A Systems Theory of Multimodality.

UTE. Revista de Ciències de l'Educació
2018 núm. 2. Pàg. 6-22
ISSN 1135-1438. EISSN 2385-4731
http://revistes.publicacionsurv.cat/index.php/ute

https://doi.org/10.17345/ute.2018.2.2489

James Paul Gee

Rebut: 05/12/2018 Acceptat: 21/12/2018

Abstract

This paper offers a tentative theory of how to approach the analysis of multimodal "texts" (using the word loosely for any integrated set of symbols that resemble written texts in communicative power). Much work on multimodality is new and many key issues are not yet broached, let alone settled. The theory I produce treats modes as semiotic systems (sets of social conventions about meaning) and multimodality as a system of systems. A theory in a new domain helps us understand what we take the "things" we are going analyze to be (our "ontology"). In turn, our choice of "things" determines the sorts of tools we will need for analysis and the sorts of questions we will ask, a number of which are delineated in this paper. The theory developed here is based on the correspondences between linguistic systems and non-linguistic modal systems. The paper closes with a discussion of the implications of the approach I propose for work on design and learning.

Key words: multimodality, semiotic systems, communiative modes, design, learning.

Resumen

Este artículo ofrece una propuesta de teoría de cómo abordar el análisis de los "textos" multimodales (usando la palabra de forma amplia para cualquier conjunto integrado de símbolos que se asemejan a los textos escritos en poder comunicativo). Mucho trabajo sobre la multimodalidad es nuevo y hay cuestiones clave que aún no se han abordado, y menos resuelto. La teoría que produzco trata los modos como sistemas semióticos (conjuntos de convenciones sociales sobre el significado) y la multimodalidad como un sistema de sistemas. Una teoría en un nuevo dominio nos ayuda a entender qué tomamos como las "cosas" que vamos a analizar (nuestra "ontología"). A su vez, nuestra elección de "cosas" determina el tipo de herramientas que necesitaremos para el análisis y el tipo de preguntas que haremos, algunas de las cuales se describen en este documento. La teoría desarrollada aquí se basa en las correspondencias entre los sistemas lingüísticos y los sistemas modales no lingüísticos. El documento concluye con una discusión sobre las implicaciones del enfoque que propongo para el trabajo sobre el diseño y el aprendizaje.

Palabras clave: Multimodalidad, sistemas semióticos, modos de comunicación, diseño, aprendizaje.

1. Introduction

While we have thousands of years of experience analyzing written texts and have long traditions in the analysis of painting and film, we have barely begun to understand the "multimodal texts" that are ubiquitous in our digital world. Such texts combine spoken and/or oral words, images (static or moving), and sounds, often in complex ways. Today, we have whole new media based on such multimodality, media such as video games, apps of all different sorts, and many different types of simulations, not to mentions advertisements.

Of course, we humans have always combined words, images, and sounds. So what is new is the diverse ways in which digital media allow us to take this very much further than we ever have done before, including creating ways for the "reader" to interact and even sometimes modify the multimodal texts.

Multimodal analysis is new today and current work rarely specifies any relationship between language as a system ("grammar") and other modes like images and sounds. Further, we get little discussion as to whether images and sounds enter into systems themselves, akin to grammar, or not. I will take up these issues in this paper, though only as tentative start, certainly not as a final answer.

2. Modes

Our understanding of how humans process and interpret multimodal "texts" is just beginning (Jewitt 2009; Jewitt & Kress 2003; Kress 2010; Kress & Van Leeuwen 1996, 2001). While we know a good deal about how humans process and interpret oral and written language, we know much less about what happens when sound, image, and words are mixed and matched. All I can offer here is an initial tentative theory about how to approach multimodal texts.

Theories are important in new domains. A theory helps us understand what we take the "things" we are going analyze to be (our "ontology"). In turn, our choice of "things" determines the sorts of tools we will need for analysis and the sorts of questions we will ask. For example, the theory that evolution happens at the level of genes requires different methods and questions than the theory that evolution happens at the level of individual bodies. The ontology of one theory is genes and the ontology of the other is bodies.

In this paper, to simplify matters, I will deal, for the most part, only with messages composed of images and language, leaving aside other modalities. Even with this restriction we face a problem in how to define "mode". A mode is connected to a way of signaling meaning using a distinctive type of material (e.g., sound waves, language, images, smells, movement). However, while language and images are considered different modes, printed language is an image, not sounds. An image (of, say, an apple) juxtaposed to a written word (say, "apple") is two images.

Communicational modes need to be defined by two things, not just material substance alone, but social conventions, as well. A mode is a way of signaling meaning defined by a particular set of social conventions and a distinctive type of material. Here print (verbal images) and pictures (iconic images) are different because they are interpreted in term of different systems of social conventions. Oral language, music, and oral poetry are all sound waves, but subject to different social conventions, thus different modes. Both aspects are important. Different material substances have different affordances for meaning making and different social conventions recruit those affordances differently.

3. System

A mode is defined by a set of social conventions being used to make a distinctive material mean in a certain way. But what are these social conventions? They are socially shared ways of creating and using a distinctive semiotic system (Halliday 1978; Saussure 1986).

To deal with any mode awe first need to delineate what makes something a semiotic system (just "system" for short). Any system is defined by conventions about what constitutes a part and a whole (Gee 2014, 2017a). Different human languages define parts and wholes (morphemes, words, phrases, and sentences) differently and, thus, are different systems. Mathematical language defines them differently than any natural language and is its own distinctive system (or several different systems since there are different branches of mathematics such as algebra and geometry).

4. System: parts and wholes

A system involves "rules" (social conventions) about what count as parts and wholes (units) in the system. Language is a system composed of parts and wholes. Morphemes are parts of words; words are parts of phrases; phrases are parts of sentences; and sentences are parts of larger "texts". Images also, of course, have parts and wholes. A picture of an apple tree is composed of branches, a tree trunk, and apples. How we humans see images as sub-parts, parts, and wholes is determined by the "grammar" of the human visual system.

Wholes have meanings in terms of their sub-parts. These meanings are determined by the system of which they are part. For example, in English many combinations of an adjective and noun (like "red flower") mean adjective + noun (e.g., "red flower" = something that is both red and a flower). But there are combinations of adjective and noun that do not follow this rule. For example, a large mouse is not something that is both large and a mouse, since no mice are big, say compared to an elephant of any size. "A small mouse" means "a mouse which is small for a mouse".

The conventions by which we determine the part-whole relations in images are very often parasitic on our everyday experiences in the world or specialized practices shared by different groups (Halliday & Matthiessen 1999). For example, given our experience of the world, we interpret the first picture below as "man with cane", the second as "rain cloud"; and the third as cactus with two arms. The last is the sort of image that some group might use for specific graphing practices.

Figure 1: Pictures often allow different interpretations which are based on our experience of the world.

It is interesting to note that while the adjective + noun pattern found in "red flower" (red + flower) is easy to depict in an image (just a picture of a red flower), it is harder to depict the pattern "adjective for a noun" (e.g., "small for a mouse") without any verbal annotations.

5. Basic units: words

In any system there are certain units which we can call "basic units" (Gee 2017a). Basic units are the pivots around which part-whole relations are defined. For example, in language, morphemes cannot stand alone, but must be parts of words. Words are the smallest units that can stand alone, but they can also combine into bigger units (phrases and sentences). In language, words are basic units. Basic units-at the level of meaning-define the ontology of a system, the basic things the system takes to exist and about which it can communicate.

What makes a word mean something? Let's start with proper names. I can use any sounds I want to name an individual thing. So, say I call my pet goat "Lucy". "Lucy" picks out, for me and anyone else who wants to follow my practice, one specific goat. Proper names are not yet really words.

Something quite different happens when we use common nouns like "goat" or "boy". Such words can be used to refer to a set of things and not just one specific thing. So, in "Goats are interesting animals" the word "goats" refers to all goats, the set of goats. Here we have not a name, but a general meaning.

Such general terms raise the issue of how we know, when we are referring to a set of things (like goats), what is in the set and what is not. Such words refer to a set of things because we have socially shared criteria for what is in the set. Let's call such criteria "socio-conceptual meaning" (SCM). SCM can be different for different social groups. For instance, scientists pick out what is a goat by genetic information, but everyday people do it based on visual and behavioral information. The two groups have different SCMs for the word "goat" and, thus, the word could pick out a different set of things for each group.

A word need not have rigid criteria for what it applies to, though some words do. The word "game", for instance, has criteria, for most people, in terms of which games do not all share the exact same features, but share more or less features with each other (as in the members of a genetically related family). The boundaries of the set of "games" is fuzzy. There can be cases where we are just not sure whether something should be called a "game" or not, or where we are tempted to say something like, "Well, its sort of a game".

6. Baisic units: images

So general words (not names) refer to a set of things (more or less clearly) based on socially shared criteria, a type of social convention (shared among a certain group of people, small or large). Now, interestingly, images work the same way. We first must identify which system the images are a part of to what counts as parts and wholes and basic units around which they are defined. Images, like words, can enter into different systems.

Imagine a picture of a tree (it could be a realistic photo or a line drawing). We can treat the picture like a name: We can treat the picture as the picture of a specific tree (the one that inspired the photo or the drawing). Or, we can use the image to signify something more general. We could use the image to mean all trees like the tree in the image (e.g., a species of tree) or all trees. What the image will mean depends on the social convention shared by people who thereby know what it is being used to pick out, what set of things it is being used for.

Images, as we have said, can enter into different systems. Remember, a system determines what constitutes parts and wholes and what constitutes basic units. Realistic images, printed words, graphs and diagrams, maps, emoticons, and many others, constitute different image systems.

7. Choice

Any system builds meaning around the notion of choice (Gee 2017a; Halliday 2013). In language, grammar (syntax and basic meaning) sets up the choices available to a speaker as to how to say what she wants to say. For example, consider the sorts of choices below:

1a. Mistakes were made

1b. The company's president made a mistake

1c. To err is human

1d. Unforeseen circumstances intervened

1e. Mistakes happen

1f. It was a real blunder by the boss

2a. Hornworms sure vary a lot in how well they grow

2b. Hornworm growth exhibits a significant amount of variation

2c. Hornworms come in lots of different sizes

2d. Manteca sexta larva grow up to 70 millimeters in length, but can vary significantly.

3a. Could you please help me?

3b. I need help

3c. I hate to ask, but could you possibly help me

3d. Get a move on and help me

4a. They are freedom fighters

(said of people who use terror to attack our enemies)

4b. They are terrorists

(said of people who use terror to attack us or our allies)

4c. They are guerillas engaged in guerilla warfare

4d. They are mujahideen engaged in jihad

Imagine that in the case of the utterances in (1) a company spokesperson has been asked why something bad has happened. The spokesperson must choose among all the alternative choices the grammar of his or her language makes available. The available choices are determined by grammar (a few of which are listed in 1). The actual choice made is language-in-action-discourse-is determined by a human being in real time. Saying "mistakes were made" allows the spokesperson to leave out the person or people responsible for the mistakes. Saying (1f) might be a good way for the spokesperson to get fired.

Choices have meaning not just by themselves, but also in relation to all the other choices that were available, but excluded, once a choice was made (Saussure 1986). If all neckties were black, wearing a black tie would just mean you chose to wear a tie. If there are many colors of ties, then wearing a black one means you did not choose to wear other colors (e.g., brighter ones) for the occasion. And we can then ask why you didn't. So, too, with language.

Choices can allow us to try to capture the truth as we see it; to lie effectively; or to shape how people think without directly lying to them. They allow us to express what we want to say in ways that can reach people's emotions and minds and even encourage them to act.

Images enter into systems of choices as well, though one and the same image can be part of different systems. It all depends on how they are being used, to what purpose they are being put. There are different ways to graph information, so each way is a choice. The rain cloud above could be a choice among differ weather patterns, with lightening come out of a cloud, or a sun with no clouds, or just clouds being choices. Even in an abstract painting we have to ask why the artist chose the colors (or, perhaps, on black and white), the lines and shapes, the amount of realism (or total lack of it) she did.

8. Affective expressitivity

As we said early, oral and written language are two different modes and so juxtaposing an image with oral language is not two images, but an image and sound waves (Halliday 1985). Oral language has one key property that written language does not have: intonation, the pattern of tones, pitch, and stress across an utterance.

Intonation plays does different things in English (Halliday & Greaves 2004). It can distinguish between questions and declaratives (a grammatical function). It can signal that information is old, new, or emphatic (a discourse function). And, it can, express different emotions, moods, and attitudes (an affective function). It is this latter function I want to discuss now, the ways in which we can use intonation to express things like surprise, irony, insult, sympathy, sadness, excitement, confidence, and so on. Let's call this "affective intonation". Active intonation, coupled with all aspects of voice (tone, pitch, loudness, whisper, and more), is one way we project feeling-tone into communication.

Affective intonation is an example of a more general feature all systems can have, a feature I will call affective expressivity. Images, like oral language, have affective expressivity as well, though the matter has been much less studied. The various visual properties of color and shape can communicate-or induce-feelings, moods, and sensations beyond the literal meaning of the image. In the case of written language, affective expressivity is expressed visually (e.g., shapes of letters) and not by intonation.

I will use here the video game Thomas Was Alone as an example of affective expressivity can be used in a multi-modal setting (Gee 2014b). Thomas Was Alone is a platform game where the player controls different characters. Each character is a simple (faceless) colored shape, with a name (e.g., Thomas) and with its own distinctive abilities (ways of moving and combing with other characters). The characters/shapes move across a very simple background of two-dimensional line, blocks, and spaces.

Thomas Was Alone has a story narrated by a narrator whose narration is both voices and printed on the screen. Within the story each shape has a name and something of a backstory as an artificially intelligent agent inside a computer whose programming has gone awry. The shapes are trying to escape the system. Each shape/character has certain unique abilities and limitations (determined by the game's game mechanics) that fit with the character's personality and role in the story.

For example, Thomas, a red rectangle, has an up-beat attitude and can do an average jump. John, the yellow rectangle, is arrogant and eager to show off and can jump quite high. Claire, the blue square, who starts off feeling bad about herself but comes to see herself as a super-hero, cannot jump well or move fast, but she can float and move in water and thereby save others by giving them rides across water.

Since the words of the narration are printed on the screen, this, in a way, subtracts the words from the oral narration and means that the oral narration mainly functions to carry the expressive intonation of the narrator's voice, the musical and affective part of speech. This affect is created in part because we

can read much more quickly than we can hear, so the player has often read the all the (short) material on the screen before the narrator has finished saying it. The player has the "meaning" but still must pay attention to the intonational contours. The narration in Thomas Was Alone is in a British accent that is amazingly good at indicating the emotions of the characters (rectangles though they be), emotions like fear, self-loathing, loneliness, liking and love, caring, arrogance, humility, and trust.

At one point in the game the player can readily see that John needs to help Chris and Thomas. Chris must jump on top of Thomas and then jump from Thomas to John and, finally, jump up to a ledge too high for her to reach by herself. Thomas can then jump on John and then up to the ledge. And only then can John jump up by himself. If John, who is tall enough to make it to the ledge on his own, does not help and goes on on his own, the game would be over, because Chris and Thomas could never escape.

John could can easily go on without Chris and Thomas, since he is tall enough to reach the ledge. We know from earlier in the game that John thinks more highly of himself than he does of the others. Forced to go back and help, he has to rationalize this as not a weakness, but as a strength. This strength is not only that his help is essential to mitigate Thomas's and Chris's weaknesses. It is also that John will look good in the act and others will see how special he is.

At this point the narrator says: "John was happy to keep helping. He felt it was important to his image that he was seen to help the little guys". Players have finished reading this before they have finished hearing it, so they pick up the literal meaning first and then savior the British intonation that wonderfully captures Thomas's attitude, his arrogance, but also his underlying innocence.

This communication of the character inner emotional lives was seen as the most powerful part of the game by many reviewers and is achieved in part by the affective intonation of the narrator, especially as it is juxtaposed on graphics that are look like childhood drawing composed of simple colors, lines, and shapes. Here we have three modes: oral language with its affective intonation, written language with its literal content, and simple lines and shapes with primary colors.

The effect affective expressivity in Thomas Was Alone, however, is a combination of both the "feeling-tone" of oral language and the way the graphic style (primary colors and simple lines and shapes) of the game looks and feels like a child's drawing. Because the affective expressivity in language and in images go so well together in this game, they can serve as a powerful pairing. The characters (just colored shapes) come across as innocent and child-like creatures, each with their own feelings, fears, and personalities, and the act of seeking to be free becomes a sort of journey of growing up.

9. Style

A system is defined by norms (social conventions) about what count as units, parts and wholes, sets of choices, and ways to engage in affective expressivity. Very often a given system is put to use in different ways by different groups of people with different interests (Gee 2014a, 2017b). So, for example, both sentences in (1) below are part of the system of English, but (1a) is in a vernacular style and (1b) is in an academic style of language:

1a. Hornworms sure vary a lot in how well they grow.

1b. Hornworm growth exhibits a significant amount of variation.

Different styles (in language, these are such things as dialects, registers, and degrees of formality) use the same elements from a shared system. However, they use these elements to different degrees and in different patterns and combinations. Thus, the academic style in (1b) uses nominalizations (e.g.,

"growth", "variation") and Latinate vocabulary (e.g., exhibits, significant, amount) more than does-and combines them together more often-than does the vernacular style.

Image systems also come in different styles. Thomas Was Alone uses a 2D style of illustration that is based on simple lines and shapes and colors. The game could be remade with more realistic characters in 3D realistic environments. The game itself-as a game in terms of game mechanics-will not have changed, but the feel, affect, and effects of the game on the player would change, as would the meanings players attribute to the game and their experience of playing it. This is because style signals identity and purpose (intention).

10. Discourse

Units in a system can be strung together across space or time to communicate larger meanings (Gee 2014a, 2014b, 2017a). In language, grammar determines parts and wholes (the type and order of words in phrases and the type and order of phrases inside sentences). However, discourse conventions determine how sentences can be sequenced in what we can call "texts".

Texts come in types (genres) such as narratives, arguments, definitions, explanations, jokes, lectures, conversation, and many more. Furthermore, different genres of text use different features (or patterning of features) to tie sentences together into larger patterns (stanzas, paragraphs, topic units, and so forth) with a text. These features are cohesion devices (Halliday & Hasan 1976). Images, too, can be sequenced in space (as language is in writing) or time (or both) to create genres of meaningful text types.

Consider the text below:

Bead: Are you really dead

Allele: Yes, did you get the heart?

Bead: I got the heart-another guy was helping

Allele: Good

Bead: I am standing over your body mourning

Allele: I died for you

Bead: So touching

Allele: It's a long way back

Bead: I know-I've done it

This is written language and, for those who do not know where it came from, odd language indeed.These are two players communicating via chat in World of Warcraft, a massive multiplayer video game. As discourse, it is a conversation, though in a style only well-known in video games. Note, though, that this conversation is taking place after these two players have seen (on a screen) the events happening, though the sequence of events has contained the mystery as to whether Bead also died or lived and accomplished the goal Allele was helping him to accomplish. The events the players have observed, partly together and partly separated, have a narrative logic that is itself situated in games like this one, and not exactly like the narrative logic of everyday life.

We can not here that as a multimodal experience there are actually four different modes going on: images depicting objects and actions; sounds in the game; written chat; and the players' tactile and mental experience of choosing and doing as they play the game (thus, affecting the images and actions).

Discourse is about sequence or juxtaposition of units from a system and connectors (cohesive devices) that help interpreters tie elements of the sequence together. These sequences are interpreted in terms if genres that tell us what type of "texts" they are (e.g., conversation, narrative, list, examples, definition, and so forth).

When we place images together-whether print or pictures-we can sequence them in space and time. A printed page lays out the print in the space of a page in certain ways (e.g., two columns versus one) and the print also "flows" across time (line by line and page by page) in terms of how it is interpreted. A child's picture book can array different images or sets of images on the page in a certain way (e.g., consider comics) and can sequence images across time, as well, as the book is decoded in in steps across time.

11. Situational meaning

Regardless of mode, the units and sequences of units in any communication are subject to a set of practices the heart and soul of communication and interpretation. These practices we will call "situational meaning" (Gee 2004, 2014a, 2017a). Situational meaning is about how people can adapt, adjust, situate, or modify their meanings in specific situations of use that require more nuanced meaning than just reference to a general set of things. So, for example, while the word "cat" refers to the set of cats, in actual situations of use the word can refer to different cat-related things: "Cats have been depicted in art for thousands of years" (cat = image); "Big cats are endangered in Africa" (cat = lions and tigers) ; "I go to cat shows all the time" (cat = house cats); "Cat is not something I would eat" (cat = food).

So, too, an image like that below can be taken to refer to the set of apples, but once we put it in a specific context of use, it can take on more specific or nuanced meanings in the given situation. In a grocery store it might signal the way to the fruit section. In a Mac Store it will signal a brand. In a classroom, it will mean teachers. On a farm it might mean apple picking.

Figure 2: Apple exemplifying situational meaning

Thus, words and images both have three properties that make them "meaningful" in semiotic terms: reference, socio-conceptual meaning, and situational meaning (the last two are types of social conventions).

The examples I have given so far show how situational meaning works for basic units and combinations thereof. The conversation between Bead and Allele above show how situational works at the level of discourse. Both the words of the conversation and the images, actions, and choices the players have experienced are situated within the special context of video games where there are two actors, the player and a surrogate body in the virtual word (the character the player manipulates) and where the both the player and the surrogate can "die" in different but related senses, which is why players can say both "I died" meaning the player lost and the character "died" but can "come back".

So, the two modalities, words and images, at the level of general meanings (not names or specific things) work pretty much the same way. But, then, what makes them different? Consider the multimodal message below:

Apple

Figure 3: Apple represented in two different modalities

Both the picture of the apple and the written word are images. They differ not in being images, but in terms of the systems through which we interpret them. Print is part of the linguistic system. In that system the image ("apple") is arbitrarily related to its socio-conceptual meaning (just as the sound of the word "apple" is arbitrarily related to its meaning. The apple picture is not arbitrarily related to its meaning (apples). The image of the apple is part of an iconic system. The image (of an apple) is not "looks like" an apple. While these two different systems are processed in different ways in the brain (and, thus, "feel" different), they relate to the exact same socio-conceptual meaning (the criteria for being an apple). Thus, the "text" above is, all by itself, fully redundant. It just says/apple/twice.

Obviously, we do not usually engage in useless redundancy. The "text" above alerts us to a fourth property that words and iconic images share: genre. People have yet another type of social convention in which they take any "text" to be part of a specific genre. By "genre" I mean a socially shared way of relating (words and/or iconic images) together to constitute a specific practice meant to accomplish a specific goal (or goals). The "text" above could, for example, be interpreted in terms of the practice of children's early books meant to teach literacy. If we modified the "text" as below, then we might well readily interpret the text in terms of the practice of a phonics lesson (since making part the "a" bigger calls attention to it):

Apple

Figure 4: Figure 3: Apple represented in two different modalities with additional information in the written mode of communication.

If we cannot imagine some genre (practice) within which to place the "text", then it remains redundant.

12.Studying a mode

The proposal I am making here is that the study of any mode of communication needs to be organized into the following questions. These questions are meant to be a uniform framework for studying different modes and eventually their combination into multimodal systems.

1. What are the "rules" (conventions) that establish parts and wholes in the system? This is the "grammar' of the system.

2. What the basic units of the system and how do these basic units give rise to the ontology of the system (the set of assumptions about what exists from the point of view of the system)?

3. How does the 'grammar' (parts, wholes) give rise to choices the choosing of which communicates both in regard to what was chosen and in respect to what was not?

4. How does affective expressivity work in the system, that is, how does the forms of expression from the system put to use in context give rise to feelings, emotions, affect, and various aesthetic effects?

5. How can structures from the system (basic units, and parts combined into wholes) be placed in sequences across time or laid out across space to give rise to discourse (the accumulation of information across a bounded amount of time or space)?

6. How does the pattern of structures (parts, wholes, basic units) and sequence of such structures give rise to a style that is used to for particular purposes and tasks with regard to users taking on certain identities?

7. How do producers and users of the system adapt, change, transform, create, and negotiate specific situational meanings relevant to contexts of use and the purposes being served in them and the practices being carried out?

13. Multimodality

Our discussion of systems above has been background so that we can now deal with mixing systems into multimodal "texts". A multimodal text is not just a combination of systems. The parts (the different systems being combined) combine into a whole that is more than the sum of its parts, the very definition of a "system". Thus, a multimodal text is a system (of systems) itself. Just as with single systems, this larger system is subject to social conventions through which units, parts and wholes, affective effectivity, choices, style, and sequence are given meaning.

There are some well-studied multimodal systems. One example is manga, Japanese anime comic books, an art form with a long history going back to the 19th century (Napier 2005). Manga involves a great many different sub-systems such as boxes, color, characters and action, narrative, speech balloons, and reading direction. Some of the complexity is captured in the quote below from a site that gives tips for creating "an authentic manga comic strip":

While the style and finish of manga art is relatively minimalist in comparison to other types of comics, this apparent simplicity is deceptive. Every line is a choice made by the artist – the thinking is to never use 10 strokes to depict something if just a single, well-placed line would suffice.

This principle of concentrating on the essentials permeates throughout manga art creation. Every panel is an exercise in choice: size, zoom, camera angle, speech bubble positioning, and type of background. Every page works as a whole to control the reader's experience, particularly in pacing.
https://www.creativebloq.com/how-to/how-to-create-a-manga-comic-strip

Multimodal systems like manga (and manga comes in different styles) have been well worked out thanks to the amount of overt teaching available in books, websites, and other media, deeply devoted fans, and a great many budding anime artists. These people have done as much or more than scholars to explicit the underling "grammar" and social conventions that constitute a well-developed multimodal system. There are other such well-developed complex multi-modal systems (such as real-time-strategy video games) that have been the subject of enough discussion to constitute key places to begin the study of multimodality as an emerging discipline.

Manga and real-time-strategy games are genres of multi-modal systems. Just as genres exist at the level of activities (e.g., a phonics lesson) and at the level of "texts" (e.g., narrative), they exist at the level of multimodal systems as systems of systems (e.g., manga, real-time strategy games).

I do not have the space here to show each step of our approach to systems in action in a particular analysis of a multimodal text. That would require another paper. What I can do is turn to an example that will explicate how multimodal analyses can offer important applications and implications.

14. User Interfaces: An Example

What I have tried to develop in this paper is the beginnings of a discourse analysis approach to multimodality. However, I believe that discourse analysis-and other sorts of interpretive research-should not be devoted just to description, but also to applications to important issues and problems. I want now to give an example of multi-modal discourse analysis that bears on important issues to do with learning. The example will look at user-interfaces in video games, though the discussion will eventually have wider implications.

Games of the sort that I will discuss here are active experiences that are well-design to help players solve problems. They are a form of problem-based learning and, as such, have important implications for learning inside and outside school (Gee 2007, Gee 2014b)

Consider the screenshot below. What mode is it in? We cannot know until we know what system gives it meaning. There are several possibilities. Here is one: This could be a screen from a simple video game, say a game in which the player is the white dot and the enemies are the red dots. The cones of light represent where enemy characters are looking. The point of the game might be to move and hide and get out of the room without being seen by the enemies. This screenshot looks like early video games when graphics were quite limited.

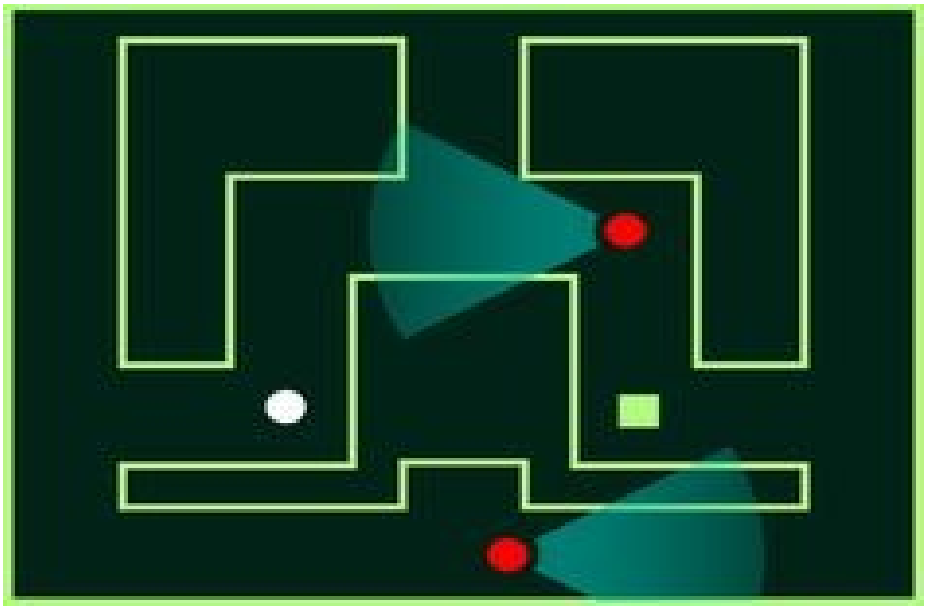

Figure 5: Radar map of the game Metal Gear Solid.

Here is another possibility. This screenshot could be (and, in fact, is) a "radar map" that appears in the top right corner of the game Metal Gear Solid (produced by Konami Computer Entertainment Japan and released for the PlayStation in 1998). In this game, the player plays the famous "Solid Snake", a master soldier of fortune who engages in stealth to infiltrate enemy bases. The map tells Solid Snake and the player what the spaces they are about to enter are like, where enemies and cameras are positioned, and where they are looking. It is sort of a way for Solid Snake and the player to look through walls. It becomes a device to help players form strategies before they move into action.

This screenshot is part of two different systems when it is a screen in a game and when it is a small radar map in a realistic-looking game world. These are two different modes. One is a simple game. The other is part of the user interface in a game. It works quite differently in the two cases. When it is a radar map in a realistic looking game world, we have a multimodal system (of systems); we have combined a user interface system with a 3-D action game system, two systems that are part of one bigger one.

When we consider a game like Metal Gear Solid as a multimodal system (of systems), it is important to realize the full range of complexity we face. The game itself is composed of two multimodal systems, the game world the player acts on and in (with a surrogate body, that is, the character he or she manipulates in the virtual world) and the game's User Interface. We will call the first system the "game world" ("GW" for short) and the second the User Interface ("UI" for short). The combination of GW and UI we will call the "game", so game = GW + UI.

The GW is composed of the following modalities, each of which is a system in its own right:

- Graphics
- Written Language
- Spoken Language
- Characters, Actions, Environments
- Game Mechanics
- Sound (like footsteps, alarms, gun fire, and so forth)
- Music
- Haptic

There are several different types of UI devices. I will mention just two here. One type is composed of the methods a player uses to play the game, for example, keyboard control or mouse control (or both) on a computer, handheld controllers on a computer or game platform, movements on a touch screen, and so forth. We will not be concerned with these types of UI devices (though they are important). I will call these types of UI devices "access devices".

The other type of UI devices are various interfaces like map screens, directional indicators, inventory screens, health bars, weapon wheels, damage meters, and information about enemies. These devices, the ones that I will be concerned with here, are devices to aid the player's strategies and decision making. They are, in a sense, cognitive aids. I will call these types of UI devices "cognitive devices".

And UI device may can contain its own, possibly multimodal, sub-elements. For example, a status bar system (on UI device) may be composed of juxtaposed linear measures of health, stamina, and "mana" (magical power), each color coded, perhaps with numbers attached to each.

Here I will consider just the relationships between the GW and the UI in a game, two multimodal systems that themselves compose a bigger system (the game). The basic units and their socio-conceptual meanings are different in each system (GW and UI). Of course, each modal sub-system in the GW (e.g., graphics, sound, game mechanics, etc.) has its own basic elements. However, each sub-system in the GW and its basic units are all devoted to creating an experience of action, choice, and problem solving for the player. They create a playable (virtual) world. The GW is an experiential base for learning akin to the real world.

Each device (e.g., a map or health bar) that is part of a UI has basic units that are informational in the sense that they are meant to inform, assess, and guide decision making and problem solving. They create

an information system that helps gamers navigate the game world, accomplish goals, and assess their progress.

A GW is interpreted in terms of an activity genre. In the case of Metal Gear Solid, the genre is stealth-based action game, a genre that sets expectations and strategies of gamers. A UI is interpreted in terms of an activity genre, as well. One way to think about UI activity genres is to see a UI as an exterior cognitive model of a certain type of thinking, planning, and goal accomplishment. In this sense, a UI is a set tools that help the gamer to look at the world (the GW) in certain way, in terms of affordances for certain types of decisions, actions, and problem solving.

The UI genre of Metal Gear Solid is a model of looking at the world (the GW) in terms of opportunities for infiltration and espionage. So, too, the UI of SWAT4 models how SWAT team members look at the world when they do their job. Portal's UI models how to think about the GW's physics so as to move and escape a set of prison like spaces.

User interfaces in modern games can go much further. In a game like World of Warcraft players can engage in raids where a dozen or players enter a dungeon and have to collaborate to defeat the enemies in the dungeon. Such raids take a good deal of pre-planning and on the spot strategizing and collaboration (using written and/or voice chat). In raids, players have on their screens a great many UI devices like the radar map above. These are images, lists, and graphs that tell then things like how much damage each player is doing, when to heal or otherwise "buff" another player, whether every player in a group is doing their assigned task and doing it well, how the statistical underpinnings of the game are working to their advantage or disadvantage, and many other things. Indeed, these user interface elements-sometimes called "mods"-often take up so much room on the screen that it is hard to see the action.

Each mod is an information tool that allows players to plan, think, evaluate, and strategize before, in, and after action. The various mods constitute a "theory" of how the game works, how best to play, and how to form and change strategy on the move, collaboratively with others. For example, a damage meter shows how well each character in a group of players is using his or her particular skills. Damage meters help players to evaluate their own performances and that of other players in their party. They also handle the "free rider problem", ensuring no one can slack off and let others do the work. The mods are made not just by the company that designed the game, but a great many are made by players themselves in web-based affinity spaces where players engage in production, discussion, teaching and learning.

Below are two examples of screens from World of Warcraft. Each screen has UI mods overlaid on the game world. Players can use many different mods, pop some off and on, and arrange them on the screen in various ways. Guilds often require players to use certain mods for raids. The mods are an information interface explicitly meant to be used for reflection in and on practice (play), strategizing, collaboratively planning, coordination, and assessment of progress. They are a collaborative socio-cognitive system designed for problem solving and learning to problem solve. They are often the source-on fan-based websites out of the game-of discussion, critique, and modding, and "theory-crafting".

The UI in World of Warcraft becomes a source of theory and theorizing over-laid on experience (play). Each informs each. More experience gives rise to more theory and more theory changes experience and the virtuous cycle goes on, much as it does in science.

Figure 6: Screen from World of Warcraft.

If you return to the questions in the section "Studying a Mode", you will see that those questions are not just a way to study a multimodal system like World of Warcraft (as GW + UI). These questions are part and parcel of what designers ask when they design both the GW and the UI. They are part and parcel of the discussion and work players do when they debate how best to set up the UI on their screen and when they mode UI devices to improve the UI for their purposes. These questions then are both the basis for analysis, but for design as well.

15.Implications of our example

Our discussion of World of Warcraft as a multimodal system composed of a GW and a UI (both of which are themselves multimodal systems) is of a great deal more general relevance than being applicate to a game or games in general. This system we have been discussing is a core example of a particular kind of learning system. This type of learning system is a collaborative problem-based learning system. In such systems the UI serves as a growing and changeable theory of practice, the practice of solving certain sorts of problems.

Such systems would appear to be much deeper-if we are interested in collaboration, problem solving, and continual improvement-than the sorts of learning systems we find in school. They are, interestingly, far deeper, more effective, and more creative and engaging than the derivative multimodal systems we too often find in the e-learning so popular today, much of which simply copies more patterns from traditional schooling.

The study of multimodal has the potential to have very real applications to how people learn and work together in society. The pressure, as always, will be to push for innovation, deep learning, and proactive productive participation and not give in to the ever-present pressure to standardize and commoditize in the name of short-term profit (Gee 2017b).

At a more general level, we need better understanding of multimodality-and of the relationships between language and other modes-because if schools, and society as a whole, do not make learning based on these news "ways with meaning" widely and equitably available, we will deny many young people some of the most powerful tools available today for learning and active participation in society. At the same time, new digital media of all sorts can have bad (even immoral) effects, good (even moral) ones, and trivial (even time wasting) ones. We educators cannot just take digital media-almost all of

which is multimodal today-as it comes. We must shape it for the good of young people, society, and the world. And we cannot shape what we do not understand.

References

Gee, J. P. (2004). *Situational language and learning: A critique of traditional schooling*. London: Routledge.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy*. Second Edition. New York: Palgrave/Macmillan.

Gee, J. P. (2014a). *An Introduction to discourse analysis: Theory and method. Fourth Edition*. London: Routledge, 2014

Gee, J. P. (2014b). *Unified Discourse Analysis: Language, Reality, Virtual Worlds, and Video Games*. New York: Routledge, 2014

Gee, J. P. (2017a). *Introducing discourse analysis: From grammar to society*. London: Routledge.

Gee, J. P. (2017b). *Teaching, learning, literacy in our high-risk high-tech world: A framework for becoming human*. New York: Teachers College Press.

Halliday, M. A. K. (1978), *Language as social semiotic: the social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M. A. K. (1985). *Written and spoken language*. Geelong, Victoria, Australia; Deakin University.

Halliday, M. A. K. (2013). *Halliday's introduction to functional grammar*. Fourth Edition. Revised by Christian M.I.M. Matthiessen. New York: Routledge.

Halliday, M. A. K. & Greaves, W. S. (2004). *Intonation in the grammar of English*. London: Equinox.

Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K. & Matthiessen, C. M. I. M. (1999). *Construing experience through meaning: A language-based approach to cognition.* New York: Continuum.

Jewitt, C. (ed.) (2009). *The Routledge handbook of multimodal analysis*. London: Routledge.

Jewitt, C. and Kress, G. (2003). *Multimodal literacy.* New York: P. Lang.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. London: Routledge.

Kress, G. and Van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London: Routledge.

Kress, G. and Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication.* New York: Oxford University Press.

Napier, S. L. (2005). *Anime from Akira to Howl's Moving Castle, Updated Edition: Experiencing contemporary Japanese animation*. New York: Palgrave/Macmillan.

Saussure, F. de (1986). *Course in general linguistics*. Chicago: Open Court. Originally published in French in 1916.