**Temporal Succession and Aspectual Type in Visual Narrative**

DORIT ABUSCH  
*Cornell University*

**Abstract**

We examine the aspectual type of pictures, and how it interacts with temporal progression in pictorial narratives. Some examples suggest that pictures can have stative aspect and that such pictures do not advance time, paralleling the situation in natural language. However, we show on the basis of formal semantics that all pictures have stative informational content, so that a construction rule for pictorial discourse representations can not be sensitive to aspect. This suggests that the temporal interpretation rule for pictorial narratives involves invariant temporal progression.

1 Introduction

This paper looks at the issue of temporal progression in visual narratives such as comics and manga. The idea is to apply a semantics for pictures that is similar to a standard propositional semantics for sentences, so that a visual narrative, like a verbal one, is a sequence of elements that denote propositions. I will examine similarities and differences between linguistic and pictorial media in the way they create a temporally ordered semantic representation. The central issue is whether the temporal structure of visual narrative is sensitive to aspectual properties of literal semantic content, as is often maintained for natural language narratives. Other issues are the

---

*Presentations of this work were given in June 2012 at the UCLA Workshop on Visual Narrative, in July 2012 at Frankfurt and at Göttingen, in November 2013 at the Cornell Workshop on Linguistics and Philosophy, in December 2013 in the Rutgers-Princeton Graduate Seminar Philosophy of Language: Semantics-Pragmatics, in March 2014 at Yale, and in June 2014 at St. Andrews in the workshop Tense in Semantics and Philosophy of Language. I would like to thank the audiences for their comments. Special thanks to Mats Rooth, Ede Zimmermann, Angelika Kratzer, Gabriel Greenberg, Magda Kaufmann, Regine Eckardt, Sam Cumming, and Cécile Meier. The final version benefitted from the comments of a reviewer for this volume. I am solely responsible for mistakes and misconceptions. Part of this research was supported by a Humanities Research Grant from the Cornell Society for the Humanities. Cited images are owned by their respective owners and are used purely for analytical, critical, and scholarly purposes.*
balance between grammar, semantics, and pragmatics in the construction of temporal sequencing
for visual narrative, and differences in the semantic content of pictures and sentences. The overall
plot of comparing visual and natural-language narratives in a uniform setting is similar to the one
of my earlier paper Abusch (2012), which looked at co-reference (or indexing).

My framework for the semantics of pictures is the geometric semantics for pictures
from Greenberg (2011, 2013), which uses varieties of geometric transformations to obtain a
propositional semantics for pictures. In diagram (1), which is quoted from Greenberg (2011), a
scene containing a cube is projected to a picture by drawing lines to a viewpoint, and marking the
points where the projection lines cross a picture plane. Greenberg suggests a formal representation
where a viewpoint encodes the configuration of the convergence point relative to a picture plane
and scene. Then the semantics for a picture is obtained by inverting projection: the semantic value
\[[A]\] of picture A is the set of pairs \(\langle v, \sigma \rangle\) of viewpoints \(v\) and scenes \(\sigma\) that project to \(A\) via the
picture plane and convergence point encoded in \(v\). This is similar to a centered proposition (Lewis
1979). An alternative way of proceeding is to say that the viewpoint is a parameter of evaluation,
so that a picture has content with respect to a viewpoint. In this case \([A]_v\) is the set of scenes \(\sigma\)
that project to \(A\) relative to the picture plane and convergence point encoded in \(v\).

Using this semantics for pictures, a sequence of pictures (such as the sequence of pictures in a
comic, manga, or an airplane security pamphlet) can be converted into a sequence of propositions.
The same can be done with a verbal narrative consisting of a sequence of sentences, by converting
each sentence into a proposition, using a semantics for the natural language. From this point
of view pictorial and verbal narratives are isomorphic in the information that is supplied by the
semantics.

In order to focus on the pure case of pictorial narrative, I will look at “silent” comics, which
are comics and manga without speech bubbles, thought bubbles, or captions, and at silent passages
in classic manga. The Gon books of Masashi Tanaka are silent mangas depicting the adventures of
Gon, a small powerful dinosaur, in the world of modern animals. Example (2) shows two facing
pages from Gon Episode 4, a story where Gon joins a family of golden eagles. The pages are
read in Japanese order, with the right page read first, and the left page read second. Each page is
in turn divided into three panels, which are intended to be read linearly, in an order determined
by the parsing procedure reviewed in Section 2. So these pages reduce to an ordered sequence of
pictures \(\langle P_3, P_4, P_5, P_6, P_7, P_8 \rangle\) that maps to an ordered sequence \(\langle [P_3], [P_4], [P_5], [P_6], [P_7], [P_8] \rangle\)
of centered propositions. The passage in (2) is understood with temporal progression. For instance,
the situation satisfying panel \(P_8\), which depicts some fish in the air next to the mother eagle, is
understood as temporally following the situation that satisfies panel \(P_7\), which depicts a couple of
fish in the beak of the mother eagle. The model-theoretic object that satisfies the sequence must
be some kind of temporally structured object, with parts that stand in temporal relations. And
there must be principles for introducing these temporal relations. I will approach this issue using
methods of temporal discourse representation theory (Kamp and Reyle 1993; Reyle, Rossdeutscher and Kamp 2007; Altshuler 2010). An interesting issue introduced in Section 3 is that in some cases, described situations for successive panels appear to be in a relation of temporal overlap, rather than temporal succession.

The paper is organized as follows. Section 2 describes the principles by which a linear order of panels is obtained from the two-dimensional layout of a page in mangas such as Gon Episode 4 and the classic mangas of Osamu Tezuka. Section 3 looks at temporal succession and temporal overlap in the interpretation of sequences of panels. Section 4 is concerned with temporal discourse representation for pictorial narratives, and introduces an aspectually sensitive construction principle for visual narratives. That principle is then criticized, and Section 5 proposes an invariant construction principle. Section 6 wraps up the discussion.

2 Panel sequencing

Panels in comics are laid out two-dimensionally on a page, but are usually read linearly in a particular intended order. Obtaining this linear order is a necessary prerequisite to reasoning about the relation between sequencing of panels and temporal sequencing of events in the described situation. There is work by T. Tanaka et. al. (2007) on an algorithm that creates a linear order of panels from the two-dimensional layout of panels in a comic. See also Cohn (2013), which includes experimental work on reading-order for panels in comics. I will present the construction in terms of a labeled tree that has one interpretation as a description of a page layout, and another interpretation as a linear sequence of panels. Suppose we start with two panels (panels 27 and 28) and combine them in a binary tree labeled H. Interpreted as page layout, the tree $[H_{28} 27]$ indicates that panels 28 and 27 are combined horizontally into a block, with the vertical dimensions scaled to match—see the top box in (3). The same tree interpreted as order of reading indicates the right child 27 is to be read before the left child 28, using the right-to-left convention of manga. Panels 29 and 30 are combined in the same way into binary-branching tree $[H_{30} 29]$. 

![Panel examples](image)
In the next step, the two trees \([H_{28 \ 27}]\) and \([H_{30 \ 29}]\) and the atomic panel 31 are combined into a ternary-branching tree labeled V. In the tree at the right in (3), the parts involved in the last step are in bold. Interpreted in page layout, the V node \([V_{31 \ H_{30 \ 29} \ H_{28 \ 27}}]\) stacks the blocks denoted by the three children vertically, with the right child on top, and with the horizontal dimensions scaled to match. In reading order, the sequences contributed by the three children are concatenated, resulting in the reading order (27, 28, 29, 30, 31) for the sub-tree.

Panels 32, 33, and 34 are combined into a ternary-branching tree labeled H. Panels 35 and 36 are combined into a binary-branching tree also labeled H. Then these two H sub-trees are combined into a binary-branching node labeled V. In (4), the parts involved in the last step are boxed at the left, and marked in bold in the tree at the right.

Finally, the two sub-trees labeled V are combined into a binary-branching tree labeled H:

Generalizing this procedure, each leaf is interpreted as a unit sequence. An internal node is interpreted as the concatenation of the sequences that interpret its children, respecting the order of the children in the tree (as I do it, in reverse, to make the tree have manga panel ordering, with earlier panels on the right). The result in this case is the sequence of panels 27, 28, 29, 30, 31, 32, 33, 34, 35. Note that this linear interpretation ignores the H and V labeling. In the two-dimensional interpretation, each sub-tree is interpreted as a block-shaped picture. The
interpretation of $[H_{i}A_{1}...A_{n}]$ is obtained by scaling the vertical dimensions of the $A_{i}$’s so that they match, and stacking together the pictures horizontally, separated by channels, and respecting the order of the children in the tree. The interpretation of $[V_{i}A_{1}...A_{n}]$ is obtained by scaling the horizontal dimensions of the $A_{i}$’s so that they match each other, and stacking together the pictures vertically, separated by channels, respecting the order of the children in the tree. Following this scheme, the tree above is interpreted in panel sequencing as the sequence (27, 28, 29, 30, 31, 32, 33, 34, 35). In page layout, the same tree is interpreted as the page layout seen at the left in (3)-(5).

The tree interpretations create a correspondence between page layouts and reading orders. In some cases, a single page can correspond to more than one linear order. Below, because of the plus junction, the block with four panels can be composed in two ways, resulting in linear orders $a,b,c,d$ or $a,c,b,d$. M. Tanaka does not use plus junctions in Gon.

(6)

It is common for the spine of the book to provide the last division in a pair of facing pages, without a visible channel. But this is not always so. In example in (7) from Gon episode 4, panel 129 is a single panel that crosses the spine. In the tree, the last step is horizontal stacking of two blocks, panel 129 and $[V_{i}132\ 131\ 130]$, with division by a channel in the layout interpretation. This is a source of ambiguity, because the spine could in principle provide a division into panels 129a and 129b. This is an ambiguity of another kind because it depends on the identity of the atomic panels in the two-dimensional rendering. There are similar cases where a long white area might be a channel, or a part of the picture in one panel.

(7)
The parsing scheme is observed in many classic manga. But there are exceptions. The panel on the left in (8) from O. Tezuka’s Ode to Kirihito, Part 2, functions as a montage of doctors working in the lab and is not composed of recursively divided blocks. The page in the middle from the same work, depicting Izumi being raped by Dr. Urabe, shows vertical stacking with an interpretation of sequencing, but the jagged divisions convey violence.

There is a final step in the tree scheme, where pages or pairs of facing pages are assembled into a book. I represent this with an additional tree level, with a label B and ordered children that correspond to the individual pairs of facing pages. B has the same linear interpretation of concatenation of sequences.

Tanaka’s *Gon* series can be interpreted as following the stacking scheme completely faithfully. The result is that the panels in a *Gon* episode can be construed as having a unique linear order, similar to the order of words or sentences in a linguistic narrative. The procedure reviewed in this section derives this linear order, given a two-dimensional layout of panels. Despite the trees that are employed in defining the interface, this derivation is more similar to the phonetics-phonology interface in natural language than to syntax, since it has to do with the physical realization of units.

As with linguistic narratives, there is a default interpretation of temporal progression for the sequence. This establishes the plot for this paper. As reviewed above, I assume a propositional semantics for pictures, so that the semantics for a panel in a comic and the semantics for a sentence in a verbal narrative are the same or nearly the same. This makes verbal and pictorial narratives formally comparable—each consists of units (sentences or panels) that have a propositional semantics, and the units are linearly ordered.

3 Temporal progression

I will look now at temporal progression in comics and manga. There is some work by Scott McCloud in his book *Understanding Comics* that is interesting from the linguistic perspective, because it is suggestive of the aspectual distinctions between states and events that are claimed to influence temporal progression in verbal narratives (McCloud 1994). McCloud gives a typology of temporal relations between linearly adjacent panels. In one common case, there is an agent depicted in the two panels in distinct but related actions. This is McCloud’s Type 2, and is exemplified by the first pair on the left below where the ball is approaching and the player hits
it. Another case is where the agent is different but the activities are in the same location and the actions are related. This is exemplified by the pair to the right with two people in a conversation, where each panel shows just one of them. The examples are quoted from McCloud.

These are examples of panel pairs that are understood as implying temporal progression. They are analogous to linguistic narratives where two successive sentences each describe events, and the events are understood as temporally sequenced.

In (10) are examples from McCloud of panel sequences that lack temporal progression. In the first pair we see the sun shining down in the first panel, and a man on his back with sunglasses in the second panel. Rather than being temporally sequenced, the described situations seem to be understood as temporally overlapping—or as McCloud cautiously says, there is little feeling of temporal movement. In the second pair, the first frame shows drink containers and ashes on a table at a party, and the second frame shows two guests. It seems open whether the described situations overlap or are sequenced in one order or another.

What is interesting from the perspective of linguistic semantics and pragmatics is that the panel information in these cases appears to have a stative nature. “The sun is shining” is a stative sentence and in linguistic narratives, stative sentences typically do not induce temporal progression. For instance in (11), the darkness situation is by default understood as containing the waking up situation, rather than following it. Similar effects are observed for an eventive subordinate when-clause, and a stative main clause in (12).

(11) Justin woke up. It was dark.

(12) When Justin woke up, it was dark.

So in sequences of this kind (McCloud’s Type 3), the information in the panels is stative, and we don’t have a feeling of temporal progression.

Now I will give examples of temporal progression and overlap from *Gon* and other manga. First some information about the general temporal structure of *Gon* Episode 4, which covers the time up to the first flight of some baby eagles. The age of a golden eagle at first flight is about two months, so the total time covered in a realistic model is a little less than this. The story breaks into half a dozen tightly structured passages: first the little birds and Gon eat some fish in panels 1-11.
In panels 12-26, a bobcat attacks the birds. A baby eagle falls down a cliff in panels 27-59. The birds and Gon eat a big fish in panels 60-63. In panels 64-97, the birds learn to fly, and the bobcat attacks again. Panels 97-133 show the first flight, including the attempts to fly of the dinosaur Gon. Within each sequence, the action is closely structured and not much time passes between the panels. Between the different event sequences, a lot of time can pass. We already saw in (2) an example of temporal progression. Here is another one. Example (13) shows panels 31 and 32 from the story. In panel 31 at the right, a dozing Gon kicks a baby eagle in the nest. In panel 32, the bird bounces down a cliff. The bouncing is unambiguously understood as following the kick.

Some other panel transitions in *Gon* have an apparent interpretation of temporal overlap, rather than temporal succession. (14) shows panel 131 at the top, depicting a bobcat on its back on the ground, and panel 132 at the bottom, depicting the young eagles flying overhead. Just as in McCloud’s example with the sun and the man in the sunglasses, the described situations for these panels are understood as identical or overlapping. There is no feeling of temporal succession.

Example (15) shows the opening sequence of *Gon* Episode 17. In the two panels at the right, we see some nests in the top of a tree, and a nest is starting to shake. In the middle panel, we see Gon at the bottom of the tree kicking it. At the left, the whole tree is shaking. The last three panels have a process-like interpretation, and the processes overlap rather than being in succession. The nest is shaking and shaking. Gon is kicking and kicking. And the whole tree is shaking and shaking. We understand these processes to overlap.
shows examples of overlap from Osamu Tezuka’s *Ode to Kirito*, Part 2. In the big panel at the top of the page at the left, we see villagers surrounding and looking at a hut. At the bottom, there are three eventive panels involving the hero inside the hut. We understand that the villagers are still surrounding the hut. In the second example, on the left we have a sequence of four panels. A character walks down some steps, reaches the water, and then in the water he sees that the hero has the face of a dog. He gets out of the water. The initial panel on the right shows the scene for the whole action—a bathhouse below some mountains. We understand the events depicted on the right as temporally and spatially embedded in the scene.

The phenomenon of overlapping interpretations in visual narratives is very similar to what is seen in natural language. Consider this brief narrative:

Events of waking up, getting out of bed, and groaning are described, and they are understood as being in temporal succession. But the state of being dark is understood as extending over the event of John’s waking up, and probably also the getting up and groaning. This can be taken to indicate that there is a distinct temporal interpretation for the second sentence, triggered by the fact that this sentence is aspectually stative. McCloud’s examples and the examples from Tanaka and Tezuka raise the question of whether there are panels with “stative aspect”, whose described situations overlap the described situations of adjacent panels, rather than being in a relation of temporal succession with them. If the answer is yes, are there alternative basic temporal interpretations for the juxtaposition of panels? Or is the overlapping interpretation the consequence of a pragmatic process? I take up these questions in the next section.
4 Temporal discourse representations

In what is called temporal discourse representation theory, a syntactic representation of temporal ordering among events or situations is built up. The basic approach is to introduce discourse referents or variables for events or situations, and to include formulas that describe temporal ordering among the discourse referents. Creating such a representation is accomplished by an interface rule called a discourse representation construction rule (Kamp and Reyle 1993). Let us look at a sequence of two eventive sentences:

(18) Justin woke up.
     He got out of bed.

We associate each sentence with a situation variable:

(19) \( s_1 \) is a situation where Justin wakes up.
     \( s_2 \) is a situation where Justin gets out of bed.

Situations are conceived of as spatio-temporally bounded entities that can be ordered temporally, with \( s_1 \prec t s_2 \) read “\( s_1 \) temporally precedes \( s_2 \)”. Here is the straightforward construction rule for a pair of eventive sentences:

(20) Given two consecutive eventive sentences \( \phi_1, \phi_2 \) with respective described situations \( s_1 \) and \( s_2 \), add a constraint \( s_1 <_{t} s_2 \).

Now look at the sentences in (21), where the second sentence is stative rather than eventive. Here there is the suggestion that the time of Justin waking up is embedded in the time when it was dark, rather than preceding it. This could indicate that a distinct construction rule should apply, which introduces a relation of temporal embedding or temporal overlap.

(21) Justin woke up. It was dark.

Asher and Lascarides (1993) pointed out that where the scenario is not compatible with temporal overlap, but instead suggests temporal succession, a pair with an eventive sentence followed by a stative one can be odd:

(22) ? Max won the race. He was home with the cup.

So the phenomenon is robust: even when the scenario strongly suggests temporal progression, the stativity of the sentence triggers an overlapping interpretation. In (22) the result is oddity, because the overlapping interpretation conflicts with the content.

Such data have motivated the hypothesis that a distinct temporal discourse representation construction rule is triggered by stative sentences (Kamp and Rohrer 1983, Hinrichs 1986, Kamp and Reyle 1993, Lascarides and Asher 1993, Asher and Lascarides 2003). In the case where the second sentence is stative, a temporal predication of embedding (\( s_1 \subseteq_t s_2 \)) is added, rather than one of temporal succession (\( s_1 \triangleleft_t s_2 \)). Or a predication of temporal overlap (\( s_1 \circ_t s_2 \)) can be used. (23) is a simple construction rule for temporal discourse representations that is sensitive to aspectual type.

(23) Given two consecutive eventive sentences \( \phi_1, \phi_2 \) with respective described situations \( s_1 \) and \( s_2 \), add a constraint one of the following:
     - \( s_1 \prec_{t} s_2 \)
     - \( s_1 \subseteq_{t} s_2 \)
     - \( s_1 \circ_{t} s_2 \)

(23) is a simple construction rule for temporal discourse representations that is sensitive to aspectual type.
Given two consecutive sentences $\phi_1, \phi_2$ with respective described situations $\sigma_1$ and $\sigma_2$, if $\phi_2$ is eventive, by default add a constraint $\sigma_1 \preceq \sigma_2$, otherwise if $\phi_2$ is stative, add a constraint $\sigma_1 \odot \sigma_2$.

“By default” is included to allow for the possibility that other information, including considerations of plausibility of the scenario, can override the temporal ordering that is stipulated. In adapting this to visual narrative, I will use a symmetric formulation.

Given two successive panels $\phi_1, \phi_2$ with respective described situations $\sigma_1, \sigma_2$, if $\phi_1$ or $\phi_2$ is stative (by default) add a temporal constraint $\sigma_1 \odot \sigma_2$. Otherwise (by default) add a temporal constraint $\sigma_1 \preceq \sigma_2$.

In the case where either panel is stative, the described situations for the two panels are required to be in a relation of temporal overlap. Otherwise, covering the eventive case, temporal succession is stipulated. Assuming the bathhouse and hut panels are stative, the rule will apply to produce temporal overlap rather than temporal succession in these examples.

But what is it for a sentence or a panel to be stative? The general term for categories such as stative or eventive is aspectual type, aktionsart, or temporal constitution. Linguistic work has evolved to the hypothesis that distinctions of aspectual type are semantic distinctions that can be defined in terms of the semantic content of sentences (Taylor 1977, Dowty 1979, Krifka 1989). To make distinctions in aspectual type, one looks at the semantic content of sentences, and checks certain axioms. One of these is temporal cumulativity. Suppose $\sigma_1$ is a situation where the Yurari bathhouse is below Mt. Fuji, and $\sigma_2$ is a temporally contiguous situation where the Yurari bathhouse is below Mt. Fuji. Then assuming that $\sigma_1$ and $\sigma_2$ can be joined into a larger situation $\sigma_1 \cup \sigma_2$, it is plausible to assume that the larger situation is also a situation where Yurari bathhouse is below Mt. Fuji. Suppose $\sigma_1$ is a situation where a big tree is shaking, and $\sigma_2$ is a temporally contiguous situation where the big tree is shaking. Then it is plausible that $\sigma_1 \cup \sigma_2$ is a situation where the big tree is shaking. This inference from a property holding at $\sigma_1$ and at $\sigma_2$ to the property holding at the fusion $\sigma_1 \cup \sigma_2$ is an example of temporal cumulativity.

In research about aspect in natural languages, it is common to assume axioms about semantic properties of verbs and other lexical items, and then to derive properties including cumulativity for clause meanings. To do this, one has to assume that clause meanings are some kind of entity for which cumulativity can be assessed. We can say that clauses are true or false relative to a time interval of evaluation (Dowty 1979). Or nearly equivalently, it can be assumed that the extensional denotation of a clause is a set of time intervals. Or one can say that clauses denote sets of events, and that there is a part-whole structure and a join operation on events (Krifka 1989). Or we can say that clauses denote sets of situations, and that there is a part-whole structure and a join operation on situations.

In English, cumulativity holds for stative clauses such as ‘the bathhouse is below the mountain’ and for activity sentences such as ‘the big tree shakes’. Cumulativity does not hold for eventive sentences. Suppose $\sigma_1$ and $\sigma_2$ are situations where sentence (25a) is true. Then it is not plausible that $\sigma_1 \cup \sigma_2$ is a situation where the same sentence is true. Similarly for (25b). This statement must be hedged, because checking cumulativity is not just a matter of intuition, but of reasoning about axioms and families of models (Taylor 1977, Dowty 1979, Krifka 1989, Abusch 2005).

(25)a. Justin walk exactly one kilometer.
   b. Justin turned 90 degrees to his left.
By the way, an additional way to distinguish stative propositions from activity propositions is required, if it is desired to make this distinction. The usual way to do it is to say that stative propositions (and not activity propositions) can be true at time points, while activity sentences can not be true at time points (Dowty 1979). This can be applied to situations by assuming that some situations have point temporal projections, while other situations have temporally extended temporal projections. In the first part of the argument below, I will not distinguish statives from activities.

Now I want to apply the same aspectual distinction to the semantic content of pictures. As I described it, stativity is a semantic feature. If panels in comics have a semantics similar to the semantics of sentences, as in the geometric construction of the content of pictures, it should be possible to apply the stativity distinction. So I would like to ask which panels are stative, namely have stative informational content according to the formal semantic model we are working with.

I will check cumulativity in a simple version first. We assume that time is discrete, and that $\sigma_1$ and $\sigma_2$ are instantaneous situations in immediate succession. We assume that both satisfy the same picture $P$. This means that w.r.t. the viewpoint and the picture frame encoded in $v$, $\sigma_1$ projects to $P$. Similarly for $\sigma_2$. $\sigma_1 \cup \sigma_2$ is the combined situation whose temporal projection is an interval consisting of two successive instants. To check cumulativity, we have to check if $\sigma_1 \cup \sigma_2$ satisfies the same picture. Presumably it does, because each instantaneous slice of $\sigma_1 \cup \sigma_2$ satisfies the picture with respect to $v$. The derivation is recorded in (26).

(26) \begin{align*}
\text{Premises} & \quad \sigma_1 \text{ satisfies w.r.t. } v \quad \sigma_1 \cup \sigma_2 \text{ satisfies w.r.t. } v \\
\text{Conclusion} & \quad \sigma_2 \text{ satisfies w.r.t. } v
\end{align*}

Consider some options for $\sigma_2$. First, assume $\sigma_1$ is a specific situation where a man is on the stairs. $\sigma_{2a}$ is a subsequent situation that is a copy of $\sigma_1$: the man is on the stair in the same location and posture. $\sigma_{2b}$ is another option: the man has changed into an acrylic sculpture in exactly the same location and posture. $\sigma_{2c}$ is a third option: the man is on the stairs in the same location and posture, but a new lump has appeared on the back of his leg that was not there in $\sigma_1$. Because the lump is on the back of his leg, the projection does not change. $\sigma_{2d}$ is the last option. The man, the stairs and the wall have swelled by ten percent, and also moved away from the viewpoint, so that the projection is the same. In each case, $\sigma_1 \cup \sigma_{2x}$ satisfies $P$ with respect to $v$. The point of the last three options is that a temporally extended situation that satisfies a stative proposition does not have to be a static situation. $\sigma_1 \cup \sigma_{2c}$ is not static, because a lump appears. $\sigma_1 \cup \sigma_{2d}$ is not static, because the man moves and swells. This point is the same in natural language. The stative proposition denoted by ‘he be on the stairs’ can be true in $\sigma_1 \cup \sigma_{2x}$ even if he is moving his arm, or moves to another stair, or grows a lump, or starts thinking about the beach.

The same argument applies with a general version of cumulativity. Suppose $\sigma_1$ and $\sigma_2$ are two situations (possibly temporally extended), with $\sigma_1$ immediately preceding $\sigma_2$. We suppose that $\sigma_1$ satisfies $P$ w.r.t. $v$, and so does $\sigma_2$. I assume this amounts to each instantaneous slice of $\sigma_1$ satisfying $P$, and each instantaneous slice of $\sigma_2$ satisfying $P$. We have to check whether $\sigma_1 \cup \sigma_2$
satisfies $P$ w.r.t. $v$. Any instantaneous slice of $\sigma_1 \cup \sigma_2$ is either a slice of $\sigma_1$ or a slice of $\sigma_2$. Therefore each instantaneous slice of $\sigma_1 \cup \sigma_2$ satisfies $P$. So the content of $P$ obeys cumulativity.

So, it is a mathematical fact that the informational content of the picture is cumulative, and in fact stative, because it can be satisfied by an instantaneous situation. So unlike in language, where there are sentences that are eventive and non-stative by semantic criteria that are applied to literal contents, for pictures there are no non-stative literal contents. The method followed here is the one used for classification of the aspectual type of propositions in formal semantics of natural language—models of lexical and compositional semantics are constructed explicitly or axioms are assumed about lexical semantics. Then properties such as cumulativity are checked formally (Taylor 1977, Dowty 1982, Krifka 1989, Abusch 2005).

An alternative argument is based on the observation that the content of a picture $P$ in Greenberg’s construction is a complex proposition about the geometric configuration of the scene and viewpoint. Sentences about the geometric position of objects such as the sentences in (27) are linguistically stative. So if the content of a picture is equivalent to the content of such a geometric sentence, then the content of a picture must be stative.

(27)a. The sign is above the doorway.
   b. The rod has a circular cross-section.
   c. There is a dodecahedron on the table.
   d. The front surface of the filing cabinet is a parallelogram but not a rectangle.

Taking this further, consider the family of distributional tests for stativity in English. As illustrated in (28), stative clauses but not eventive or activity clauses can be modified by temporal still. (28b) is okay when “still” means nevertheless, but is not good on a temporal reading. As illustrated in (29), stative clauses but not eventive or activity clauses can occur in an unmarked way in the present tense. And as illustrated in (30), stative clauses but not eventive or activity clauses can occur as the complement of believe in the “raising to object” or “exceptional case marking” complement structure.

(28)a. Justin is/was still asleep.
   b. ?Justin still eats/ate an apple.

(29)a. Justin admires Keisha.
   b. ?Justin eats a cookie.

(30)a. Keisha believes Justin to be asleep.
   b. I believe the front surface of my filing cabinet to be a parallelogram but not a rectangle.
   c. ?Keisha believes Justin to eat a cookie.

Two computer graphics students are considering a scene with a physical geometric figure that can be changed in various ways. They have in mind a particular vantage point $v$, and can figure out that the scene projects to a particular picture $P$. The picture is not part of their physical situation. After they change the geometric figure a bit, one of them says:

(31) Interesting, a lump appeared on the back of the cube, but the scene still projects to $P$.

(32)a. The scene projects to $P$.
   b. Keisha believes the scene to project to $P$, and since she’s an expert, it probably does.
The possibility of modification by still in (31) indicates that clause (32a) is stative. And this sentence is in the simple present tense, another indication of stativity. Finally, (32b) shows that this clause can be embedded as the raising-to-object complement of believe, another indicator of stativity. So, the linguistic tests and the fact that geometric propositions are typically stative indicate that the proposition [[the scene project to P]] is stative. What is this proposition? It is very similar to the picture denotation [[P]]. Since we are using a situation semantics, propositions are properties of situations. The proposition is true of σ iff σ projects to P via the geometric transform that was described in section 1. Here σ is the whole situation that is being geometrically projected. Sentence (32a) contains definite reference to scene—presumably, this scene is part of a situation σ that the proposition [[the scene projects to P]] is true of. Let us assume that the projected scene is picked out from using a definite description and a contextually given predicate S, which should pick out an area of the lab in which the students are standing. Then the sentence denotation is along the lines of (33a). So, the sentence denotation can be described as a mereological inflation of the picture denotation, where a situation that the picture proposition is true of is inflated to a containing situation with respect to which the original situation is picked out by S.

(33)a. λσ.σ.the(λσ'.σ' ≤ σ ∧ S(σ, σ'), λσ'.G(σ', v) = P)
b. λσ.G(σ, v) = P

At least in basic cases, it is clear that the inflation will not affect cumulativity of the propositions in the temporal dimension. Let p be the original proposition, and q be the inflated proposition. Suppose p is cumulative, and that q is true of σ₁ and σ₂, and that these two situations are in temporal succession. By definition of inflation, σ₁ has a part σ₁ picked out by S that p is true of, and σ₂ has a part σ₂ picked out by S that p is true of. By cumulativity of p, p is true of σ₁ ∪ σ₂, and this joined scene serves as a witness for q being true of σ₁ ∪ σ₂. While this is just a sketch of an idea for a proof, I will regard it here as indicative of the stativity of the sentence denotation correlating with the stativity of the picture denotation. If this is so, it is an additional reason to think that the picture denotation is stative, since linguistically the sentence denotation (32a) is stative.

Caution is appropriate about the linguistic argument for the stativity of the content of pictures. First, there is a worry that mathematical statements are in general stative, threatening to trivialize the observation that the truth condition obtained from a geometric semantics of pictures is stative. However, the contents [[the scene projects to P]] and [[P]] as they are used in the argument are not time-invariant propositions. These propositions can shift in truth value in a single course of events, as is illustrated by a version of the graphics student scenario where the projection changes when the lump is introduced.

More problematically, the argument requires that [[the scene projects to P]] and [[P]] are the same proposition, down to the technical details of mathematical modeling, or if they differ, the difference is one which does not affect cumulativity or other criteria for stativity. There is enough flexibility in semantic modeling that it seems doubtful that this can be established. For one thing, it is known from linguistic examples that propositions with equivalent informational content can differ in aspectual type. In Emmon Bach’s example (34), the two clauses have the same informational content, but the first version has the aspectual properties of an activity proposition, while the second version is stative. I think the most that can be maintained is that linguistic and pictorial semantics can be set up so that [[the scene projects to P]] and [[P]] are the same proposition, modulo the inflation mentioned above. This still leaves the linguistic tests with some force.
(34)a. The satellite orbit the earth.
b. The satellite be in orbit around the earth.

Summing up, this section has argued that a picture contents as defined by geometric projection is semantically stative. These arguments were independent of the picture. It is not that there are stative pictures and eventive ones—all pictures are stative in semantic content. This is fatal for the discourse representation construction rule (23), which makes the choice of temporal relation (temporal succession \( \leq_t \) vs. temporal overlap \( \circ_t \)) dependent on stativity of picture content.

5 Invariant temporal construction

Look again at the two big panels in (16), one with villagers surrounding a hut, and the other with a bathhouse below the mountain. According to the empirical picture introduced in Section 3 and the formalization in Section 4, the described situations for the big panels are temporally extended and temporally overlap the described situations for the adjacent panels. There is a problem with this. In the villager picture, the heads of the villagers are shown at specific orientations. If the described situations are temporally extended and if each temporal slice satisfied the content of the big picture, as assumed in Section 4, this implies that the heads of the villagers are not moving (or rather, that they are not moving in a way that changes the projection of their heads). But we do not understand the panel sequence as implying that the villagers are motionless. Similarly, the bathhouse picture shows some vapor above the bathhouse. We do not interpret the passage of four panels in the bathhouse sequence as implying that the vapor is static.

(35) is another case, from Tezuka’s *Buddha*. In the big picture, a sage is shown in a snowstorm in a mountain pass, supported by a long walking stick. The first small picture shows the sage having dropped his stick. (The reading order in this page, from a flipped English edition, is left-to-right.) The wind blows it away, and then the sage continues walking. We understand the action depicted in the small pictures as taking place in the pass. But in this case, it is clear that the described situation for the big picture does not temporally contain the described situation for the first small picture. It could not, because the stick is shown in incompatible orientations—vertical in the big picture, and near horizontal in the small picture.

This problem of not all the content of the big pictures extending over the times for the small pictures is connected to the fact that the content of a picture is informationally rich compared to the
content of a simple sentence. The first sentence of the linguistic sequence (36) describes Keisha moving from one step on a stairway to another, and then to a third step. The second sentence, which is stative, describes Justin being on step 1. The natural way to understand the sequence is that Justin was on step 1 when Keisha moved from step 2 to step 3, and also when she moved from step 3 to step 4. This does not entail that Justin was motionless, or had an unchanging geometric projection from some vantage point. He could have moved around a bit on step one, and he could have been moving his arms or head or lips. For the second sentence in (36) to be true with respect to some temporally extended situation requires nothing more than Justin being on step 1 throughout the situation.

(36) Keisha moved from step 2 to step 3, and then to step 4. Justin was on step 1.

There is no way of replacing the second sentence in (36) with a picture that has an equally weak informational content. To depict Justin on step 1, one has to depict his body in a particular orientation. He cannot be depicted as being on step 1 without producing a propositional content that has additional entailments. In the same way, the villagers cannot be depicted as surrounding the hut without including additional entailments about the orientations of their bodies. Here is a final argument against the aspectually sensitive construction rule. Suppose the panels in the bathhouse sequence came with time stamps that indicate temporal progression by thirty seconds for each frame, as if they were shot with digital camera that writes a time stamp into each picture. See (37). I think this would hardly affect the implication about the action described by the small pictures being embedded in a larger situation with a bathhouse below a mountain. This indicates that to capture these implications, it is not necessary to assume a discourse-representation construction rule that introduces a predication of temporal embedding or overlap for the time intervals or situations that satisfy the pictures. Given the time stamps, these situations are stipulated to be in succession, separated by thirty seconds.

(37)

Summing up, there are strong reasons to reject the hypothesis of aspectually sensitive temporal construction rules for pictorial narrative. On a geometric semantics, all pictures have stative content, so it is not possible to make a construction rule sensitive to the aspectual type of pictures. Second, in the hut, bathhouse and sage sequences, not all of the content of the “stative” panel is understood as being temporally extended. Finally, the time-stamp argument indicates that even if temporal progression is stipulated, some of the content of the scene-setting picture is understood as extending over the adjacent panel.
As an alternative, I suggest we must adopt an invariant temporal construction rule for pictorial narratives, and attribute the effects described in Section 3 to some kind of pragmatic process. An invariant construction rule introduces the same temporal relation for any successive panels, regardless of their content. The obvious possibility is a relation of temporal succession. (38) gives the construction rule, which stipulates that the temporal ordering of described situations is isomorphic to the linear ordering of panels.

(38) Given two consecutive panels $\phi_1, \phi_2$ with respective described situations $\sigma_1$ and $\sigma_2$, add a constraint $\sigma_1 \preceq \sigma_2$.

For the bathhouse example, we obtain a linear temporal ordering $\sigma_1 \preceq \sigma_2 \preceq \sigma_3 \preceq \sigma_4$, where the last three are discourse referents for the described situations of the three small eventive panels, and $\sigma_1$ is the discourse referent for the described situation for the large panel depicting the bathhouse and mountain. This temporal ordering captures the understood ordering among the three “eventive” panels, but does not represent any of the content of the first panel holding also at $\sigma_2$, $\sigma_3$ and $\sigma_4$. For the sage passage, the temporal ordering correctly represents the understood linear temporal relation for all four panels. It does not represent any of the content of the initial scene-setting panel holding also at $\sigma_2$, $\sigma_3$ and $\sigma_4$.

The additional implications must be the result of pragmatic enrichment of content. It is important in this context that the literal semantic content of pictures and picture sequences is in several respects surprisingly weak. We understand the panel in (39), from Yoshihiro Tatsumi’s *A Drifting Life*, to depict three boys who are within a meter or so of each other. But many other kinds of scenes satisfy the content that is obtained with geometric semantics. In a scene projecting the picture, the individuals on the left could be not a boy, but an acrylic statue of a boy. Or the middle boy could be three times as tall as the other two, and correspondingly distant from them and the basket.

(39)

According to the most standard view in linguistic semantics, if an individual $x$ satisfies the predicate $[\text{[boy]}]$, then $x$ is a real boy—a human and not a statue. There is no way of expressing this strong a proposition pictorially. Nevertheless, and especially in the context of an entire pictorial narrative such as a comic, readers do draw implications such as the depicted objects in (39) being real boys.

In Abusch (2012), I looked at coreference or “indexing” in pictorial narratives. I argued that in literal content, depicted objects were in effect existentially quantified. It follows that in panels 6 and 7 of the Gon episode in (2), it is an open question whether the eagle-shaped individual in panel 6 is the same individual as the eagle-shaped individual in panel 7. In this respect the literal content of panels 6 and 7 is comparable to the disjointed linguistic sequence (40). I argued that predications
of identity were added to the content of pictorial narratives in the shape of formal identities between formally constructed pictorial discourse referents. These predications of identity have the status of an enrichment of literal content that is intended by the author and reconstructed by the reader.

(40) An eagle swooped in. An eagle had two fish in its beak.

So, enrichment of a weak literal content is “all over the place” in pictorial narrative. In several independent dimensions, readers piece together an enriched content using default assumptions, considerations of plausibility, and considerations of what makes a coherent story. And authors intend for them to do so. In this background, it does not seem problematic to claim that temporal predications of overlap, and classifying panels as eventive, are introduced in the same way, by free enrichment of content, without this following from construction rules that refer to aspectual properties of the literal content of pictures.

By the way, piecing together the bathhouse, hut, and sage passages is as much a matter of adding spatial predications as temporal ones. It is inferred that the stairs and bath are inside the bathhouse. Then since buildings and mountains typically stay in a fixed spatial relation to each other, it follows (for instance) that the bathhouse was below the mountain when the protagonist got into the bathing pool.

Some approaches to lexical semantics in linguistics question whether words such as boy really have strong literal contents (Carston 2008). On the alternative view, the content of the word boy is a skeletal meaning that is compatible with human boys, statues of boys, robot boys, and other things. In specific discourse situations, the basic content is enriched to some stronger content. One of the options is what is usually taken as the literal content (i.e. real boy). The choice of enrichment is driven by considerations like those mentioned above—fit with context, reasoning about intensions of authors or speakers, defaults, and considerations of plausibility. This kind of theory, with a weak or indeterminate literal content that is enriched to a much stronger understood content is what I have advocated in this section. I am pushed in this direction by the semantics that I started with, the geometric account of the content of pictures. Since this is a weak content, much of the understood content of pictorial narratives must come from enrichment.

The fact that the experienced content of pictures is enriched content recreates the possibility of aspectually sensitive construction rules for temporal discourse representations. Suppose the content of each frame $P_i$ is enriched to a stronger propositions $[P_i] \land q_i$. Even though $[P_i]$ is stative, $[P_i] \land q_i$ could have non-stative aspect. It could be eventive, just like the content of eventive English sentences. If temporal discourse construction rules operate after enrichment, there is the possibility of them being aspectually sensitive after all.

Assuming that there is an invariant narrative construction principle like (38) for pictorial narratives, does it have to use temporal succession? We can consider substituting some relation that allows for temporal overlap or identity, such as the non-precedence relation (41), which excludes the described situation for the second panel preceding the described situation for the first.

(41) Given consecutive panels $\phi_1, \phi_2$ with respective described situations $\sigma_1$ and $\sigma_2$, add a constraint $\neg[\sigma_2 \leq_\tau \sigma_1]$.

In cases where the content of $\phi_1$ and the content of $\phi_2$ are incompatible, (41) is equivalent to the strict succession option (37). For instance in panel 7 in (2), there are fish in the beak of the eagle, but in panel 8, there are not. This rules out the time projection for situations that satisfy the two panels being the same, as long as the eagle discourse referents have been identified.
One case where the possibility of temporal overlap seems welcome is where the second panel zooms in on part of the space shown in the first panel. Panel 6 in (2) shows the eagle from a distance, while panel 7 zooms in on the eagle’s upper body. At least at the level of detail that a human reader can easily see, it is unclear whether the orientation of the eagle’s body in the two frames is compatible or not. Intuitively it seems open whether time is supposed to progress between the two panels or not. Another case is McCloud’s example at the left in (10), with the sun shining down in the first panel, and the young man on his back in the second one. The pair in (14) from Gon, with the bobcat on its back in the first panel, and the eagles flying overhead in the second one, is similar. Here again it seems open whether the times for the scenes are the same, or in succession.

For the linguistic case, a rule of strict temporal succession was proposed by Dowty (1986). Dowty argued that although (42a) and (42b) appears to differ in that time “moves forward” in (42a) but not in (42b), it is not necessary to state an aspectually sensitive rule that captures this. Instead, invariant temporal succession can be assumed. Then inference that the president was behind the desk when John entered is a pragmatic one. The extension of the time of being behind the desk to contain the time of entering is possible because stative sentences can be true at nested pairs of intervals. (Examples are quoted from Dowty.)

(42)a. John entered the president’s office. The president walked over to him.
b. John entered the president’s office. The president sat behind a huge desk.

The analysis of pictorial narrative in this section supports the feasibility of Dowty’s analysis, because seemingly identical effects come up in pictorial and linguistic media, and I argued that it was not possible to draw a semantic stativity distinction in the pictorial one.

As far as I know, the phenomenon of part of the content of the second sentence being inferred to overlap the time for the first sentence has not been discussed for the linguistic case. But it is easy enough to replicate it. In (43), the greeting is understood to follow the entering. But just as much as in (42b), the president is understood to already be behind the desk when John enters. If discourse representation construction rules see the second sentence as a monolithic proposition, this is comparable to pictorial cases I looked at (the sage, bathhouse, and hut examples).

(43) John entered the president’s office. The president greeted him from an armchair behind a huge desk.

For this case, because the stative predication in the second sentence is embedded, it seems doubtful whether a construction rule that is triggered by syntax could work. And as I already said, the proposition denoted by the second sentence as a whole is not stative. This could indicate that for language too, overlap implications for stative components of meaning are introduced indirectly.

6 Conclusion

This paper has gone through a process of trying to analyze temporal progression in pictorial narrative using formal tools borrowed from linguistic semantics and temporal discourse representation theory. The semantic basis is a propositional semantics for pictures using geometric projection. I presented a construction for parsing a two-dimensional manga into a linear sequence, using a tree structure that has dual interpretations of stacking two-dimensional blocks, and concatenating sequences of panels. With these assumptions, a comic or manga is formally identical
to a linguistic narrative: in each case, we have linearly arrayed objects, each with a propositional interpretation. Section 3 reviewed data where in my linguistically-influenced interpretation, apparently eventive frames trigger temporal succession, but apparently stative frames do not. Section 4 went on to examine the formal-semantic aspectual type of the propositions denoted by pictures. The conclusion was that all propositions contributed by pictures are formally stative. Therefore, a stativity distinction cannot be used to trigger differences in temporal relations between the described situations for successive panels. An independent point was that usually, not all of the content of an apparently stative panel is understood as overlapping the described situations for adjacent panels. Section 5 concluded that the best model of temporal succession in comics is one which stipulated invariant temporal relations between successive panels, either temporal succession or temporal non-precedence.

The arguments and proposals made here should be regarded as preliminary. Clearly there is much more to say. Something that is missing both from the semantic basis and my discussion of discourse construction is an analysis of motion lines, which are often present in Gon—see (13) and (14). The second panel in (13) also includes an impact corona. I expect that if a semantics for motion lines were added to the semantic basis, some panels would prove to have non-cumulative propositional content. This could revive the possibility of an aspectually sensitive construction rule.

What are the implications of these results for the temporal structure of natural language? The pictorial example indicates that phenomena of temporal succession and overlap that are similar to what we see in language can arise in a case where literal content does not support aspectual distinctions. This could lead us to reconsider the hypothesis that phenomena of temporal discourse structure in language are triggered by aspectual distinctions in literal semantic content.

References


Dowty, David. 1986. The affect of aspectual class on the temporal structure of discourse: semantics or pragmatics?


Lewis, David. 1979. Attitudes de dicto and de se. Philosophical Review.