A Visual Lexicon[1]

Neil Cohn
neilcohn@emaki.net
www.emaki.net

Abstract
One of the most recognizable graphic components of the visual language of "comics" is the "panel," a demarcated frame of image content put into discrete sequences, thereby seeming to be the primary unit of expression. However, meaningful visual elements do exist that are both smaller and larger than this encapsulation of image and text. Spoken languages also have variation in sizes of lexical items above and below their primary sequential unit of the "word." This paper will address these varying levels of representation in visual language in comparison to the structural make-up of verbal language, to aim toward at what it means to have "visual lexical items."

Keywords: visual language, comics, lexicon, panels, construction grammar, morphology

Introduction

The units of language come in many sizes. Some pieces are the size of words, such as *coffee*, *jump*, and *fantastic*. Pieces smaller than words are morphemes like *re-*, *-ing*, and *un-*. There are also formalized patterns of words put together, including idioms like *kick the bucket*, *miss the mark*, and *hung out to dry*, or even grammatical constructions such as *What this X doing Y?* manifested as *What's this fly doing in my soup?* or *What is this scratch doing on the table?* (Kay and Fillmore 1999). Ray Jackendoff (2002) has proposed that all of these "constructions" can be included in the mental lexicon. This breaks step with previous approaches to the lexicon that insist on maintaining the level of the "word" as the sole purview of "lexical items," lying in wait to be pulled out into various types of grammatical patterns. The change from this view reflects the sentiments of the "construction grammar" movement in linguistics, which examines form-meaning pairings in language of varying sizes (Goldberg 2003). Here, Jackendoff (2002) departs from traditional models of grammar by denying a separate "lexicon" that exists outside other grammatical structures such as syntax and phonology. Rather, in this model the lexicon emerges out of the mutual interfacing of parallel structures of grammar: phonological, conceptual, and syntactic structures.

While Jackendoff deals with structures of verbal language (with sign language implicitly accepted), similar issues of size variations can be addressed in the visual-graphic modality as well. Indeed, drawing images joins the vocal creation of sound and gesticulation to form the only three channels of expressing propositional concepts available to the human animal. While semiotic expression can follow from other senses like taste or smell, they cannot produce conceptual information in any comparable capacity. Extending this observation, this project hypothesizes that whenever *any* of these conceptual expressing modalities takes on structured rule-bound sequences (a grammar), that form becomes a "language." Thus, we have a verbal language of sound, a signed language of body movements, and a *visual lan-*

---

[1] All images are copyright © 2006 Neil Cohn, except those cited throughout the text. Cited images are copyrighted their respective owners, and are used purely for analytical, critical and scholarly purposes.

guage (VL) of sequential images.[2] This visual language appears most commonly in the social objects of "comics" — essentially the *parole* to the visual *langue* (Cohn 2005b). Moreover, both sequential and non-sequential forms of these modalities can unite in multimodal combinations, as in speech-gesture (McNeill 1992) and text-image relationships (McCloud 1993).

Note that this hypothesis for "language" involves exactly the three structural features found in Jackendoff's Parallel Architecture: modality, concepts, and syntax. In this case, for a *visual* language, the modality component to Jackendoff's Parallel Architecture becomes "photological structures" to account for the principles necessary for recognizing and constructing visual representations as opposed to verbal ones. The addition of such a structure should already be crucial for the grammar anyhow, since "writing" requires stored memory of graphic representation that must link to the other aspects of grammar. Following the constructional definition of a "lexical item" as *a meaningful unit or combination of units of form-meaning pairing,* this paper will address the varying levels of representation in visual language to arrive at a general understanding of what it means to have "visual lexical items."

Attention Units

The most obvious unit of representation in visual language comes in the form of a "panel" or "frame," which is most often clearly demarcated by some sort of encapsulated border, be it a drawn frame or empty white space. As the primary components of the sequence, panels are the essential unit of *syntax* in this visual language. Following the linguistic definition, syntax is here conceived of as *a system of rules that govern the ordering and arrangement of units*. Though this definition might not have been followed, or even known of, various other approaches have addressed the topic of a "visual grammar."

Perhaps the most well known "visual grammar" is Kress and van Leeuwen's (1996) semiotic approach. As Forceville (1999) notes, Kress and van Leeuwen's model suffers from its strict orientation to 'social semiotics,' suggesting that they embrace a more cognitive stance. However, the importation of a notion of "grammar" from linguistics itself leads to problems. Though Kress and van Leeuwen's model outlines the compositional elements relationally juxtaposed by force dynamic vectors, in no way is it a "grammar" in any real linguistic sense. Indeed, they acknowledge the metaphorical quality of their use of the term on first page of the book. Nevertheless, it is worth examining why such a metaphorical usage leads to problems compared to what a real linguistic notion of visual grammar might offer.

Outright, their model lacks the requisite Saussurean (1972 [1916]) paradigmatic and syntagmatic relationships of substitutable elements that has grounded the field for nearly a hundred years, instead drawing its power only through spatially arranged semiotic components. True syntactic categories are assigned by distributional regularity within a discrete array, not simply by being semantic objects. That is, a "noun" is only a noun because it falls into certain distributional (*syntactic*) positions within a larger sentence; for instance, in English a noun can potentially follow a determiner (ex. *the, some*) and adjective (ex. *big, smelly*). A noun does *not* get its grammatical category because it represents a (*semantic*) "person, place,

_____

[2] It should be noted that writing systems are not considered "visual languages," though they might be considered "visualized (verbal) languages." "Writing" is essentially the importation of the natively verbal structures mapped into the visual modality (see Cohn 2005a).

thing, or idea," of which there are innumerable exceptions, including *redness, concert, millennia,* and *finesse* (Jackendoff 2002).

This leads to the second issue with Kress and van Leeuwen's "grammar": it entirely lacks syntactic categories that outline specific roles played by individual units in relation to the sequential whole. This "visual grammar" only contains observable semantic components and their spatial relations, not rule-bound categories determined by their distributional arrangement (though they eschew the need for rules upfront, they dismissively admit that it is central to the notion of grammar). Indeed, their sense of "Actor" aligns well with the notion found in linguistic *semantics* of an Actor or Agent: an entity that carries out an action (e.g. Jackendoff 1990, "Agonist" in Talmy 2001). In this sense, Kress and van Leeuwen's approach is (admittedly by them) merely syntactic by metaphor, and should be treated as such. It is more useful for its commentary on compositional qualities of individual arrays than as a "visual grammar."

While not overt, the belief that syntax lies within a singular image assumes that an individual image is equal to a "sentence," since this is where syntax operates in verbal language. This equation is motivated again by semantic concerns: an image is as much if not more densely filled with information as a sentence. However, information structure does not necessitate syntax — only semantics. As stated previously, the approach herein takes panels to be the primary syntactic unit, though this does not equate to panels being a "word" in any sense of information structure. Quite clearly, pictures contain more conceptual content than words (a thousand or so, as the saying goes). Nevertheless, this "visual language" approach to panels as syntactic units acknowledges that they are subject to distributional regularities within a syntagmatic sequence in the same way that words are. That is, panels are not the visual equivalent of words (and neither are elements of an individual image). Rather, both words and panels play similar structural roles within the confines of their own systems of grammar. Any visual display that lacks the requisite sequence needed to have distributional regularities thereby does not have any qualitative *syntax*, and is thereby also not *linguistic* either.

Perhaps most well known of the direct approaches to visual syntax pertaining to "comics" sequences is McCloud's (1993) taxonomy of "panel transitions," which actually did attempt to define specific roles played by panels, though he limited these relations to apply only to immediate constituents. Both Saraceni (2000, 2001, 2003) and Stainbrook (2003) attempt to expand on McCloud (1993) by associating it to concepts in applied linguistics and discourse theory. However, these approaches again assume the "sentence level" for panels,[3] hence pursuing the track of discourse theory, but leaving the same problems intact. Indeed, by expounding broader descriptions of coherence between panels, they skirt the advantages McCloud gained through an explicit formulation of relational panel roles.

McCloud's own approach suffers from a variety of problems as well. However, addressing these more syntactic theory oriented issues and proposing an alternative model of visual grammar are ancillary to the concerns needed for exploring a "visual lexicon" alone. For the present discourse, visual grammatical categories will be glossed as "narrative states." The important point here is what *visual syntax* of

---

[3] Saraceni (2000:96) uses the Kress and van Leeuwen model overtly, and provides a chart weighing the pluses and minuses of comparing *paragraphs, sentences, clauses,* and *phrases* to panels. Conspicuously absent is the level of the "word," though he no doubt assumes it less worthy than a *phrase* for which he states, "the information level never equates with that of the panel."

a visual language must entail. Following its linguistic meaning, visual syntax is taken as *a system of rules that govern the distributional ordering and arrangement of units within a syntagmatic whole,* thereby requiring a *discrete array* of units and excluding the compositional qualities of individual images. With this broad foundation for visual grammar, let us return to examine the properties of its primary unit, the panel.

Within panels there are two distinguishable characteristics with regard to their relationship to the overall grammatical sequence: positive and negative entities. "Positive" or "active" elements make up the figures and focal action of a panel, while "negative" or "passive" elements are the background information (Natsume 1997). Since visual syntax is concerned with the sequential relations of panels, active elements become the "grammatical entities" involved in the actions or events of the visual sequence. A similar notion is provided by Talmy's (2001) Figure/Ground distinction, drawing upon Gestalt psychology. He describes the Figure as "a moving or conceptually movable entity" contrasted by the Ground which is "a reference entity… that has a stationary setting relative to…the Figure's site, path, or orientation" (Talmy 2001: 184). Due to the difference between iconic and symbolic expression, in visual form, the Figural entities repeat across sequential units rather than become presented as an isolated unit within the sequential array. As a result, it is out of this sequence that active entities find their definition. While active and passive elements seem to prototypically correspond respectively to foreground characters and the environment that they are in, ultimately such assignment comes through the *sequence* itself, and not compositional arrangement. For instance, in a sequence like this, the environment is positively charged and the person becomes negatively charged:
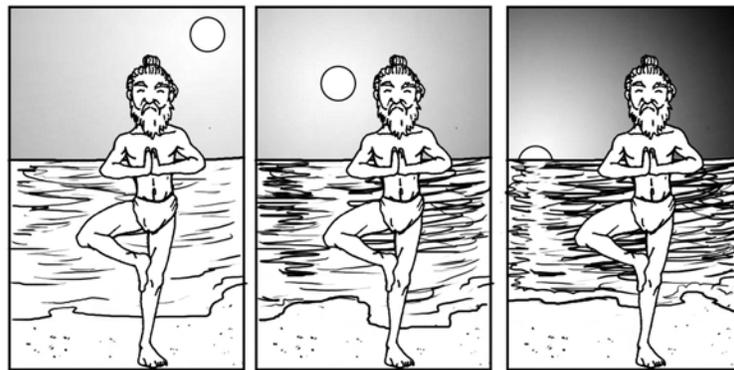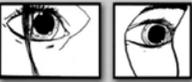

(1)

While the sunset changes the environment, the ascetic stands still. Granted, like facets of most passive features, the man is not completely negligible for his *semantic* value. The sequence does convey his resolve to stand in one place over a long period of time, and that is undeniably an important aspect of the sequence's *overall* meaning. However, despite this importance semantically, the man in this sequence does not affect the *syntax,* which must deal with the functional relations of panels to the whole. The event in these panels depicts the sun setting. Thus, the syntax is determined wholly by the *movement of the sun* and the effect it has on the surrounding environment, becoming the positively charged Figure to the negatively charged Ground of the man. Though a predisposition might exist for considering

compositionally foreground elements as the prototypical positive elements of the scene, such distinctions are not absolute.

Moreover, because of the necessity for recognizing relationships across sequences, a single panel must not be overloaded with positive elements. If too many positive elements exist in each panel, it becomes more and more difficult to parse the syntactic change. This is not necessarily a structural restriction per se, but can be likened to a maxim of conversational Quantity (Grice 1967), where only the sufficient information is required to achieve communicative success. Panels do not intrinsically limit including more positive elements than are necessary, though they may burden the efficacy of the visual communication. In sum, while active and passive elements add up to create the semantic whole for the sequence, it is only the active elements that engage in the syntax.

Based on the amount of positively charged entities they depict, paneled representations can be categorized in a *Lexical Representational Matrix* (LRM).[4]

	BASE	FRAMING	
		DIVISIONAL	INCLUSIONARY
ACTIVE — Polymorphic			
Macro			
Mono			
Micro			
PASSIVE — Amorphic			

The highest level features *Polymorphic* panels, which allow for event representation to exist within the boundaries of a singular frame through the *repetition of*

---

a single entity at different stages of an action. Below this are *Macros*, containing more than one grammatical entity — a positively charged element determined by its role in the sequence. *Monos* are one level lower, with panels that depict only a singular entity. At the bottom of the Active tier of the LRM are *Micros*, which feature less than one entity and often come in the form of "close-ups." Finally, at the very bottom of the LRM is another Passive tier, holding *Amorphic* panels that have no active entities whatsoever. These are commonly views of environmental features, though they can also include animate objects depending on the context of the sequence.

The vertical axis shows a progression from full actions to scenes, to one entity, to less than one entity, and finally down to no entities at all. The Framing Tier set to the right takes these Base distinctions and applies varying paneling options to them. *Divisional* panels divide a single image into image constant parts, while *Inclusionary* panels use frames within frames. Since these framing devices break up Base assignments their componental parts might belong to categories lower than the whole category. Thus, a Divisional Macro might end up featuring two Monos, while a Divisional Mono might feature multiple Micros.

While they might intersect, the rankings within the LRM should be clearly distinguished from filmic notions of "framing" (see for instance Bordwell and Thompson 1997). Though cinematic framing such as "long shots" and "close-ups" might correspond prototypically to Macros and Micros respectively, ultimately these categories are not a one-to-one mapping. Notions of filmic framing certainly can apply to the depictions in panels, and function to crop information within the enclosed space of a cinematic screen. However, the determination of ranking within the LRM — and a "grammatical entitiy" —is based not only on the quantity of information within a panel, but also on the relational qualities of a panel to its sequence. Filmic shots might specify *how* the elements of a frame are shown, but the LRM measures *what* in the frame is important to the broader sequence in the first place. In this way, a close-up may also be a Macro because it has more than one acting entity, and a long shot could be a Mono by showing a single acting entity. While a close-up of an eyeball might be a Micro for a sequence of a *person studying*, it could be a Mono for a sequence of an *eye blinking*. Again, LRM rankings are wholly relative to content of a panel and its sequence. Take for instance this chase sequence from Scott Chantler's *Northwest Passage* (2005:2-3):

Figure 2.

This sequence shows a Native American running from angry frontiersmen. In the first panel we see multiple frontiersmen, the second just the Native American, and in the third we see both frontiersmen and (very small in the upper middle) the running chief. Throughout the latter two panels, the forest setting serves as the passive entity to the active people. The final panel can easily be identified as a Macro, because it shows both interacting entities: the Native American and the frontiersmen. The second panel contains one lone active entity in it, the Native American, and is thereby considered a Mono, though it uses a long shot to show

the action. Now, despite the fact that there are multiple "characters" in it, the first panel is also considered a Mono because the interacting "entity" is the *group* of frontiersmen as a whole. As dictated by the *sequence*, the frontiersmen function as a unified collection of individuals[5] in relation to the truly lone Native American chief. If another hypothetical sequence showed members of that group individuated and interacting (say, talking to one another about their chase), then they would become distinct entities unto themselves. Thus, like the assignment of active entities and LRM position, the whole of what constitutes an entity in the first place is also determined by the context of the sequential structures.

While syntax does occur to fuse the understanding of linear panels, they are not necessarily minimal "syntactic units" unto themselves. Take for instance adjacent Mono panels featuring different entities, which can potentially be combined into an "environment" into the same functional "narrative state" (i.e. "grammatical category"):

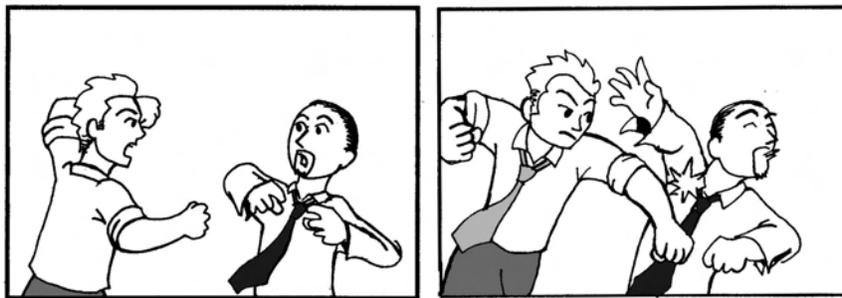**Figure 3a. Mono - Mono - Macro**

Figure 3b. Marco - Macro

Though two Monos are used in (3a), the amount of information here equates to that of a Macro, as evidenced in the distributional equivalency in (3b). To accomplish the single environment that (3b) shows in one panel (3a) must somehow fuse the first two panels together in order to connect with the final panel. I have named this process "E(nvironmental)-Conjunction" since it unites disparate units into a common conceptual environment. Further examples of this phenomenon occur below, with panels engaging in E-Conjunction bracketed for clarity:

---

[5] In Jackendoff 's (1991) terms, this entity of "group" would be a bounded concept with internal structure: it has no boundary limits, and distinct internal components could be separated from it yet retain the same concept. In contrast, the chief (and individual members of that category "group") would be bounded *without* internal structure: a person has a boundary limit and is not divisible into smaller parts of itself.

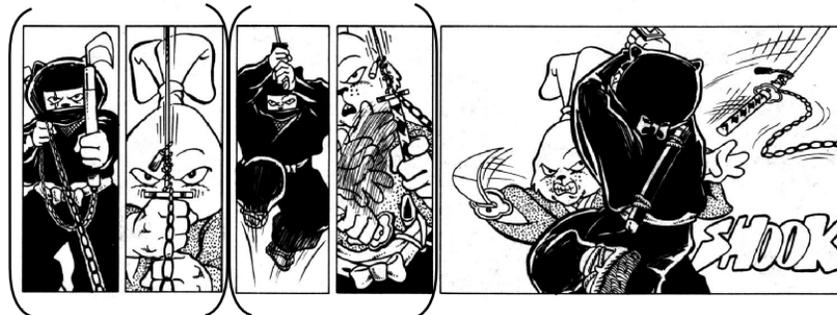**Figure 4a. (Samura 2004:26)**

Figure 4b. (Sakai 1987:28)

In (4a), the first two panels serve to set up the interaction that occurs in the final two panels. They both reflect a common function of establishing the context of the overall event. A similar effect is achieved by the first two panels in (4b), but continues into the second set of panels where the action is initiated. The final panel in both features a Macro where the entities unite to fulfill the predication of the interaction.

Because Mono panels such as these join within a singular narrative function, panels cannot be considered as minimal syntactic units alone. E-Conjunction shows that panels can be grouped into functional constituents that interact with the larger sequential whole.[6] Rather, panels seem to play a role as "attention units" (AU)[7] for the overall schema of the interrelation, since they *focus the attention* of the reader on particular elements of the sequence. A Macro panel focuses the attention on larger displays, while Monos and Macros hone that focus to more precise elements of the interrelation. In this light, E-conjunction breaks up a singular narrative "moment" (i.e. grammatical category) into multiple AU to achieve certain representational intents. This could be helpful, for instance, in that E-conjunction can aid in upholding the maxim of Quantity that prevents the overloading of a single panel with too many active elements by breaking up narrative segments into smaller more manageable parts per unit.

Beyond E-conjunction, Polymorphic panels allow for grammatical structures to occur within a single panel, potentially carrying syntax of a phrasal level or higher. The attention unit in these panels becomes cast much wider, to show the pieces

---

[6] Incidentally, this grouping of panels into functional "chunks" also provides evidence against McCloud (1993) and others (e.g. Saraceni 2000 and Stainbrook 2002) transitional models of visual syntax, since those larger sensitutans must connect to each other, transcending the direct linear relationships.

[7] Thanks go to David Wilkins for contributing this term.

of an action or event all at once. Note this example where a figure jumps from building to building in a singular panel (Dixon and Johnson 2003:11):

Figure 5.

Here, an event structure unfolds in full within a singular Polymorphic panel by repeating the singular entity of the martial artist multiple times to show movement. For both E-Conjunction and Polymorphics, the level of the panel cannot be assumed as equivalent to a single narrative segment. In this way, panels serve to facilitate what Leonard Talmy (2001) calls the "windowing of attention." While certain elements in verbal sentences will be considered at the core of the interrelation, others may be pushed to the periphery. By highlighting different parts of the conceptualization, speakers "window" aspects of the overall event frame. A "maximal windowing" allows the full conceptualization of an event to be included into a sentence, though different portions can be "gapped," as shown in these of Talmy's examples (2001: 269):

a.   *With maximal windowing*

   i.   My bike is across the street from the bakery.
   ii.   Jane sat across the table from John.

b.   *With medial gapping*

   i.   My bike is across from the bakery.
   ii.   Jane sat across from John.

c.   *With initial gapping*

   i.   My bike is across the street.
   ii.   Jane sat across the table.

Polymorphic panels can serve to maximally window event frames, while the selection of other levels vary based on the intended representation. For instance, breaking up of a single environment into multiple panels through E-Conjunction can bring focal attention to each of the entities involved, rather than to the scene as a whole. This windowing can be exemplified quite literally by the use of Inclusionary panels, which embed a panel into another panel. While these can be used for many

grammatical purposes, including E-conjunction, marking off a section of a whole image focally distinguishes that element from the larger scene, as in this example (Miller 2000):

Figure 6.

The enclosed panel in this instance literally "windows the attention" by demarcating a space within the representation to focus upon. However, it could equivalently been drawn as a separate Micro panel modifying the initial Macro scene. The enclosure of this modifier within a larger panel — instead of separated in its own panel — again shows how single panels can contain more than a single unit of syntax. In this case, the modifier and "modified" use two panels on a single image.

Thus, while they form the level of analysis for syntax, panels themselves do not represent isolated syntactic units. Regardless of the grammatical role they play — whether as segments of a scene, modifiers, or whole events — panels serve to focus attention on various parts of the conveyed information.

Smaller Than Syntax

Panels may be the most noticeable unit of encapsulation in VL, but very rarely are panels maintained as fossilized wholes that repeat in usage the way that we consider words to be units of a vocabulary. That is, a "visual dictionary" listing of panels might seem impossible to create, since most of them are distinct and unique. By and large, the internal structure of panels seems to change constantly, though productive elements within them might stay the same.

While this creative capacity for panels is dominantly true, some consistent panel forms do exist. This systematization is most exemplified by *Wally Wood's 22 Panels that Always Work,* a "cheat-sheet" of panel compositions created by the legendary comic artist for making "boring" scenes of lengthy dialogue more visually interesting (Johnson 2006):
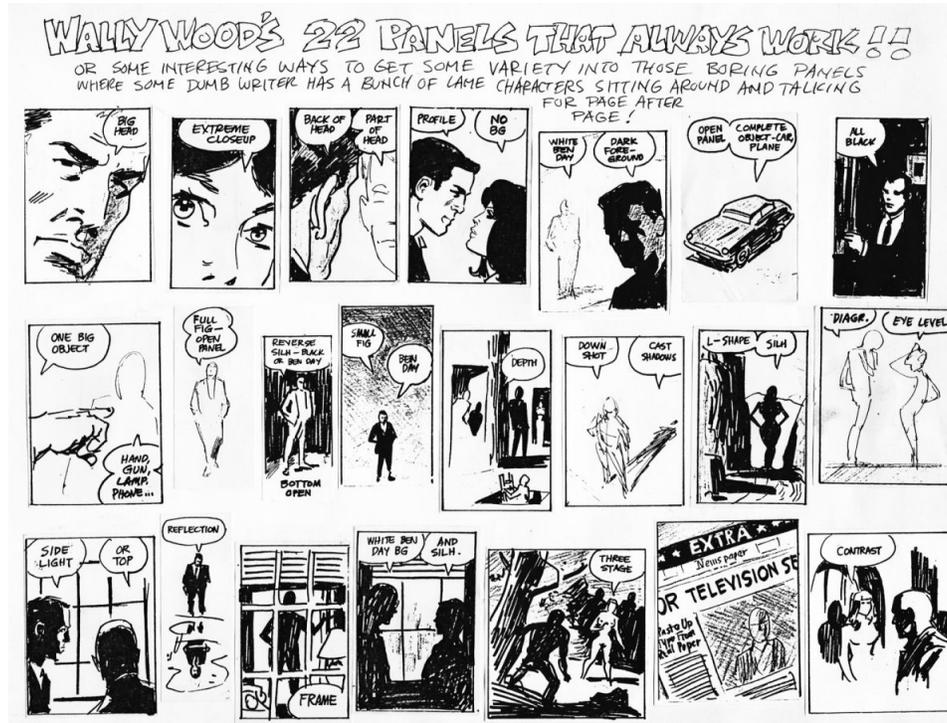
Figure 7.

Years after its creation, an editor at the Marvel Comics company made a paste up of Wally Wood's originals to disseminate to other artists, resulting in countless copies floating around the industry (and now Internet) for several decades (Johnson 2006). While no formal studies have confirmed the reach of these schemas, the spread of this cheat-sheet has led to an acknowledged pervasive use of these panel compositions across authors' works.

While *Wally Wood's 22 Panels that Always Work* provide systematized panel sized units of expression, most remain unconventional in their make-up. As discrete syntactic units, the internal structure of panels is unlike *analytic* languages like English, where morphology — the internal structure of "words" — plays a fairly small role and word forms are both consistent and enter syntax. Rather, VL panels can be regarded somewhat akin to *synthetic* languages like Turkish or West Greenlandic where smaller productive elements combine to form units that enter syntax in various ways. This is not to say that paneled visual languages *are* synthetic or analytic, but that they exhibit a similar method of chunking information into workable units rather than letting meaningful information stand alone as units unto themselves. These two strategies run on a gradation, from those that feature conventionalized syntactic units (*analytic*) to those that use smaller combinable parts to create larger unconventional syntactic units (*synthetic*) (Haspelmath 2002).

Here again emerges the usefulness of not thinking about a lexicon comprising its own structure in the grammar, because parallel processing allows meaningful units to depart in size and be assembled productively in a variety of ways based on the features of the system. This is especially useful for an iconic lexicon, which can vary the representation of entities across panels, though visual features will remain constant. For instance, in this example the same characters persist through many of the panels, and most of the graphic linework for each of them is consistent though it changes in each panel with different perspectives, sizes, and poses (Kibushi 2004, excerpted):

Figure 8.

Even though the overall AUs vary — the unit at the level of syntax — there is still a consistent representational structure depicting the internal parts. These parts have a level of *productivity*, allowing for *creative alteration from a base form* that can then *combine with other forms into a compositional whole* (Haspelmath 2002). These malleable schemata are what seem to be stored in long-term memory, as opposed to full panel units that seem to be constructed online. Perhaps this is one of the reasons that consistent costumes have been favored in superhero comics, because they conventionally schematize an aspect of the character into long-term memory that still allows for variable productivity with regards to the rest of the representation (not to mention across different drawers). This free-form variability departs greatly from limitedly productive signs such as heart symbols or dollar signs, which have relatively little flexibility in their representations (to be discussed shortly).

Not all visual languages are like this though. In the sand narratives of the central Australian Arrernte community (Wilkins 1997), very little additive morphology seems to exist, and most visual signs appear in fixed representations. For instance, because their system maintains a consistent aerial view, a person is consistently drawn in an upside-down U-shape to show the iconic shape of an individual's imprint in the sand. The main variation to this sign occurs when depicting a person lying down, shown instead with a narrow oval (Wilkins 1997:141):
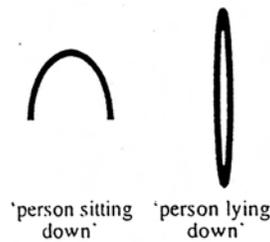
'person sitting   'person lying
down'   down'

Figure 9.

Because the sand narratives are drawn in real-time, each sign is created and used on its own. From all indications, no synthetic-like conglomerations into attention units seem to exist in Arrernte, and individual signs represent lexical items. Indeed, in this regard Arrernte is closer to English-type morphology than the visual languages that use panels. Based on productive time demands alone, this makes sense. Given the print cultures that panel-using VLs exist in, no demands on interactivity exist for the "visual speaker" to communicate quickly with the "visual listener" (Grice's maxim of Manner), allowing them to create as detailed representations as they wish. In Arrernte, the conventionality and simplicity of the signs aligns with the speed burdens enforced by real-time interactivity, not to mention adapting to the canvas of sand, which does not allow for high degrees of detailed representation anyhow. In contrast, pencil and ink on the portable surface of paper facilitate a vastly different relationship between producer and receiver. These aspects of time demands and media of expression bring up important concerns regarding the ecological and pragmatic contexts affecting the structure of the visual lexicon.

At the same time though, sand narratives are not wholly restrictive to the possibility that larger concatenations of signs can occur. Anthropologist Nancy Munn (1986) reports that the Australian Walpiri community use a very similar system to that of the Arrernte. She describes that certain element combinations occur at great frequency. For instance, while elements such as the U-shaped person might be used on their own, they also might be consistently paired with an object to create what Munn calls an "actor-item" (1986: 81). While these pairings might be as simple as a man with a spear, others become more complex to convey a large amount of narrative information. Sometimes, particular combinations of elements are highly idiomatic with specific fixed meanings, such as a specific way of drawing a man throwing a spear at a kangaroo, while other patterns on their own are ambiguous to their broader meaning without the context provided by the multimodal narrative. These complex patterns and basic actor-item pairs hint at some degree of morphology and idiomaticity in sand drawings.

Additionally, visual languages contain less malleable signs that cannot enter into syntax directly at all. These visual signs range from word balloons and thought bubbles to stars or hearts hovering above heads to show pain or love respectively, to sweat drops to show exasperation. Since these types of signs are often highly conventionalized, they often vary per culture (McCloud 1993:131, Shipman 2006), though they also might connect to deeper level cognitive processes (Talmy 2001:125, Forceville 2005).

While some focus has been given to identifying these conventions (see McCloud 1993, Walker 1980), little work has probed how these signs interact with and modify others. While most productive signs simply combine in ways that reflect iconic scenes, like other linguistic aspects of morphology, many of these conventional

signs alter an already existing sign, either through replacement or attachment. For instance, path lines affix to objects, appearing most often to show the progression of motion as "speed lines." These are "bound morphemes" since they cannot exist independently of a root object that they are modifying:

Figure 10.

Without connecting to a "moving root," speed lines could not convey the meaning of movement. In some cases, this depends on the depiction though. Path lines placed in the middle of a panel with no object might seem unusual, but those extending into the side of the frame might index that the root has gone out of view of the panel.

Path lines represent unseen aspects of the visual representation, and can range from depicting a trajectory attached to a moving object, to the fictive representation of smelly objects with wavy lines, to lines emerging from a mouth to show the path of air traveled in a breath. All of these elements are "invisible" in any "realistic" visual sense, emerging graphically only as conventionalized symbols (McCloud 1993).

Other "invisible" bound morphemes include types of "Carriers" such as speech balloons or thought bubbles, which link to a Root "speaker" or "thinker" through a Tail (Cohn 2003). These types of interfaces between word and image integrate the content of the Root and the Carrier to create a unified semantic bundle.

Figure 11.

Indeed, since Carriers can convey the expressive power of an entity's thoughts or speech, they are able to distribute animacy to anything they attach to. A thought bubble connected to a rock or chair immediately makes that object a "thinking being." This is different from interfaces that use Carriers unattached to any Root in the image, appearing as "narrative captions," and therefore are "free floating morphemes."

In contrast, heart symbols have much greater flexibility than Carriers in the way they enter into representations, though they present non-perceptual abstract concepts. Hearts can float around people to convey the general emotion of love or they can serve as the shape of an entire panel as an overarching semantic modifier. They can also be substituted into the eyes of a character to reflect desire felt for the object in vision, yet the syntactic component is still the entire figure, as in these examples by Derek Kirk Kim:
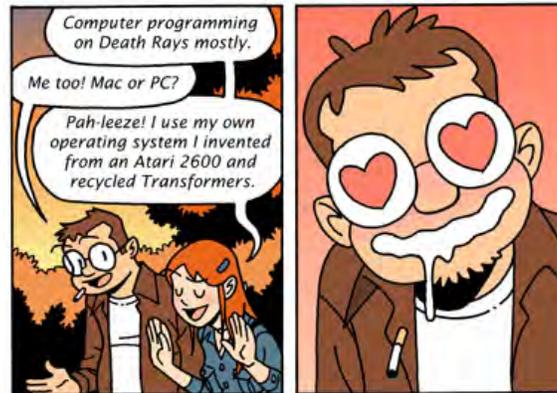

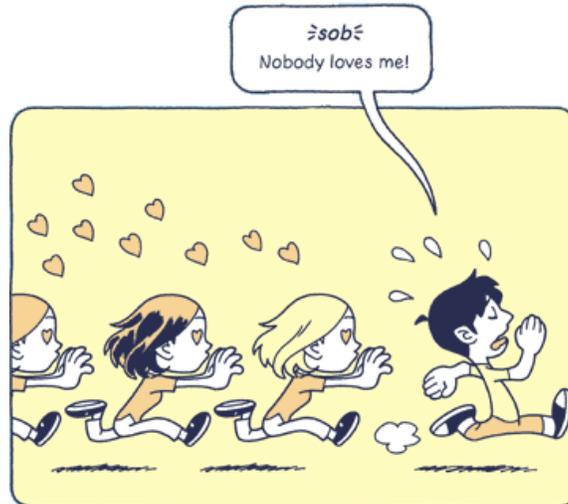**Figure 12a. (Kim 2001)**

Figure 12b. (Kim 2004)

In all cases, the heart symbols contain semantic information that is important to the overall meaning yet does not directly influence the overall structure of the scene, which is still dominated by the iconic features. Of course, this could change if the *whole* positive element was a heart that underwent some sort of predication (such as a heart in one panel getting an arrow shot through it in the next), yet this seems to be an exceptional case to regularized usage. Note that this sort of "grammaticizing" of a morpheme would seem very odd if applied to an abstract and fully bound morpheme such as path lines. Turning speed lines into "characters" would be far more difficult than creating a grammatical entity out of a heart symbol.

The productive sign of the human body often allows several places into which parasitic signs can provide extended meanings. Besides hearts, "suppletion" into the eyes can use various signs, including Xs (lack of consciousness or pain), spi-

rals (hypnotism), stars (desire of fame), and dollar signs (greed). The space above the head also allows several attached signs beyond hearts, like stars (pain), gears (thinking), exclamation marks (surprise), question marks (curiosity), circling birds (wooziness), dark scribbles or rain clouds (bad mood), bubbles (drunkenness), or light bulbs (inspiration). All of these signs use a specific place to modify the meaning of the base sign of a person, and none of them could do so without being attached to that root. Indeed, the distribution of where signs are put can change the meaning of the sign. While hearts retain the meaning of love or lust no matter where they are placed, stars mean different things based on whether they are in the eyes (desirous of fame) or above the head (feeling pain).

This distinction between productive signs (like human figures) and conventional symbols (like heart symbols and speed lines) can be likened to the linguistic distinction between open and closed class lexical items (Talmy 2001). Morphemes that are considered to be open are usually in a large class that is augmentable, while a closed class is generally limited and fixed. The difference here is usually drawn between "lexical" morphemes such as nouns and verbs, which are open and productive, while "grammatical" morphemes like prepositions belong to a closed class that is small and unchanging. While they do not necessarily play the same roles grammatically as Talmy's observed categories, broadly conceived, productive signs clearly seem to belong to an open class of visual signs while conventional symbols occupy a closed class, making the VL lexicon similar to other forms of language.

By and large, in the context of VL, these two classes of lexical items seem to take on semiotic peculiarities as well. For instance, closed class items such as hearts, speed lines, and word balloons all contain a higher degree of symbolism than far more iconic images of the human figure. It should be unsurprising that more iconic (and productive) elements tend to fall into an open class, since perceptual input can provide an unlimited array of potential objects and/or variations on those objects. Though icons can allow conventionality (such as the *smiley face* ☺, or many Arrernte signs), symbolic signs *must* be conventional.[8] Indeed, altering symbolic signs would be far harder than altering iconic ones, since they draw their meaning from communally agreed upon conventions. As a result, symbolic signs are forced to be more entrenched, and thus fall into a closed class category of morphological items.

Again, these components of individual images are not proposed as the equivalent to verbal words or morphemes, but the signs that construct panels and those that build words both constitute meaningful units below the level of larger syntactic units within their respective systems. In both cases, these signs might attach within or outside of other signs to alter the overall meaning. In the case of the visual signs, these elements contribute to the construction of the panel sized attention units (with limited productivity in the Australian signs), while in and of themselves they are below the level of syntactic analysis. Yet, as form-meaning pairings that contribute to the meaningful expressions of visual language they still remain a part of the visual lexicon.

---

[8] See Peirce's (1931) distinction between Legisigns and Sinsigns for more on the distinctions in conventionality of symbols versus icons and indexes.

Constructions

Constructions are form-meaning patterns in language that vary in size, and can include lengths longer than individual words. For example, the productive construction *verb - Noun Phrase - away* licenses both a verb and a direct object, manifesting in sentences such as *Bill slept the afternoon away* and *We're twistin' the night away* (Jackendoff 1997). Constructions can even reach the size of full length sentences, such as *The more you think about it, the less you understand*, which has an awkward syntactic pattern that seems to be stored in long-term memory (Goldberg 2003).

At present, not enough is known about visual language grammar to be able to identify any visual-only constructions similar to those in spoken language. While Polymorphic panels do enter the grammar at a higher level of syntax than Monos or Macros, they are still not constructions in the same way as idioms or other patterns since they are still generally built productively. That is, Polymorphic panels are not entrenched patterns. However, this does not mean that the potential for constructions does not exist in visual language, and we now turn to examining some contexts that herald this likelihood.

One consistent pattern across bimodal text/image syntagms seems to have emerged in what comic artist Neal von Flue has coined as the "set-up – beat – punchline" (SBP) pattern for comic strips (von Flue 2004). It begins with one or two panels "setting up" the humorous dialogue or situation, only to then give a "beat" or "pause" with a panel that has no text in it. Finally, the last panel delivers the punchline of the joke:

Figure 13a.

Figure 13b. (Cham 2004)

Figure 13c. (Pérez and Coughler 2003)

By all indication, it is difficult to state outright that the SBP pattern matches any pairing just between VL grammatical categories and an overarching constructional frame, because of the heavy meaningfulness of the text. However, it does still seem subject to certain syntactic principles. For example, in (13c) the beat segment is broken up into three separate panels for each of the different characters in the scene. Here E-conjunction seems to function with regards to this bimodal narrative pattern. Indeed, it would be difficult to identify visual syntactic categories since the first, second, and last panels are nearly identical — thereby lacking any *visual* syntax through change between them — which allows the text to dominate the semantics completely (see Cohn 2003). Moreover, the construction itself relies on the text for its effectiveness: the beat being the distinguishing characteristic of the construction and *defined* by the *absence* of text. This intertwining of the narrative pattern with syntactic phenomena and bimodal expression hints to close connections between these structures, and bears investigating in future research.

Though constructions dominated by visuals have yet to be discovered across a broad usage, the potential for their creation is certainly apparent in local contexts. For instance, in early 2005 the *Chicago Tribune* launched an advertising campaign that utilized several comic strips to convey the usefulness of different sections of their newspaper. All of these strips followed the same pattern, with the first panel proposing an initial state, the second panel showing the character reading the newspaper (which is marked with the only text in the sequence), and the final panel providing some alteration to the first image. A small sample of these include the following (Chicago Tribune 2005):
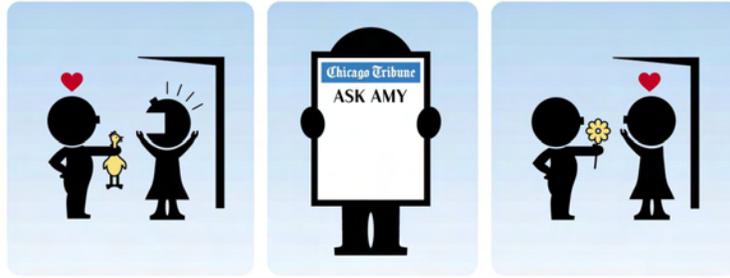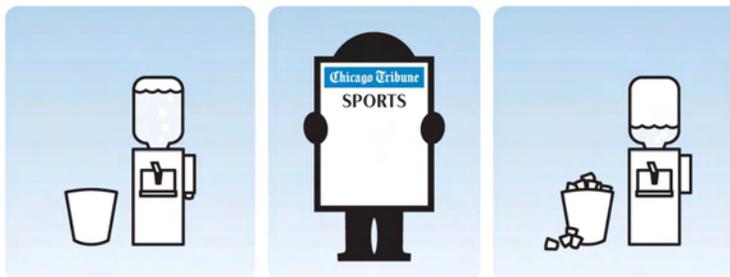
Figure 14a.

Figure 14b.

Figure 14c.

The constructional makeup of these examples should be clear. The first panel sets up the situation, the second represents a causative force, and the third the resultant effect of the causation. Schematized, it could look like this:

Initial state — Causative [reading of paper section] — Resultant state

Although the strip does contain text that is essential to its overall meaning, it is still dominated by the visual syntax. Once the pattern is understood, familiar readers can make expectations about the relationship between the first and last panels, knowing that the second panel always expresses some causative force based on the section of the newspaper. This becomes evident just in these examples. While (14a) and (14b) depict a clear narrative progression with individuals and their actions permeating every panel, (14c) does not have as transparent a reading. The character with the newspaper in panel two only appears in *that* panel, and the watercooler serves entirely as a metonymic representation for the overall concept conveyed in the strip: "if you want to have something to talk about at work, read the sports section." Here, the second panel only has a causative meaning to it, giving the strip as a whole a conceptual rather than narrative basis of semantics. Without that causative meaning, the apparent narration is that a man reads the newspaper while the watercooler becomes emptied — hardly a connected event. Truly, the figure in panel

two does not represent an individual either; it stands for a conception of people in general who could read the paper, especially since more than one person is required for watercooler chatting. If constructions are possible, these strips might hint at the type of routinization necessary for such entrenchment to occur.

Based on these examples, the potential for constructions in visual language and across bimodal visual/verbal language seems quite evident. Indeed, since VL in actual usage most often occurs with writing, it would make sense that bimodal constructions might be possible, yet bears further investigation. No matter what though, they show that patterned representations beyond individual panels does exist in the graphic form.

Conclusion

In sum, like spoken language, visual language contains a variety of sizes of "lexical items" that combine across several levels of grammar to create meaningful units and constructions. This approach to visual language has strived to avoid stating that graphic structures are likened to surface features of verbal language, instead attempting to note the functional similarities in base structure within each respective system. As such, nowhere has this visual language been directly equated with the verbal constructs of "word" or "morpheme." Rather, a "lexical item" is defined as *a meaningful unit or combination of units of form-meaning pairing* that can be either productive or non-productive. Note that in "form-meaning" pairings, there is no restriction on the semiotic quality of the signs. A lexical item can potentially be symbolic, indexical, or iconic, all of which occur in visual (as well as verbal and sign) language, and motivate inclusion into either open or closed classes of morphemes based on their potential for manipulation. As would be expected, productive signs create a far larger class of lexical items than those that are less malleable.

In most visual languages of the world, panels are attention units built out of a large amount of rich productive morphology that can combine in various ways, though this is not absolute. Systems like Australian sand narratives feature highly conventional signs that seem to stand on their own as syntactic units. Finally, like patterns pointed out in construction grammar approaches to linguistics, VL also seems to show the potential for form-meaning pairings of lengths greater than individual formatives.

The comparison of the graphic form to language has often grappled with how best to equate one to the other. However, perhaps more fruitful than searching for words and sentences within images is to examine how words and sentences function as structural elements *within their own system* and compare *that* to the graphic form. Doing so not only could reveal correspondences between forms that appear to have very different semiotic characteristics on the surface, but might also provide windows to broader functioning of the human cognitive system to which language, graphics, and semiotics all belong.

References

([*] contains visual reference)

Bordwell, David and Thompson, Kirstin. 1997. *Film Art: An Introduction (Fifth Edition)*. New York, NY: McGraw-Hill

Cham, Jorge. 2004. Piled Higher and Deeper. www.phdcomics.com. Posted 4/7/2004, accessed 5/1/2005 [*]

Chantler, Scott. 2005. *Northwest Passage, Vol. 1*. Portland, OR: Oni Press [*]

Chicago Tribune. 2005. Advertising campaign strips. http://classified.tribune.com/whatsinitforyou/, accessed 5/1/2005 [*]

Cohn, Neil. 2003. Interfaces and Interactions: A Study of Bimodality. www.emaki.net:Emaki Productions. Posted 11/2003

Cohn, Neil. 2005a. ¡Eye ♥ græfIk Semiosis!: A Cognitive Approach to Graphic Signs and "Writing". Masters Thesis. University of Chicago.

Cohn, Neil. 2005b. Un-Defining 'Comics.' *International Journal of Comic Art* 7.2

Dixon, Chuck, and Johnson, Jeff. 2003. *Way of the Rat, Vol. 1: The Walls of Zhumar*. Florida: CrossGen Entertainment [*]

Forceville, Charles. 1999. Educating the eye? Kress & van Leeuwen's *Reading Images: The Grammar of Visual Design*, (1996). Review article. *Language & Literature*, 8.2, Pp.163-78.

Forceville, Charles. 2005. Visual representations of the idealized cognitive model of *anger* in the Asterix album *La Zizanie*. *Journal of Pragmatics* 37, Pp. 69-88

Goldberg, Adele. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Science* 7.5, Pp. 219-224

Grice, H. Paul. 1967. Logic and conversation. In Cole, P. and Morgan, J.L. 1975: *Syntax and Semantics 3*. New York: Academic Press, Pp. 41-58

Haspelmath, Martin. 2002. *Understanding Morphology*. New York, NY: Arnold

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press

Jackendoff, Ray. 1997. Parts and Boundaries. *Cognition* 41, Pp. 9-45

Jackendoff, Ray. 1997. Twistin' the night away. *Language* 73.3, Pp. 534-559

Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York, NY: Oxford University Press

Johnson, Joel. 2006. Wally Wood's 22 Panels That Always Work: Unlimited Edition. http://joeljohnson.com/archives/2006/08/wally_woods_22.html. Posted 8/18/2006, accessed 8/20/2006. [*]

Kay, Paul and Charles Fillmore. 1999. Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. *Language* 75.1, Pp. 1-33

Kibushi, Kazu. 2004. Copper: Fall. www.boltcity.com/copper_016_fall.htm. Posted 2/2004, accessed on 5/1/05 [*]

Kim, Derek Kirk. 2001. Dave's Blind Date. www.narbonic.com/070701.html. Posted on 7/7/01, accessed on 5/1/05 [*]

Kim, Derek Kirk. 2004. The 10 Commandments of Simon. http://www.lowbright.com/Comics/10Commandments/10Commandments.htm. Accessed on 8/17/2006 [*]

Kress, Gunther and Theo van Leeuwen. 1996. *Reading Images: The Grammar of Visual Design*. London: Routledge

McCloud, Scott. 1993. *Understanding Comics: The Invisible Art*. New York, NY:Harper Collins Inc

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press

Miller, Frank. 2000. *Hell and Back: A Sin City Love Story*. Milwaukie, OR: Dark Horse Comics [*]

Munn, Nancy. 1986. *Walbiri Iconoggraphy: Graphic Representation and Cultural Symbolism in a Central Australian Society*. Chicago: University of Chicago Press

Natsume, Fusanosuke. 1997. *Manga wa naze omoshiroi no ka [Why are Manga Fascinating?: Their visual idioms and grammar*. Tokyo: NHK Library

Peirce, Charles Sanders. 1931. *Collected Papers of Charles Sanders Peirce: Vol. 2: Elements of Logic*. Cambridge, MA: Harvard University Press

Pérez, Ramón and Rob Coughler. 2003. Butternutsquash. www.butternutsquash.com, Posted on 11/05/03, accessed on 4/20/05 [*]

Sakai, Stan. 1987a. *Usagi Yojimbo: Book One*. Seattle, WA: Fantagraphics Books [*]

Samura, Hiroaki. 2004. *Mugen no Junin*, Vol. 17. Tokyo, Japan: Kodansha [*]

Saraceni, Mario. 2000. Language Beyond Language: Comics as verbo-visual texts. Doctoral Dissertation. University of Nottingham

Saraceni, Mario. 2001. R elatedness: Aspects of Textual Connectivity in Comics. In Baetens, Jan (Ed.). *The Graphic Novel*. Leuven: Leuven University Press

Saraceni, Mario. 2003. *The Language of Comics*. New York, NY: Routelage

de Saussure, Ferdinand. 1972 [1916]. *Course in General Linguistics*. Harris, Roy. (Translator). Chicago and LeSalle, IL: Open Court Classics

Shipman, Hal. 2006. Hergé's *Tintin* and Milton Caniff's *Terry and the Pirates*: Western Vocabularies of Visual Language. Paper Presented at the 2006 Comic Arts Conference.

Stainbrook, Eric J. 2003. Reading Comics: A Theoretical Analysis of Textuality and Discourse in the Comics Medium. Doctoral dissertation. Indiana University of Pennsylvania.

Talmy, Leonard. 2001. *Toward a Cognitive Semantics*. Cambridge, Mass.: MIT Press von Flue, Neal. 2004. Set-up, (beat), Punchline. http://ape-law.com/hypercomics/beat. Posted 10/11/2004, accessed 10/11/2004

Walker, Mort. 1980. *The Lexicon of Comicana*. Port Chester, NY: Comicana, Inc

Wilkins, David P. 1997. Alternative Representations of Space: Arrernte Narratives in Sand. In Biemans, M., and van de Weijer, J. (Eds), *Proceedings of the CLS Opening Academic Year '97 '98*. Center for Language Studies [*]