguistics to include spoken, visual, spatial, audio, gestural, haptic, olfactory, and other forms of meaning-making. Similarly, students today cannot escape, even very early in life, immersion in a plurality of texts of globalized communication environments. Children and youth in contemporary ecologies require advanced critical and technological skills and expertise with a broadened array of materials or media for meaningful knowledge, recreation, and communication (Mills, 2013a).

A proliferation of classroom studies, particularly since the beginning of the 21st century, have examined the application of new pedagogies in schools for multimodal literacy learning. Classroom pedagogies for teaching multimodal literacy have been researched in culturally diverse classrooms (McGuinnis, 2007; Stein, 2007), including the analysis of place and multimodality in children's filmmaking (Mills, Unsworth, Bellocchi, Park, & Ritchie, 2014), Indigenous multimodal pedagogies (Menezes de Souza, 2004; Mills, Davis-Warra, Sewell & Anderson, 2016), and English as a second language learning classrooms (Ajayi, 2009).

Across levels of education, there are studies of multimodality and higher education discourses (Mills, 2013b), multimodality and technology in schools (Jewitt, 2006), adolescent multimodal composition (Mills, 2010c), multimodal teacher learning (Miller, 2007), student multimodal responses to teacher learning (Bezemer, 2008), and default pedagogies in multimodal teaching contexts (Mills, 2005).

While multimodal literacy research has aimed to account for a widened array of modes that are taught to students as "schooled literacies," there are particular modes, such as olfactory, taste, haptics, and kinesics, and even sound, that have received less attention than the teaching of the verbal-visual couplet. This may be partly a consequence of ocular centrism—the dominance of the visual mode over other forms of communication—and with the Western empiricist prioritizing of truth or knowledge that can be observed with the eyes (Mills, 2016). For example, Howes (2009) has noted that there is a privileging of audio-olfactory communication in Melanesian contexts, compared to the elaboration of audio-visual communication, such as in film and television, in the Western context.

While there may be historical and pragmatic dimensions to the emphasis on logocentric forms of communication in the school curriculum, there is potential for a new paradigm shift in the way multimodal literacy research more seriously accounts for the role of the so-called "lower senses"—touch and movement, smell and taste—and their associated tactile, kinesic, olfactory, and gustatory modes of communication in literacy learning.

References

- Adami, E. (2009). "We/You Tube": Exploring sign-making in video-interaction. *Visual Communication*, 8(4), 379–399.
- Ajayi, L. (2009). English as a second language learners' exploration of multimodal texts in a junior high school. *Journal of Adolescent & Adult Literacy*, 52(7), 585–595.
- Alvermann, D. E., Unrau, N. J., & Ruddell, R. B. (2013). *Theoretical models and processes of reading* (Vol. 978). Newark, DE: International Reading Association.
- Andrews, R. (2004). Where next in research on ICT and literacies. *Literacy Learning: The Middle Years*, 12(1), 58–67.
- Australian Curriculum and Assessment Authority. (2015). The Australian Curriculum: English 8.2. Retrieved from <http://www.australiancurriculum.edu.au/copyright>.
- Bakhtin, M. (1986). *Speech genres and other late essays*. Austin, TX: University of Texas Press.
- Barthes, R. (1967). *Elements of semiology* (A. Lavers & C. Smith, Trans.). London, U.K.: Jonathan Cape.
- Barton, G., & Unsworth, L. (2014). Music, multiliteracies and multimodality: Exploring the book and movie versions of Shaun Tan's *The Lost Thing*. *Australian Journal of Language and Literacy*, 37(1), 3–20.
- Bateman, J. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Abingdon, U.K.: Taylor and Francis.
- Bernstein, B. (1971). *Class, codes and control*. London, U.K.: Routledge and Kegan Paul.
- Bezemer, J. (2008). Displaying orientation in the classroom: Students' multimodal responses to teacher instructions. *Linguistics and Education*, 19(2), 166–178.
- Bezemer, J., & Kress, G. (2009). Visualizing English: A social semiotic history of a school subject. *Visual Communication*, 8(3), 247–262.
- Birdwhistell, R. (1952). *Introduction to kinesics: An annotation system for analysis of body motion and gesture*. Louisville, KY: University of Louisville.
- Bogatyrev, P. (1976). Costume as a sign. In L. Matejka, & J. R. Titunic (Eds.), *Semiotics of art*, (pp. 12–20). Cambridge, MA: Prague School Contributions.
- British Columbia Ministry of Education. (2009). *English language arts kindergarten to grade 7: Integrated resource package*. Vancouver, Canada: Ministry of Education, Province of British Columbia.
- Browne, A. (1983). *Gorilla*. London, U.K.: Julia MacRae Books.
- Browne, A. (1989). *The tunnel*. London, U.K.: Julia McRae Books.
- Browne, A. (1994). *Zoo*. London, U.K.: Random House.
- Burn, A., & Parker, D. (2003). Tiger's big plan: Multimodality and the moving image. In C. Jewitt & G. Kress (Eds.), *Multimodal literacy* (pp. 56–72). New York, N.Y.: Peter Lang.
- Burningham, J. (1984). *Granpa*. London, U.K.: Penguin/Puffin.

- Callow, J. (2013). *The shape of text to come: How image and text work*. Newtown, N.S.W: Primary English Teaching Association of Australia.
- Chan, E., & Unsworth, L. (2011). Image-language interaction in online reading environments: Challenges for students' reading comprehension. *Australian Educational Researcher*, 38(2), 181–202.
- Clark, A. (2011). Multimodal map making with young children: Exploring ethnographic and participatory methods. *Qualitative Research*, 11(3), 311–330.
- Cohn, N. (2013a). Beyond speech balloons and thought bubbles: The integration of text and image. *Semiotica*, 2013(197), 35–63.
- Cohn, N. (2013b). Navigating comics: An empirical and theoretical approach to strategies of reading comic page layouts. *Frontiers in Psychology*, 4, 186.
- Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (2014). *Handbook of research on new literacies*. London, U.K.: Routledge.
- Dyson, A. H. (2001). Donkey Kong in Little Bear country: A first grader's composing development in the media spotlight. *The Elementary School Journal*, 101(4), 417–433.
- Exley, B., & Mills, K. A. (2012). Parsing the Australian English curriculum: Grammar, multimodality and cross-cultural texts. *Australian Journal of Language and Literacy*, 35(1), 192–205.
- Fairclough, N. (1989). *Language and power*. London, U.K.: Longman.
- Gilje, O. (2010). Multimodal redesign in filmmaking practices: An inquiry of young filmmakers' deployment of semiotic tools in their filmmaking practice. *Written Communication*, 27(4), 494–522.
- Goffman, E. (1959). *The presentation of self in everyday life*. New York, N.Y.: Doubleday.
- Haas, C., & McGrath, M. (2018). Embodiment and literacy in a digital age: The case of handwriting. In K. A. Mills, A. Stornaiuolo, A. Smith, & J. Z. Pandya (Eds.), *Writing, literacies and education in digital cultures*. New York: Routledge.
- Hall, E. T. (1959). *The silent language*. Garden City, New York: Doubleday.
- Halliday, M. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London, U.K.: Edward Arnold.
- Halliday, M. A. K., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd ed.). London, U.K.: Arnold.
- Hamston, J. (2006). Pathways to multiliteracies: Student teachers' critical reflections on a multimodal text. *The Australian Journal of Language and Literacy*, 29(1), 38–51.
- Hepple, E., Sockhill, M., Tan, A., & Jennifer, A. (2014). Multiliteracies pedagogy: Creating claymations with adolescent, post-beginner English language learners. *Journal of Adolescent & Adult Literacy*, 58(3), 219–229.
- Hodge, B., & Kress, G. (1988). *Social semiotics*. London, U.K.: Polity Press.
- Honzl, J. (1976). Dynamics of the sign in the theater. In L. Matejka, & J. R. Titunic (Eds.), *Semiotics of art*, (pp. 74–93). Cambridge, MA: Prague School Contributions.
- Hood, S. (2008). Summary writing in academic contexts: Implicating meaning in processes of change. *Linguistics and Education*, 19(4), 351–365.

- Howes, D. (2009). Anthropology and multimodality: The conjugation of the senses. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (2nd ed., pp. 225–235). London, U.K.: Routledge.
- Hull, G., & Nelson, M. (2005). Locating the semiotic power of multimodality. *Written Communication*, 22(2), 224–261.
- Iedema, R. (2003). Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice. *Visual Communication*, 2(1), 29–57.
- Jeffers, O. (2009). *Heart and the bottle*. Hammersmith: Harper Collins.
- Jewitt, C. (2006). *Technology, literacy and learning: A multimodal approach*. Abingdon, U.K.: Routledge.
- Jewitt, C. (2014). An introduction to multimodality. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis*. London, U.K.: Routledge.
- Kress, G. (1997). Visual and verbal modes of representation in electronically mediated communication: The potentials of new forms of text. In I. Snyder (Ed.), *Page to screen: Taking literacy into the electronic era* (pp. 53–79). Sydney: Allen and Unwin.
- Kress, G. (2000). Multimodality. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures* (pp. 182–202). South Yarra, VIC: Macmillan.
- Kress, G. (2003a). Genres and the multimodal production of “scientificness.” In C. Jewitt & G. Kress (Eds.), *Multimodal literacy* (pp. 173–186). New York, N.Y.: Peter Lang.
- Kress, G. (2004). *Literacy in the new media age*. London, U.K.: Routledge.
- Kress, G., & Bezemer, J. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication*, 25(2), 166–195.
- Kress, G., & van Leeuwen, T. (1995). Critical layout analysis. *Internationale Schulbuchforschung*, 17(1), 25–43.
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). London: Routledge.
- Leeds-Hurwitz, W. (1987). The social history of the natural history of an interview: A multidisciplinary investigation of social communication. *Research on Language and Social Interaction*, 20, 1–51.
- Lemke, J. (1998). Metamedia literacy: Transforming meanings and media. In D. Reinking, M. McKenna, L. Labbo, & R. Kieffer (Eds.), *Handbook of literacy and technology: Transformations in a post-typographic world* (pp. 283–302). New Jersey, N.J.: Erlbaum.
- Luke, A., Sefton-Green, J., Graham, P., Kellner, D., & Ladwig, J. (2018). Digital ethics, political economy and the curriculum: This changes everything. In K. A. Mills, A. Stornaouiolo, A. Smith, & J. Z. Pandya (Eds.), *Writing, literacies, and education in digital cultures*. New York, N.Y.: Routledge.
- Luke, C. (2003). Pedagogy, connectivity, multimodality and interdisciplinarity. *Reading Research Quarterly*, 38(10), 356–385.
- Martin, J. R. (1992). *English text: System and structure*. The Netherlands: Benjamins.

- Martin, J. R. (2008a). Innocence: Realisation, instantiation and individuation in a Botswanan town. In N. Knight & A. Mahboob (Eds.), *Questioning linguistics* (pp. 27–54). Cambridge, U.K.: Cambridge Scholars Publishing.
- Martin, J. R. (2008b). Intermodal reconciliation: Mates in arms. In L. Unsworth (Ed.), *New Literacies and the English curriculum* (pp. 112–148). London, U.K.: Continuum.
- Martin, J. R. (2008c). Tenderness: Realisation and instantiation in a Botswanan town. In N. Nørgaard (Ed.), *Systemic functional linguistics in use* (Odense Working Papers in Language and Communication 29) (, pp. 30–62). Odense, DE: University of Southern Denmark, Institute of Language and Communication.
- Martin, J. R. (2010). Semantic variation: Modelling system, text and affiliation in social semiosis. In M. Bednarek & J. R. Martin (Eds.), *New discourse on language: Functional perspectives on multimodality, identity and affiliation* (pp. 134). London, U.K.: Continuum.
- Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause* (2nd ed., Vol. 1). London, U.K.: Continuum.
- Martin, J. R., & White, P. (2005a). *The language of evaluation: Appraisal in English*. London, U.K.: Palgrave/Macmillan.
- Martinec, R. (2004). Gestures that co-occur with speech as a systematic resource: The realization of experiential meanings in indexes. *Social Semiotics*, 14(2), 193–213.
- Martinec, R. (2013). Nascent and mature uses of a semiotic system: The case of image-text relations. *Visual Communication*, 12(2), 147–172.
- Martinec, R., & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication*, 4(3), 337–371.
- McCloud, S. (1994). *Understanding comics: The invisible art*. New York, N.Y.: Harper Collins.
- McGuinnis. (2007). Khmer rap boys, X-Men, Asia’s fruits, and Dragonball Z: Creating multilingual and multimodal classroom contexts. *Journal of Adolescent & Adult Literacy*, 50(7), 570–579.
- McKee, D. (1987). *Not now, Bernard*. London, U.K.: Arrow.
- Menezes de Souza, L. (2004). The ecology of writing among the Kashinowa: Indigenous multimodality in Brazil. In S. Canagarajah (Ed.), *Reclaiming the local in language policy and practice* (pp. 73–98). New York, N.Y.: Psychology Press.
- Miller, S. (2007). English teacher learning for new times: Digital video composing as multimodal literacy practice. *English Education*, 40(1), 61–83.
- Mills, K. (2010a). *The multiliteracies classroom* (Vol. 21). Bristol, U.K.: Multilingual Matters.
- Mills, K., & Unsworth, L. (2016). The literacy curriculum: A critical review. In D. Wyse, L. Hayward, & Z. Pandya (Eds.), *The Sage handbook of literacy, pedagogy and assessment* (Vol. 2, pp. 621–637). Thousand Oaks, CA: SAGE.
- Mills, K. A. (2005). *Multiliteracies: Remnant discourses and pedagogies*. Paper presented at the Australian Literacy Educator’s Association/Australian Association of the Teaching of English National Conference 2005: Pleasure, Passion, Provocation, Broadbeach Queensland. Retrieved from <https://eprints.qut.edu.au/2966/>.
- Mills, K. A. (2006a). *Critical framing in multiliteracies*. Paper presented at the Australian Literacy Educator’s Association/Australian Association of the Teaching of English National Conference

- 2006: Voices, Vibes, Visions, Darwin, Northern Territory. Retrieved from <https://eprints.qut.edu.au/4844/>.
- Mills, K.A. (2006b). Mr. Travelling-at-will Ted Doyle: Discourses in a multiliteracies classroom. *Australian Journal of Language and Literacy*, 28(2), 132–149.
- Mills, K. A. (2006c). We've been wastin' a whole million watchin' her doin' her shoes: Situated practice within a pedagogy of multiliteracies. *The Australian Educational Researcher*, 33(3), 13–34.
- Mills, K. A. (2007). Have you seen Lord of the Rings? Power, pedagogy and discourses in a multiliteracies classroom. *Journal of Language, Identity, and Education*, 6(3), 221–241.
- Mills, K. A. (2008). Transformed practice in a pedagogy of multiliteracies. *Pedagogies: An International Journal*, 3(2), 109–128.
- Mills, K. A. (2010b). Filming in progress: New spaces for multimodal designing. *Linguistics and Education*, 21(1), 14–28.
- Mills, K. A. (2010c). Shrek meets Vygotsky: Rethinking adolescents' multimodal literacy practices in schools. *Journal of Adolescent and Adult Literacy*, 54(1), 35–45.
- Mills, K. A. (2011a). "I'm making it different to the book": Transmediation in young children's print and digital practices. *Australasian Journal of Early Childhood Education*, 36(3), 56–65.
- Mills, K. A. (2011b). *The multiliteracies classroom*. Bristol, U.K.: Multilingual Matters.
- Mills, K. A. (2013a). CUOL—See you online. *Screen Education*, 70, 52–57.
- Mills, K. A. (2013b). *Multimodal and monomodal discourses of marketization in higher education: Power, ideology, and the absence of the image*. Paper presented at the Education and Poverty: Theory, Research, Policy and Praxis: Proceedings of AERA Annual Meeting 2013, San Francisco, CA. Retrieved from <https://eprints.qut.edu.au/52711/>.
- Mills, K. A. (2014). *Cultural flows in an Aboriginal school: Deterritorialising textual production through a socially mediated Indigenous heritage*. Paper presented at the American Educational Research Association: The Power of Education Research for Innovation in Practice and Policy, Philadelphia, PA. Retrieved from <https://eprints.qut.edu.au/68220/>.
- Mills, K. A. (2015a). Doing digital composition on the social web: Knowledge Processes processes in literacy learning. In B. Cope & M. Kalantzis (Eds.), *A pedagogy of multiliteracies: Learning by design* (pp. 172–185). New York, NY: Palgrave/, Macmillan.
- Mills, K. A. (2016). *Literacy theories for the digital age: Social, critical, multimodal, spatial, material and sensory lenses*. Bristol, U.K.: Multilingual Matters.
- Mills, K. A., Chandra, V., & Park, J. (2013). The architecture of children's use of language and tools when problem solving collaboratively with robotics. *Australian Education Researcher*, 40(3), 315–337.
- Mills, K. A., & Comber, B. (2013). Space, place, and power: A spatial turn in literacy research. In K. Hall, T. Cremin, B. Comber, & L. Moll (Eds.), *International handbook of research on children's literacy, learning and culture* (pp. 412–423). Oxford, U.K.: Wiley Blackwell.
- Mills, K. A., & Comber, B. (2015). Socio-spatial approaches to literacy studies: Rethinking the social constitution and politics of space. In K. Pahl & J. Rowsell (Eds.), *Handbook of literacy studies* (pp. 91–103). London, U.K.: Routledge.

- Mills, K. A., Davis-Warra, J., Sewell, M., & Anderson, M. (2016). Indigenous ways with literacies: Transgenerational, multimodal, placed and collective. *Language and Education*, 30(1), 1–21.
- Mills, K. A., & Exley, B. (2014). Time, space, and text in the elementary school digital writing classroom. *Written Communication*, 31(4), 368–398.
- Mills, K. A., Unsworth, L., Bellocchi, A., Park, J., & Ritchie, S. M. (2014). Children's multimodal appraisal of places: Walking with the camera. *Australian Journal of Language and Literacy*, 37(3), 171–181.
- Mills, K. A., Unsworth, L., & Exley, B. (2018). Sensory literacies, the body and digital media. In K. A. Mills, A. Stornaoulo, A. Smith, & J. Z. Pandya (Eds.), *Writing, literacies and education in digital cultures*. New York: Routledge.
- Morgan, S., & Bancroft, B. (1996). *Dan's grandpa*. Freemantle, W.A.: Sandcastle.
- Mukarovsky, J. (1976). Art as semiotic fact. In L. Matejka, & J. R. Titunic (Eds.), *Semiotics of art*, (pp. 3–9). Cambridge, MA: Prague School Contributions.
- New London Group. (1996). *A pedagogy of multiliteracies: Designing social futures*. Harvard Educational Review, 66(1), 60–92.
- New York, State Education Department, (2012). *New York State P–12 common core learning standards for English language arts & literacy*. Retrieved from <http://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards-for-english-language-arts-and-literacy>.
- Norman, L., & Young, N. (1998). *Grandpa*. Sydney, N.S.W.: Margaret Hamilton Books.
- Norris, S. (2011). Modal density and modal configurations: Multimodal actions. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 78–90). London: Routledge.
- O'Halloran, K. (2004). *Multimodal discourse analysis: Systemic-functional perspectives*. London, U.K.: Continuum.
- O'Halloran, K. (2005). *Mathematical discourse: Language, symbolism and visual images*. London, U.K.: Continuum.
- O'Halloran, K. (2008). Systemic functional-multimodal discourse analysis: Constructing ideational meaning using language and visual imagery. *Visual Communication*, 7(4), 443–475.
- O'Halloran, K. L. (2009). Historical changes in the semiotic landscape: From calculation to computation. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 98–113). London, U.K.: Routledge.
- O'Toole, M. (1994). *The language of displayed art*. London, U.K.: Leicester University Press.
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Organisation for Economic Co-operation and Development.
- Painter, C., & Martin, J. R. (2011). Intermodal complementarity: Modelling affordances across image and verbiage in children's picture books. In F. Yan (Ed.), *Studies in functional linguistics and discourse analysis* (pp. 132–158). Beijing: Education Press of China.
- Painter, C., Martin, J. R., & Unsworth, L. (2013). *Reading visual narratives: Image analysis of children's picture books*. Sheffield: Equinox Publishing

- Peirce, C. S. (1934). *Collected papers: Volume V. Pragmatism and pragmaticism*. Cambridge, MA: Harvard University Press.
- Pike, K. (1954). *Language in relation to a unified theory of the structure of human behaviour*. Glendale, CA: Summer Institute of Linguistics.
- Ranker, J. (2009). Redesigning and transforming: A case study of the role of the semiotic import in early composing processes. *Journal of Early Childhood Literacy*, 9(3), 319–347.
- Roth, W., Pozzer-Ardhenghi, L., & Han, J. (2005). *Critical graphicacy: Understanding visual representation practices in school science*. Dordrecht, The Netherlands: Springer.
- Rowell, J., Kress, G., Pahl, K., & Street, B. (2013). The social practice of multimodal reading: A new literacy studies-multimodal perspective on reading. In D. Alvermann, N. Unrau, & R. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 1182–1207). International Reading Association.
- Ruesch, J., & Bateson, G. (1951). *Communication: The social matrix of psychiatry*. London, U.K.: Transaction Publishers.
- Scollon, R., & Scollon, S. W. (2011). Multimodality and language: A retrospective and prospective view. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 170–180). London, U.K.: Routledge.
- Sebeok, T. A. (1976). *Contributions to the doctrine of signs*. Bloomington, IN: Indiana University Press.
- Semali, L. M., & Fueyo, J. (2001). Transmediation as a metaphor for new literacies in multimedia classrooms. *Reading Online*, 5(5).
- Siegel, M. (2006). Rereading the signs: Multimodal transformations in the field of literacy education. *Language Arts*, 84(1), 65.
- Singapore, Ministry of Education. (2008). *English language syllabus 2010 primary & secondary*. Singapore: Ministry of Education.
- Stein, P. (2007). *Multimodal pedagogies in diverse classrooms: Representation, rights, and resources*. Abingdon, U.K.: Routledge.
- Sweden, N. A. F. E. (2009). *Syllabuses for the compulsory school*. Stockholm, Sweden
- Tan, S. (2000). *The lost thing*. Sydney, N.S.W.: Hachette.
- Tan, S. (2013). *Rules of summer*. Sydney: N.S.W.: Hachette.
- Turner, K. C. N. (2012). Multimodal hip hop productions as media literacies. *The Educational Forum*, 76(4), 497–509.
- Unsworth, L. (2001). *Teaching multiliteracies across the curriculum: Changing contexts of text and image in classroom practice*. Buckingham, U.K.: Open University Press.
- Unsworth, L. (2013). Interfacing comprehension of image-language interaction in state-wide reading texts and semiotic accounts of image-language relations. In C. Gouveia & M. Alexandre (Eds.), *Languages, metalanguages, modalities, cultures: Functional and socio-discursive perspectives*. (pp. 177–198). Lisbon, Portugal: Books on Demand/Instituto de Linguística Teórica e Computacional (ILTEC).

- Unsworth, L. (2014). Multimodal reading comprehension: curriculum Curriculum expectations and large-scale literacy testing practices. *Pedagogies: An International Journal*, 9(1), 26–44.
- Unsworth, L. (2015). Persuasive narratives: Evaluative images in picture books and animated movies [<http://dx.doi.org/10.1177/1470357214541762>]. *Visual Communication*, 14(1), 73–96.
- Unsworth, L. (2017). Image-language interaction in text comprehension: Reading reality and national reading tests. In C. Ng & B. Bartlett (Eds.), *Improving reading in the 21st century: International research and innovations*. Dordrecht, The Netherlands: Springer.
- Unsworth, L., & Chan, E. (2008). Assessing integrative reading of images and text in group reading comprehension tests. *Curriculum Perspectives*, 28(3), 71–76.
- Unsworth, L., & Chan, E. (2009). Bridging multimodal literacies and national assessment programs in literacy. *Australian Journal of Language and Literacy*, 32(3), 245–257.
- Unsworth, L., & Cleirigh, C. (2009a). Multimodality and reading: The construction of meaning through image-text interaction. In C. Jewitt (Ed.), *Handbook of multimodal analysis* (pp. 151–164). London, U.K.: Routledge.
- Unsworth, L., & Cleirigh, C. (2009b). Towards a relational grammar of image-verbiage synergy: Intermodal representations. In S. Dreyfus, S. Hood, & M. Stenglin (Eds.), *Semiotic margins*. University of Sydney: Australian Systemic Functional Linguistics Association [<http://www.asfla.org.au/>].
- Unsworth, L., Meneses, A., Ow, M., & Castillo, G. (2015). Analyzing the semiotic potential of typographic resources in picture books in English and in translation. *International Research in Children's Literature*, 7(2), 117–135.
- van Leeuwen, T. (1999). *Speech, music, sound*. London, U.K.: Macmillan.
- van Leeuwen, T. (2005). *Introducing social semiotics*. London, U.K.: Routledge.
- van Leeuwen, T. (2006). Towards a semiotics for typography. *Information Design Journal*, 14(2), 139–155.
- White, P. R. R. (2014). The attitudinal work of news journalism images: A search for visual and verbal analogues. *Quaderni del CeSLiC Occasional Papers del CeSLiC*, 6–42.