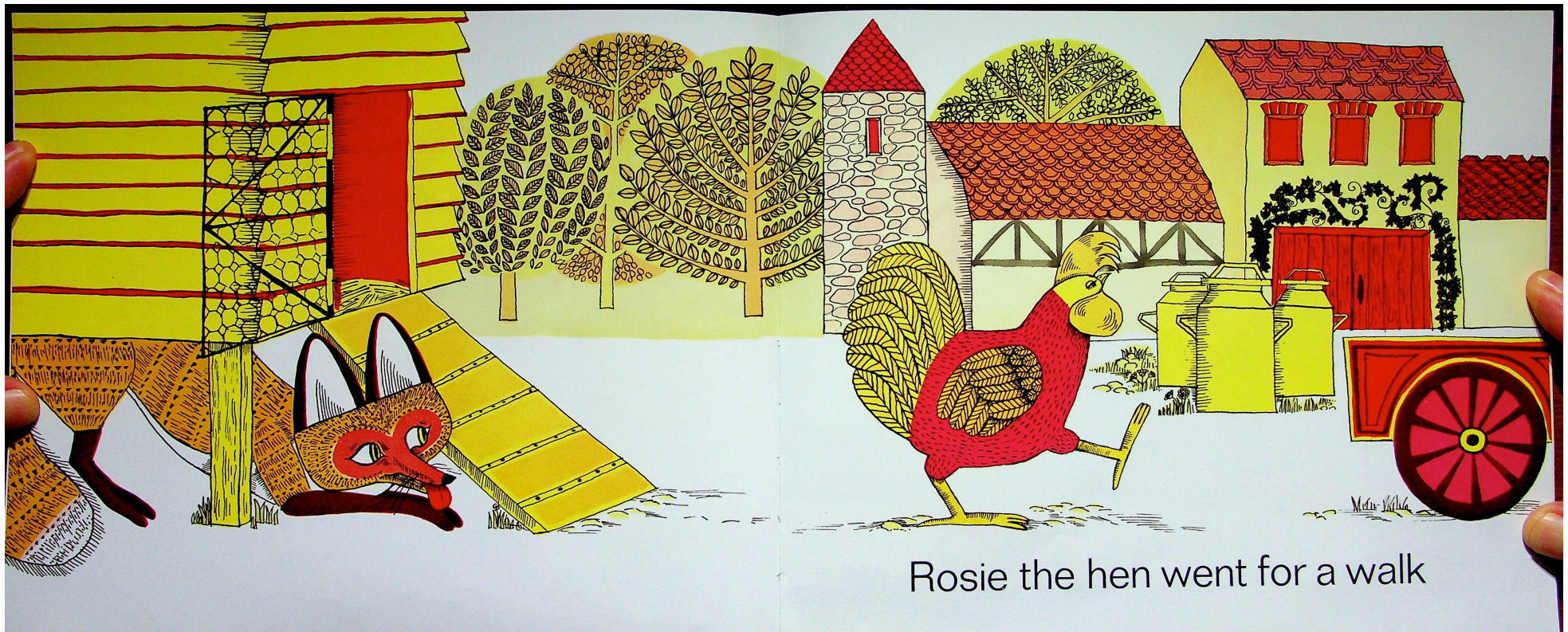


Implicatures from Text with Pictures

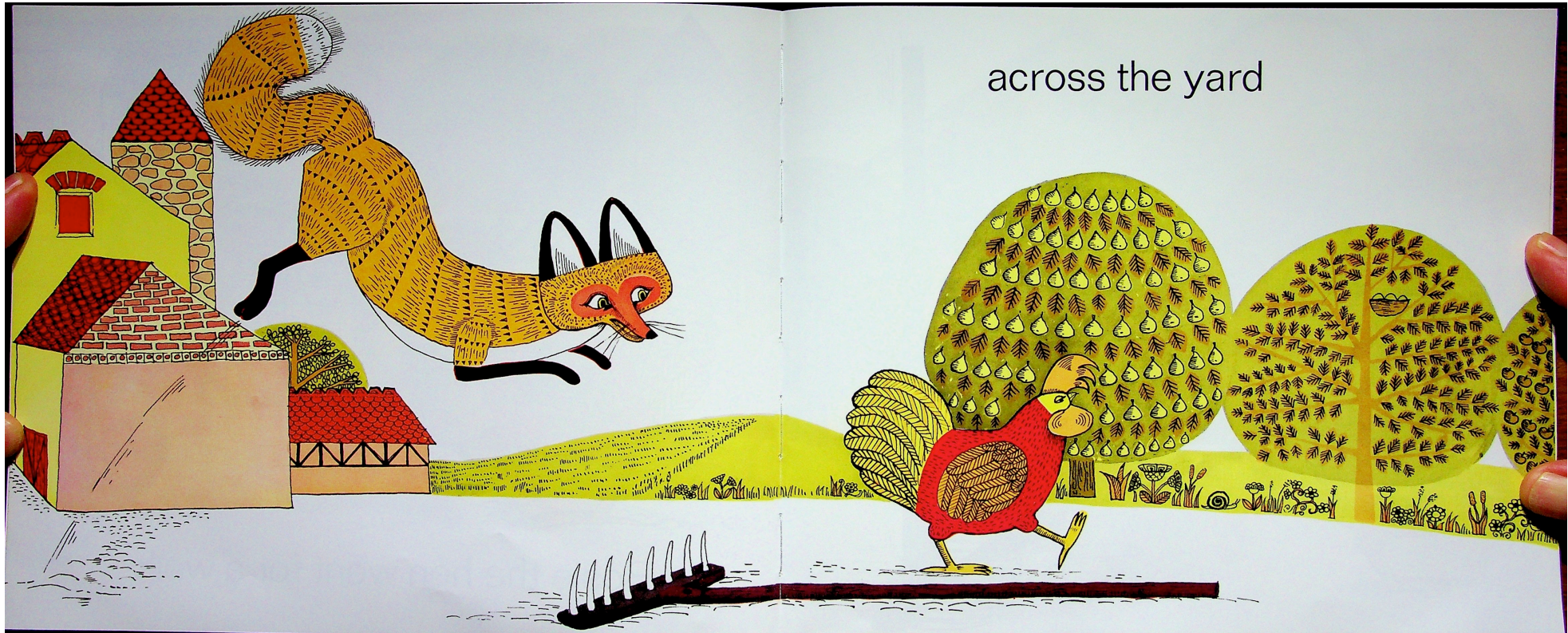
Dorit Abusch
Cornell University

Rosie's walk

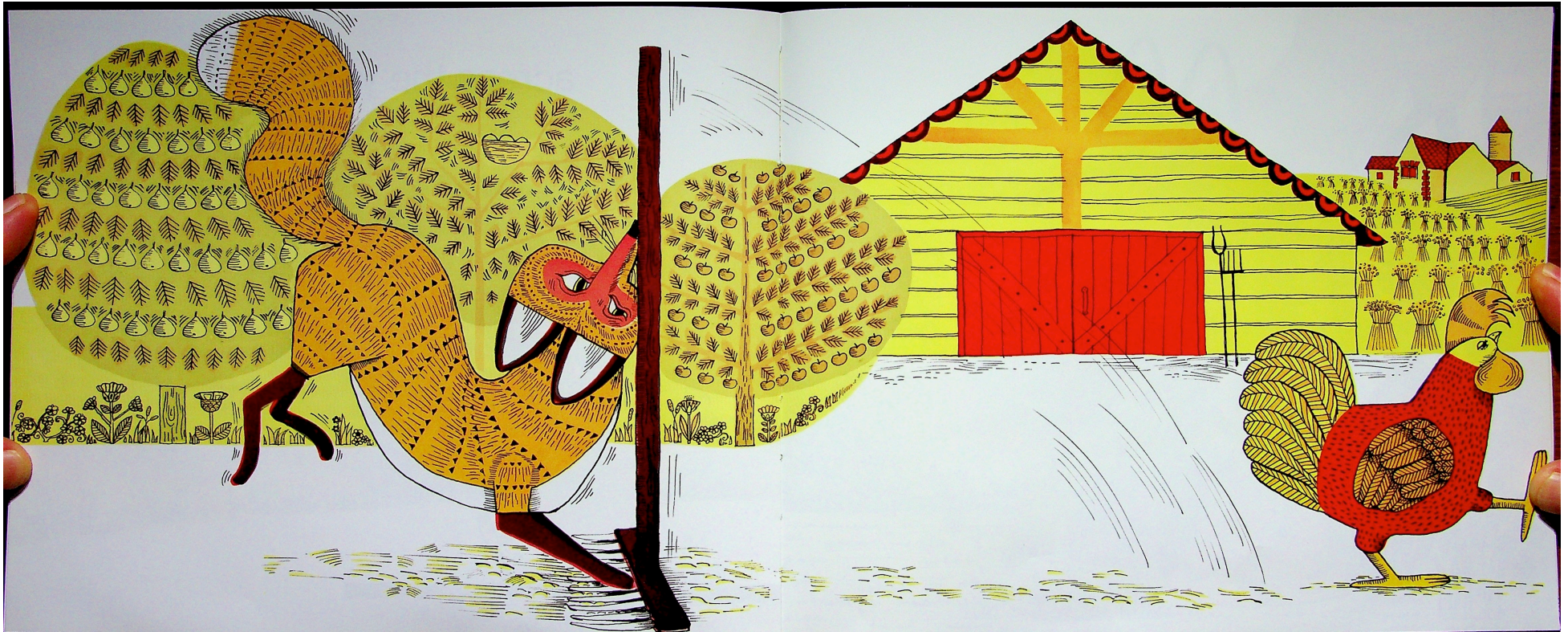
Pat Hutchins



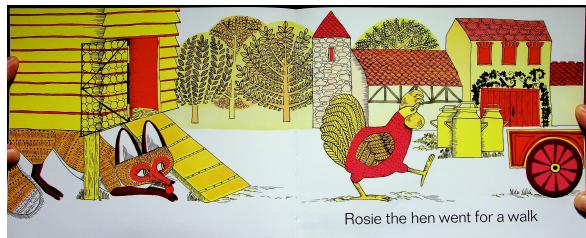
Rosie the hen went for a walk



across the yard



The basic discourse relation between language and image in picturebooks is *co-temporal juxtaposition*. This means that the events and states described by the language on a single page or two-page spread temporally overlap the eventualities described by the accompanying picture. Typically some events are described by both of them, and typically some individuals are described by both of them.



For example, Rosie is depicted in the picture, and mentioned in the text.

Basic problems in the semantics of pictorial+linguistic narratives

1. How to combine information from the two media
2. How to index across the two media –
coindex the nominal phrase *Rosie*
with the depiction of the hen.

Solution from earlier work (part 1)

Use the same possible worlds toolkit for the semantics of pictures and the semantics of sentences.

This makes the two media uniform at the semantic level.
Information from the two media is combined conjunctively.

Abusch 2012

Maier 2019

Rooth and Abusch 2019

Greenberg 2019

Abusch 2021

Abusch and Rooth 2022

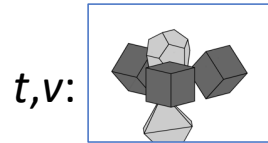
Solution from earlier work (part 2)

Include discourse referents (drefs) in the semantics of the linguistic part.


Introduce drefs in the semantics of the pictorial part.

Index between the two media by equating discourse referents.

Projective model of the semantics of pictures



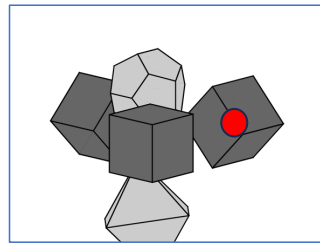
The above is part of a discourse representation.

The semantics is that a described situation (world at a time) looks like picture  at time t from viewpoint v .

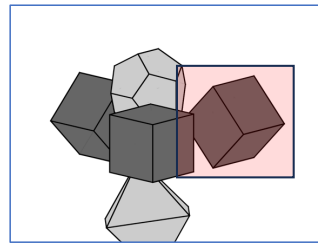
Margaret Hagen, ed. The Perception of Pictures. Vol 1.
Alberti's Window: the Projective Model of Pictorial
Information

G. Greenberg's 2011 PhD dissertation initiated the current
"Supersemantic" project of treating the semantics of pictures
as parallel to possible-worlds semantics for language.

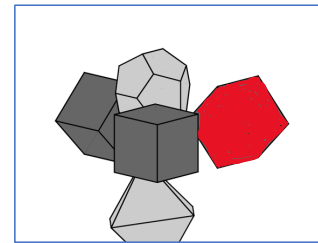
Discourse referents for depicted individuals



point



bounding box



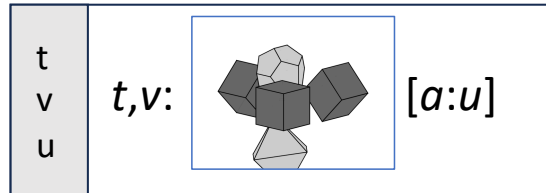
segmentation
map

Use a point in a picture, or
a bounding box, or
or a segmentation map
to introduce a discourse referent for a depicted individual.

Standard practice in AI and machine learning

In supersemantics: Abusch 2012

Notation in a Discourse Representation



Logical syntax	Description
a	A specific point in the picture (the red dot in the previous slide)
u	Dref for the cube
v	Dref for viewpoint determining a point in space and an oriented picture plane
t	Dref for time at which the described world looks like the picture from the viewpoint

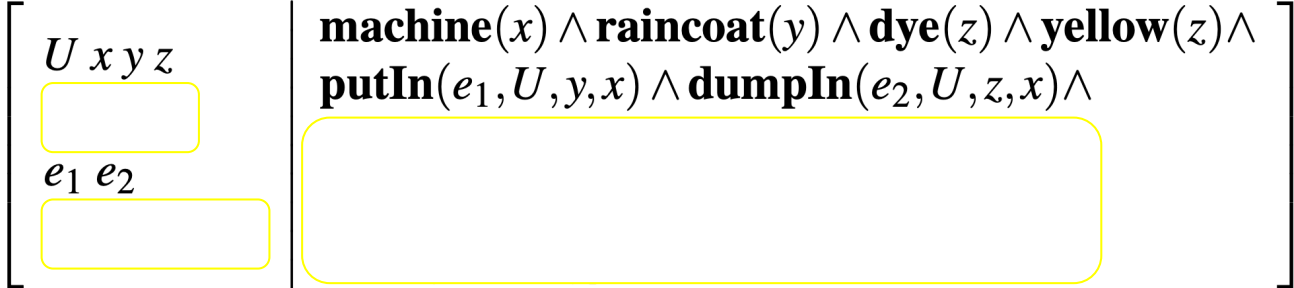
Example



We put the raincoat in the machine and dumped in some yellow dye.

From *Gaspard and Lisa's Christmas Surprise*
(Gutman 1999)

Discourse representation of the linguistic part

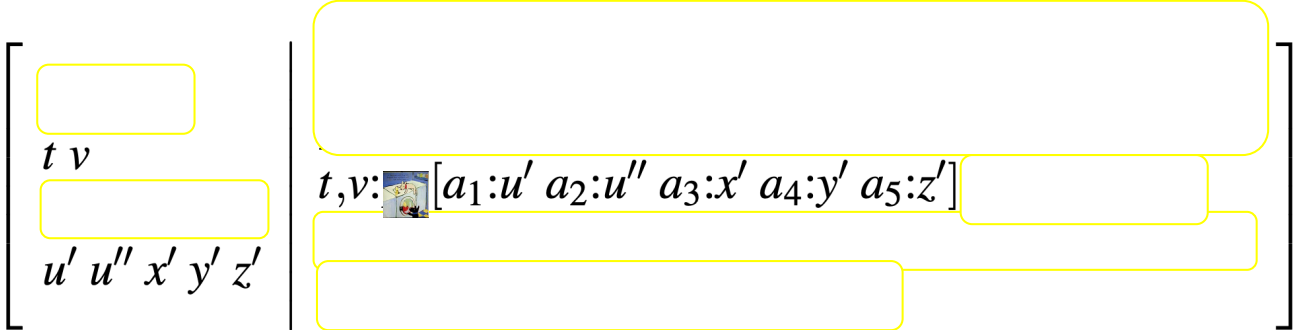


x washing machine from text
 y raincoat from text
 z dye from text

e_1 putting event e_2 dumping event

U we (Gaspard and Lisa)

Discourse representation of the pictorial part



(empty box)

u' Lisa as depicted

(empty box)

v viewpoint for picture

U we (Gaspard and Lisa)

x' washing machine as depicted

y' raincoat as depicted

z' dye as depicted

u'' Gaspard as depicted

(empty box)

t projection time

(empty box)

Combine the parts syntactically

$$\left[\begin{array}{l} U \ x \ y \ z \\ t \ v \\ e_1 \ e_2 \\ u' \ u'' \ x' \ y' \ z' \end{array} \middle| \begin{array}{l} \mathbf{machine}(x) \wedge \mathbf{raincoat}(y) \wedge \mathbf{dye}(z) \wedge \mathbf{yellow}(z) \wedge \\ \mathbf{putIn}(e_1, U, y, x) \wedge \mathbf{dumpIn}(e_2, U, z, x) \wedge \\ t, v: \img alt="A small icon of a machine with a yellow light" data-bbox="330 375 355 405"/> [a_1:u' \ a_2:u'' \ a_3:x' \ a_4:y' \ a_5:z'] \wedge \\ U = u' \oplus u'' \wedge x = x' \wedge y = y' \wedge z = z' \\ t \sqsubset \tau(e_1 \oplus e_2) \end{array} \right]$$


... and coindex across the media

$x = x'$ *equate the mentioned machine with the depicted machine*

$y = y'$ *equate the mentioned raincoat with the depicted raincoat*

$z = z'$ *equate the mentioned dye with the depicted dye*

$$\left[\begin{array}{l} U \ x \ y \ z \\ t \ v \\ e_1 \ e_2 \\ u' \ u'' \ x' \ y' \ z' \end{array} \middle| \begin{array}{l} \mathbf{machine}(x) \wedge \mathbf{raincoat}(y) \wedge \mathbf{dye}(z) \wedge \mathbf{yellow}(z) \wedge \\ \mathbf{putIn}(e_1, U, y, x) \wedge \mathbf{dumpIn}(e_2, U, z, x) \wedge \\ t, v: \text{👤} [a_1:u' \ a_2:u'' \ a_3:x' \ a_4:y' \ a_5:z'] \wedge \\ U = u' \oplus u'' \wedge x = x' \wedge y = y' \wedge z = z' \\ t \sqsubset \tau(e_1 \oplus e_2) \end{array} \right]$$

The picture  is a syntactic part of the DRS. It is not an atomic symbol --- rather the interpretation of the third line in the DRS pays attention to the appearance of the picture.

A useful semantics for the DRS is a relation between a world and witnesses for the drefs --- in this case nine individuals, two events, a time, and a viewpoint. If the drefs are existentially quantified, a proposition (property of of worlds) results. So, a propositional semantics is obtained that combines information from the two media.

Summary

Use the possible worlds toolkit to model meaning of both pictures and language.

Introduce discourse referents in both.

Combine information from the two sources conjunctively.

Express co-indexing across media via equations between discourse referents.

The result is a unitary relation or proposition, with contribution from the two media.

Coming up: the combining of information *goes too far*, because pragmatics needs to access pictorial and linguistic information separately.

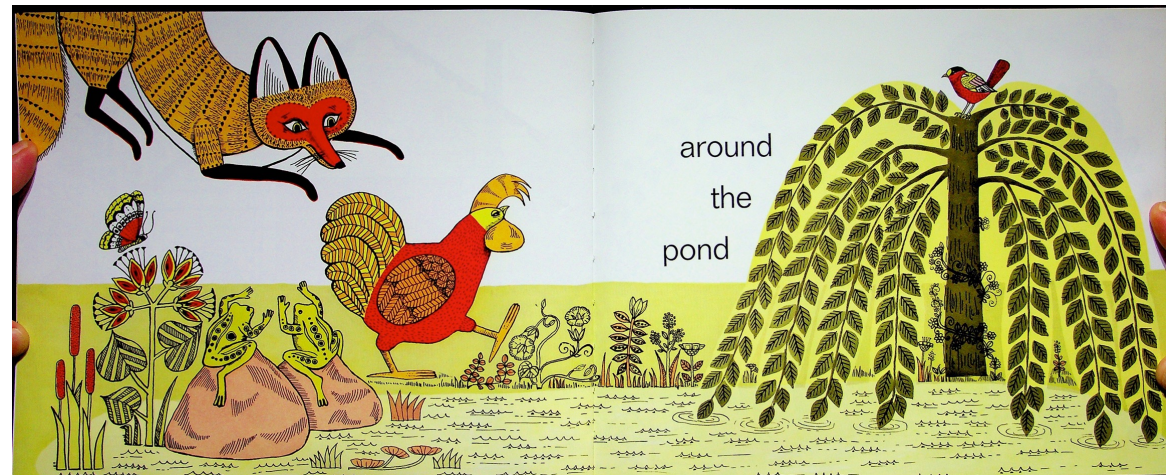
How Picturebooks Work

Nikolajeva and Scott give numerous examples of picturebooks where the pictures and the text tell markedly different stories. The text is understated or incomplete compared to the pictures, leaving out something notable.



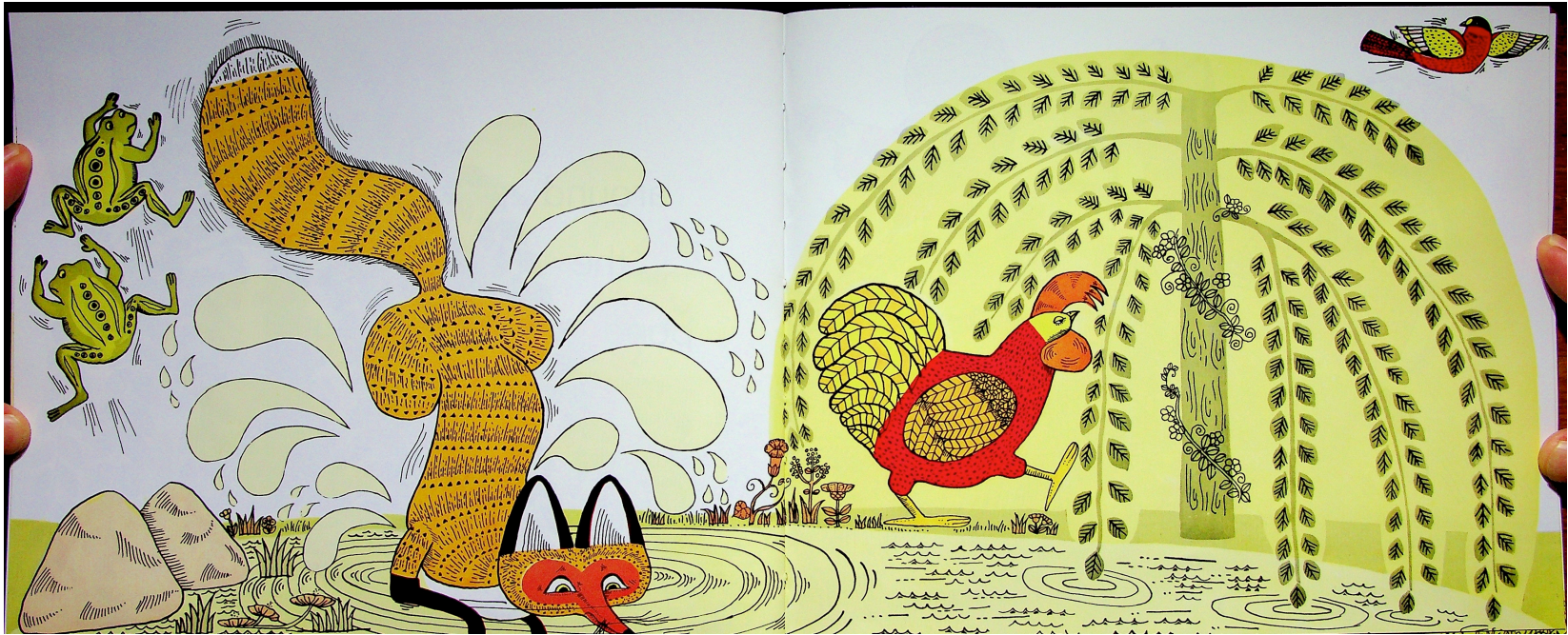
Maria Nikolajeva and Carole Scott

Rosie's Walk
Pat Hutchins



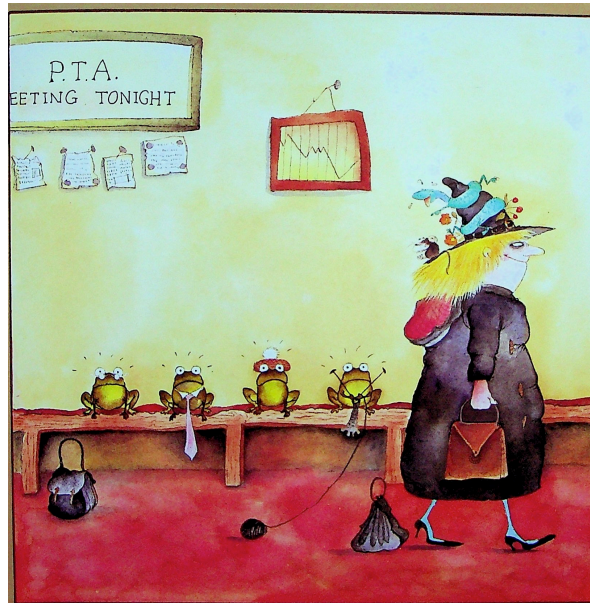
Text describes a hen Rosie taking a walk through a farmyard

Pictures depict a hen taking a walk through a farmyard, and a fox following her



The Trouble with Mum

Babette Cole

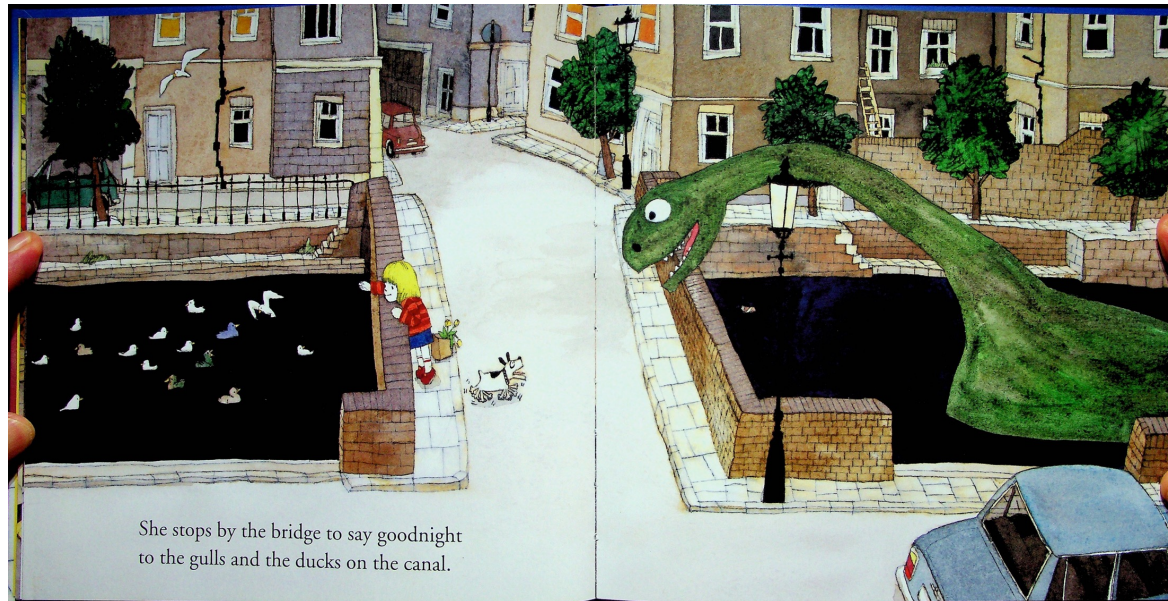


Text describes prosaic interactions and makes prosaic generalizations about a mother and her child's school, the parents of other children, and so forth

She doesn't get along with the other parents.

Pictures depict extreme, unusual and alarming events such as the mother turning other parents into frogs at a parents-teacher meeting

Lily Takes a Walk
Satoshi Kitamura



Text describes a girl Lily and a dog Nicky taking a walk through a city
Pictures depict a girl and a dog walking through a city, and show as well monsters and other odd and alarming things.

In each of the books, the pictures compared to the text have additional dramatic, interesting and alarming information

Rosie

stalking fox

Lily

monsters

Trouble

turning other parents into frogs

Nikolajeva and Scott

“In *Rosie’s Walk*, words and pictures contradict each other. The visual narrative is more complicated and exciting than the verbal one, which comprises a single, twenty-five-word sentence.”

The contradiction is not semantic – certainly there are worlds that are compatible with both the pictorial and the verbal information.



She doesn't get along with the other parents.

As analyzed here, the contradiction is pragmatic --- a certain kind of quantity implicature.

To theorize about this, it is necessary for the pragmatics to have *separate* access to linguistic information.

Separating linguistic and pictorial content



Certainly as readers/viewers we can pay attention to the picture and not the language, or pay attention to the language and not the picture.

Child readers may have access to the picture and not the language. Or have better access to the picture, because language is supplied verbally by somebody else.

Method I

Stick with a DRS that syntactically pools linguistic and pictorial information

$$\left[\begin{array}{l|l} U \ x \ y \ z & \mathbf{machine}(x) \wedge \mathbf{raincoat}(y) \wedge \mathbf{dye}(z) \wedge \mathbf{yellow}(z) \wedge \\ t \ v & \mathbf{putIn}(e_1, U, y, x) \wedge \mathbf{dumpIn}(e_2, U, z, x) \wedge \\ e_1 \ e_2 & t, v: p_1[a_1:u' \ a_2:u'' \ a_3:x' \ a_4:y' \ a_5:z'] \wedge \\ u' \ u'' \ x' \ y' \ z' & U = u' \oplus u'' \wedge x = x' \wedge y = y' \wedge z = z' \\ & t \sqsubset \tau(e_1 \oplus e_2) \end{array} \right]$$

We define separate linguistic, pictorial, and joint contents for it.

- $\llbracket \Phi \rrbracket^L$ Linguistic content of DRS Φ
- $\llbracket \Phi \rrbracket^P$ Pictorial content of Φ
- $\llbracket \Phi \rrbracket$ Pooled content of Φ

Method I continued

$\llbracket \Phi \rrbracket^L$ Linguistic content of Φ

$\llbracket \Phi \rrbracket^P$ Pictorial content of Φ

$\llbracket \mathbf{machine}(x) \rrbracket^P \triangleq \text{True}$ In the pictorial content, trivialize the interpretation of information coming from language. In effect, ignore formulas coming from language.

$\llbracket t, v: \img alt="A small icon of a person sitting at a desk with a computer monitor." data-bbox="198 573 223 603"/> \rrbracket^L \triangleq \text{True}$ In the linguistic content, trivialize the interpretation of information coming from pictures. In effect, ignore formulas coming from pictures.

Summary of Method 1

Recursively build up $\llbracket \Phi \rrbracket^P$ and $\llbracket \Phi \rrbracket^L$, trivializing linguistic formulas in the pictorial content, and trivializing pictorial formulas in the linguistic content.

Method II

Syntactically separate the linguistic content from the pictorial content in the DRS

$$\left[\begin{array}{l|l} U & \mathbf{machine}(x) \wedge \mathbf{raincoat}(y) \wedge \\ x & \mathbf{dye}(z) \wedge \mathbf{yellow}(z) \wedge \\ y & \mathbf{putIn}(e_1, U, y, x) \wedge \\ z & \mathbf{dumpIn}(e_2, U, z, x) \wedge \\ \hline e_1 & \\ e_2 & \end{array} \right] \oplus \left[\begin{array}{l|l} u' & \\ u'' & \\ \hline x' & \\ y' & \\ z' & \end{array} \middle| t, v: p_1[a_1:u' \ a_2:u'' \ a_3:x' \ a_4:y' \ a_5:z'] \right] \oplus \left[\begin{array}{l|l} U = u' \oplus u'' \wedge x = x' \wedge \\ y = y' \wedge z = z' \\ \hline t \sqsubset \tau(e_1 \oplus e_2) \end{array} \right]$$

linguistic content
pictorial content with dref introductions
coindexing

\oplus is dynamic conjunction.

This move provides for a separate DRS constituent for the linguistic content.

Characterizing Understatedness

In the examples gathered by Nikolajeva and Scott, there is a systematic phenomenon of the linguistic material being weak in comparison with the pictorial information. In *Rosie*, the pictures show a fox following the hen, and the words do not mention a fox. In *Trouble*, the pictures show a witch and extreme events including parents being turned into frogs, while the text does not describe such events. In *Lily*, the pictures show the monsters, while the text does not mention them.

So, the text is weak and understated compared to the combination of the text and the pictures. The effect for the adult reader is dry humor.

Overt understatement

Amy and Bill know each other and know that they share aesthetic standards about architecture. They are touring an embarrassingly banal real estate development.

Amy: The architecture is not distinguished.

Implicature: The architecture is banal.

Pictorial understatement

This is a lithograph by the Arctic explorer Nansen. A polar bear is sniffing some ski tracks. We can work out that in the distance, there is a person on skis who is in danger from the bear. Nansen crossed the Arctic on skis, so it could be him.

Many viewers find it humorous in a dry or morbid way. It would be less humorous if the picture showed the skier.



Schematization

There is a weak content W that is presented in a direct way.

There is additional content Q that is presented in a different way.

The combined content $W \wedge Q$ is extreme in a way that W by itself is not, either because the combined content is alarming, or because it expresses a strongly negative sentiment.

As a result, W is understated by comparison with $W \wedge Q$.

	W	Q	$W \wedge Q$
Real estate	it's not distinguished	<i>exclude middle</i>	it's banal
Polar bear	pictorial content	inference about skier	scary pooled information
Rosie's walk	linguistic content	pictorial content	pooled content

In *Rosie's Walk*, the text and pictures are consistent or semantically compatible because we can describe a sequence of events that satisfy both. What is said in the text and what is depicted can happen in one world, where Rosie is walking and the fox follows her. In general, in the examples, W is consistent with Q .

There is however a way of deriving a contradiction at the pragmatic level. The linguistic parts of *Rosie* and *Lily* are prosaic in that they describe an unremarkable sequence of events in which a hen walks through a farmyard, or a girl walks through a town. These prosaic stories can be held to implicate by a process of relevance and quantity reasoning that nothing very remarkable happened during the walk. Let \hat{F} be some additional linguistic information that describes a following fox, while \hat{W} is the original linguistic part. Then $\hat{W} \wedge \hat{F}$ is a linguistic LF that competes with \hat{W} . Given that $\hat{W} \wedge \hat{F}$ was not narrated, this generates the negation of \hat{F} as a quantity implicature.

Since the corresponding content $\neg F$ (entailing that there was no fox) is inconsistent with $W \wedge Q$ (the combined content including the pictorial information about a fox), there is a contradiction at the pragmatic level, when the no-fox or nothing-remarkable quantity implicature is computed from the linguistic part of the story.

Can the informational status found in the *Rosie, Lily, and Trouble* stories be reversed, with the pictures being understated compared to the language?

Verso



Recto



Text

1. *Ray the fox spotted a hen in the farmyard.*
2. *He followed and lunged toward her ...*

Pictures don't depict the fox.

The language does mention the fox.

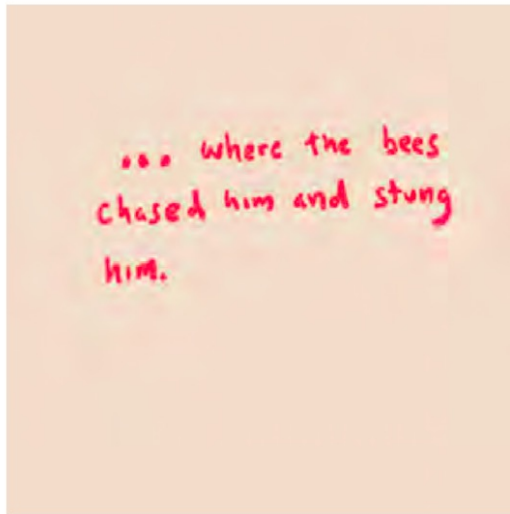
... and landed
on the rake and
banged his head.



3. ... and landed
on the rake
and banged
his head.

4. He followed her
along the pond
and jumped
again ...

Pages about the mishaps have text only.



11. ... where the bees attacked and stung him.

12. The hen is called Rosie. She made it back in time for dinner.

Ray's Chase is symmetric to Rosie's Walk , with the language and not the pictures having information about the following fox.

Ray's Chase is symmetric to *Rosie's Walk* , with the language and not the pictures having information about the stalking fox.

Intuitively I don't think *Ray's Chase* comes across as humorously understated.

In the real estate and polar bear examples, the information W is literal content, and Q is implicated, yielding a stronger conveyed content $W \wedge Q$. The information W is primary because it is literal content. The results for *Rosie's Walk* and *Ray's Chase* suggest the linguistic part of a picturebook is in some sense primary, and the pictorial content secondary, so that only the linguistic part can generate a quantity implicature. Why can't the pictorial part in *Ray's Chase* generate a nothing-dramatic implicature? Because the pictorial part is secondary.

A reason for language being primary is found in the situation of a parent reading a picturebook to a child, where the linguistic material is read out, making it common ground that worlds consistent with the story satisfy the linguistic content. The status of the pictorial information is not the same, because the child needs to seek out pictorial information by looking. The child does not know whether the parent looked at the picture at all, or what parts the parent looked at. Also, children can have different perceptual abilities than adults, so that it cannot be assumed that they will extract the same information when they look. For both reasons, what pictorial information has been picked up by the child and what pictorial information has been picked up by the parent is not common ground between them. This gives the pictorial information a secondary status, comparable to the implicated information in the other cases.

More complex Discourse Representations: Narrated language



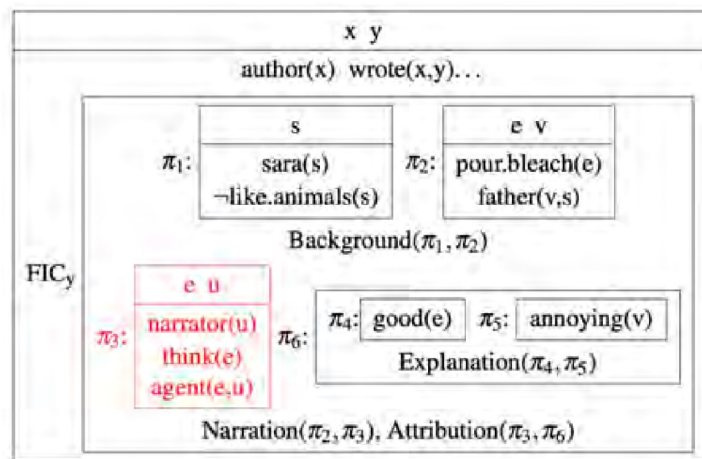
The text part of *Gaspard and Lisa* is construed as first-person narrative. It can be worked out that the narrator is Lisa, the white dog. This is seen in the use of the plural first-person pronoun *we* in the washing machine spread, and elsewhere of first-person pronouns.

The language is in past tense, as if the story were being related retrospectively.

Lisa is as well a character who is referred to with nominal phrases in the linguistic part, and who is depicted in the pictorial parts. In terminology of narrative theory, Lisa is an intradiegetic narrator, a narrator who is an individual who exists in worlds consistent with the narrative.

A standard way of treating this is to introduce narration events in the discourse representation, of which the intradiegetic narrator is the agent. This was developed in a DRS framework by Altshuler and Maier in their study of imaginative resistance (Altshuler and Maier 2022).

Sara never liked animals ... she poured bleach in the big fish tank ... Good thing that she did, because he was really annoying.



From Altshuler and Maier (2022)

$$\left[\begin{array}{l|l} e_4 & \mathbf{narration}(e_4, l, q_4) \\ q_4 & q_4: \left[\begin{array}{l|l} U \ x \ y \ z & \mathbf{machine}(x) \wedge \mathbf{raincoat}(y) \wedge \mathbf{dye}(z) \wedge \mathbf{yellow}(z) \wedge \\ e_1 \ e_2 & e_1:\mathbf{putIn}(U, y, x) \wedge e_2:\mathbf{dumpIn}(U, z, x) \end{array} \right] \\ l & \end{array} \right]$$

e_4 is an event of agent l narrating content q_4

q_4 is described with an embedded box similar to the unembedded box seen earlier

This applies to *The Trouble with Mum*. The ginger-haired narrator is picked out with a first-person pronoun on an early page. The *Mum* of the title implies a first-person perspective.

The DRS representation is parallel to the one on the previous slide, with discourse referents for narration events e_1, e_2, \dots , and for the corresponding linguistic contents q_1, q_2, \dots

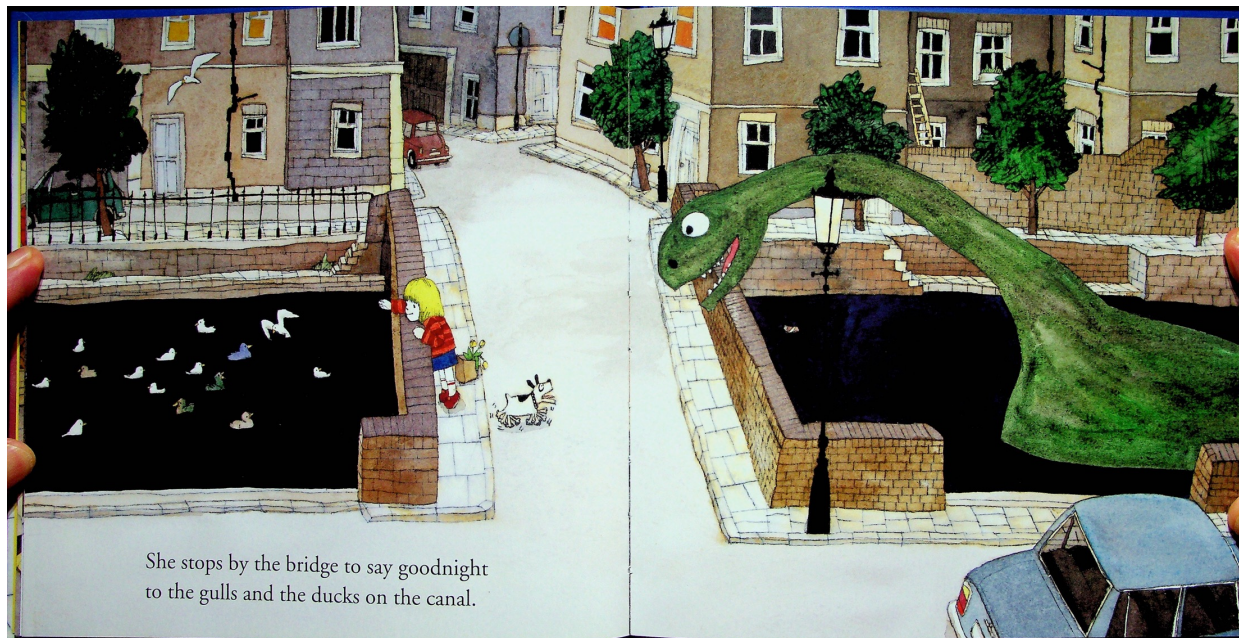
Since the linguistic content is given in discourse referents q_1, q_2, \dots it is possible to reason pragmatically about the linguistic content.

What about books where there is no indication of narration? This is the case with *Rosie's Walk*. Here one can either posit narration anyway, or proceed as earlier with Method I or Method II.

Narratologists usually think that all narratives have narrators. In part it is because there is evidence for it. But in part it could be because it is held that meaning comes from agents. An account using syntax and compositional semantics does not need to say that meaning comes from agents. At the technical level it comes from compositional semantics.

Suppose the method is adopted of always including narration events in the DRSs of fictions. Then should visual information be treated in the same way? Just as events of narrating the linguistic parts are included, events of “narrating” or displaying the pictures could be included, as if the narrator were presenting a slide show and narrating it. (This presents the worry of where the slides come from in worlds where the linguistic material is narrated truthfully.)

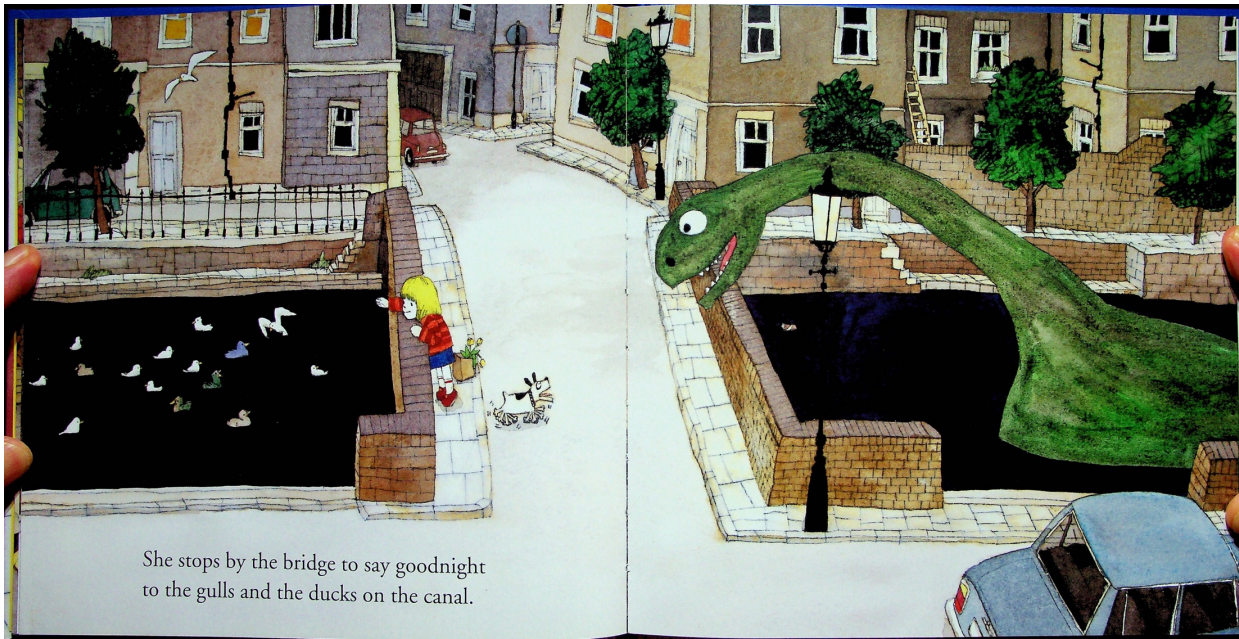
More complex discourse representations: Pictorial embedding



Lily Takes a Walk

If it has a DRS like the one discussed for Rosie's Walk, then there are dinosaurs or monsters in the described situations. In the DRSs discussed earlier, pictures provide extensional information.

Lily Takes a Walk



Alternatively it can be claimed that in a described situation for the panel, the dog Nicky is imagining or non-veridically seeing a dinosaur.

There is literature on this.

From *Gravity*, Cuarón (2013).

Sandra Bullock (Stone) and George Clooney (Kowalski).



Abusch and Rooth (2022): in the discourse representation, the shot is embedded under an imagination or dreaming operator, with agent Stone.



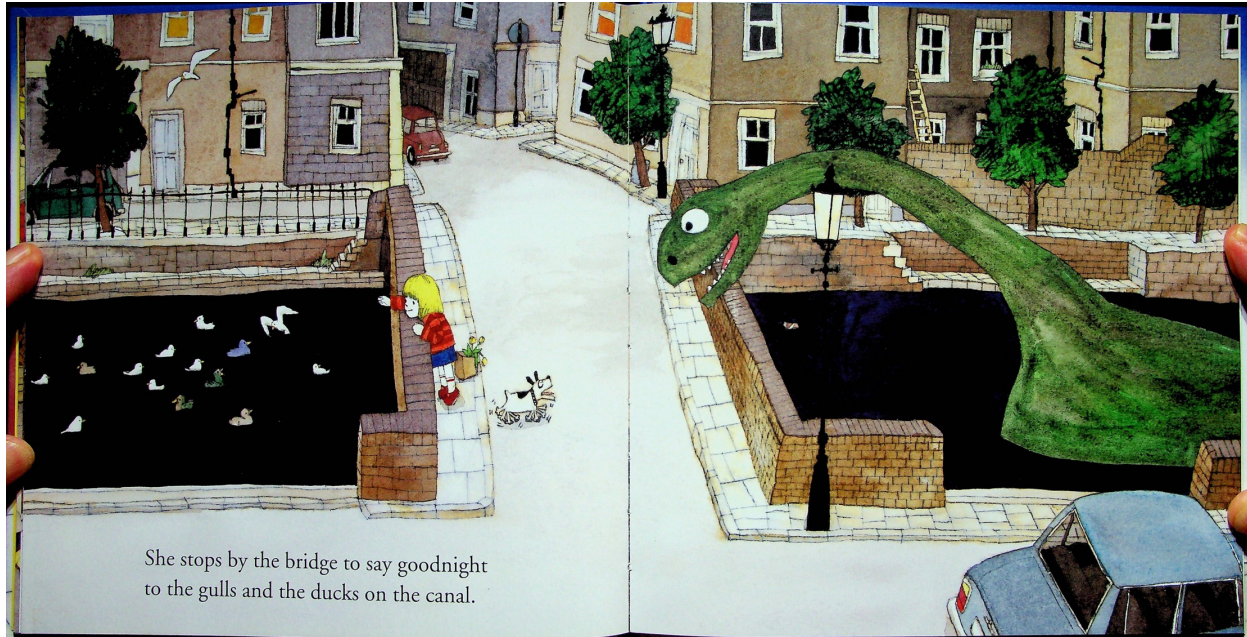
Examples and analyses where apparently *part* of a panel is intensional, and part is extensional. Joe imagines his toys coming to life. From *Joe the Barbarian*, discussed in Bimpikou 2018.

Either split the 2D panel geometrically (Bimpikou 2018) or splice together two 3D scenes geometrically (Abusch and Rooth 2022).



Blood Curse of the Fairies. Discussed in Maier and Bimpikou (2019) and picked up in Abusch and Rooth (2022).

The cone of space in Bart's field of vision comes from an intensional world (Bart's hallucination), while the rest of the space is the base world.



She stops by the bridge to say goodnight
to the gulls and the ducks on the canal.

The monsters are in the dog Nicky's field of view. This makes splitting attractive.

In this case the extensional part of the image is prosaic, together with the language. The intensional part of the image (showing Nicky's perception or hallucination) is extreme.

	Language	Image
<i>Rosie's Walk</i>	Extensional and prosaic	Extensional and alarming
<i>Trouble with Mum</i>	Narrated and prosaic	Extensional and alarming
<i>Lily Takes a Walk</i>	Extensional and prosaic	Extensional and prosaic
		Intensional and alarming

It's interesting that the phenomenon of prosaic language and extreme pictures cuts across the different discourse structures. A child or even an adult reader/viewer might not recover the more complicated discourse structure, so that the books are equivalent as experienced.

Summing up

Information from language and image are combined in a possible-worlds framework, including discourse referents for co-indexing.

For implicatures based on just the language, it is necessary to have separate access to the linguistic content, together with the pooled content. This is straightforward (Method 1 and Method 2).

In *Rosie*, a “nothing alarming” quantity implicature is generated from the prosaic language.

This conflicts pragmatically with the story told by the pictures with the following fox.

But the language and pictures are consistent in the basic semantics.

