Parallel and Differential Contributions from Language and Image in the Discourse Representation of Picturebooks[1]

Dorit ABUSCH — *Cornell University*
Mats ROOTH — *Cornell University*

Abstract. This paper proposes an account in Discourse Representation Theory of children's picturebooks, combining language and image. The focus is on works where the language and image have a different pragmatic status, with the linguistic part of the book being prosaic and understated by comparison with the pictorial part. The effect of wryness and incongruity is analyzed in pragmatic terms.

Keywords: children's literature, discourse representation theory, event semantics, implicature, narration, picturebooks, possible worlds semantics, superlinguistics, temporal relations, understatement

1. Introduction

Children's picturebooks combine language and images, and have narrative structure that involves temporal progression and identification of discourse referents across language and images. This paper formulates discourse representations (DRSs) for common discourse structures in picturebooks, with emphasis on works where the language and the images have a different pragmatic status. To combine information from pictorial and linguistic media, we rely on earlier work that uses a uniform dynamic possible world semantics for language and image (Abusch 2012; Maier 2019; Rooth and Abusch 2019; Greenberg 2019; Abusch 2021; Abusch and Rooth 2022). The two media contribute information that is represented using the same possible-worlds toolkit, and pictorial and linguistic information have nearly the same semantic type. Hence information from the two sources can be combined conjunctively. The dynamic part of the framework includes a mechanism for discourse referents, and so it is possible to index individuals and events across the media. A basic discourse relation between language and image in picturebooks is *co-temporal juxtaposition*, where the eventualities (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.

Differential informational status for language and image in children's picturebooks was studied in Maria Nikolajeva and Carole Scott's *How Picturebooks Work*, referring to a rich variety of examples (Nikolajeva and Scott 2006). Here are three of them. Pat Hutchins's *Rosie's Walk* describes and illustrates a hen Rosie walking around a farmyard (Hutchins 1967). Exceptionlessly, language and image on a page or two-page spread are in co-temporal juxtaposition. The language mentions no threatening events, while images show a fox stalking the hen. See the middle column of (1) for the text, and (2) for examples of complete two-page spreads, sometimes with text and image, and sometimes with an image only. Nikolajeva and Scott comment: "In *Rosie's Walk*, words and pictures contradict each other. The visual narrative is more compli-

[1] We thank participants in Sinn und Bedeutung 27 in Prague for their comments. Thanks also to Masha Esipova and Radek Šimík for comments and assistance connected with the document.

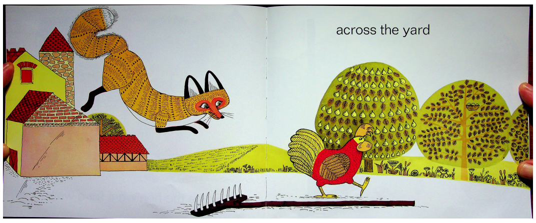
cated and exciting than the verbal one, which comprises a single, twenty-five-word sentence." The impression of the pictures telling a different or markedly extended story is enhanced by every other two-page spread in the central part of the book having no text, and those pages showing the fox suffering some mishap after leaping at the hen, such as in the third spread being banged by a rake.

(1)

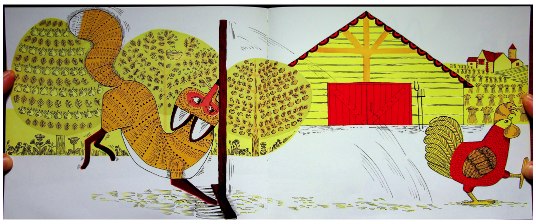
spread	text	mishap for fox
1	Rosie the hen went for a walk	
2	across the yard	
3	*none*	banged by rake
4	around the pond	
5	*none*	lands in pond
6	over the haystack	
7	*none*	sinks in haystack
8	past the mill	
9	*none*	covered by flour
10	through the fence	
11	*none*	lands in wagon
12	under the beehives	
13	*none*	chased by bees
14	and got back in time for dinner.	

(2)

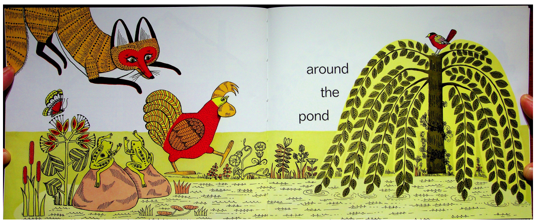
2-page spread



2



3

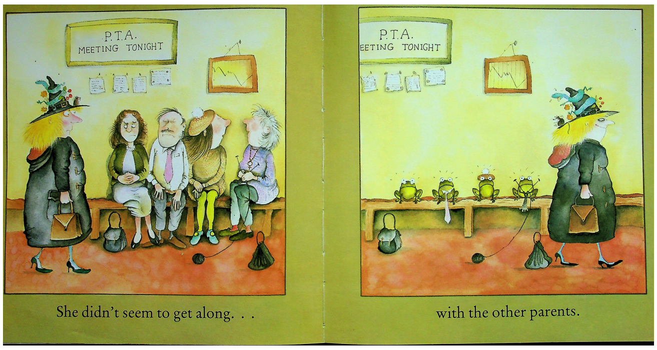


4

Babette Cole's *The Trouble with Mum* is a story with a first-person narrator whose mother is a witch (Cole 1983). This fact is evident in the pictures throughout the book, but not in the language, with the effect that the language is wryly understated by comparison with the images.

In the spread (3), the picture shows Mum having turned the other parents into frogs, while text merely mentions not getting along.

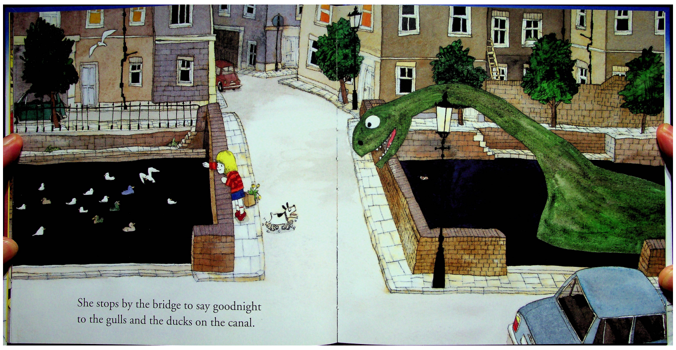
(3)



She didn't seem to get along
with the other parents.

Lily Takes a Walk describes a girl Lily and a dog Nicky taking a walk through a city (Kitamura 1987). The language is prosaic, but the images veer into hallucination, with Nicky seeing monsters, see (4).[2]

(4)



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

The point of the current paper is to formulate the examples brought up by Nikolajeva and Scott in the framework of super-semantics, which applies techniques of possible worlds semantics and discourse representation theory that were developed in linguistic semantics to materials such as comics and film. It will come out that the contradiction in *Rosie's Walk* is pragmatic rather than semantic. Moreover, the three examples introduced above have substantially different discourse representations, which however share the feature of textual information being understated compared to the pictorial information.

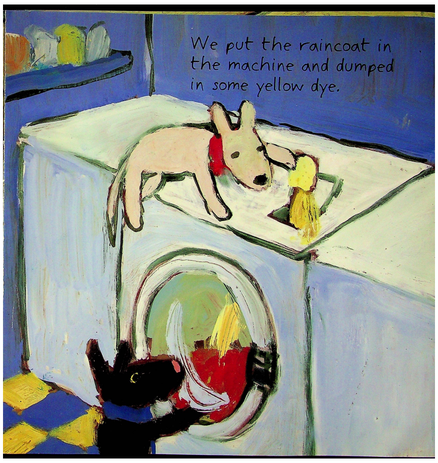
2. DRS framework

Previous super-semantic research on multimodal materials uses a unitary discourse representation based on semantic primitives of worlds, individuals, and viewpoints. Some of this literature uses a linear representation where hidden material is interleaved into a sequence of pictures, in order to cover anaphora, and sometimes with hidden embedding operators included (Abusch and Rooth 2017; Abusch 2021; Abusch and Rooth 2022). Other literature uses the box notation of discourse representation theory (Abusch 2012; Maier and Bimpikou 2019; Maier

[2] Images that are quoted from the cited works are used for educational and critical purposes, and are property of the respective owners.

2019; Schlöder and Altshuler 2022). We follow the second strategy here, formulating logical forms in the formal language of discourse representation theory (Kamp and Reyle 1993). Logical forms are thus discourse representation structures (DRSs), which are syntactic objects that are associated in a grammatically formalized way with information-bearing units such as sentences in story, sequences of shots in film, and juxtapositions of language and pictures in a picturebook. (6) is a simplified DRS for the page (5) from *Gaspard and Lisa's Christmas Surprise* (Gutman 1999). There are discourse referents for two characters, two objects, and two events. Discourse referents coming from language are handled in the standard way of discourse representation theory: nominal phrases introduce discourse referents such as x, y and z, and constraints on them such as $\textbf{raincoat}(y)$. The verbs *put* and *dump* introduce event discourse referents e_1 and e_2, which are incorporated as arguments of the basic relations in the atomic formulas $\textbf{putIn}(e_1,U,y,x)$ and $\textbf{dumpIn}(e_2,U,z,x)$.[3] Turning to visual information, the picture enters into the DRS syntactically, as the picture p_1. In the notation t,v:p_1, picture p_1 is accompanied by a discourse referent t for a time, and a discourse referent v for a geometric viewpoint. The intended interpretation is that t,v:p_1 constrains a described world to look like p_1 from viewpoint v at time t.

(5)



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

(6)

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

x	washing machine from text	x'	washing machine as depicted
y	raincoat from text	y'	raincoat as depicted
z	dye from text	z'	dye as depicted
u'	Lisa as depicted	u''	Gaspard as depicted
e_1	putting event	e_2	dumping event
v	viewpoint for picture	t	projection time
U	we (Gaspard and Lisa)		

[3] An alternative is notation such as e_1:$\textbf{putIn}(U,y,x)$, where an event dref is juxtaposed with a formula that describes it, potentially a non-atomic one. This is what is found in Chapter 5 of Kamp and Reyle (1993).

The complex of information repeated in (7) introduces discourse referents u', u'', x', y', and z' for depicted individuals. For instance x' is a discourse referent for the washing machine as depicted in picture p_1, and u' and u'' are discourse referents for the protagonists Gaspard and Lisa as depicted in picture p_1. Following the approach suggested in Abusch (2012), discourse referents for depicted individuals are introduced geometric points that are within the depiction of the individual.[4] So for instance a_3 in the DRS is a specific geometric point in the two-dimensional picture p_1 that is within the projection of the washing machine in the picture. This constrains a witness for the discourse referent x' to look like the depiction of the washing machine at time t from viewpoint v. In the notation $a{:}d$, a is a specific geometric point such as (0.5,0.5), and d is the discourse referent it constrains. See Abusch (2021) for the formulation in possible worlds semantics of this way of introducing discourse referents for depicted individuals.

(7) $t,v{:}p_1[a_1{:}u'\ a_2{:}u''\ a_3{:}x'\ a_4{:}y'\ a_5{:}z']$

With discourse referents for depicted individuals introduced, they can be equated with discourse referents introduced by language. For instance the equality $x = x'$ at the bottom of the DRS expresses that a witness for the machine mentioned in the linguistic part is constrained to be identical to a witness for an individual depicted in the vicinity of a_3 in the picture. The full set of equalities, repeated in (8), match up the depicted dogs with the group of mentioned dogs, the depicted dye with the mentioned dye, the depicted machine with the mentioned machine, and the depicted raincoat with the mentioned raincoat.

(8) $U = u' \oplus u'' \wedge x = x' \wedge y = y' \wedge z = z'$

A feature of this analysis is that "indexing is analyzed at the semantic level, where the media are not distinguished" (Rooth and Abusch 2019). As a result there is no puzzle of how indexing can cross the boundary between linguistic and pictorial media.[5] More generally, an approach using a unitary DRS for pictures and language integrates information from the two sources. The semantic content of the DRS (6) is a multi-place relation with some argument slots for individuals, some slots for events, one slot for a world, one slot for a time, and one slot for a viewpoint. At this semantic level, there is no distinction between pictorial and linguistic information.

3. Separating linguistic and pictorial content

Given the observations about the differential status of language and image in *Rosie's Walk*, *The Trouble with Mum*, and *Lily Takes a Walk*, the semantic approach from the previous section seems to go too far. If information coming from language and information coming from pictures are integrated into a single DRS, the interpretation of which is a relation constructed in possible worlds semantics, how is it possible to discuss the phenomenon of linguistic information being understated, and how is it possible to analyze it pragmatically? This problem is a

[4] Research in AI and machine vision typically uses bounding boxes or segmentation maps in place of points. See for instance Wang et al. (2016). Abusch (2012) in fact referred to segmentation maps.

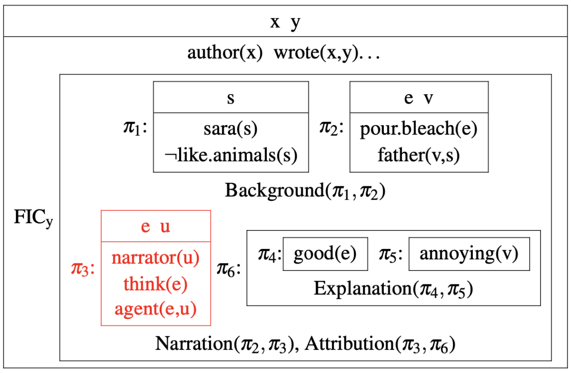
[5] The study in Greenberg (2019) of pictures with localized linguistic tags makes the same point, using a different technical construction.

genuine one. We will end up addressing it by making available separate interpretations of the discourse representation that correspond to linguistic and pictorial content. Before getting into this, it is in order to point out some differences among the picturebooks being discussed.

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.[6] 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. The narrator 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 (Pier 2014). 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). They introduced the DRS (10) for passage (9). The sub-DRS on the lower left that is shown in red describes a thinking event, the agent of which is the narrator u. The sub-DRS on the right describes the content of the thinking event. The two are linked by the formula $Attribution(\pi_3, \pi_6)$. The point of this in Altshuler and Maier's discussion is that the jarring evaluation "good thing that she did ... annoying" is attributed to the narrator, rather than being characterized as true in a described world.

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

(10)



From Altshuler and Maier (2022)

For our purposes the central point is that narration events are included in the DRS. For the linguistic part of example (5), an event discourse referent e_4 is included, the agent of which is the narrator Lisa, as expressed by the formula $\mathbf{narration}(e_4, l, q_4)$. The dref l is the discourse referent for Lisa. q_4 is the narrated content, which is described with an embedded DRS.[7] This results in the structure indicated in (11).[8]

[6] Evidence for identifying the narrator with an individual named "Lisa" is that quoted speech uses the phrases "I" and "Gaspard", but not the phrase "Lisa". Evidence for identifying the name "Lisa" with the white dog is more indirect. Pages with quoted narration where only one dog is depicted tend to show the white dog.

[7] Alernatively one can apply the notation $\mathbf{narration}(e_4)$, $\mathbf{agent}(e_4, l)$, and $\mathbf{theme}(e_4, q_4)$.

[8] This takes the theme of narration q_4 to be semantic. One could alternatively claim that the story presents the linguistic information as being narrated with a specific syntax, including an LF that takes the form of a DRS, say K_4. In this case the theme of the narration should be the syntactic object K_4. The equivalent of q_4 is still available

$$(11)\quad \left[\begin{array}{l|l} e_4 & \textbf{narration}(e_4,l,q_4) \\ q_4 & q_4{:}\left[\begin{array}{l|l} U\,x\,y\,z & \textbf{machine}(x)\wedge\textbf{raincoat}(y)\wedge\textbf{dye}(z)\wedge\textbf{yellow}(z)\wedge \\ e_1\,e_2 & e_1{:}\textbf{putIn}(U,y,x)\wedge e_2{:}\textbf{dumpIn}(U,z,x) \end{array}\right] \\ l & \end{array}\right]$$

On this account, the linguistic part is treated as narrated, in a way that is made explicit in the discourse representation. What about the pictorial part? Nikolajeva and Scott (2006) suggest in one passage that there is an essential difference between pictorial and linguistic parts of picturebooks: "The function of pictures, iconic signs, is to describe or represent. The function of words, conventional signs, is primarily to narrate." Our basic approach does not make this distinction, since the semantics of multi-media constructs is designed to map pictures and language to the the same kind of information. Nevertheless, in works where the linguistic part is narrated, as it clearly is in *Gaspard and Lisa's Christmas Surprise*, it has to be determined whether the pictorial part is narrated as well. If it is not, pictorial information should be entered in the DRS without embedding. (12) adds pictorial information on the second line at the top level, without embedding via an event of narration (or displaying) of the picture. As before, the part beginning with $t,v{:}p_1$ uses the picture p_1 to place a constraint of the appearance of the described world from viewpoint v at time t. This is accompanied by introductions of discourse referents u', u'', x', y' and z' for depicted objects, exactly as before.

$$(12)\quad \left[\begin{array}{l|l} e_4 & \textbf{narration}(e_4,l,q_4) \\ q_4 & t,v{:}p_1[a_1{:}u'\ a_2{:}u''\ a_3{:}x'\ a_4{:}y'\ a_5{:}z'] \\ l & \\ u'\,u'' & q_4{:}\left[\begin{array}{l|l} & \textbf{machine}(x)\wedge\textbf{raincoat}(y)\wedge\textbf{dye}(z)\wedge\textbf{yellow}(z)\wedge \\ U\,x\,y\,z & e_1{:}\textbf{putIn}(U,y,x)\wedge e_2{:}\textbf{dumpIn}(U,z,x)\wedge \\ e_1\,e_2 & x=x'\wedge y=y'\wedge z=z'' \\ & U=u'\oplus u'' \end{array}\right] \\ x'\,y'\,z' & \end{array}\right]$$

Clearly this DRS separates out the linguistic content via the discourse referent q_4. Given this representation, it is possible to reason about the linguistic content separately from the combined content. The linguistic content is simply the information q_4.

On this scheme, the DRS of a picturebook with intradiegetic narration includes a sequence of narration events $e_1,...,e_n$, each of which occurs in any world consistent with the content of the book. These events occur in the same world as the one where (in the Gaspard and Lisa story) there is an event of pouring dye into a washing machine, and which from a certain viewpoint at a certain time, looks like the picture of Lisa pouring dye into a washing machine. The circumstances of these narration events, and where they occur in a timeline, is only weakly constrained. They could be events of Lisa narrating the text to some listener. They could be events of Lisa narrating internally. This indeterminacy is not a defect—the picturebook is simply non-committal about the circumstances and time for the narration events. All that can be said is that the narration events follow the events related in the story, since the narration is in past tense, and that the narration events follow one another in the order $e_1 < ... < e_n$, since default axioms for ordering events apply. This possibility of interpreting tense counts as motivation for the setup with explicit narration events.

The picturebook *The Trouble with Mum* also has intradiegetic narration. The ginger-haired nar-

as the content of K_4.

rator is shown on the first page, introducing the narrator, and the second page is picked out with a first-person pronoun. The way these combine with visual and textual information indicate that the narrator is the child of the depicted woman. The DRS representation is parallel to (12), with discourse referents for narration events $e_1, e_2, ...$, and for the corresponding contents $q_1, q_2, ...$ As before, the linguistic content is separated out with the discourse referents $q_1, q_2, ...$ So it is possible to reason pragmatically about this linguistic content being understated.

What about books where there is no indication of intradiegetic narration? This is the case with *Rosie's Walk*. Here there is the option of positing a DRS without narration events, parallel to (6). If the DRS is like this, the linguistic content is not separately available. We consider two solutions to this. One is to provide separate linguistic and pictorial interpretations of the unitary DRS. The method for this is straightforward. The linguistic interpretation uses True as the interpretation of $t, v : p$, thus trivializing the content of picture p in the linguistic content of the DRS, in effect removing the pictorial content from the linguistic interpretation. Symmetrically, the pictorial interpretation uses True as the interpretation for atomic formulas such as **machine**, thus trivializing the condition derived from language in the pictorial content, in effect removing the linguistic content from the pictorial interpretation. See (13). The semantics $[\![\cdot]\!]^L$ and $[\![\cdot]\!]^P$ above the atomic level is standard. Then $[\![\phi]\!]^P$ is the pictorial content of a DRS ϕ, and $[\![\phi]\!]^L$ is the linguistic content. These are available together with a combined content $[\![\phi]\!]$.

(13) $[\![\mathbf{machine}(x)]\!]^P \triangleq \mathit{True}$
$[\![t, v{:}p]\!]^L \triangleq \mathit{True}$

An alternative is to structure the DRS into several conjunctive parts from the beginning. Consider again the simplified DRS (6) for the page from *Gaspard and Lisa*. It includes some discourse referents and formulas coming from the language, some discourse referents and a pictorial condition coming from the picture, and some equalities that identify discourse referents across language and image representations. (14) divides these structurally into three component DRSs. The parts are combined with an operation written "⊕" of dynamic DRS conjunction, which is treated as semantic. So, instead of a single DRS for a multi-modal page, we postulate a structured DRS of the form $\phi_i \oplus \psi_i \oplus \xi_i$, where ϕ_i is the linguistic DRS, and ψ_i is the pictorial one. This does not change the overall semantics $[\![\phi_i \oplus \psi_i \oplus \xi_i]\!]$. But we postulate that not just the conjoined content $[\![\phi_i \oplus \psi_i \oplus \xi_i]\!]$ is available to pragmatic interpretation, but also the purely linguistic content $[\![\phi_i]\!]$, and the purely pictorial content $[\![\psi_i]\!]$.

(14)
$$\left[\begin{array}{l|l} & \mathbf{machine}(x) \wedge \mathbf{raincoat}(y) \wedge \\ U\ x\ y\ z & \mathbf{dye}(z) \wedge \mathbf{yellow}(z) \wedge \\ e_1\ e_2 & \mathbf{putIn}(e_1, U, y, x) \wedge \\ & \mathbf{dumpIn}(e_2, U, z, x) \wedge \end{array}\right] \oplus$$
$$\left[\begin{array}{l|l} u'\ u'' & \\ x'\ y'\ z' & t{,}v{:}p_1[a_1{:}u'\ a_2{:}u''\ a_3{:}x'\ a_4{:}y'\ a_5{:}z'] \end{array}\right] \oplus$$
$$\left[\begin{array}{l|l} & U = u' \oplus u'' \wedge x = x' \wedge \\ & y = y' \wedge z = z' \\ & t \sqsubset \tau(e_1 \oplus e_2) \end{array}\right]$$

Summing up, whether the linguistic part of a picturebook is represented using narration events or not, it is possible to get access in the DRS formulism to a separate linguistic content. This will be used in the following section.

Above was considered a DRS representation for *Rosie* that included no narration events. An alternative is to include narration events also in stories that are not intradiegetically narrated. This accords with the common assumption in narrative theory that works of fiction are always narrated. And in the philosophy of language, Lewis (1978) proposed that the described worlds for works of fiction include events of the same content being narrated accurately. On this account, the DRS for *Rosie* should also include narration events.[9] This assimilates the DRS of any fiction to the DRS of fictions with intradiegetic narration, with the DRS including narration events.

Suppose the stance is adopted of systematically including narration events in the DRSs of fictions. Is there then justification for treating visual information differently in the DRS of multi-modal artifacts such as picturebooks? 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 with verbal accompaniment. This presents the worry of where the slides come from in worlds where the linguistic material is narrated truthfully.

We prefer to allow for discourse representations where the pictorial information is merely information about the appearance of the described worlds at certain times and from certain viewpoints, and does not imply the inclusion in those worlds of anything like events of subsequently displaying that information, as in a slide show. And for stories as simple as *Rosie*, we are inclined to extend this, with DRSs not representing narration events for linguistic material either.

4. 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 stalking the hen, and the words to 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 of the dog Nicki's imagination, while the text does not mention them.

We reason in this section with the assumption that the semantics makes available a pictorial content $[\![\phi]\!]^P$, a linguistic content $[\![\phi]\!]^L$ and a combined content $[\![\phi]\!]$ for the DRS ϕ mapped from a picturebook, in the way described in the Section 3. $[\![\phi]\!]^L$, $[\![\phi]\!]^P$, and $[\![\phi]\!]$ all have the status of literal semantic contents. We aim at characterizing the pragmatic effect of *Rosie* as being one of understatement in the linguistic part. Here is a case that is in some ways parallel. (15) is a scenario of overt understatement. Suppose A and B know each other and know that they share aestheic standards pertaining to architecture. Let W_{15} be the literal content

[9] Lewis's account is stated as a semantics of the construction "In fiction f, ϕ":

> A sentence of the form "In the fiction f, ϕ" is non-vacuously true iff some world where f is told as a known fact and ϕ is true differs less from our actual world, on balance, than does any world where f is told as a known fact and ϕ is not true.

The worlds referenced in the definition include narration events for f. In our construal, these are split up into narration events for the individual LFs of f.

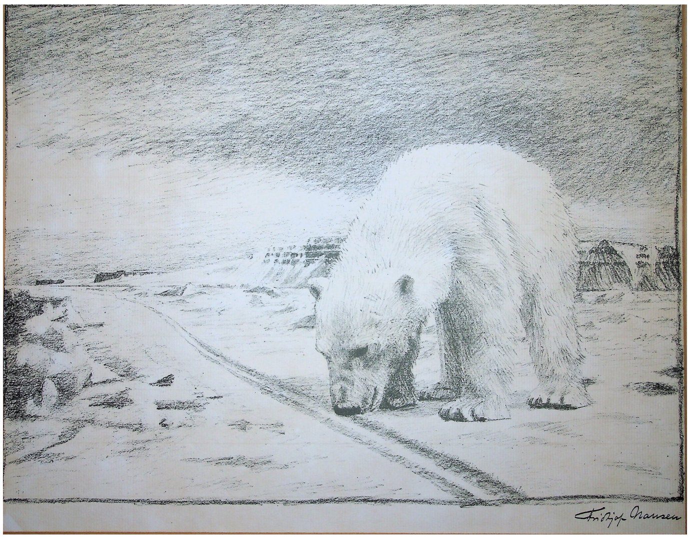
of A's utterance. Asserting W_{15} generates by R-implicature an implicature along the lines of the architecture of the development being painfully banal. Let Q_{15} be this implicature. In this scenerio, the information that the development is banal is available to the speakers from their environment, and Q_{15} is not new information. Instead A's utterance merely thematizes the strengthened information. The effect of wryness is related to the literal content W_{15} being unremarkable, the strengthened content $W_{15} \wedge Q_{15}$ expressing a negative sentiment, and Q_{15} not being directly asserted.[10]

(15) (A and B are touring a blatantly banal real estate development.)

A:	The architecture is not distinguished.	W_{15}
Implicature:	The architecture is banal.	Q_{15}

Here is another case, which is topically and pragmatically similar to the Rosie story, while being purely pictorial. (16) is a lithograph of a polar bear in a snowy landscape, sniffing some parallel tracks in the snow. A viewer works out that the polar bear is stalking or beginning to stalk the human on skis who made the tracks. The effect is ominous, and is more wry and humorous than would be the case if the skier were depicted directly. This is amplified by the information that the artist is the explorer Fridtjof Nansen, who crossed parts of the Arctic on skis, and who therefore can conjecturally be identified with the individual being stalked.

(16)



Fridtjof Nansen
Lithograph, 1922
Nansen International
Children's Centre, Oslo

The information about the skier in (16) is implicated. It is recognized by viewers, and the artist intended for viewers to recognize it. Call this implicated information Q_{16}, and let W_{16} be the basic content of the polar bear lithograph. The combined content $W_{16} \wedge Q_{16}$, is alarming, since it describes a situation where a human is threatened with injury and death. This parallels the alarming nature of the pooled linguistic and pictorial content in the Rosie story, where a hen is threatened with attack by a fox. While the basic pictorial content W_{16} is not prosaic, it is not alarming in the same way.

The examples suggest this schematization. There is a weak content W that is presented in a

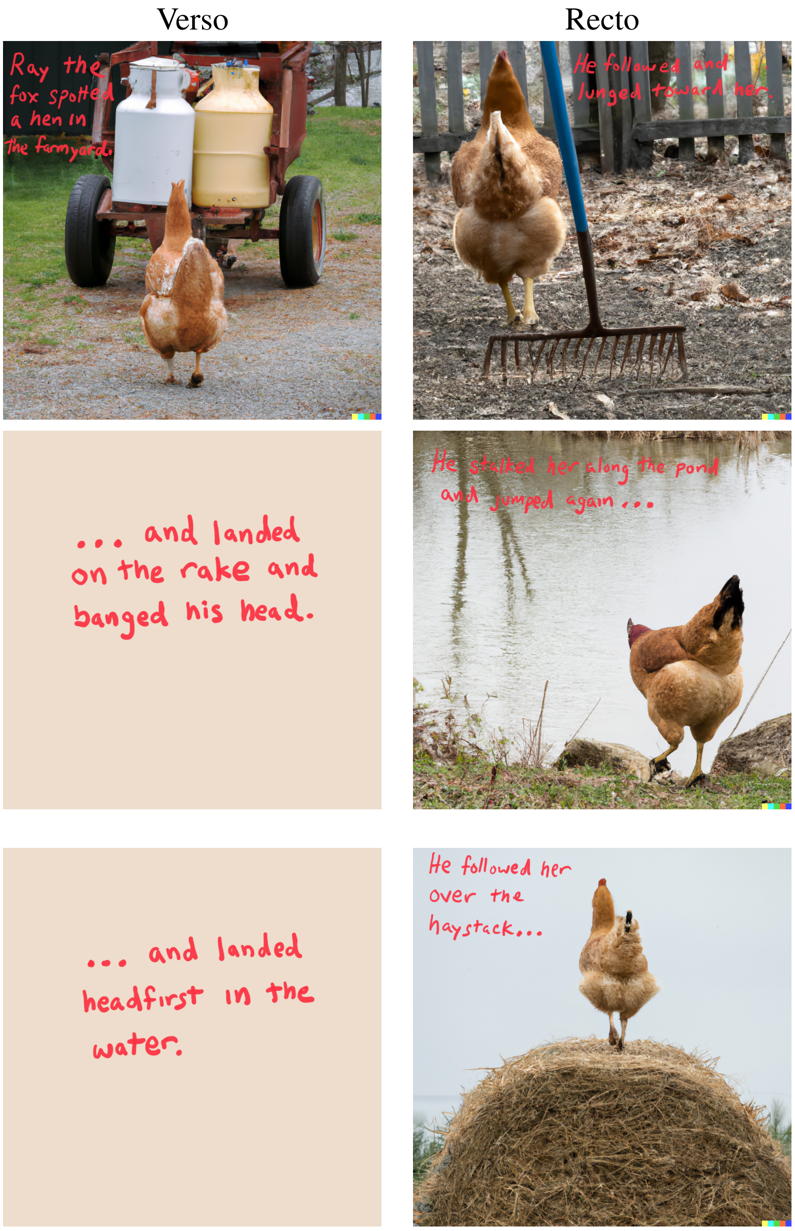
[10] As discussed in Horn (1984, 1989), R-implicature has additional pragmatic functions, including hedging and politeness. In this case A does not wish to hedge the assertion, or to be polite.

direct way, in the picturebooks by the linguistic material, in (15) by A's utterance, and in (16) as the content of the drawing. There is additional content Q that is presented in a different way, in the picturebooks as pictorial content (the fox), in (15) as implicature (the banality) and in (16) as implicature (the skier). 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 $W \wedge Q$.

The passage from Nikolajeva and Scott quoted earlier states that in such cases the linguistic content contradicts the pictorial content. This is not a matter of contradiction in the semantic sense, which would entail that no possible world satisfies both the linguistic content $[\![Rosie]\!]^L$ and the pictorial content $[\![Rosie]\!]^P$ in the case of *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 $\hat{P}$ be some additional linguistic information that describes a stalking fox, while $\hat{W}$ is the original linguistic part. Then $\hat{W} \wedge \hat{P}$ is a linguistic LF that competes with $\hat{W}$. Given that $\hat{P}$ was not narrated, this generates the negation of $\hat{P}$ as a quantity implicature. Then since the corresponding content $\neg P$ (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 nothing-remarkable quantity implicature is computed from the linguistic part of the story.

In the examples (15) and (16), 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, while the additional information Q is implicated. This raises the question whether the linguistic part of a storybook is in some sense primary, and the pictorial content secondary. A reason for this might be found in the situation of a parent reading a storybook 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. Also, children can have different perceptual acuity 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 (15) and (16).

This discussion raises the question whether the informational status found in the *Rosie*, *Lily*, and *Trouble* stories could be reversed, with the pictorial information being understated compared to the linguistic information. We do not know of any examples of this in children's picturebooks. But Figures 1 and 2 present *Ray's Chase*, a constructed inverted version of *Rosie's Walk*, where information about the fox is found in the text and not the pictures. The rhythm is retained, with alternate pages containing only text, and describing the mishaps of the fox, just as alternating two-page spreads of *Rosie* show the mishaps of the fox purely pictorially. Intuitively we think

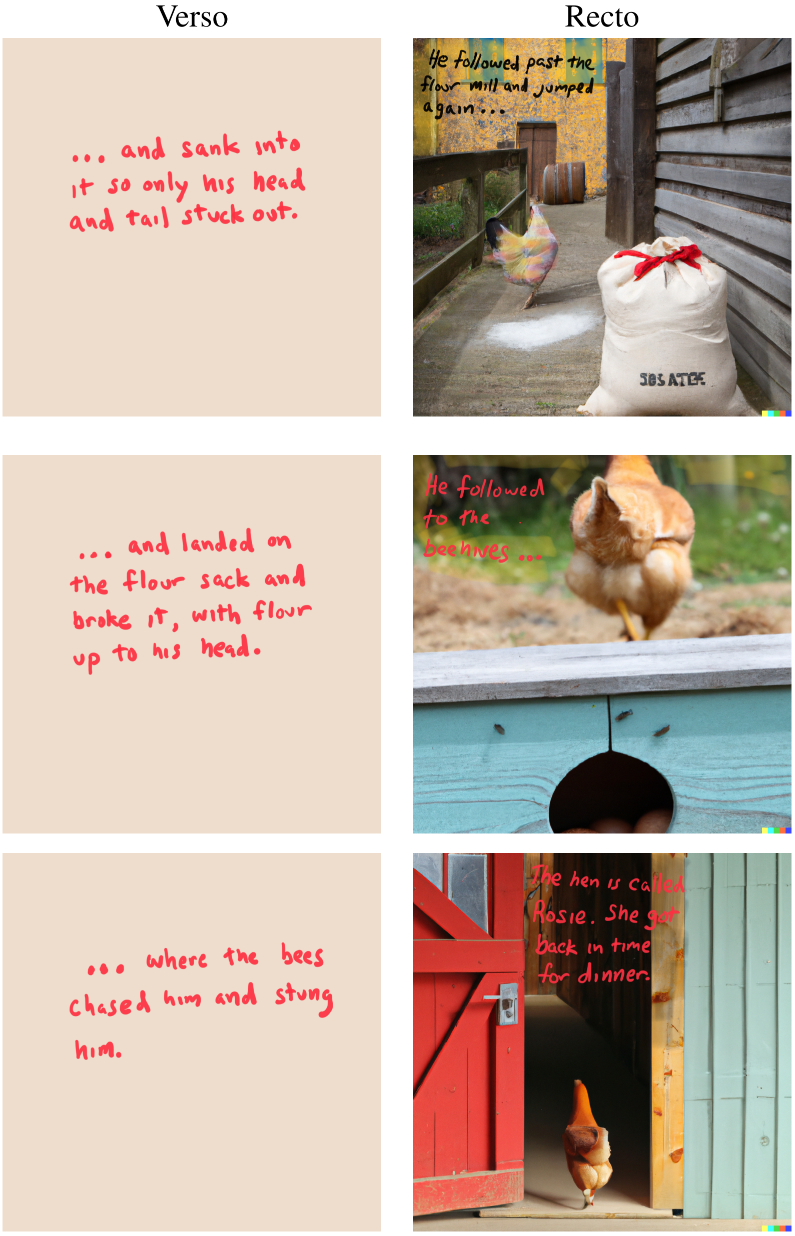

Verso
Recto
Ray the fox spotted a hen in the farmyard.
He followed and lunged toward her.
... and landed on the rake and banged his head.
He stalked her along the pond and jumped again ...
... and landed headfirst in the water.
He followed her over the haystack...

Text

1. Ray the fox spotted a hen in the farmyard.
2. He followed and lunged toward her ...
3. ... and landed on the rake and banged his head.
4. He followed her along the pond and jumped again ...
5. ... and landed headfirst in the water.
6. He followed her over the haystack ...

Figure 1: Pages 1-6 of *Ray's Chase*, a picturebook that reverses *Rosie's Walk* by putting information about the fox exclusively in the text. Starting with page 2, alternating pages have text only, and describe the mishaps of the fox.

that *Ray's Chase* coheres as a narrative. But we think it does not exhibit the wryness and understatement observed for *Rosie*. The pictures function as relatively low-information additions to the verbal narrative, but the low information (not depicting the fox) does not generate an implicature that conflicts with the linguistic narrative. To formalize this, we suggest that while the linguistic content $[\![\phi]\!]^L$ is available by itself for generating implicatures, the pictorial content $[\![\phi]\!]^P$ is not. This might follow from the pictorial information by itself being secondary, in the way discussed above. In *Rosie's Walk* one can get a no-fox implicature from $[\![\textit{Rosie}]\!]^L$, in *Ray's Chase* one cannot get a no-fox implicature from $[\![\textit{Ray}]\!]^P$. This does not stop implicatures from being generated from the *combined* content. In *Rosie* there is an implicature that the hen


Verso
Recto
... and sank into it so only his head and tail stuck out.
He followed past the flour mill and jumped again ...
SBS ATEK
... and landed on the flour sack and broke it, with flour up to his head.
He followed to the beehives ...
... where the bees chased him and stung him.
The hen is called Rosie. She got back in time for dinner.

Text

7. And sank into
 it so only
 his head and
 tail stuck out.

8. He followed past
 the flour mill
 and jumped
 again ...

9. ... and landed
 on the flour
 sack which broke,
 with flour up
 to his head.

10. He followed to
 the beehives ...

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

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

Figure 2: Pages 7-12 of *Ray's Chase*. Images were generated with DALL-E.

does not notice the fox. This is an implicature, because it can be cancelled: a final page could be added that shows Rosie turning around, and announcing "You silly fox, I saw you the whole time. You are wasting your time trying to catch me." The Rosie-did-not-know implicature is generated by the same kind of quantity and relevance reasoning that is outlined above, from the combined content ⟦*Rosie*⟧. While the combined content has information about the fox, it does not have the information that the hen is aware of the fox. Since Rosie being aware of the fox is not narrated, an implicature is generated that she was not aware of the fox.

A complicating factor is that pictures in *Ray's Chase* can be parsed as point of view shots, assuming the geometric visual point of view of the fox. If the pictures are point-of-view shots, this might undermine the relevance of the example. On the analysis from Abusch and Rooth

(2022), point of view shots include in their LF a discourse referent for the viewing agent. The LF from that paper is as in (17), where the picture is embedded under a seeing predicate, and x is the agent. With this LF, pictorial part of the LF includes a discourse referent for the viewing agent. The pictorial part does not identify the viewing agent as a fox, but it does carry the information that the hen is being observed. This is not true of the linguistic part of *Rosie*.

(17) $S_x(p)$

5. Characterizing temporal juxtaposition

This section looks at the discourse relation of co-temporal juxtaposition in picturebooks. In this construction, the language and the image on a single page or a two-page spread describe the same events, in a sense that needs to be clarified. In the DRS, the notation $t,v{:}p$ includes a time dref t, which is taken to be a time point. It is a time when the described world looks like picture p from viewpoint v. Let $e_1,...,e_n$ be the event discourse referents introduced by the linguistic part of a page or spread. Each event e_i has a temporal projection $\tau(e_i)$. Often the interpretation is such that the time t is within one of the temporal projections. An example of this is the page (5) from *Christmas Surprise*, where the discourse representation (6) has an event dref e_1 of putting a shower curtain into a washing machine, and an event dref e_2 of dumping yellow dye into the machine. The picture portrays a time point that is construed as falling within the temporal projection of e_2. This is expressed by the formula $t \sqsubseteq \tau(e_2)$. Another example from the same book is the initial two-page spread, which includes the text “it was almost Christmas”, and shows a street scene with the two dogs near a Christmas display in a shop window. The linguistic material introduces a state discourse referent s, and the time t for the picture is construed as falling within it, $t \sqsubseteq \tau(s)$.

Where $p_1,...,p_n$ are the pictures in the picturebook, they are accompanied in the DRS by projection times $t_1,...,t_n$. All the books we are analyzing seem to satisfy strict temporal progression, $t_1 < t_2 < ... < t_n$. Temporal progression is part of the interpretation of the construction of incrementing an initial sequence of pages or spreads with an additional page or spread.[11] There is an interesting complexity in *Rosie* in the syntax-semantics interface. That story has fourteen pictures $p_1,...,p_{14}$, see the overview in (1). They enter into the DRS as in (18). To this should be added a formula $t_i \sqsubseteq \tau(e_i)$, for the pages 1,2,4,6,8,10, and 14 where there is text. However the linguistic part of the Rosie story has only a single verb, which is on the first two-page spread. It is not implausible though that in the path motion predication, the conjoined path PPs *across the yard* through *under the beehives* introduce motion sub-events $m_1...m_{14}$, together with drefs for component paths of motion $r_1...r_{14}$.[12] Then individual time alignments $t_i \sqsubseteq \tau(m_i)$ can be included.

(18) $$\left[\begin{array}{l|l} p_1\ t_1\ v_1\ ...\ p_{14}\ t_{14}\ v_{14} & t_1 < t_2 \wedge ... \wedge t_{13} < t_{14} \wedge \\ x'_1\ y'_1\ ...\ x'_{14}\ y'_{14} & t_1,v_1{:}p_1[a_1{:}x'_1\ b_1{:}y'_1] \wedge ... \wedge t_{14},v_{14}{:}p_{14}[a_{14}{:}x'_{14}\ b_{14}{:}y'_{14}\ ] \\ u'\ u''\ x'\ y'\ z' & x'_1 = x'_2 \wedge ... \wedge x'_{13} = x'_{14} \wedge \\ & y'_1 = y'_2 \wedge ... \wedge y'_{13} = y'_{14} \end{array}\right]$$

[11] See Abusch (2021) for such a principle applied to linear pictorial narratives. There, strict progression is weakened to temporal non-regression, $t_i \leq t_{i+1}$.

[12] Compare Abusch (2005); Zwarts (2005).

The DRS framework introduced so far does not involve discourse referents for depicted events. For the cases mentioned so far, one can claim that the picture introduces a discourse referent for an event, and that this gets equated with one of the event drefs introduced by the language. This parallels the treatment of individuals, where discourse referents are introduced by both language and image, and the drefs from the two sources are linked up with equalities in the DRS. Many of the events referenced in picturebooks are concrete physical ones, and for these a spatial projection at a time could be postulated. Then discourse referents for events can be introduced by the same syntax as that which introduces discourse referents for individuals. The notation (19) introduces a discourse referent x_e of individual type, and a discourse referent e_v of event type.[13] The interpretation of the second part is that at time t the directed line from v through point a_2 in the picture plane passes through the volumetric spatial projection of event e_v.

(19) $t,v : p[a_1{:}x_e,\ a_2{:}e_v]$

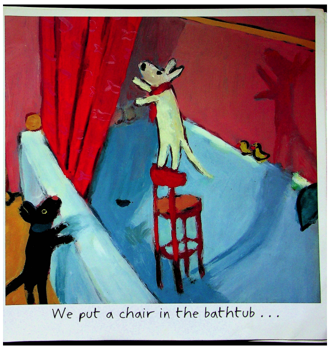
This approach referring to depicted events seems unobjectionable for concrete events. It does involve the complication of the model structure having to specify spatial projections of events. And this specification is potentially redundant. For a concrete event such as Lisa pouring dye into a washing machine, the spatial projection presumably bears a close relation to the sum of volumes of space occupied by Lisa, the dye, and the washing machine at the same time. Already discourse referents for individuals are introduced and identified across language and image, and the event dref with a linguistic source is a co-argument with individual drefs with a linguistic source, in formulas such as $\textbf{dumpIn}(e_2,U,z,x)$ from (6). If z' (the depicted dye) and x' (the depicted machine) are equated with z (the mentioned dye) and x (the mentioned machine) respectively, then z' and x' are co-arguments of event e_2, and are depicted in the picture. This is hard to distinguish from positing an event e_2' that is depicted and equated with e_2.[14]

(20) is another page from *Christmas Surprise*. While the language describes an event of putting a chair in the bathtub, the picture shows the chair in the bathtub. The picture depicts the post-state/result-state of the mentioned putting event. In literature on discourse representation theory, it is common to include result states in the DRSs of sentences with main predicates describing change. A simple move is to include a state dref as an additional argment of the basic relation in a formula like $\textbf{putIn}(e,s,U,y,x)$. Here e is the event argument, s is the result state, and U, y, and x are the individual arguments. This results in a DRS like (21). The temporal dref for the picture, rather than being temporally embedded in the putting event, is embedded in the result state s.

[13] v is the type label for events. This has nothing to do with the discourse referent v for viewpoints. Just as the event discourse referent e has nothing to do with the type label e.

[14] But again, postulating spatial locations for concrete events seems innocuous. On a Davidsonian analysis, it is in fact involved in prepositional location modifiers as in (i).
(i) Lisa danced in the courtyard.

(20)



We put a chair in the bathtub.

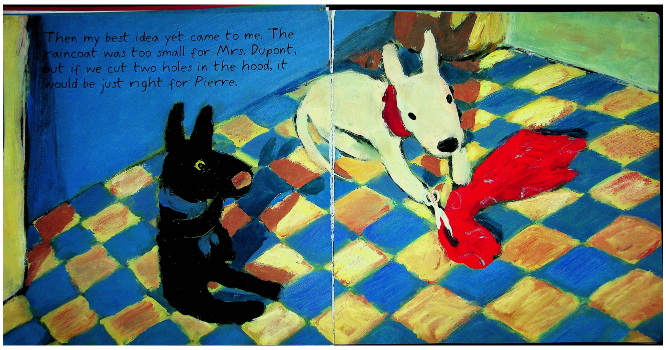
(21)
$$\left[\begin{array}{l|l} U\ x\ y\ z & \mathbf{bathtub}(x)\wedge\mathbf{chair}(y)\wedge\mathbf{dye}(z)\wedge\mathbf{putIn}(e,s,U,y,x)\wedge \\ t\ v & t,v{:}p_1[a_1{:}u'\ a_2{:}u''\ a_3{:}x'\ a_4{:}y'\ a_5{:}z']\wedge \\ e\ s & U=u'\oplus u''\wedge x=x'\wedge y=y'\wedge z=z' \\ u'\ u''\ x'\ y'\ z' & t\sqsubseteq\tau(s) \end{array}\right]$$

So far this section has developed the hypothesis that on a page or spread of a picturebook, the time dref for the picture is temporally embedded in one of the event drefs introduced in the text. How should this be formulated? One option is to include mechanics (perhaps in a feature constraint formalism) for collecting the eventuality drefs $e_1,...,e_n$ introduced by the linguistic material, and to mechanically require that the time dref t for the picture is temporally embedded in one of the e_i. We prefer to state this in a more general way, which anticipates what comes below. The temporal constraint is treated as a presupposition. It involves the time for the picture, and an eventuality dref that, rather than being chosen from the collection of eventuality drefs projected from the linguistic syntax, is an eventuality pronoun that comes with a requirement to find a salient antecedent. Salience is assumed to be modeled as in centering theory, where the context provides a ranked list of available, typed antecedents. The analysis is summarized informally in (22). The notation $e_?$ indicates an eventuality pronoun that needs to find a salient antecedent. In (6), the antecedent is the main event of dumping. In example (21), the antecedent is the result state of the chair being in the bathtub.

(22) $[t,v|t,v{:}p[...]\wedge t\sqsubseteq e_?]$

(23) is another spread from the same book. While the text mentions cutting, this is in a conditional context, and arguably it is in an embedded context of free indirect discourse. This is so because the passage beginning with "the raincoat was too small ..." describes Lisa's thought or statement. It follows that the DRS for the linguistic material does not make available an accessible antecedent for a cutting event. However the passage is interpreted as accommodating the information that Lisa formed the plan to cut holes in the hood, and then executed it. This accommodated information does make an event dref available, which is the antecedent $e_?$ in (22). Here the general formulation has an advantage, since it allows for accommodated material to provide the event antecedent. In fact the presupposition can be seen to contribute to triggering the accommodation.

(23)



Then my best idea yet came to me. The raincoat was too small for Mrs. Dupont, but if we cut two holes in the hood, it would be just right for Pierre.

.

This discussion is continuous with theorization about temporal relations in purely linguistic narratives (e.g. Kamp and Rohrer 1983, Lascarides and Asher 1993, Bittner 2014). What has been said in this section is only a small step in investigating how this should be extended to the case of juxtapositions of language and image in picturebooks. The area of inquiry is fascinating because of the way it ties in with the analysis of language. Importantly, narrative language and images are more in balance than they are in comics and film, so that the linguistic interpretation of narrative can be expected to make as much of a contribution as the pictorial material. Another dimension of the enterprise is the application of discourse relations in the framework of segmented discourse representation theory to pictorial materials and to mixed materials such as picturebooks, an issue studied in Schlöder and Altshuler (2022).

6. Conclusion

This paper has looked at the semantic and pragmatic interpretation of children's picturebooks, in a framework where the information content of both language and pictures is expressed in discourse representation structures that are interpreted in possible worlds semantics. While the discourse representation of a picturebook integrates linguistic and pictorial information, it was argued that the linguistic information was accessible independently to pragmatic interpretation. This requirement was met in a couple of technically straightforward ways. An effect of understatement was attributed to the combined content of a picturebook conflicting with a 'nothing-remarkable' implicature of the text part. Section 5 looked at the discourse relation of co-temporal juxtaposition between information coming from language and image on a single page or two-page spread of a picturebooks. It is common for the time constrained by the picture to be construed as temporally contained in the temporal projection of one of the event discourse referents introduced by the language. But temporal relations can also be mediated by accommodated information.

The semantics and pragmatics of children's picturebooks is an exciting arena for investigation using supersemantic methodology, comparable to comics and film. As illustrated here, semantic modeling using possible worlds, individuals, times, and events, and discourse representation structures is applicable. There is a superb basis of empirical observation and theorization in research on children's literature. To this, supersemantic methodology can contribute formalization of the interface to semantics and pragmatics, modeling of content in possible worlds semantics, careful attention to the distinction between literal semantic content and implicated content and other aspects of pragmatics, and a methodology for representing modality and intensionality.

References

Abusch, D. (2005). Causatives and mixed aspectual type. In G. N. Carlson and F. J. Pelletier (Eds.), *Reference and Quantification: the Partee Effect.* CSLI.

Abusch, D. (2012). Applying discourse semantics and pragmatics to co-reference in picture sequences. In *Proceedings of Sinn und Bedeutung 17.*

Abusch, D. (2021). Possible-worlds semantics for pictures. *The Wiley Blackwell Companion to Semantics*, 1–31. Circulated in 2015.

Abusch, D. and M. Rooth (2017). The formal semantics of free perception in pictorial narratives. In *Proceedings of 21st Amsterdam Colloquium.*

Abusch, D. and M. Rooth (2022). Pictorial free perception. *Linguistics and Philosophy*, 1–52.

Altshuler, D. and E. Maier (2022). Coping with imaginative resistance. *Journal of Semantics 39*(3), 523–549.

Bittner, M. (2014). *Temporality: Universals and Variation.* John Wiley & Sons.

Cole, B. (1983). *The Trouble with Mum.*

Greenberg, G. (2019). Tagging: semantics at the iconic/symbolic interface. In *Proceedings of the 22nd Amsterdam Colloquium*, pp. 11–20.

Gutman, A. (1999). *Gaspard and Lisa's Christmas Surprise.*

Horn, L. (1984). Towards a new taxanomy for pragmatic inferences: Q-based vs. R-based implicatures. In *Georgetown Round Table on Languages and Linguistics.*

Horn, L. R. (1989). *A Natural History of Negation.* CSLI.

Hutchins, P. *Rosie's Walk.* Bodley Head.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic.* Springer.

Kamp, H. and C. Rohrer (1983). Tense in texts. In R. Bäuerle, C. Schwarze, and A. von Stechow (Eds.), *Meaning, Use, and Interpretation of Language*, pp. 250–269. de Gruyter.

Kitamura, S. (1987). *Lily Takes a Walk.*

Lascarides, A. and N. Asher (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy 16*(5), 437–493.

Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly 15*(1), 37–46.

Maier, E. (2019). Picturing words: the semantics of speech balloons. In *Proceedings of 22nd Amsterdam Colloquium.*

Maier, E. and S. Bimpikou (2019). Shifting perspectives in pictorial narratives. In *Proceedings of Sinn und Bedeutung 23*, pp. 1–15.

Nikolajeva, M. and C. Scott (2006). *How Picturebooks Work.* Routledge.

Pier, J. (2014). Narrative levels. *Handbook of Narratology*, 547–564.

Rooth, M. and D. Abusch (2019). Indexing across media. In *Proceedings of 22nd Amsterdam Colloquium.* ILLC, University of Amsterdam.

Schlöder, J. J. and D. Altshuler (2022). Super pragmatics of (linguistic-) pictorial discourse. *Linguistics and Philosophy.*

Wang, M., M. Azab, N. Kojima, R. Mihalcea, and J. Deng (2016). Structured matching for phrase localization. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, pp. 696–711. Springer.

Zwarts, J. (2005). Prepositional aspect and the algebra of paths. *Linguistics and Philosophy 28*, 739–779.