Applying Discourse Semantics and Pragmatics to Co-reference in Picture Sequences

Dorit Abusch
Cornell University

Abstract This paper looks at co-indexing in pictorial narratives such as comics. Using a formal-semantic model of the content of pictures, it is argued that depicted objects are existentially quantified, and are identified post-semantically. A DRT model for pictorial narratives is proposed where discourse referents are constructed as areas of a picture.

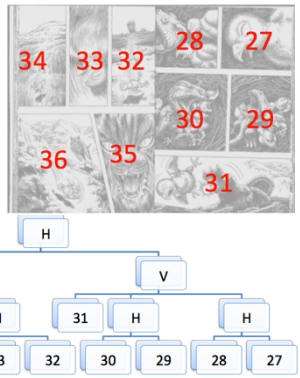
1. Introduction

The database for this discussion consists of comics and manga without words, namely without speech bubbles thought bubbles or captions. Those are wordless or silent comics, or sourds in French. I am going to bring out analogies between semantics and pragmatics of indexing (or coreference) in comics and the semantics and pragmatics of indexing in natural language. Many of my examples will be drawn from *Gon* by Masashi Tanaka, a manga series that portrays the adventures of a small powerful dinosaur in the world of modern animals. There are twenty-three episodes in *Gon*, and the images on the right are from the start of Episode 4, *Gon Goes Flying*, where Gon joins a family of golden eagles. In the first frame we see the mother eagle flying toward the cliff where her nest is located. The next image shows the nest with four baby eagles and Gon, who has decided to be an eagle. Then we see the baby eagles opening their beaks and the mother eagle flying in, carrying fish in the mouth and claws. She hovers over the nest and drops the fish into the mouths of the babies. On the next page (not reproduced here) the babies eat the fish, fall asleep, and the mother flies off.

The drawings of Tanaka are realistic drawings of modern animals and their habitat. Both the eagles and the nest in these drawings look similar to what is seen in photos of young golden eagles and their nests. The stories are wordless except for the title page and the final page which names some of the species that participate in the story, in this case the golden eagle and the bobcat.

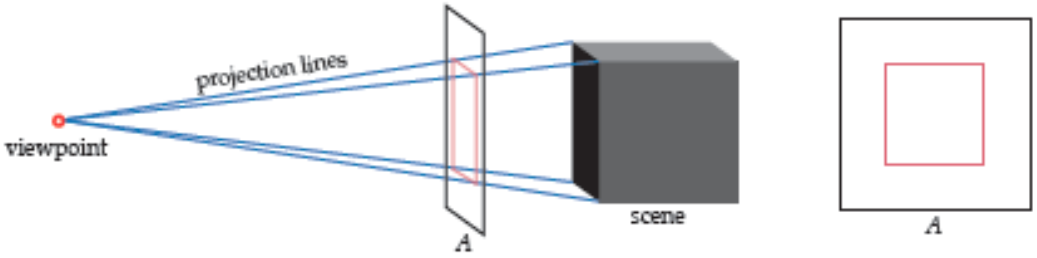
Comics consist of frames or panels which are laid out two-dimensionally on the page. The images at the top right on the next page illustrate for frames 27-36 of Episode 4 that the two-dimensional layout is parsed into a linear sequence. The parsing (which is arguably analogous to phonological parsing in language) can be formalized as a process recursive horizontal and vertical division (T. Tanaka et. al. 2007, Cohn 2008).

2. Geometrical semantics for pictures

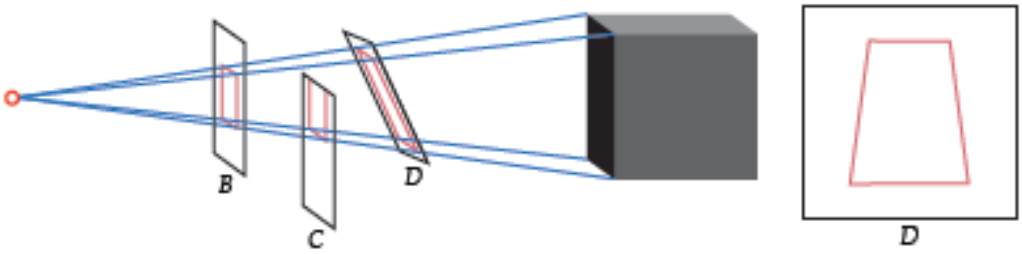


In philosophy there is a debate about the semantics of pictures. In an approach developed by Nelson Goodman in *Languages of Art*, pictures are made up of symbols with conventional arbitrary content, similarly to the lexicon, syntax, and semantics of language (Goodman 1968). According to C.S. Peirce, a picture *P* accurately depicts a scene σ if and only if *P* is similar to σ with respect to a certain set of features. According to the geometrical account of the semantics of pictures, a picture *P* accurately depicts a scene σ if and only P is obtained from σ via a geometrical transformation, $P = G(\sigma)$. An example of *G* is linear perspective, but there are other possible transformations. The geometrical approach is familiar from studies of perspective projection (e.g. Hagen 1986, Bärtschi 1994). In philosophical research it has been elaborated as a formalized semantics for pictures (Greenberg 2011).

The diagram below from Greenberg illustrates perspectival geometrical transformation. We have a scene which containts a gray cube that we want to project to a picture. We do it by choosing a viewpoint, which is the red circle, and a picture plane A. To make the drawing, one draws projection lines from points in the scene to the viewpoint, and make red marks where the projection lines cross the picture plane. The result in this case is the picture on the right in the diagram: the cube is depicted as a square.



The next diagram (also quoted from Greenberg) shows that if one puts the picture plane in different places, one gets different pictures. The square can be smaller as in B, shifted up as in C, or by tilting the plane you can get the trapezoid shape D. Just like the square A, all these are perspective drawings of a cube.



I mentioned that linear perspective is only one kind of transformation. Some painters draw what looks like realistic pictures without using linear perspective. On the next page is a painting of Yehuda Halevi street in Tel Aviv by Shalom Flash. The building to the right curves in towards the top. This curving effect results when the scene is projected onto a part of a sphere, rather than a plane. Then the sphere drawing is flattened out. Greenberg (2011) discusses spherical projection as an argument against a similarity account of the semantics of pictures. It also tends to support the applicability of geometrical semantics to a variety of drawing and painting styles, including Flash's.

An interesting example of formalized geometrical transformations is what is called non-realistic rendering in computer graphics. Algorithmic models of image generation use three-diminensional models encoded in data types, and varieties of projection to generate images. They may include transformations that produce styles of painting and drawing such as the Monet's haystack on the left below, or pen-and-ink drawing at the right. Even if the added transformations are defined at a two-dimensional level, they can be inverted to obtain the semantics for a picture as a set of pairs of viewpoints and situations. This supports the applicability of a geometrical formal semantics also to images such as comics. (These images are by Barbara Meier on the left and by Jörg Hamel on the right. Quoted from Reynolds 2003).

The statement of geometric projection above refers to a viewpoint in addition to a described situation. This can be introduced into the semantics in several ways; one is to take the semantics of a picture to be a set of pairs of a viewpoint and a situation. See Greenberg (2011) for discussion of additional options. Formula (1) uses a geometric transformation G, an artistic transformation A, and a viewpoint parameter v to define a semantic value of a picture. v is assumed to encode also the picture plane.

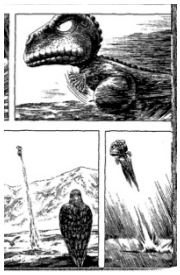
(1) $$[\![P]\!] = \{\langle v,\sigma\rangle \mid A(G(\sigma)) = P\}$$

The semantic value of a picture P is the set of pairs of viewpoints v and situations σ such that σ projects to picture P with respect to viewpoint v. This semantic value is similar to the centered propositions that are used in Lewis's account of de se attitudes (Lewis 1979). The construction does not have to do with agents or

attitudes though, and the pictorial modality can be characterized as circumstantial rather than epistemic.

If we want to compare the discourse pragmatics of natural language to the discourse pragmatics of silent comics, this is a nice setup. The statement of semantic values allows us to reason formally about the semantics and pragmatics of pictures. Because the basic semantic values are so similar to what is used in natural language semantics, we can compare what happens in the two domains.

There are other things which are represented by comic images which are not covered by projection semantics. In *Gon* we often see impact coronas like the one in the picture on the right where Gon kicks a little eagle and the corona represents the occurrence of an impact. Here projection is involved in determining the location of the impact, but the outline of the impact picture is not determined by projection. Tanaka also makes frequent use of various kinds of motion lines, such as the ghost lines in the kicking picture, and the lines in the direction of motion in the three panels on the right. These devices seem to involve projection, but at more than one time point. Clearly impact coronas and motion lines require extensions in the denotational framework. But the point is orthogonal to the topic of this paper, I will not discuss it further here.



3. Indexing and (in)definiteness

I now turn to the topic of indexing. In Episode 4 of *Gon* there is a passage where Gon kicks a little eagle which starts bouncing down a cliff, as depicted in the first four panels on the right. Then we see a bobcat opening his mouth, and the bobcat jumps towards the little eagle. These are panels 31 through 36 in the episode. In denotional models, we get satisfaction conditions along the following lines:

(2) σ_{31} atisfies P_{31} only if in σ_{31} a small dinosaur kicks a small eagle.

This is not complete satisfaction condition, because the picture places further constrains on the geometric configuration of the dinosaur and eagle. This is why I say "only if". There is another issue which I will mainly gloss over here. The literal content of the picture according to projection theory does not entail that there is a real little eagle in the described situation. It could be a statue of an eagle, or a picture of an eagle, or many other things. The satisfaction condition (3) is somewhat more correct.

(3) σ_{31} satisfies P_{31} only if in σ_{31} there is an impact between the moving leg part of a dinosaur-shaped thing and a small moving eagle-shaped thing.

For the next frame 32, we get a condition along these lines:

(4) σ_{32} satisfies P_{32} only if in σ_{32} a small eagle-shaped thing bounces down a cliff-shaped surface.

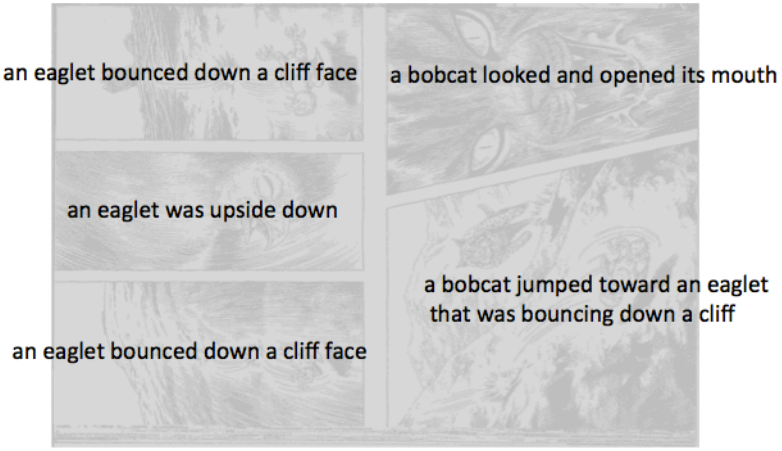
In these satisfaction conditions that come out of the denotational semantics, the eagle, dinosaur, and cliff surface are existentially quantified. As a result, nothing in the semantics tells us that the eagle in σ_{31} is the same as the eagle in σ_{32}. So in the basic semantic model, the panel sequence P_{31}, P_{32} does not carry the information that the eagle that is kicked is the eagle that bounces down. To see this point, it is helpful to be more specific about how the described situation σ_{31} relates to the described situation σ_{32}. We can say that a described situation for the sequence of pictures P_{31},P_{32} is a situation that contains σ_{31} and σ_{32} as subparts, perhaps with an additional constraint that σ_{31} temporally precedes σ_{32}. Then we get this partial satisfaction condition:

(5) σ satisfies P_{31},P_{32} only if σ has a part σ_{31} such that in σ_{31} there is an impact between the moving leg part of a dinosaur-shaped thing and a small moving eagle-shaped thing, and σ has a part σ_{32} such that in σ_{32} an eagle-shaped thing bounces down a cliff-shaped surface.

Thus the information in P_{31} and P_{32} gets combined by summing described situations and conjoining conditions on the situations. In (5), there are two existential quantifiers for eagles, and the witnesses for these quantifiers could be different. So in a situation that satisfies the combined condition, the eagle that is kicked could be different from the eagle that bounces down. Clearly though, Tanaka intends for us to understand that they are the same, and readers understand the story in this way.

It is interesting to compare the comic passage with paraphrases in natural language for the satisfaction conditions that come out of the picture semantics. (6) is the kick sequence decorated with English paraphrases. I use indefinite descriptions for the eagle, bobcat, and cliff because there are existential quantifiers in (5).

(6)



As an English passage, this sequence of sentences is disjointed, and if anything we infer that the eagle that was kicked is different from the eagle that bounces down. Yet the semantics of (6) is identical in relevant respects to the semantics (5) of the picture sequence. The difference must come from the fact that in English, a passage with indefinites is in completion with one with definites. To paraphrase the comic as it is naturally understood, we must use definite descriptions:

(7) A bobcat looked and opened its mouth. The bobcat jumped toward the eaglet that was bouncing down the cliff.

So the fact that the English passage (6) does not convey a meaning with intended coreference emerges from scalar conversational logic. Because a competor (7) that expresses coreference is available and not used, it is inferred that the encoder of the passage intends to convey a meaning with non-coreference (i.e. where the eaglet that was kicked is different from the eaglet that bounced down).

If we go back to the picture semantics (5), the reader does fill in the information that the eagle that bounced down is the same as the eagle that was kicked, and the author indends for her to do so. This is not blocked because there is no competing picture that indicates coreference---there is no option of adding "morphology" that expresses definiteness, at least in Tanaka's manga practice.

The identities that are "filled in" have the status of pragmatic enrichment. I understand this to be information that is added conjunctively to literal meaning in constructing the discourse representation for a passage. A linguistic example is the implication in (8) that the key was used to open the door. This is standardly held to be a pragmatic enrichment, rather than an entailment of literal content.

(8) He took out a key and opened the door.

For a closer analogy in natural language semantics, we can look to grammatical categories for which there is no definiteness distinction, such as tensed verbs. Consider (9) as analyzed in event semantics, where an existentially quantified event variable is introduced for each verb. This results in a discourse representation along the lines of (10), where there are two stripping-down event discourse referents. Let us look first at what happens with the discourse referent x_3 that corresponds to the pronoun that is the object of the second occurrence of *stripped*. In the standard version of DRT that is employed in (10), this pronoun introduces a fresh discourse referent, but it is part of the conventional meaning of a pronoun that discourse referent must be equated with some antecedent, here the discourse referent x_1 for the engine (Kamp and Reyle 1993). There is also a fresh discourse referent e_3 for the event argument of the second occurrence of *stripped down*. Because this verb is not morphologically definite, there is nothing in the morphology or syntax that prompts identifying e_3 with another event discourse referent. But we understand the stripping down event that is a witness for the second sentence to be the same as the stripping down event that is a witness for the first. This incremental information has the status of a pragmatic enrichment. In a model where such enrichments are written into the discourse representation, the

equation in (11) should be added. This process is parallel to what I said happened with the manga kick sequence, because in the literal semantics (10) the discourse referents e1 and e3 are distinct and independently quantified. The information (10) is compatible with the stripdowns being different or the same, just as (5) is compatible with the two eagles being different or the same.

(9) An engine was stripped down and rebuilt. Justin stripped it down.

(10) x_1 e_1 e_2 y x_3 e_3
engine(x_1)
stripdown(e_1) theme(e_1,x_1)
rebuild(e_2) theme(e_2,x_1)
stripdown(e_3) agent(e_3,y) theme(e_3,x_3)
y=Justin x_3=x_1

(11) e_3=e_1

As an aside, it is interesting to check what happens when we fill in an indefinite in (9) in the place of the pronoun. In the resulting sentence (12) one can hardly understand the passage as conveying that the engines were the same, though I guess one can understand that it is an open question whether they are. So even though the pair of discourse referents x_3,x_1 in (13) are isomorphic to the event discourse referents e_1,e_3 in (10), the identity x_3=x_1 can *not* be added as an enrichment. As I already stated, this is to be attributed to the avaibality of a competing sentence with a definite nominal.

(12) An engine was stripped down and rebuilt. Justin stripped an engine down.

(13) x_1 e_1 e_2 y x_3 e_3
engine(x_1)
stripdown(e_1) theme(e_1,x_1)
rebuild(e_2) theme(e_2,x_1)
stripdown(e_3) agent(e_3,y) theme(e_3,x_3)
y=Justin engine(x_3)

I think in (14) the engine discourse referents can be understood with enriched coreference. Perhaps somehow the counterpart (15) with a definite does not block coreference by enrichment in (14), because the discourse structure is different. In (14), the second sentence is perceived as a restatement and strengthening of the first.

(14) An engine was stripped down and rebuilt. Justin stripped an engine down, and Keisha rebuilt it.

(15) An engine was stripped down and rebuilt. Justin stripped it down, and Keisha rebuilt it.

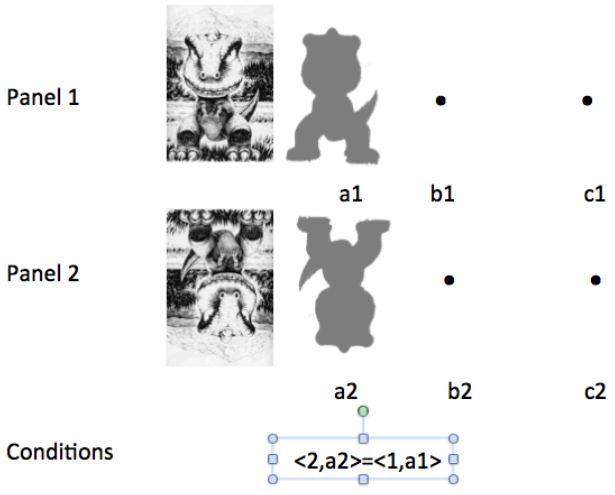
Returning to the main argument, we have seen that in the satisfaction conditions that fall out of the geometric semantics for pictures, as extended in a simple way to a semantics for picture sequences, variables for agents and objects in the described

situations are existentially quantified. When there is understood co-reference across panels, this is a matter of pragmatic enrichment. On this analysis co-reference in comics is "purely pragmatic". This is different from nominals in English, but is parallel to what is seen for tensed verbs in English. The pictorial and natural language data are both covered by the hypothesis that existentially quantified variables or discourse referents in the literal meaning can be enriched with added identities among discourse referents, as long as this is not blocked by a competing representation with a definite.

4. A DRT formalism for pictures

One can take some steps toward formalizing these ideas by designing a DRT-like notation for the discourse semantics of pictures. The main issue is how to introduce discourse referents. In discourse representation theory for natural language, discourse referents are variable-like objects that are projected from syntax, and that at the DRT level serve as arguments for predicates mapped from natural-language content words such as nouns and verbs (Kamp 1981, Kamp and Reyle 1994). In the semantics for DRT, discourse referents are mapped to individuals using assignment functions. It is hard to state an analogue to any of this in the geometric semantics for pictures. For instance, in the geometric semantics for pictures, there are no predicates that a discourse referent could be an argument of.

My strategy is to introduce discourse referents at the level pictures. Discourse referents will be constructed out of areas in a picture. The image at the right is a panel from *Gon* with a contiguous area shaded in pink. The image was obtained in a photo manipulation program by tracing out a closed curve. Areas are used in photo-manipulation software to make semantically significant changes, such as putting someone in a different environment, or changing the color of a model's shoes. Here areas of pictures are used in a construction of discourse referents. To construct a discourse representation, one distinguishes some areas within each picture in the sequence of pictures that constitutes the pictorial narrative. Identities between discourse referents are then formal identity predications between the areas. The image at the right represents the process schematically. Some areas a_1 , a_2,..., a_1, b_2,... in the sequence of panels are distinguished. These areas are geometric. To assure that discourse referents for different panels are distinct, discourse referents are constructed as *pairs* of a panel index and an area. Then co-references are stated as syntactic identities between such discourse referents, for instance $\langle 2, a_2 \rangle = \langle 1, a_1 \rangle$. These conditions are of the



same nature as the identity conditions used in DRT for natural language. But aside from these conditions, there are no formulas in the discourse representation.

In the semantics for DRT, discourse referents are mapped to the model with assignment functions. On this account the identity $c = d$ in a discourse representation structure is satisfied with respect assignment g if and only if $g(c) = g(d)$. Here the objects that are assessed for identity are individuals in the model. The trick now is that in pictorial discourse representations, the projection relation already provides a kind of mapping to the model. In Panel 1 in the example, if with respect to viewpoint *v* the area a_1 corresponds to an individual *x* in scene σ, it is because lines drawn from the viewpoint through the area a_1 in the picture plane intersect *x* before they intersect any other object. We want to say roughly that the object picked out in σ by the area *a* in the picture plane of *v* is the object *x* such that any ray drawn from the viewpoint of *v* through *a* instersects *x* before any other object.

There are a couple of problems with this. Consider a shoe box (including a cover) that is tipped towards the viewpoint, so that only the cover is visible from the viewpoint. Then a ray from the viewpoint through a certain point in the picture plane that intersects the cover also intersects the shoe box. Is it the shoe box or the shoe box cover that is picked out by the discourse referent? Consider a picture of Gon underwater. A ray in the direction of Gon intersects the water before it intersects Gon. So how can Gon be picked out by a discourse referent constructed as an area? Consider a picture of a polar bear in a snowstorm. The polar bear can be made out, but for most points in the picture within the outline of the polar bear, one can't tell whether they are within a projection of some fur of the polar bear, or of a snowflake.

In addition, there are issues of cross-identification of individuals through time. The water and air in a tornado is exchanged continuously, but the tornado persists through time. To assess the truth of an equality condition on discourse referents constructed as above, it matters whether we assess identity of tornadoes, or identity of masses of water and air.

I think these problems indicate that when a reader infers identity between objects depicted in different frames of a visual narrative, she is not simply inferring identity between the objects that project to certain areas of the different frames. To determine what objects are picked out in this way, and to make sense of the identities, it is necessary to refer to a predicate such as 'shoe box', 'shoe box cover', 'polar bear', or 'tornado'. While it is problematic to ask whether the object depicted in area a_1 of picture 1 is the same as the object depicted in area a_2 of picture 2, one can un-problematically ask whether the shoebox depicted in area a_1 of picture 1 is the same as the shoebox depicted in area a_2 of picture 2. To implement this idea, predicates could be introduced as part of the identity predication, or as part of the discourse referents:

(16) $\langle 2, \mathbf{animal}, a_2 \rangle = \langle 1, \mathbf{animal}, a_1 \rangle$

(17) $\langle 2, a_2 \rangle =_{\textbf{animal}} \langle 1, a_1 \rangle$.

Either way, it is the triple that takes on the role of a discourse referent---something that maps to an individual in the world. I will assume the first representation.

This completes my formalization of a discourse representation notation for comics. Returning to the issues from the previous section, when we add formulas like (16) or (17) to a picture sequence, we get a meaning-bearing representation that is true only if certain individuals that are mapped from the discourse referents are identical. As before, I maintain that while the identities are not part of the literal content of the comic, the representation including the identities is enriched "reading" of the comic that is intended by the author and recovered by the reader.

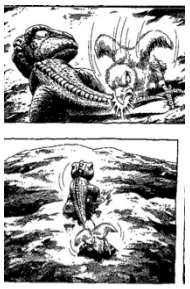
5. Hypotheses about indexing

On the account discussed so far, co-reference in comics is post-semantic. In the basic semantic model, the individuals that witness the truth of different panels are permitted to be different, and unlike for nominals in English, there is no morphological phenomenon of definiteness that prompts co-indexing. I record all of this as a hypothesis.

(18) Indexing in comics is not part of the basic representational or denotational mechanism. It is added post-semantically (pragmatically).

Now I am going to look at an alternative hypothesis that is inspired by work on indexing in vision. Z. Pylyshyn has proposed that in human vision, some indexing is performed by a low-level system, and that an image is presented to the higher cognitive system as already indexed (Pylyshyn 2003). The indexing is performed by geometrically-based algorithms. For instance if within a short time span, a dot is presented in one position and then a dot is presented in a nearby position, this is perceived as a moving dot, rather than a dot disappearing and another dot appearing. As Pylyshyn has it, the output of the low-level system includes an "index" that is instantiated at the two time points. While low-level and algorithmic, the processes involved encode aspects of the geometry of vision, such as the possibility of occlusion of one object by another.

Could indexing in comics relate to low-level properties of the image, rather than being post-semantic? On the right we see adjacent panels from *Gon* 4. Ignoring meaning, the images of Gon look similar as two-dimensional patterns. Maybe they get matched up because of their similarity at this level, disregarding semantic interpretation. Some similar things are done in computer vision. Garg et. al. (2012) describe a system that finds picture-parts depicting the same person in a set of images of a crowd. In the image on the right (quoted from Garg et al.'s paper), the system has found images of a girl in a gray tank top sitting in different postures at different times, and viewed from different vantages. The system in part uses picture-level areas that correspond to body parts and pieces of clothing, and pixel-level features that capture color and texture.



Suppose it were possible to evaluate pictorial discourse referents as pictures, using low-level non-semantic features. Then we could state a discourse representation construction rule along the lines of 'if drefs *x* an *y* are similar as pictures, add the identity *x*=*y*.' This is subtly different than the previous hypothesis because it has to do with similarity of the two-dimensional pictures rather than with a pragmatic process---or at any rate it refers to properties of the picture, rather than their semantic interpretation. I am going to restate it. Suppose we have a mathematical function *s* which allows us to evaluate the similarity of two pictures. $s(x, y)$ is close to 1 if *x* and *y* are similar, and is close to 0 if *x* and *y* are not similar. Then we can state this default rule.

(19) Given drefs *x* and *y* in different panels such that $s(x, y) > 0.9$, by default add the DRS condition x =y.

If there was such a function *s* and if readers assumed such a coreference convention, it would be useful because the comic author can use it to express his intensions about coreference. If two drefs are supposed to be coreferent, he draws them so that their similarity according to *s* is high. If they are not supposed to be coreferent, he draws them so that their similarity according to *s* is low.

This procedure is actually similar to some default axioms used in the theory of discourse structure. Take the default axiom (20) governing temporal succession of the events described by juxtaposed clauses.

(20) If clause B immediately follows clause A, by default add a constraint that the event described by B follows the event described by A.

The relevant thing about this is that it refers to linguistic form, namely to clause A and clause B being adjacent, but the change in representation that it creates is essentially semantic. Also, the force of "by default" is to allow the enrichment to be cancelled by semantic and pragmatic information, including considerations of plausibility. Default principles like this are used by Asher and Lascarides (2003) in modeling enriched information in natural language discourse. From this point of view, the two hypotheses under discussion in this section are not incompatible. The informational increment of equating discourse referents in different panels is certainly post-semantic, but the default principles that govern it could involve features at different levels, including picture-level features.

5. Group discourse referents

Let us apply our discourse-structural theory to panel 2 of *Gon 4*, with Gon in the eagle nest with four baby eagles. To establish discourse referents, one area of the picture is distinguished for Gon, and four areas for the sibling birds are distinguished. Say the resulting discourse referents are $\langle 2, a_2 \rangle$ (Gon), $\langle 2, b_2 \rangle$ (one eaglet), $\langle 2, c_2 \rangle$ (another eaglet), $\langle 2, d_2 \rangle$ (another eaglet), and $\langle 2, e_2 \rangle$ (another eaglet).

In panel 31, the panel repeated on the right, Gon is shown kicking an eaglet out of the nest. Here two areas are distinguished, resulting in discourse referents $\langle 31, a_{31} \rangle$ and $\langle 31, f_{31} \rangle$. Unproblematically the two discourse referents for Gon can be equated, $\langle 2, a_2 \rangle = \langle 31, a_{31} \rangle$. But for the eaglet that is kicked there is a problem: any of the equations $\langle 31, f_{31} \rangle = \langle 2, b_2 \rangle$, $\langle 31, f_{31} \rangle = \langle 2, c_2 \rangle$, $\langle 31, f_{31} \rangle = \langle 2, d_2 \rangle$, or $\langle 31, f_{31} \rangle = \langle 2, e_2 \rangle$ is justified on grounds of consistency, plausibility, and simplicity of the semantic content, and on grounds of similarity of the drefs as pictures. Yet the story at this point is not perceived as incoherent or unacceptably indeterminate.

Putting the panel into English suggests a solution. (21a) is a partitive, with an embedded phrase that refers to the *plurality* of four eaglets. (21b) is the same, but with a pronoun in the partitive. (21c) is perhaps most similar to panel 31---though it is overtly indefinite, in the context of the story it is understood partitively, as equivalent to (21a).

(21) a. Gon kicked one of the eaglets.
b. Gon kicked one of them.
c. Gon kicked an eaglet.

Research on plural reference in discourse representation has suggested that plural discourse referents can be freely created (Kamp and Reyle 1993). In the story (22), suppose the first sentence sets up discourse referents *u* and *v* for the wild turkey and the Golden Retriever. In order to analyze the partitive in the second sentence, a plural discourse referent is introduced with a sum operator as in (23a). The second

sentence in (22) introduces a discourse referent x corresponding to "one", and a group-level discourse referent Z corresponding to "them". (23b) states the semantic of the partitive. Z is related to its group-level antecedent with the equation (23c).

(22) This morning a Golden Retriever fought with a wild turkey right outside our house. One of them was badly injured, and I called the Cayuga Heights police. They weren't interested, I should have called the humane society.

(23) a. $W = u+v$
b. $x \in Z$
c. $Z = W$

Creation of such group-level antecedents is entirely "free". Although in a sense the summation (23a) is an accommodation that is prompted by the need to find an antecedent for Z, there is no perception of disfluency.

I suggest that free creation of group-level discourse referents is an attribute of our cognitive machinery for representing narrative, and so is available also for pictorial narratives. Once discourse referents for pictorial narrative are in place, the formal process is the same. A plural discourse referent is created from the four bird referents in (24a). Then the eagle discourse referent in panel 31 is related to an antecedent using an element relation with the formula (24b).

(24) a. $W = \langle 2, b_2\rangle+\langle 2, c_2\rangle+\langle 2, d_2\rangle+\langle 2, e_2\rangle$
b. $\langle 31, f_{31}\rangle \in W$

To be sure, there are differences between the linguistic and pictorial media. For one thing, in the pictorial medium there are no plural pronouns that prompt the creation of a plural discourse referents. For another, in the pictorial medium it is difficult to introduce a discourse referent that can be witnessed indeterminately by a Golden Retriever and a wild turkey, because these animals look so different. But the differences can be seen as arising from the informational resources of the two media. We can still say that the mechanism of forming plural discourse referents is the same.

There is a subtly different case that comes up frequently in superhero comics. There is an impostor (say) Superboy who looks like the real one, and in a panel depicting both the reader doesn't know which is which. This is captured by equations of the form (25). The difference is that a lot of information is available about $\langle j, f\rangle$ (the real Superboy) so that it matters which is which.

(25) a. $W = \langle k, b_2\rangle+\langle k, c_2\rangle$
b. $\langle j, f\rangle \in W$

6. Further cases of default indefiniteness

Above I discussed default semantic indefiniteness in comics, and with tensed verbs in an event-semantic analysis of natural language. I claimed that an analysis is well supported where variables are semantically indefinite (existentally quantified), and are optionally identified with identity predications.

Another case of this comes up in novels with dialogue where speakers are not explicitly identified. (26) is the opening passage from William Gaddis's *JR*.

(26) Justice? –You get justice in the next world, in this world you have the law.
--Well of course Oscar wants both. I mean the way he talks about order? She drew back her food from the threat of an old man paddling by in a wheelchair,
--that all he's looking for is some kind of order?
--Make the trains run on time, that was the…
--I'm not talking about trains, Harry.
--I'm talking about fascism, that's where this compulsion for order ends up. The rest of it's opera.
--No but do you know what he really wants?

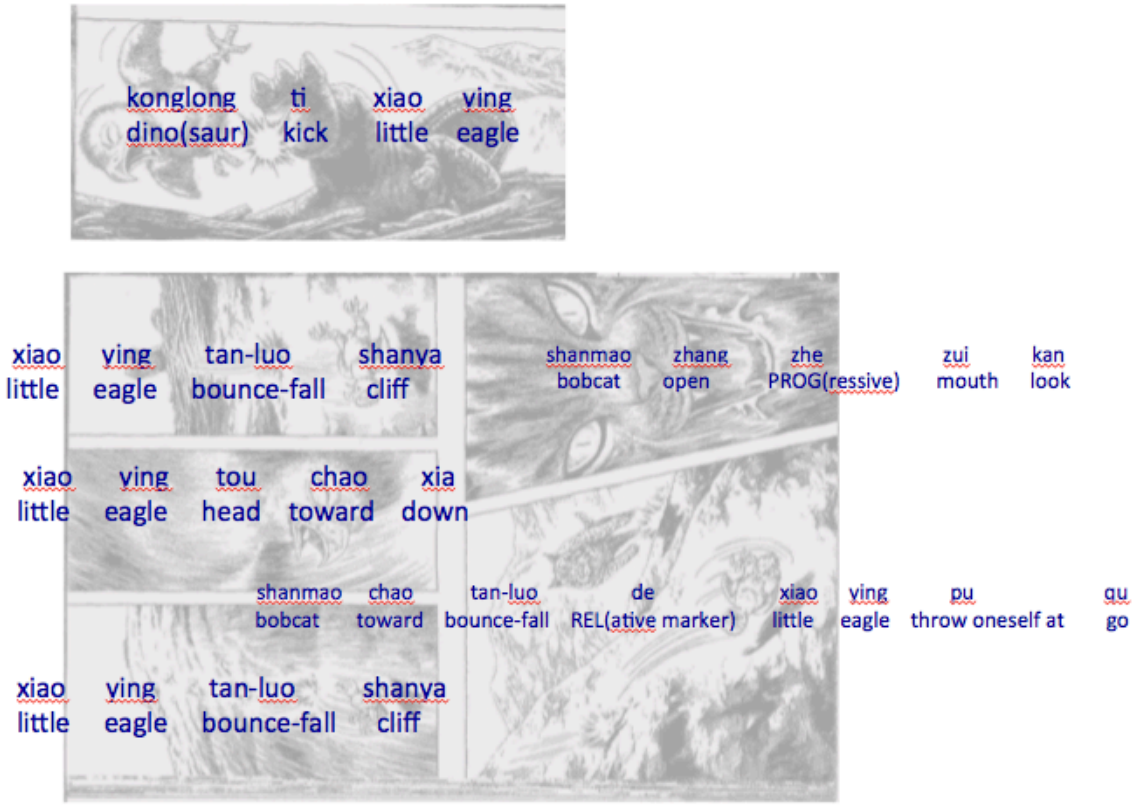
Quoted speech is marked with the dash, but speakers are not identified directly. Identifying the speakers is an optimization problem, using constraints involving alternation of speakers, characteristic diction of speakers, names used by other speakers, and much else. Readers report that they work out the cast of characters over a couple of hundred pages of the 700-page novel. The speakers in the passage above turns out to include two aged aunts whose roles as speakers here can not be distinguished.

Arguably the literal content of the passage is along the lines of (27), where the speakers are existentially quantified. Then speakers are cross-identified using predications of identity and membership.

(27) Someone said "You get justice in the next world, in this world you have the law."
Someone said "Well of course Oscar wants both. I mean the way he talks about order?", drew back her food from the threat of an old man paddling by in a wheelchair, and said "that all he's looking for is some kind of order?"
Someone said "Make the trains run on time, that was the…"
Someone said "I'm not talking about trains, Harry."
Someone said "I'm talking about fascism, that's where this compulsion for order ends up. The rest of it's opera."
Someone said "No but do you know what he really wants?"

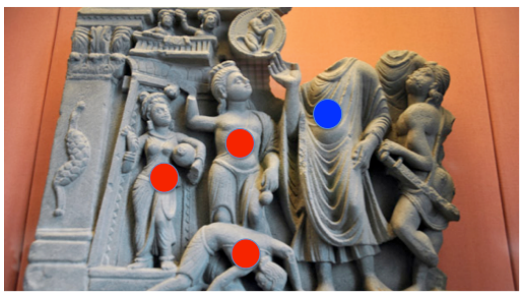
More controversial is the analysis of languages such as Mandarin Chinese without overt definitnes marking. Below is a decoration of the kicking passage with Chinese glosses. The nominals are bare, without a definite or indefinite determiners. We could say that there is a hidden definiteness distinction (in LF). But conceivably we could also say that the nominals are all indefinitely quantified, like discourse referents in comics, the events described by tensed verbs in English, or speakers in

Gaddis's novels. Judging by the cases, nothing would "go wrong" on an existential analysis, because coreference can be introduced pragmatically.



7. Conflated narrative

Dehija (1997) discusses data where contrary to what was seen above, coreference across temporally separated depicted events is stipulated in the "syntax" of narrative pictures. This comes where different panels effectively overlap in what Dehija labels "conflated" narrative.



The narrative statue on the right depicts some events in the life of Dipankara, an incarnation of the Buddha. First Sumedha (the leftmost figure marked with red) buys some lotuses. Then Sumedha (the top figure marked with red) tosses the lotuses at Dipankara (marked with blue). Finally Sumedha (at the bottom marked with red) bows down and spreads his hair on the ground for Dipankara to step on. Sumedha is depicted once in each "panel", but Dipankara is depicted just once. One can say that image of Dipankara is part of two overlapping panels, and that there is just one discourse referent for Dipankara. As a result there is no issue of equating discourse

referents (or not) across temporally separated events. In this way, coindexing across temporally separated events is encoded in the syntax of the sculpture.

In English, coreference can be stipulated in this way using relative clauses, and other constructions. Just like the conflated sculpture, sentence (28) conveys syntactically that the individual at whom the lotuses were thrown is the individual before whom Sumedha bowed down. The relative clause construction stipulates coreference between the head of the relative clause and the trace position in the relative clause.

(28) Sumedha bowed down and spread his hair before a man to whom he had thrown some lotuses.

To fill out the proportion, we can say that the Dipankara statue stands in the same relation to a pair of sculptures without overlap as the relative clause sentence (28) stands to a variant (29) with separate clauses and two indefinites.

(29) Sumedha threw some lotuses at a man. Sumedha bowed down before a man.

8. Conclusion

This paper looked at co-indexing across panels in silent visual narratives. It falls out of a geometric account of the semantics of pictures that depicted objects are existentially quantified. Nevertheless, readers identify individuals across panels, without any morphological cue. The same phenomenon shows up in event variables of tensed verbs, and in the identity of speakers for quoted dialogue in some novels. My account was formalized in an adaptation of discourse representation theory.

From the standpoint of the semantics and pragmatics of natural language, visual narratives are startlingly different and startlingly familiar. Here I emphasized that coindexing is achieved without marking in morphology or syntax, and that depicted objects are in effect existentially quantified. This raises the question whether an existential analysis might be applicable to more cases in natural language than we usually think. Just as interesting is evidence that there are principles of discourse representation that cut across multiple media.

Acknowledgements

This paper was presented at Sinn und Bedeutung 2012. Other versions were presented at the ESSLLI Workshop on Projective Meaning, Ljubljana (Aug. 2011), the UCLA workshop on Visual Narrative (June 2012), and the Cornell Workshop on Linguistics and Philosophy (Fall 2012). Colloquium presentations of the materiel were given at Toronto, Stanford and Rutgers. Thanks to the audiences on these occasions. I am grateful to Ede Zimmermann for great comments, and to Cleo Condoravdi, Regine Eckardt, Gabe Greenberg, Elsi Kaiser, Magdalena Kaufmann, Roger Schwarzschild, Yael Sharvit, Zhiguo Xie for comments and assistance. Special thanks to Mats Rooth for endless discussions of this topic.

References

Asher, N. and N. Lascarides (2003). *Logics of Conversation*. Cambridge University Press.

Bärtschi, W. (1994). *Geometrische Linear- und Schatten-Perspektive*. Wiesbaden: Vieweg.

Cohn, N. (2008). Navigating comics: reading strategies of page layouts. Ms. Tufts University.

Dehejia (1997). *Discourse in Early Buddhist Art*. New Delhi: Munshiram Manoharlal.

Goodman, N. (1968). *Languages of Art: an Approach to a Theory of Symbols*. New York: Bobbs-Merrill.

Greenberg, G. (2011). *The Semiotic Spectrum*. PhD dissertation, Dept. of Philosophy, Rutgers.

Hagen, M. (1986). *Varieties of Realism: Geometries of Representational Art*. Cambridge University Press.

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk et al. (eds.). *Formal Methods in the Study of Language*. Amsterdam: Mathematics Center, 1981.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic: Introduction to Model Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer.

Lewis, D. (1979). Attitudes de dicto and de se. Philosophical Review.

Reynolds, C. (2003). Stylized Depiction in Computer Graphics: Non-Photorealistic, Painterly and 'Toon Rendering. http://www.red3d.com/cwr/npr/.

Tanaka, M. (1992-1994). *Gon 1*. Episodes 1-4. Kodansha Limited. Current edition Kodansha Comics (2011).

Tanaka, T., K. Shoji, F. Toyama, J. Miyamichi (2007). Layout analysis of tree-structured scene frames in comic images. Proceedings of IJCAI.

Pylyshyn, Z. (2003). *Seeing and Visualizing: It's Not What You Think*. MIT Press.