2

THE MEANING OF MEANING

Is a Paragraph More than an Equation?

Patricia Freitag Ericsson

Several chapters in this collection allude to or deal briefly with issues of "meaning" in the controversy about the machine scoring of essays. This chapter's intent is to explore extensively the "meaning of meaning," arguing that, although they may appear to be esoteric, considerations of "meaning" are central to the controversy about the machine scoring of student essays and important to include as we make arguments about it. Foregrounding of the "meaning of meaning" in this chapter establishes a foundation for other chapters that may allude to the importance of meaning in the machine-scoring controversy. Discussion in this chapter can also serve as a vital, integral part of the argument when machine scoring is being considered.

The meaning of meaning is critical when advertisements of machine-scoring products make claims that their products actually can ascertain meaning. Although Knowledge Analysis Technologies' Web-based advertisements for the Intelligent Essay Assessor have recently been revised, in May 2003 their Web site proudly proclaimed that they were providing "[m]achine-learning technology that understands the meaning of text." This provocative claim has now been moved to a less prominent place in the site. In its place is the claim that the Intelligent Essay Assessor operates based on a "machine-learning algorithm that accurately mimics human understanding of language" (2004a). Although this claim may be less shocking, the contention that a machine "mimics human understanding of language" is fallacious and misleading.

IntelliMetric, another popular machine-scoring program, is similarly promoted. The August 2004 Vantage Learning Web site defines IntelliMetric as an "intelligent scoring system that relies on artificial intelligence to emulate the process carried out by expert human scorers" and describes the five main features of writing that IntelliMetric is purportedly capable of determining. One of these is "Focus and Meaning," which is described as "cohesiveness and consistency in purpose and main idea; maintaining a single point of view" (2005a). If

these are the elements of "meaning," then meaning is indeed a very simple concept.

Conceptions of meaning articulated in publications about these scoring machines are as troublesome as their Web-based advertising. For example, Thomas Landauer, the chief researcher and founder of the company that sells the Intelligent Essay Assessor, claims: "The fundamental idea is to think of a paragraph as an equation: its meaning equals a combination of the meanings of its words. Then the thousands of paragraphs in a textbook are just a huge set of simultaneous equations." Intelligent Essay Assessor is based primarily on an algorithm that equates meaning to "*meaning of word*$_1$ + *meaning of word*$_2$ + . . . + *meaning of word*$_n$ = *meaning of passage,*" working on "the basic assumption . . . that the meaning of a passage is contained in its words" (Landauer, Laham, and Foltz 2003, 88). Although Landauer and his associates do admit that this conception "is by no means a complete model of linguistic meaning," they still believe that "for practical purposes" this kind of analysis "simulates human judgments and behavior" adequately (or in their words, "quite well") (89).

Vantage Learning guards the inner workings of its scoring machines closely. Almost every publication by Vantage Technologies describes Intellimetric as a system that depends on the "proprietary" technologies of "CogniSearch" and "Quantum Reasoning." Researchers interested in finding out more about these technologies hit a dead end. Impressive-sounding names and volumes of research conducted exclusively by Vantage Learning are the only assurances that anyone has about Intellimetric. Independent researchers cannot replicate this research or verify Vantage's claims, since the technologies used are not available to anyone else. (Edmund Jones's chapter 6 in this volume provides insights into how Intellimetric works, despite having no access to the actual algorithms used in the program.) Scott Elliot of Vantage Learning says that Intellimetric is based on a "blend of artificial intelligence (AI), natural language processing and statistical technologies" and claims that through this blend of technologies the program "internalizes the pooled wisdom of many expert scorers" (2003, 71).[1] Besides this gloss of what technologies are behind the interface and the unsupported claim that the program is wise, very little about the actual workings of Intellimetric can be found. Since Vantage Learning's promotional material defines "focus and meaning" as "cohesiveness and consistency in purpose and main idea; maintaining a single point of view," we must assume that their understanding of meaning is somehow encompassed

in those fourteen words and that their technologies depend on this simplistic definition.

WHY DO WE TEACH STUDENTS TO COMPOSE?

Considering the meaning of meaning is vitally important not only because some machine-scoring promoters advertise their products as able to ascertain meaning, but also because most people believe that conveying meaning is the most important goal of the written word. During discussions of the machine scoring of writing, participants must carefully consider why we ask students to compose essays and what we expect them to gain from knowing how to compose such texts. To begin this consideration of meaning and composing, the work of Ann E. Berthoff in *The Making of Meaning* is particularly helpful. In this 1981 volume, Berthoff argues for a theory of composing as a meaning-making activity, not just composing for the purpose of regurgitating specific content-knowledge information in a predetermined form. She emphasizes the need for composing as a process fueled by imagination consisting of "abstraction, symbolization, selection, 'purposing'" (4). This process "requires or enables us to coordinate and subordinate, to amalgamate, discard, and expand; it is our means of giving shape to content" (4–5). In this kind of composing process, learners are discovering, interpreting, and coming to know—they are making meaning. Later in the same book, Berthoff argues that if we teach composition "by arbitrarily setting topics and then concentrating on the mechanics of expression and the conventions governing correct usage," our students cannot learn to write competently (19). Composing, Berthoff says, works in "contradistinction to filling in the slots of a drill sheet or a preformed outline—[composing] is a means of discovering what we want to say, as well as being the saying of it " (20). Unfortunately, using computers to evaluate and score student compositions does exactly what Berthoff claims will not teach students to write competently. When composing for a machine, students are given arbitrary topics and are judged by a machine that concentrates on mechanics and conventions, plus the addition of a few important content words—not content ideas.

Although Berthoff's understanding of composing is decades old, it still underlies the "best-practice" models in composition theory and practice. Writing in 1998, Sharon Crowley asserts that composition "focuses on the process of learning rather than the acquisition of knowledge" (3). This focus continues Berthoff's emphasis on composition as a project of discovery and meaning making rather than a project of

repeating facts and figures. Composition, Crowley maintains, "encourages collaboration" and "emphasizes the historical, political, and social contexts and practices associated with composing rather than concentrating on texts as isolated artifacts." Again, computer scoring of student compositions has nothing to do with collaboration and everything to do with texts solely as "isolated artifacts."

THE MEANING OF MEANING

If we agree that we teach students to compose so that they can make sense of isolated facts and figures, so that they can make meaning, and so that they understand the social nature of meaning making and convey meaning to others, we are obligated to consider the meaning of meaning. Even though machine-scoring promoters tout their programs as being able to discern meaning, the scholarly areas they depend on have little truck with the meaning of meaning. In the introduction to their 2003 book *Automated Essay Scoring: A Cross-Disciplinary Perspective*, Mark Shermis and Jill Burstein claim that perspectives on machine scoring should come from "writing teachers, test developers, cognitive psychologists, psychometricians, and computer scientists" (xv).[2] Although I strongly agree that writing teachers should be involved in discussions about machine scoring, I must protest the way Shermis and Burstein include this perspective in their book. The only chapter in the book that acknowledges a teacher's perspective is written by a person who has led the National Council of Teachers of English but is decades removed from teaching or researching writing. He can hardly be considered a "writing teacher." The others areas listed by Shermis and Burstein—test developers, cognitive psychologists, psychometricians, and computer scientists—are less concerned with what is at the core of writing—making meaning.

Which other "perspectives" need to be considered if making meaning is central to what writing and composing is all about? I would argue that we need to include writing teachers (real teachers in the trenches, not figureheads), composition scholars, rhetoricians, linguists, philosophers, and a host of others in such an inquiry. To remedy the shortcomings of depending only on Shermis and Berstein's limited list, I begin with Ann Berthoff, whose understanding of writing as a meaning-making project underpins this chapter. While she was writing *The Making of Meaning*, Berthoff was an in-the-trenches writing teacher as well as a scholar (which qualifies her in Shermis and Berstein's view as well as mine). Since she defines composing as a meaning-making project, Berthoff is obligated to explore what the term "meaning making" entails. She

argues that "meanings are not things, and finding them is not like going on an Easter egg hunt. *Meanings are relationships: they are unstable, shifting, dynamic; they do not stay still nor can we prove the authenticity or the validity of one or another meaning that we find*" (1981, 42; emphasis added). Berthoff warns composition teachers and others of the danger of conflating the terms *meaning* and *information.* Citing support from engineers and logicians, she argues that "*information* has nothing to do with *meaning*" and bemoans the fact that the word "information" is regularly and erroneously "used as a synonym for meaning." She urges readers to resist this conflation, arguing, "We should continually be defining the meaning of meaning, but instead we consider that there is no need since we are using a scientific term" (53).

Many machine-scoring programs, according to ETS's Jill Burstein, treat student compositions like a "bag of words" and go on a virtual Easter egg hunt to find the right words (Phelan 2003). These programs treat essays as pure information that can be mined for some abstracted set of words that, at least to their promoters, equates to meaning. The shifting, dynamic relationships that these words have to each other, to society, and to different readers is invisible to these information-seeking machines. The machines can tell users whether writers have matched the words in an essay with words in a database (or a triangulated database matrix), but they cannot assess whether this mix of words conveys any meaning.[3]

In 1988 Berthoff argued that ideas can be flattened so that any "generative power" they might have had is lost. Words fed into a scoring machine are flattened this way, stripped of their generative power; thus the possibility for "interaction" with ideas is reduced to only "transaction." A machine that equates meaning to a combination of *word* + *word* + *word* reduces the reader/word relationship to a one-dimensional "stimulus-response" connection. The machine responds to the stimuli of words, not concepts or ideas. The machine responds with limited experience—only that experience the programmers have been able to feed it. The machine has no understanding, no sense of the concepts and ideas that underlie the words, no ability to bring to the words what Berthoff claims is important in discerning meaning, "what we [humans] presuppose and analyze and conjecture and conclude"—all of this adding up to a human sense of what a text might mean (p. 43).

Considering the meaning of meaning is not a newfound intellectual pursuit, especially in philosophical circles. We have access to thousands of years of consideration of the meaning of meaning—dating at least from 360 BCE and the Platonic Dialogues. In *Theaetetus,* Socrates asks

Theaetetus (a young aristocrat) to define the meaning of the term *knowledge*, claiming (in his inimitable Socratic way) that the meaning of this term was something he could never solve to his satisfaction. Theaetetus takes the bait, answers, and (to his credit) states that if he is incorrect, he knows Socrates will correct him. Thereupon, Socrates and Theaetetus embark on a classic Socratic adventure that considers not only the meaning of knowledge but considerations of perception, true and false beliefs, the mutability of knowledge, and the meaning of meaning. In this dialogue, Socrates asks, "How can a man understand the name of anything, when he does not know the nature of it?" Later he states, "[N]othing is self-existent." Certainly Socrates would agree that a word or collection of words has no meaning without some knowledge of the nature of the social, historical, and political context within which those words are being used.

Fast-forwarding nearly 2,300 years (and countless considerations of meaning during those centuries) we find rhetorician I. A. Richards and linguist C. K. Ogden studying the meaning of meaning in a 1923 book aptly titled *The Meaning of Meaning*. In this book, Richards and Ogden explore misconceptions about meaning and coin the term "proper meaning superstition," which is the mistaken idea that every word has a precise, correct meaning. They argue convincingly that different words mean different things to different people in different situations. Computer-scoring enthusiasts fail to comprehend what Richards and Ogden understand about meaning: "Meaning does not reside in the words or signs themselves; to believe that it does is to fall victim to the 'proper meaning superstition,' the belief that words have inherent meaning" (Bizell and Herzberg 1990, 964). Richards and Ogden argued (in 1923) that "everyone now knows" that words "'mean' nothing by themselves," although that belief was once "universal" (968). Unfortunately, their proclamation was premature. The belief that words have meaning on their own still holds sway with many, as evidenced by the public claims of the machine-scoring industry. That industry ignores scholarly considerations like those of Richards, who in 1936 claimed that "the stability of the meaning of a word comes from the constancy of the context that gives it its meaning" (11).

Philosopher Mihailo Markovic's 1961 volume on meaning, *Dialectical Theory of Meaning*, illustrates just how complex meaning is. In part 3 of this book, Markovic offers four possible general definitions of meaning. Definition A is particularly appropriate to concerns in this chapter. "When a group of conscious beings, witnessing the appearance of a

material object, is disposed to think of an object (or an experience any other mental state whose external correlate is an object), and that thought (experience) may be expressed objectively using some means which all the members of the given social group can understand and use, we may say that in that case that the given material object is a sign and it has a definite meaning" (363).

Markovic limits meaning to "conscious beings" and deliberately adds the "social group" to the definition, thus making any claim that a computer could discern or understand meaning highly questionable. Although the goal of artificial intelligence might be the creation of a sentient machine, few (if any) honest experts in that field would claim that this goal has been reached. Despite fictional presentations in futuristic books and movies, the idea of an artificial-intelligence machine being part of a social group is hardly in the realistic future. Some machine-scoring companies may claim that they are simulating an artificial sort of social group when they feed volumes of words on a topic into a computer, but the database of words created this way is a far cry from even the most broadly construed definition of a "social group." It takes a huge leap to imagine that a machine fed word after word after word would have any relationship to a real social group made up of conscious beings who have experienced the word with their senses. Markovic's succinct claim that "social, practical meaning is greatly dependent on context" (1961, 365) is worth committing to memory as we carry on the debates about meaning and machine scoring.

Because of his background in both science and language studies, Jay Lemke's (1995) perspectives on meaning are particularly appropriate for consideration. Lemke earned a Ph.D. in theoretical physics in 1973 but turned to linguistics, semiotics, and language studies in the 1980s to help him better understand the teaching of science. He argues that the meanings of words, phrases, sentences, paragraphs, and bigger chunks of texts are all dependent on context and that the meanings of these parts change from one social situation to another. In an idea that has Socratic echoes, Lemke argues that "[l]anguage does not operate in isolation"; it is part of the "whole 'dance' of meaning-making (a dance that always assumes a partner, that always helps to create one)" (8).[4] In what could be taken as a caution to the computer-scoring enthusiasts, Lemke counsels, "We are not likely to understand the role of language in our culture or in our society if we divorce it from its material origins or from its integration into larger systems of resources for making meaning." He argues that "all meanings are made within communities" and that

"analysis of meaning should not be separated from the social, historical, cultural and political dimensions of these communities" (9).

THE CONSTRUCTION OF MEANING IN DISCOURSE COMMUNITIES

In addition to his perspectives on meaning, Lemke's work can help us understand the complications and disjunctions created when different research and language communities meet. According to Lemke, scientific discourse is built on the "language of truth" (an objective view, usually based on numerical proofs). Lemke argues that scientific discourse's power is "the power to compel belief in the truth of what they [scientists] are saying" (1995, 178). When other discourse communities (like the linguistic or the rhetorical) try to advance their ways of thinking—ways of thinking that "include elements of the language of feeling or of the language of action and values," ways of thinking that "argue from values or the implications of propositions for action and social consequences"—they are discounted as being nonscientific, beyond proof as *true*, and therefore not believable. In the machine-scoring world, the scientific discourse community depends on a limited, numbers-based meaning of meaning that holds currency for some in the general public. However, other discourse communities, those that do not rely on a numbers-based meaning of meaning, hold compelling views that must be brought into the discussion as a counterbalance.

As an example of this counterbalance, we can contrast information theory with semiotics. The approach to scoring student essays used by the scoring machines is based largely on an information theory that "looks for the common denominator in all forms of information and quantifies information in common units." In contrast, language studies (especially semiotics) looks at "the significant ways in which units that carry information differ from one another" (Lemke 1995, 170). This remarkably different way of looking at information (in this case the information in a student essay) at least partially explains the problems that ensue when information theory is used as a basis for finding meaning in a text. The differences between an information theory approach that tries to determine "the amount of information that a text could contain" is remarkably different from a linguistic, semiotic, or social constructionist theory that is interested in discovering "the possible meanings that a text could have in a community."

Literacy theory is also valuable in helping us bring the views of a different discourse community into discussions about machine scoring and the meaning of meaning. Drawing on the work of James Gee, literacy

scholars Colin Lankshear and Michele Knobel (2003) offer more on the community or sociocultural perspective. In contrast to an information or artificial intelligence perspective, the sociocultural perspective holds that it is "impossible to separate out from text-mediated social practices the 'bits' concerned with reading or writing (or any other sense of the 'literacy') and to treat them independently of all the 'non-print' bits." The non-print bits, which may include "values and gestures, context and meaning, actions and objects, talk and interaction, tools and spaces," are "non-subtractable parts of integrated wholes." In a sociocultural perspective, meaning cannot exist in isolation from the social and cultural milieu in which those meanings are made. Lankshear and Knobel argue that "if, in some trivial sense they [literacy bits] *can* be said to exist (e.g. as code), they do not *mean* anything" (8).

Machine-scoring programs "see" student essays as code. They take students' words, sentences, and paragraphs out of their social/cultural contexts, process them as meaningless "bits" or tiny fragments of the mosaic of meaning, and claim to have "read" these essays for meaning. They claim to be able to "mimic" the way a human reader would read them. And they base these claims on uninformed, possibly fraudulent, understandings of meaning. If we bring a broad spectrum of discourse communities into discussions about the machine scoring of student essays, perhaps we can insist that the machine-scoring industry account for the severely limited capabilities of their programs. Perhaps we can even convince them (since the industry is peopled by highly educated, and hopefully educable, researchers) of what we know about students' communication needs and of the serious disservice they are doing to students with their limited understanding of why we teach students to write and how students become better writers.

IF WRITING IS MORE THAN *WORD + WORD + WORD,* THEN WHAT?

The machine-scoring industry is misleading the public with untenable claims about what their machines can do—claims that state these machines can evaluate student writing and even help students become better writers. If we agree that we teach students to write so that they can make and communicate meaning, we need to promote an appropriate understanding of both those goals and thereby undermine the claims made by the industry. In their 2000 book, *Multiliteracies: Literacy Learning and the Design of Social Futures,* Bill Cope, a communication and culture scholar, and Mary Kalantzis, an education and language scholar, argue that students need to be taught how to be successful communicators in

a world that is marked by "local diversity" and "global connectedness" (14). This world is not one that can be virtually simulated by a computer program. Cope and Kalantzis contend that "the most important skill students need to learn is to negotiate regional, ethnic, or class-based dialects; variations in register that occur according to social context; hybrid cross-cultural discourses; the code switching often to be found within a text among different languages, dialects, or registers; different visual and iconic meanings; and variations in the gestural relationship among people, language and material objects."

For Cope and Kalantzis, language is a "dynamic representation resource" that is continually remade by writers and speakers as they endeavor to accomplish their goals in various cultural projects (2000, 5). Students who write to and for machines will not develop any sense of the dynamics of language; they will not acquire an understanding of diverse audiences and the need to adapt to those audiences; and, like those who program and promote machine scoring, they will be oblivious to and uninformed about the meaning of meaning.

Assuming that we agree with composition scholars, rhetoricians, linguists, philosophers, literacy scholars, and others that writing is a process of learning, that it is about making meaning rather than spitting out a series of facts and figures, that it is about analyzing, integrating, and understanding historical, political, and social contexts in which we are located, then we need to challenge machine scoring on these counts. Machine-scoring machines "see" texts as isolated artifacts. These machines cannot understand texts as social instruments, as organic entities that work to help writers and readers make sense of social and political environments. If composition is about making meaning—for both the writer and the reader—then scoring machines are deadly. Writing for an asocial machine that "understands" a text only as an equation of *word* + *word* + *word* strikes a death blow to the understanding of writing and composing as a meaning-making activity. Students who learn to write for these machines will see writing and composing as a process of getting the right words in the "bag of words" without a concern for a human audience or any legitimate communicative purpose. Students deserve better than this dumbed-down version of writing and composing. We need to take responsibility for getting them what they deserve.