If the central task of art history is the study of visual images, the issue of “word and image” focuses attention on the relation of visual representation to language. More broadly, “word and image” designates the relation of art history to literary history, textual studies, linguistics, and other disciplines that deal primarily with verbal expression. Even more generally, “word and image” is a kind of shorthand name for a basic division in the human experience of representations, presentations, and symbols. We might call this division the relation between the seeable and the sayable, display and discourse, showing and telling (Foucault 1982; Deleuze 1988; Mitchell 1994).

Consider, for instance, the words you are reading at this moment. They are (one hopes) intelligible verbal signs. You can read them aloud, translate them into other languages, interpret or paraphrase them. They are also visible marks on the page, or (if read aloud) audible sounds in the air. You can see them as black marks on a white background, with specific shapes, sizes, and locations; you can hear them as sounds against a background of relative silence. In short, they present a double face to both the eye and the ear: one face is that of the articulate sign in a language; the other is that of a formal visual or aural gestalt, an optical or acoustical image. Normally we look only at one face and ignore the other: we don’t pay much attention to the typography or graphic look of a text; we don’t listen to the sounds of words, preferring to concentrate on the meaning they convey. But it is always possible to shift our attention, to let those black marks on a white background become objects of visual or aural attention, as in this self-referential example. We are encouraged to do this by poetic or rhetorical uses of language that foreground the sounds of words, or artistic, ornamental uses of writing (e.g., illuminated manuscripts, calligraphy) that foreground the visual appearance of letters. But the potential for the shift “from word to image” is always there, even in the most spare, unadorned forms of writing and speech.

A similar potential resides in visual images. In the act of interpreting or describing pictures, even in the fundamental process of recognizing what they represent, language enters into the visual field. Indeed, so-called “natural” visual experience of the world, quite apart from the viewing of images, may be much like a language. The philosopher George Berkeley (1709) argued that eyesight is a “visual language,” a complex, learned technique that involves the coordination of visual and tactile sensations. Modern neuropsychologists like Oliver Sacks (1993) have confirmed Berkeley’s theory, showing that people who have been blinded for an extended period of time have to relearn the cognitive techniques of seeing, even when the physical structure of the eye has been fully repaired. As a practical matter, the recognition of what visual images represent, even the recognition that something is an image, seems possible only for language-using animals. The famous image game of the duck-rabbit illustrates the intimate and intricate interplay of words and images in the perception of a visual image. Being able to see both the duck and the rabbit, to see them shift back and forth, is possible only for a creature that is able to coordinate pictures and words, visual experience and language (Wittgenstein 1953).

“Word and image” has become something of a hot topic in contemporary art history, largely because of what are often seen as invasions of the visual arts by literary theory. Scholars like Norman Bryson, Mieke Bal, Michael Fried, Wendy Steiner, and many others (myself included) have been spotted crossing the borders from departments of literature into art history. These scholars bring along methods and terms developed initially in the study of texts: semiotics, structural linguistics, grammatology, discourse analysis, speech-act theory, rhetoric, and narrative theory (to name only a few examples).
Not surprisingly, the border police are on the alert to protect the territory of art history from colonization by literary imperialism. Even an adventurous, wide-ranging art historian like Thomas Crow gives in to a defensive “art historian's reflex” when he sees scholars from “text-based academic disciplines” moving into the study of visual art (Crow 1994, 83). This sort of defensiveness might seem strange, given the intimate relations between word and image we have just observed in a pair of textual and pictorial examples. It seems even stranger when we reflect on the intense interest of great art historians like Erwin Panofsky in philology and literature. The very name of Panofsky's science of image analysis, “iconology,” contains a suturing of the image (icon) with the word (logos). What is art history, after all, if not an attempt to find the right words to interpret, explain, describe, and evaluate visual images?

Insofar as art history aims to become a critical discipline, one that reflects on its own premises and practices, it cannot treat the words that are so necessary to its work as mere instrumentalities in the service of visual images or treat images as mere grist for the mill of textual decoding. It must reflect on the relation of language to visual representation and make the problem of “word and image” a central feature of its self-understanding. Insofar as this problem involves borders between “textual” and “visual” disciplines, it ought to be a subject of investigation and analysis, collaboration and dialogue, not defensive reflexes.

There is one dimension of art-historical defensiveness that makes good sense, then, and that is the resistance to the notion that vision and visual images are completely reducible to language. One of the more depressing sights in contemporary art history is the rush to fix on some master term (discourse, textuality, semiosis, and culture come to mind) that will solve the mystery of visual experience and representation and dissolve the difference between word and image. The maintenance or even policing of this border is a useful task when it is conducted in a spirit of respect for difference. G. E. Lessing's words bear repeating here:

Painting and poetry should be like two just and friendly neighbors, neither of whom indeed is allowed to take unseemly liberties in the heart of the other's domain, but who exercise mutual forbearance on the borders, and effect a peaceful settlement for all the petty encroachments which circumstance may compel either to make in haste on the rights of the other.

(Lessing 1766, 116)

The domains of word and image are like two countries that speak different languages but that have a long history of mutual migration, cultural exchange, and other forms of intercourse. The word/image relation is not a master method for dissolving these borders or for maintaining them as eternally fixed boundaries; it is the name of a problem and a problematic—a description of the irregular, heterogeneous, and often improvised boundaries between “institutions of the visible” (visual arts, visual media, practices of display and spectation) and “institutions of the verbal” (literature, language, discourse, practices of speech and writing, audition and reading).

The relation between words and images is an extraordinarily ancient problem in the study of the arts and in theories of rhetoric, communication, and human subjectivity. In the arts, the comparison of poetry and painting, literature and visual art has been a consistent theme since antiquity in both Eastern and Western aesthetics. The casual remark of the Roman poet Horace “ut pictura poesis” (as is painting, so is poetry) became the foundation for one of the most enduring traditions in Western painting and has served as a touchstone for comparisons of the “sister arts” of word and image ever since. Aristotle's theory of drama includes a careful gauging of the relative importance of lexis (speech) and opsis (spectacle) in tragedy. Theories of rhetoric routinely appeal to the model of word/image conjunctions to define the relation between argument and evidence, precept and example, verbum (word) and res (thing, substance). Effective rhetoric is characteristically defined as a two-pronged strategy of verbal/visual persuasion, showing while it tells, illustrating its claims with powerful examples, making the listener see and not merely hear the orator's point. Ancient theories of memory regularly describe it as a technique of coordinating a sequence of words with a structure of visible places and images, as if the mind were a wax tablet inscribed with images and words, or a temple or museum filled with statues, paintings, and inscriptions (Yates 1966).

Contemporary culture has made the interplay of word and image even more volatile, intricate, and pervasive. Whatever else movies may be, they are clearly
complex suturings of visual images and speech, sight and sound, and (especially in the silent era) image and writing. The transformation of visual and verbal identity we saw in the example of the duck-rabbit is multiplied many times in the digital manipulation of electronic images, the “morphing” which shifts rapidly through a series of racial and gender types in the videos of Michael Jackson or a Gillette shaving cream commercial. Any child nurtured on the alphanumeric soup of Sesame Street knows that letters are visible signs and words may turn into images and back again at the flash of a “silent E.” If ancient memory systems had their illustrated wax tablets and art-filled temples, modern memory technologies coordinate streams of digital and analog information within a virtual electronic architecture, converting images to texts and vice versa. Although one of the central impulses of artistic modernism in the twentieth century has been, as Clement Greenberg argued, to explore the distinctness and difference of verbal and visual media, seeking a purely optical painting and a purely verbal poetry, the larger culture has been dominated by the aesthetics of kitsch, which freely mixes and adulterates the media.

What is it about the construction of the human mind that makes the interplay of words and images seem, despite innumerable historical and regional variations, to be something like a cultural universal? One might appeal to the hemispheric structure of the brain, with its divisions between visual, spatial, intuitive functions and verbal, sequential processes of reasoning. One might adopt a psychoanalytic account of the formation of subjectivity as a progression from an imagistic “mirror stage” in infancy to a symbolic, verbally constructed self in maturity. Or one might prefer a theological explanation that looks to the recurrent accounts of the creation of the human species as both image and word of the creator, the sculpting of Adam and Eve as clay vessels from the earth, and the breathing of spirit into them, making them not only “images” of their creator, but living, speaking emanations of the Word. In my view, these are not so much “explanations” of the word/image phenomenon, as highly general, mythic instantiations of it. They are foundational cultural narratives that turn the categories of word and image into something like characters in a drama that is subject to infinite variation, historical transformation, and geographical dislocation. It is stories like these that make the relations of word and image something more than a merely technical matter of distinguishing different kinds of signs and associate them with deeply felt values, interests, and systems of power. Before we go further with these broader issues, however, it might be useful to examine a bit more closely just what the relation of words and images is, how it is usually defined, and why it plays such a pervasive and volatile role in discussions of art, media, and consciousness.

Much of the power and interest of the word/image relation comes from its deceptive simplicity. What could be more straightforward than the distinction between a picture of a tree and the word “tree”?

As a practical matter, we have no trouble in saying which is the word, which is the image. The problem comes when we try to explain the difference, to define the precise features that make one sign a word, the other an image. One common explanation would base the difference in the sensory “channel” appropriate to each kind of sign. The word is a phonetic sign: it is meant to be read aloud or subvocalized and “heard” as an acoustical event. The image is a visual sign: it represents the visual appearance of an object. The difference between word and image is simply the difference between hearing and seeing, speaking and depicting.

The clarity of this distinction is less secure than it might seem at first glance. We do, after all, see the written word “tree,” and the word refers us to a class of visible objects, the same class that the image designates. And it's not entirely clear that we simply “see” the tree represented by the image. We could easily see these marks as something else—as an arrowhead or a pointer indicating a direction. To see this as an image of a tree means assigning that label to it, giving it that name. If we were seeing this image in the context of a pictographic or hieroglyphic inscription, we might discover a whole range of symbolic connotations: the image seen as a tree could refer to a whole forest, or to associated concepts like growth and fertility; seen as an arrowhead it could be a sign for war or hunting, or for the warrior or hunter. The image might even lose all connections with the visual appearance of a tree, and become a phonetic sign, indicating the syllabic unit “tree,” so that it would be usable in a rebus like the following:
At this point the image is well on its way into the domain of language, becoming part of a phonetic writing system. This doesn’t mean there is no difference between words and images, only that the difference is not simply traceable to the difference between seeing and hearing. We can see words and hear images; we can read pictures and scan the visual appearance of texts. The difference between word and image cuts across the difference between visual and aural experience.

It might seem, then, that the difference between words and images is not built into our sensory apparatus or inherent in different kinds of symbolic forms, but has to do with different ways of coordinating signs with what they stand for. Images, we might say, signify by virtue of resemblance or imitation: the image of the tree looks like a tree. Words, by contrast, are arbitrary signs that signify by virtue of custom or convention. This is one of the most enduring accounts of the word/image difference, cropping up as early as Plato's Cratylus and recurring throughout the history of theories of representation. It has the added virtue of explaining why images are not necessarily visual, why there can be things like sound images. Resemblance is an extraordinarily general relation, one that can function in any sensory channel and connect any number of perceptual experiences.

The problem, in fact, is that resemblance applies far too generally to be of much use in picking out what is special about visual images. One tree may resemble another tree, but that doesn’t mean that one tree is the image of the other. Many things resemble each other without being images of one another. It may be that resemblance is a necessary condition for something to be an image, but it certainly is not sufficient. Something else is required: the image has to denote or represent what it stands for; merely looking like it isn’t enough. There is also the problem that many images don’t look much like anything in particular except themselves. Many things we would want to call visual images (the formal patterns in ornament, the array of shapes and colors in abstract paintings) don’t resemble things in the visual world nearly as closely as they resemble each other.

The theory, then, that images are copies of things, that they signify by resemblance fails on two counts: on the one hand, it cannot explain the existence of images that do not resemble or represent anything; on the other hand, it identifies only a necessary, not a sufficient condition for images that do both resemble and represent something. It seems that once again for images to do their work, they have to intersect with the domain of language, this time by appealing to the role of custom and convention. The image of the tree signifies a tree, not just because it resembles it, but because a social agreement or convention has been established that we will “read” this sign as a tree. The abstract or ornamental image that resembles and represents nothing is seen as an image because it functions like an image in a social practice. The image in this sense is not a representation, but a representative sample. It is a visual form that has meaning, even if it doesn’t represent anything.

The straightforward, practical difference between words and images turns out to be much more complicated than it looked at first glance. In fact, the situation threatens to become thoroughly paradoxical. We began with what looked like an obvious and clear difference, and yet the more we tried to give a theoretical explanation of that difference, the shakier it became. The sensory division of eye and ear both aligns itself with and cuts across the boundary between word and image, most notably in the phenomenon of writing or “visible language.” The semiotic distinction between signs by convention and signs by resemblance also unravels as we pull at it. Words (like “quack”) can resemble what they represent; images are riddled with convention, could not exist without conventions, and they need not represent anything.

My inability to discover a firm, unequivocal basis for the distinction between words and images doesn’t mean, of course, that there aren’t any real distinctions to be observed. And it also doesn’t mean that issues like resemblance, convention, and the visual/aural division are irrelevant. What it does suggest is that the word/image difference is not likely to be definitively stabilized by any single pair of defining terms or any static binary opposition. “Word and image” seems to be better understood as a dialectical trope. It is a trope, or figurative condensation of a whole set of relations and distinctions, that crops up in aesthetics, semiotics, accounts of perception, cognition, and communication, and analyses of media (which are characteristically “mixed”
forms, “imagetexts” that combine words and images). It is a dialectical trope because it resists stabilization as a binary opposition, shifting and transforming itself from one conceptual level to another, and shuttles between relations of contrariety and identity, difference and sameness. We might summarize the predicates that link word and image with an invented notation like “vs/as”; “word vs. image” denotes the tension, difference, and opposition between these terms; “word as image” designates their tendency to unite, dissolve, or change places. Both these relations, difference and likeness, must be thought of simultaneously as a vs/as in order to grasp the peculiar character of this relationship.

If we were to go on with the search for figures of the difference between words and images, we would have to complicate the eye/ear and resemblance/convention distinctions even further, coordinating them with Lessing's classic argument that the categories of space and time (images seen in space; words read in time) provide the most fundamental basis. We would have to take up Nelson Goodman's distinction between “dense” and “differentiated” signs, images understood as dense analog symbols in which a great many features of visual appearance have significance, words construed as differentiated, digital symbols in which many visual/aural features can be disregarded as long as a minimally legible character is presented (Goodman 1976). The binary opposition of resemblance and arbitrary designation would have to be complicated by a third term, the semiotic notion of the “index” or “existential” sign, which signifies by pointing, or by virtue of being a link in a chain of cause and effect (tracks signifying an animal; an autograph signifying an author; a graphic mark signifying the activity of the artist) (Peirce 1931–58).

The pursuit of the word/image relationship would ultimately take us back to the very notion of the linguistic sign as such. It will not have escaped the alert reader that my use of the word “tree” and its corresponding image evokes Saussure's famous diagram of the dual structure of the linguistic sign, with the word (“arbor”) standing for the signifier or sound image, and the picture standing for the concept (Saussure 1966).

The picture of the tree in this diagram is consistently “overlooked” (in every sense of this word). It is taken to be a mere place-holder or token for an ideal entity, its pictoriality a merely accidental or conveniently illustrative feature. But the rendering of the signed concept as picture or what Saussure calls a “symbol” constitutes a fundamental erosion in the Saussurean claim that “the linguistic sign is arbitrary” (67) (that is, the linguistic sign is “empty,” “unmotivated,” and without any “natural bond” between signifier and signified). The problem is that an important part of the sign seems not to be arbitrary. As Saussure notes, the pictorial tree, the “symbol” that plays the role of signified concept, “is never wholly arbitrary; it is not empty, for there is the rudiment of a natural bond between the signifier and signified” (68). The word/image difference, in short, is not merely the name of a boundary between disciplines or media or kinds of art: it is a borderline that is internal to both language and visual representation, a space or gap that opens up even within the microstructure of the linguistic sign and that could be shown to emerge as well in the microstructure of the graphic mark. In Saussure's diagram, this space or gap is itself made visible by a Peircean index: the horizontal bar that separates the (iconic) tree from the word “arbor” is neither word nor image but an indicator of their relationship in conceptual space, just as Saussure's elliptical frame and the ascending/descending arrows that flank it, convey “the idea of the whole” and the circulation of significance within its structure.

The further one goes in pursuit of the word/image distinction, the clearer it becomes that it is not simply a question of formal or technical differences between sign types. More is at stake than conceptual housekeeping or a policing of boundaries between art history and literary theory. Understood as a dialectical trope rather than a binary opposition, “word and image” is a relay between semiotic, aesthetic, and social differences. It never appears as a problem without being linked, however subtly, to questions of power, value, and human interest. It rarely appears without some hint of struggle, resistance, and contestation. The defensiveness of art history in the face of textual studies is simply a professional, disciplinary reenactment of a paragone or contest between visual and verbal art that has been going on at least since Leonardo made his famous argument for the superiority of painting to poetry. But variations on this contest are played out in all the arts and media. Lessing's Laocoon was written to defend the domain of poetry against what he saw as an invasion by the visual arts, and Clement Greenberg's aptly titled “Towards a Newer Laocoon” was an attempt to purge the pure opticality of painting from
invasions by “literature.” Ben Jonson wrote “An Expostulation with Inigo Jones” to denounce the dominance of the latter's spectacular set designs over the “poetic soul” of the masque, and Aristotle made it clear that opsis should be sacrificed to lexis in the working of dramatic art. Panofsky thought the coming of sound was corrupting the pure visuality of silent movies, and film theory, as Christian Metz has shown, “has found it difficult to avoid shuttling back and forth between two positions: the cinema as a language; the cinema as infinitely different from verbal language” (Metz 1974).

The “shuttling” of the word/image opposition is, moreover, almost invariably connected to larger social and cultural issues. Lessing's attempt to police the borders of poetry and painting was linked explicitly to his attempt to defend German literary culture from what he saw as an excessively visual French aesthetic and implicitly to an anxiety about the confusing of gender roles (Mitchell 1986). Greenberg’s attack on the blurring of genres in “literary painting” was a defense of an elitist, avant-garde culture against contamination by mass culture. The word/image difference functions as a kind of relay between what look like “scientific” judgments about aesthetics and semiotics, and deeply value-laden ideological judgments about class, gender, and race. Traditional clichés about visual culture (children should be seen and not heard; women are objects of visual pleasure for the male gaze; black people are natural mimics; the masses are easily taken in by images) are based on a tacit assumption of the superiority of words to visual images. Even in the most basic phenomenological reflections on intersubjectivity, the “self” is constructed as a speaking and seeing subject, the “other” as a silent, observable object, a visual image (Tiffany 1989). It is these kinds of background assumptions about the relations of semiotic and social difference that make deviations seem transgressive and novel: when women speak out, when blacks attain literacy, when the masses find an articulate voice, they break out of the regime that has constructed them as visual images. When mute images begin to speak, when words seem to become visible, bodily presences, when media boundaries dissolve— or, conversely, when media are “purified” or reduced to a single essence—the “natural” semiotic and aesthetic order undergoes stress and fracture. The nature of the senses, the media, the forms of art is put into question: “natural” for whom? since when? and why?

From the standpoint of the word/image problematic, then, the difficult and deeply ethical/political task of art history may be somewhat clearer. If art history is the art of speaking for and about images, then it is clearly the art of negotiating the difficult, contested border between words and images, of speaking for and about that which is “voiceless,” representing that which cannot represent itself. The task may seem hopelessly contradictory: if, on the one hand, art history turns the image into a verbal message or a “discourse,” the image disappears from sight. If, on the other hand, art history refuses language, or reduces language to a mere servant of the visual image, the image remains mute and inarticulate, and the art historian is reduced to the repetition of clichés about the ineffability and untranslatability of the visual. The choice is between linguistic imperialism and defensive reflexes of the visual.

No method—semiotics, iconology, discourse analysis—is going to rescue us from this dilemma. The very phrase “word and image,” in fact, is a way of signaling this. It is not a critical “term” in art history like the other concepts in this collection, but a pair of terms whose relation opens a space of intellectual struggle, historical investigation, and artistic/critical practice. Our only choice is to explore and inhabit this space. Unlike Mieke Bal and others who have written on this matter, I do not think we can go “beyond word and image” to some higher plane, though I respect the utopian and romantic desire to do so. “Word and image,” like the concepts of race, gender, and class in the study of culture, designates multiple regions of social and semiotic difference that we can live neither with nor without, but must continually reinvent and renegotiate.

REFERENCES AND SUGGESTED READINGS

Sacks, Oliver. 1993. “To See or Not to See.” New Yorker, 10 May.
W. J. T. Mitchell

© 1996, 2003 by The University of Chicago

Citing this entry