If pictures are stative, what does this mean for discourse interpretation?[1]

Daniel ALTSHULER — *Dept. of Linguistics, Philology and Phonetics, University of Oxford*
Julian J. SCHLÖDER — *ILLC / Dept. of Philosophy, University of Amsterdam*

Abstract. The goal of this paper is to explore the consequences of adopting Abusch (2014)'s hypothesis about pictures—that they are stative depictions of the world—for the interpretation of discourses. We focus on the phenomenon of narrative progression and reject Abusch's proposal that aspectual differences between linguistic and pictorial narrative do not factor in providing a uniform analysis across media. Based on eventive-stative sequences in linguistic narrative, we develop a new answer within Segmented Discourse Representation Theory. We state a single pragmatic algorithm that exploits the aspectual differences between linguistic and pictorial narratives to derive the correct predictions.

Keywords: aspect, temporal anaphora, discourse coherence, picture semantics, SDRT

1. Introduction

Abusch (2014) defends the following hypothesis:

(1) Abusch's Hypothesis: Pictures are stative depictions of the world.

At first blush, (1) may seem wrong. After all, there are many pictures which we would linguistically paraphrase using an event description. Indeed, artists often use conventions to indicate a change-of-state, as in the picture below, which may be paraphrased as 'Some dude threw a water balloon'.

(2)
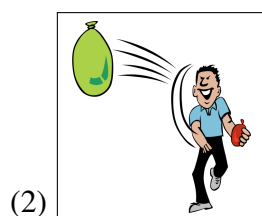
In what sense, then, is this is a *stative* depiction of the world? According to Abusch, whatever change-of-state inferences are made by the viewer, those inferences are pragmatic. For example, in the picture above, we infer that the water-balloon moved from point A to point B. However, according to Abusch, the picture doesn't *semantically* depict this. But, then, what does the picture *semantically* depict?

Following Greenberg (2013) Abusch assumes that pictures have propositional content: Given a picture A and a viewpoint v, $[\![A]\!]^v$ is the set of scenes σ that project to A relative to v (see Figure 1 below).

This assumption allows Abusch to motivate her hypothesis in (1) using linguistic tests for stativity. Here we mention one of these tests, from culminativity, which Abusch paraphrases in the form of the argument below, in Figure 2.
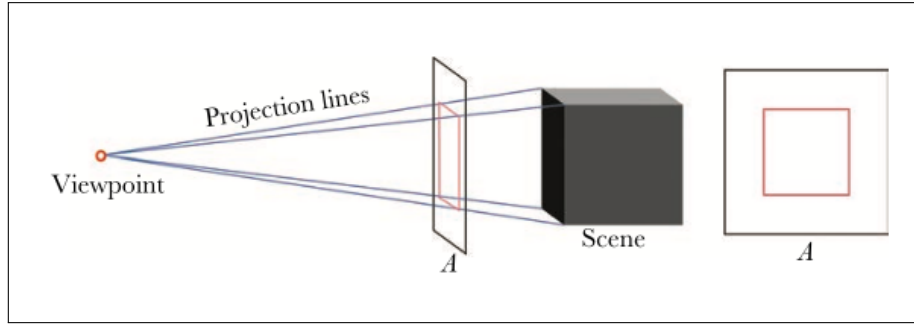
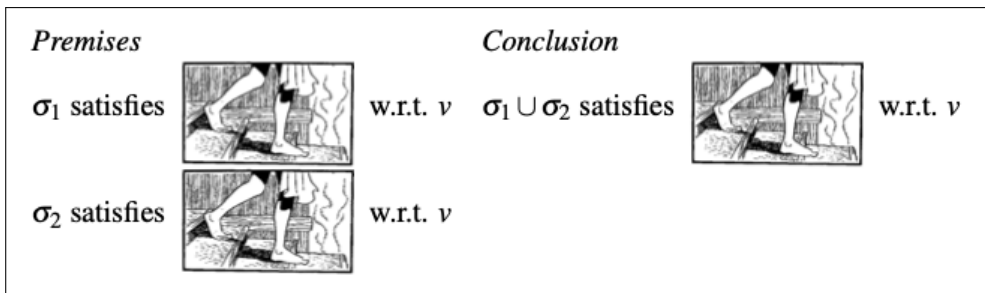Figure 1: Scenes and viewpoints. Graphic from Greenberg 2013.

Figure 2: Pictures are cumulative. Graphic from Abusch 2014.

Assume that time is discrete, and that σ_1 and σ_2 are instantaneous situations in immediate succession that satisfy the same picture A. Given this assumption, the argument shows that A is cumulative, which is a defining property of statives (Krifka 1989). Or, to put the point in simpler terms, as any depiction is compatible with there being no change of state, any interpretation involving a state change is inferred, not supplied by the semantics.

In what follows, we will be not be concerned with whether the argument above is a good one, and, indeed, whether Abusch's hypothesis is well-motivated. Instead, the goal of this paper is to assume that Abusch's hypothesis is right, and explore the question in (3).

(3) If pictures are stative, what does this mean for the interpretation of a discourse?

We will see that there are subtle differences between the interpretations of pictorial narratives and (arguably) analogous linguistic narratives. If Abusch's hypothesis is correct, we can provide a formal pragmatic algorithm to infer the proper interpretations in either medium.

This paper proceeds as follows. In the next section, we consider Absuch's answer to (3) with respect to narrative progression. If pictures are stative, she argues, then a uniform analysis of narrative progression across media does not involve aspectual information. That is, aspect does not contribute to inferences about narrative progression. Afterwards, we consider counterexamples to this answer that involve eventive-stative sequences in linguistic narrative (discussed by Altshuler 2021) and (arguably) analogous sequences of pictures. These counterexamples will motivate a new answer, which exploits aspectual difference between linguistic and pictorial narrative to provide a single pragmatic algorithm that makes the correct predictions across media. Our algorithm is buttressed on the tools of Segmented Discourse Representation Theory (Lascarides and Asher 1993; Asher and Lascarides 2003) and we describe how we will adopt

and extend those tools in §3. In §4 we show how our algorithm can explain the core data, before considering challenges for our analysis in §5 from data involving exclamatives, deverbals and evalutative statives. We provide some possible avenues for addressing these challenges and then conclude the paper in §6.

## 2. Interpretation of statives and eventives in discourse

Abusch (2014: 25) proposes that (i) pictorial narratives are subject to fixed rules that force pictures to be understood in succession and (ii) common sense pragmatics can "extend" a state backward in time to infer temporal overlap. The latter part of the proposal follows Dowty (1986)'s influential analysis of narrative progression in linguistic discourse, allowing Abusch to conclude that same rules are operative in pictorial and linguistic narrative. In what follows, we consider some examples that Abusch uses to motivate this view (§2.1), before considering potential counterexamples (§2.2).

2.1. Motivating a Dowty-style analysis

Consider the following pair of pictorial narratives, (4) and (5), also discussed by Abusch (2014). They illustrate that regardless of the order in which two pictures are presented, the cause-effect interpretation remains constant: a bird was kicked and, as result, it fell down the cliff. This is expected on a Dowty-style analysis of narrative progression, according to which eventuality sequencing is the default and common sense pragmatics determines the nature of the sequence.

(4)

(5)
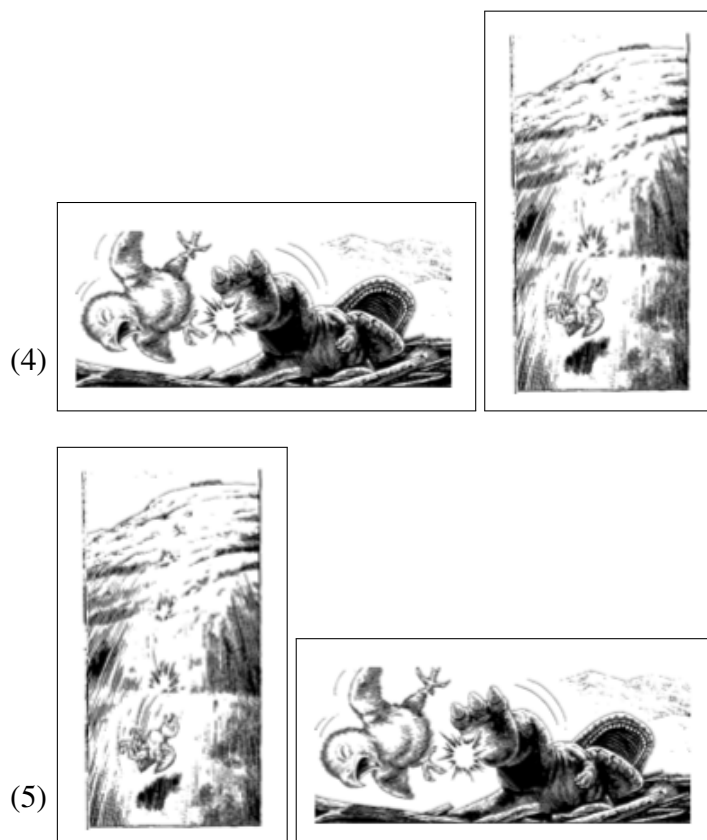
Now consider the sequence of pictures in (6), also taken from Abusch (2014), where we infer that an individual was wearing glasses *while* the sun was out. That is, the situations depicted in the two pictures temporally overlap.

(6)
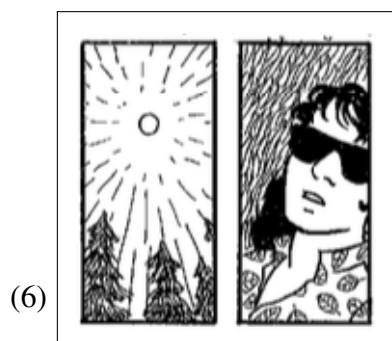
This overlapping interpretation is also expected on a Dowty-style analysis. Abusch suggests that by default, we read pictures in succession. Here this means that the sun was out before the individual wore glasses. This is compatible with the sun continuing to be out while the individual wore glasses. Since this interpretation is the most plausible one (given common sense reasoning), this is what is inferred. Note that although this overlapping interpretation is what one immediately obtains when confronted with (6), others are (in principle) compatible with the information presented in (6), e.g. that the sun set before the individual put on glasses. The possibility to assign such interpretations as well appears to underwrite the role of common sense when interpreting (6).

Of course, Dowty's analysis was originally motivated by narrative progression in *linguistic* narratives. The examples in (7), (8) and (9) illustrate temporal succession, temporal precedence and temporal overlap respectively.

(7)   I threw a water balloon at Hans. His shirt got wet.

(8)   I arrived to Zoom late. I was cooking.

(9)   I put on my sunglasses. The sun was out.

Again we can infer by common sense reasoning that throwing a water balloon precedes Hans's shirt getting wet; cooking is presented as an excuse for being late, and thus precedes the lateness; and that the sun being out temporally overlaps one putting on their sunglasses.

The parallel between such and similar data (on the one hand) and the pictorial data above (on the other hand) led Abusch to conclude that narrative progression rules are the same across media, even if one medium lacks event descriptions. This entails that Abusch, like Dowty, denies the *Aspect Hypothesis* defended by Kamp and Rohrer, 1983.

(10) Aspect Hypothesis: Aspectual information partially determines narrative progression: states are typically understood to overlap prominent discourse events.

In the next section, we see examples of linguistic narrative that suggest reviving (some version of) this hypothesis. This prompts us to rethink Abusch's uniform treatment of linguistic and pictorial narrative.

2.2. Reviving the Aspect Hypothesis
Consider the linguistic narrative below in (11). As noted by Altshuler (2021), event-state sequences (ESSs) of this kind cannot have a causal interpretation. To the extent that this sequence could be interpreted, it could only mean that the speaker threw a water balloon at Hans when his shirt is already wet.

(11) I threw a water balloon at Hans.
      %His shirt was wet. (Altshuler, 2021)

In other words, this ESS respects the hypothesis in (10) despite what common sense pragmatics tells us. To be convinced of this, compare (11) with (7) from the previous section. Based on common sense alone, both examples should lead us to infer a causal relation and in particular that the throwing of the water balloon precedes the wetness of the shirt. But while this is the natural interpretation of (7), this is not so for (11). When reading (11), one prefers an overlapping interpretation: that the shirt was already wet.

In light of this contrast, note that the overlapping interpretation is not possible in (12). The causal interpretation is the only one available, just like common sense pragmatics tells us.

(12)

Thus, there is a difference in the interpretation of linguistic and pictorial narratives. As far as common sense is concerned, the linguistic narrative (11) and the pictorial narrative (12) appear to confront us with the same information, yet we seem to interpret them differently. This puts pressure on the Dowty-style analysis of narrative progression.

An even more striking contrast is found below in (13) and (14).

(13) A cat bit a mouse while the mouse was wiggling its tail.
      #It was dead. (Altshuler, 2021)

(14)
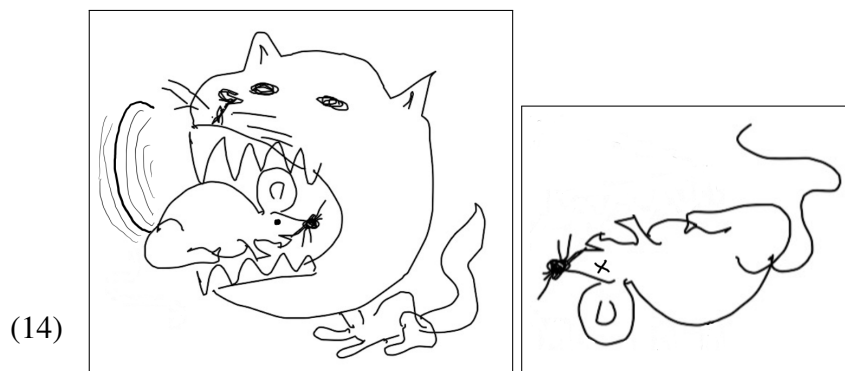
The Aspect Hypothesis in (10) can explain the infelicity of (13). If the hypothesis is correct, we infer that the mouse being dead overlaps the event of it being bitten while wiggling its tail. But then there is no pragmatic context that could make this ESS felicitous because in any such context it would (absurdly) be the case that the mouse wiggles its tail while dead. This is so despite what common sense reasoning would suggest: that the mouse died *because* a cat bit into it (i.e. a causal interpretation on which the being-dead succeeds the biting). This causal interpretation is available (and indeed the only available one) in the pictorial narrative in (14).

These contrasts may suggest that we need to posit two different pragmatic analyses across the two media. If Abusch's Dowty-style analysis of pictorial narratives is correct and the examples

in (11) and (13) do vindicate the Aspect Hypothesis for linguistic narratives, then different principles govern narrative sequencing in these media. This, however, would be unsatisfying. We know of no reason why our assessment of basically the same pieces of information should fundamentally vary with whether that information is presented linguistically or pictorially. But then how can it be that linguistic and visual narrative differ with respect to the availability of causal inferences?

The next two sections are an attempt to answer this question. In particular, we propose an SDRT analysis according to which both Abusch's Hypothesis in (1) and the Aspect Hypothesis in (10) are correct. This allows us to exploit aspectual differences between linguistic and pictorial narrative to provide a single pragmatic algorithm that makes the correct predictions for the data considered thus far.

3. Towards an SDRT analysis

Our plan crucially involves the notion of a *coherence relation* (Hobbs, 1985; Kehler, 2002; Asher and Lascarides, 2003). The guiding idea is that clauses/sentences compose to *narratives* that convey more information than their parts—just like subclausal units compose to meaningful clauses that contain more information than the sum of their parts. We will say that clauses/sentences *cohere* with one another to form a narrative. In this section we discuss a version of Segmented Discourse Relation Theory (Asher and Lascarides, 2003) that we then apply to pictorial discourse

3.1. Coherence Relations

The first component of a formal theory of coherent narrative is a vocabulary of coherence relations that specify the different ways in which clauses/sentences can cohere with one another. For present purposes it is important to note that which coherence relation is associated with a sequence also determines the temporal sequencing of the resulting narrative (Lascarides and Asher, 1993). The following are cases in point.

(15) a. Max fell,
    b. **because** John pushed him. ⎤—Explanation (temporal precedence)

(16) a. Max fell,
    b. **so** John helped him up. ⎤—Result (temporal succession)

(17) a. Max fell,
    b. **while** John was away. ⎤—Background (temporal overlap)

In all three cases, the two clauses are linked by a cue phrase that indicates a coherence relation. Consider (15). The presence of *because* establishes that the *pushing* causes the *falling*. We indicate such causal relationships with the coherence relation *Explanation*. This is to say that (15) expresses a (small) narrative in which two events are reported (the pushing and the falling) and put in a particular relation: that one event causes the other. As causes must precede effects, we know that the pushing preceded the falling. Thus although in (15) the falling is *described* before the pushing, the fact that the two clauses relate by *Explanation* entails that the falling *happened* after the pushing.

Matters are similar in (16) and (17). In (16), the cue phrase *so* indicates that the *falling* caused the *helping up* which is indicated by the coherence relation *Result*. In this case, the order of

description matches the order of events, again due to the fact that causes must precede results. In (17) we see a third option for the order of events: overlap, as cued by the phrase *while*, which indicates the coherence relation *Background*. In this case, there is no causal relation. The falling was not caused by the being-away or *vice versa*. Instead, the being-away is presented as supplemental ('backgrounded') information. The relevance of this information may only become apparent once the discourse continues.

We can formalize the just noted analyses as follows. We assign to clauses their usual dynamic meaning (i.e. a context change potential, CCP) in a Neo-Davidsonian event semantics (Davidson, 1967; Parsons, 1990). For example, (15a) and (15b) are interpreted as follows. (For simplicity, we suppress here the presuppositions triggered by proper names and the contribution of the simple past on the temporal ordering.)

$$[\![(15a)]\!] = \exists e_a, t_a \, fall(e_a) \wedge at(e_a, t_a) \wedge agent(e_a, m) \wedge Max(m).$$

$$[\![(15b)]\!] = \exists e_b, t_b \, push(e_b) \wedge at(e_b, t_b) \wedge agent(e_b, j) \wedge theme(e_b, m) \wedge John(j).$$

We will refer to such interpreted forms of individual clauses in a narrative as the *segments* of the narrative (sometimes these are also called *discourse units*). Observe that these logical forms do not yet contain the information contributed by *because*, namely that e_b is the cause for e_a (and hence in particular that e_b precedes e_a). This information is obtained by assigning a dynamic *meaning postulate* to the coherence relation *Explanation*. Meaning postulates map clauses to a CCP that is defined in terms of the CCPs of the clauses.

$$[\![Explanation]\!](c_1, c_2) = [\![c_1]\!] \wedge [\![c_2]\!] \wedge cause(e_2, e_1) \wedge t_2 \prec t_1.$$

Where c_1 and c_2 are clauses, e_1 denotes the main eventuality described in c_1, t_1 denotes the time index of the main eventuality in c_1 and analogous for e_2, t_2 and c_2. In (15), the main eventualities/time indices of the two clauses are e_a/t_a and e_b/t_b, respectively, so interpreting (15) as *Explanation*(15a,15b) delivers the desired interpretation as the event described in the latter clause preceding and causing the event described in the former.

Coherence relations fall into two broad classes: *subordinating* relations and *coordinating* ones.[2] The intuitive distinction is that subordinating relations add further information to an event that is already under discussion (e.g. by introducing a sub-event or adducing further properties of the event, its agents, themes or sub-events) whereas coordinating relations 'move the narrative onwards' to a new event under discussion. Roughly put, coordinating relations move the narative to a new scene whereas subordinating relations flesh out the current scene.

Of the relations seen so far, *Explanation* and *Background* are subordinating (they add an explanation or supplemental information) and *Result* is coordinating (it moves the narrative to a new event: the effect of the current one. Other subordinating relations include *Elaboration* (adducing a sub-event) and other coordinating ones include *Narration* (moving to a new event that is temporally close to the previous one) and *Continuation* (moving to a new event that is thematically related to the previous one). See the Appendix of Asher and Lascarides, 2003 for a list of coherence relations and their meaning postulates.

---

[2]This distinction goes back to Hobbs (1985) and is adopted in various domains, e.g. to explain extraction phenomena (see, e.g. Deane 1991 and Culicover and Jackendoff 1997). However, Hobbs' notion is distinct from the SDRT notions that are central here. For more discussion, see Asher and Vieu 2005; Altshuler and Truswell 2021.

As a usual shorthand, it is customary to graph the coherence structure of a narrative by letting horizontal lines represent coordinating relations and vertical lines represent subordinating relations. The following toy example illustrates this.

(18) a. John had a great meal.
    b. He particularly liked the cheese platter.
    c. Then he went dancing.

(18a) —— *Narration* —— (18c)

$\vert$ *Elaboration*

(18b)

As observed by Hobbs (1985) the distinction between coordinating and subordinating relations is important for the interpretation of anaphora. When extending a narrative with a clause containing an anaphor, its only available attachment sites are in those segments that accessible from the last segment in the narrative by traversing subordinating relations—but binders behind coordinating relations are inaccessible. If one uses the above notational conventions, then the accessible segments are exactly the segments on the right-most branch of the graphed narrative structure. Hence this constraint is known as the *Right Frontier Constraint* (Polanyi 1985; see Hunter and Thompson 2021 for recent discussion).

3.2. Inferring Relations

The examples discussed in the previous subsection (in (15)–(18)) all contain cue phrases that allow us to determine an associated coherence relation (e.g. *because* cuing *Explanation* and *particularly* cuing *Elaboration*). But in many cases we need to determine the correct coherence relation without explicit cuing. For example, dropping the cue *because* from (15) results in the example (19) which is still most naturally interpreted with *Explanation*.

(19) a. Max fell.
    b. John pushed him.]-Explanation (temporal precedence)

The coherence relation makes (again) visible that the *pushing* happened before the *falling*, despite the *falling* being described earlier in the discourse. Note, however, that this interpretation of (19) is merely the most natural one, by which we roughly mean the interpretation produced as a first-glance assessment of the discourse. What is the most natural interpretation is subject to revision by further context. One can, for example, continue (19) with *But this is not why he fell* (cancelling the reading as *Explanation*) or extend (19b) with *while he was on the ground* to establish temporal succession.

Moreover, there is not always *the* most natural interpretation. The following example appears to be multiply ambiguous.

(20) a. Amy left.
    b. Lisa cried.]-Result? Explanation? Narration?

Depending on what is known about Amy and Lisa, it could be that Amy leaving caused Lisa to cry (*Result*) or that Lisa's crying caused Amy to leave (*Explanation*) or that Amy left and later Lisa cried for unrelated reasons (*Narration*).

SDRT (Asher and Lascarides, 2003) takes into account such facts about cancellation, revision and ambiguity to provide a model of how and why particular coherence relations are inferred. The idea is to formalize *principles for pragmatic enrichment* expressing the commonsense

reasoning patterns leading to the 'most natural' interpretations. A guiding idea in phrasing these principles is that they should state the *most plausible and most coherent* interpretations given imperfect information. For example, if there is a salient way to read one event as causing another (e.g. that pushing someone might result in them falling), one interprets them as causally connected by assigning the relation *Explanation* or *Result* (Schlöder, 2018: ch. 7).

Asher and Lascarides (2003) formalize such principles in a *default logic* in which one can phrase defeasible conditionals $p > q$ (paraphrased: 'if p then normally q'). Their logic has the following properties that make it appropriate for the task at hand.

- If p and $p > q$, then infer q only if $\neg q$ is not the case ($\neg q$ *defeats* the conditional).

- If p and $p > q$, then infer q only if there are no r and s such that r, $r > s$ and $q, s \models \bot$ are the case ($r > s$ *clashes* with $p > q$).

- But more informative premises win clashes, i.e. if p, $p > q$, r and $r > \neg q$ all are the case and also $p \models r$, but $r \not\models p$, then infer q.

Phrasing the pragmatic principles using the conditional $>$ ensures (i) that their conclusions can be overridden by additional information (defeating a conclusion); (ii) that when there are multiple conflicting principles in play, the interpretation remains ambiguous (no conclusions are drawn in clashes); and (iii) more detailed information can sway an ambiguity. See Asher and Lascarides (2003); Lascarides and Asher (2009) for details on and further justification of this logic.

For example, we can now formalize the principle that possible causes are typically interpreted as being causes as follows.[3]

(21) a. $R(\alpha, \beta) \land \Diamond cause(e_\beta, e_\alpha) > R = Explanation$.
b. $R(\alpha, \beta) \land \Diamond cause(e_\alpha, e_\beta) > R = Result$.

Where Greek letters range over discourse segments, the predicate *cause* describes a causal relation between two events and $\Diamond$ is alethic possibility.

Such principles are typically still not sufficient to determine the full coherence structure of a discourse. In SDRT, one proceeds as follows: consider all possible assignments of coherence relations that are compatible with the information inferred by the pragmatic principles. From these possible assignments, select the *most coherent* ones via a mechanism that grades coherence; this is known as the principle to *maximise discourse coherence* (Asher and Lascarides, 2003).

Moreover, one can also include principles that *constrain* the possible interpretations. Note that it is possible to interpret two segments by assigning *multiple* relations.

(22) a. I painted the barn.
b. It was an ugly red. $\Big]$ Explanation + Background

---

[3]These are our versions of these principles, which we note to be deviating from the axioms for *Explanation* and *Result* suggested by Asher and Lascarides (2003). They add additional premisses to make it so that causal information overrides aspectual information. We have seen that in some ESSs potential causes are not interpreted to be causes, so we make a modification here.

The most natural interpretation here is that the speaker painted the barn because (previously) it was an ugly red, but *also* that the *being an ugly red* state overlapped the *painting* event (i.e. that the barn had no colours in between and the speaker painted *over* the ugly red). The temporal consequence of *Explanation* (that causes precede effects) is that the barn being red extends in time to sometime before the time index of the painting event.

However, not all coherence relations can be paired up. For example, the same two segments cannot be both connected by *Explanation* and *Result*, as causation ever only goes in one direction. In fact, there is something else wrong with pairing these two relations: *Explanation* is subordinating and *Result* is coordinating. But one cannot pair a subordinating with a coordinating relation, as it makes no sense for the same segment to add to a scene and also move to a new scene . Txurruka (2003) expresses this principle for pragmatic interpretation as in (23), where the predicates *coord* and *subord* describe a coherence relation to be coordinating and subordinating, respectively.

(23) $\neg(R(\alpha, \beta) \wedge R'(\alpha, \beta) \wedge coord(R) \wedge subord(R'))$.

Adopting this principle ensures that the principles for *Explanation* and *Result* always clash. That is, if there is equally good reason to believe that α can cause β and that β can cause α one infers neither *Explanation* nor *Result*. Arguably, this is the case in (20).

However, it is not *always* desirable for certain principles to be clashing, as sometimes we want some principle to take *precedence* over another one. To achieve this, it is useful that more informative premisses win clashes. We exploit this when stating our principle for interpreting ESSs.

4. ESSs and Pictorial Narratives

We now state natural and general pragmatic principles for narrative progression that one can take to apply regardless of the medium in which a narrative is interpreted. We outline how our principles can be put to work to derive the right interpretations for the data discussed earlier. We begin by discussing an appropriate pragmatic principle for the interpretation of eventive-stative sequences.

There is no single coherence relation that is distinctively associated with ESSs (*pace* Asher and Lascarides, 2003 who associate ESSs with *Background*). The examples (8) and (9), repeated here with annotation, show that ESSs can at least support interpretations as temporal precedence and temporal overlap. We annotate these examples now with the appropriate coherence relations.

(8)   I arrived to Zoom late.
      I was cooking.   ⎤-Explanation

(9)   I put on my sunglasses.
      The sun was out.   ⎤-Background

These examples also show that there also is no particular temporal order that is distinctively associated with ESSs (*pace* Kamp and Rohrer, 1983). However, inspection of the data reveals that the natural interpretations all correspond to one of the subordinating relations (like *Explanation* and *Background*), whereas reading an ESS as *Result* (a coordinating relation) sounds wrong even if there is, in principle, a potential causal reading of the event and the state (recall

(11) and (13)). This leads us to suggest the following generalization, formalized as a pragmatic principle in SDRT.

(24)   ESSs typically subordinate the state:
    a. $R(\alpha,\beta) \wedge ev(\alpha) \wedge st(\beta) > subord(R)$.
    b. $R(\alpha,\beta) \wedge ev(\alpha) \wedge st(\beta) \wedge \Diamond cause(e_\alpha, e_\beta) > subord(R)$.

Adding an axiom like (24b) to the generalization in (24a) ensures that the contribution of the aspectual information in (24a) takes precedence over any potential causal information. Specifically, (24a) says that ESSs are typically subordinating the state and (24b) says that causal information cannot by itself override this default. This is because the premiss of (24b) is more informative than just $\Diamond cause(e_\alpha, e_\beta)$. In particular, then, (24b) wins clashes with the principles (21) to infer *Explanation* or *Result*.

This is particularly important for ESSs in which one may see a plausible causal relation between the described event and the described state. Consider again (11) and (13), repeated here.

(11)   I threw a water balloon at Hans. %His shirt was wet. (✓ got wet)

(13)   A cat bit into a mouse that wiggled its tail. #It was dead. (✓ died)

In these examples, a *Result* interpretation (water balloon causing wetness; biting causing death) seems highly plausible on the face of it, but the stativity of the second part of the sequence seems to conflict with such an interpretation. Formally, letting α and β label the eventive and the stative, respectively, we take this to mean that $\Diamond cause(\alpha,\beta)$ is a premiss available for computing the interpretation of these examples. According to our pragmatic principle for *Result* (21), this would normally allow us to infer $Result(\alpha,\beta)$. However, according to our pragmatic principle for ESSs (24), it follows from this that $subord(Result)$ which is not the case, as *Result* is coordinating. Thus, the two principles clash.

Due to the fact that more informative premisses win clashes and the antecedent of (24b) is more informative than the one of (21), we infer that whatever coherence relation joins α and β must be subordinating. This means that in the most natural interpretations of the ESSs in (11) and (13), the eventive coheres with the stative by a subordinating relation. Now also taking into account the principle (23) stating that two segments cannot be connected by both subordinating and coordinating relations, it follows that in the most natural interpretations, *Result* is ruled out. When the second part of a sequence is another eventive, however, as in the alternatives 'got wet' and 'died', the principle (24) does not apply and nothing stands in the way of interpreting the sequences as *Result*.

Note that the foregoing does not mean that in an ESS the eventive and stative always *have* to cohere with a subordinating relation. As the principles in (24) are also phrased as default conditionals, they can be cancelled by defeating or clashing information. One salient way to do so is to add *explicit* cuing to the discourse that defeats the defaults in (24). For example, in the following modifications of (11) and (13):

(25) a. I threw a water balloon at Hans. Therefore his shirt was wet.
    b. A cat bit into a mouse that wiggled its tail. As a result, it was dead.

These cases are naturally and unproblematically interpreted as *Result*, in particular as the wetness temporally succeeding the throwing and the death succeeding the bite. The explicit cuing

with the phrases *therefore* and *as a result*, respectively, enforces this interpretation and simply cancels the application of (24).

Similarly, the pictorial narrative (12), repeated here, is naturally interpreted as *Result*.

(12)

To explain this, one can now appeal to Abusch's hypothesis (1). If the hypothesis is correct, both segments of the discourse are stative descriptions of scenes. This means that the principle (24) does not apply. Thus—since the first picture depicts a possible reason for the state in the second picture—nothing prevents us from applying the principle for *Result* in (21).

Now, in contrast to *Result*, the coherence relations *Explanation* and *Background* are subordinating, so they are not in similar conflict with the principle (24). As a matter of fact, we agree with the observation of Asher and Lascarides (2003) that ESSs are typically read as *Background* (i.e. one typically uses a stative to describe the situation in which an event unfolds). A paradigm example is (26).

(26) a. I went for a walk.
     b. It was raining.   ⎤-Background

To infer *Background* in such and other examples, we use the following principle.

(27) $R(\alpha, \beta) \wedge subord(R) > R = Background$.

That is, when interpreting an ESS we first infer subordination by the principle (24) from which we may infer *Background* by (27). This suffices to obtain the desired interpretations of (11) and (13) as *Background* (i.e. as event and state overlapping).

Together with our principle for *Explanation* in (21), we can now also derive the correct interpretation of (22), repeated here.

(22) a. I painted the barn.
     b. It was an ugly red.   ⎤-Explanation + Background

This is an ESS, so by (24) we infer that the eventive coheres with the stative by subordination. As above, this rules out an interpretation as *Result* (the ugly red was not the result of the painting). Conversely, something being an ugly color is a possible reason to paint it, so the principle for *Explanation* in (21) allows us to infer *Explanation*. Finally, the principle for *Background* (27) applies as well, allowing us to infer *Background*. Thus the most natural interpretation of (22), according to our pragmatic principles, is indeed *Explanation* and *Background* meaning that the ugliness of the previous coat of paint was the speaker's reason to paint over it.

This is how our pragmatic principles vindicate the Aspect Hypothesis in (10). We infer *Background* (and hence, temporal overlap) by a two-step process that first applies (24) and then (27). That these are distinct steps has a subtle but important upshot. When the inference to

Background is cancelled (by defeat or clash), this need not mean that the inference to subordination is defeated as well.[4] This is the case in examples where the most natural reading is only *Explanation* and does not include *Background* (i.e. there is strict temporal precedence between cause and effect).

(28) a. I took a shower.
    b. I was out jogging. ⎤-Explanation

Again, this is an ESS, so by the principle (24) we infer subordination. However, the principle (27) to infer *Background* is defeated here, as commonsense knowledge entails that one cannot simultaneously be out jogging and taking a shower. However, similar knowledge entails that exercising is a possible reason for later taking a shower. So the principle for *Explanation* in (21) licenses the interpretation as *Explanation*.

Note that it is important for the interpretation of (28) that (24) applies and subordination is inferred. Otherwise, interpretations with coordinating relations would compete with the inference towards *Explanation* here; e.g. the interpretation as *Narration* where one went jogging *after* the shower. To see this, compare (28) with an analogous eventive-eventive sequence.

(29) a. I took a shower.
    b. I went jogging. ⎤-Explanation? Narration?

In this case, the interpretations as *Explanation* (jogging being the reason for showering) and *Narration* (the speaker showering and then going jogging) are, arguably, equally natural. In (28), however, there appears to be a clear preference for the *Explanation* reading. This distinction is explained by the principle (24) applying to (28)—ruling out the coordinating relation *Narration*—but not to (29).

5. Event structural triggers of *Result*

Altshuler (2021) discusses the examples below which, at first blush, seem like a challenge to the analysis just proposed. These data show that (even without explicit cueing) we can understand ESSs with a *Result* reading when an an exclamative intervenes the eventive and stative description:

(30) I painted the Hampshire barn. Voilà! It was red.

(31) Ava pushed Justin. Smash! He was on the ground.

(32) Anna got the top score on the entrance exam. Poof! She was a student at Harvard.

(33) I threw a water balloon at Hans. Splat! His shirt was wet.

(34) A cat bit into a mouse while it was wiggling its tail. Oof! It was dead.

Altshuler further shows that emotive adjectives like *happy* and *upset* are also possible triggers of *Result*:

(35) Xacho gave Narine a dozen roses. She was (really) happy.

---

[4]This is an important difference to how Asher and Lascarides (2003) treat ESSs. They assign to ESSs a defeasible inference to *Background*, which means that when *Background* is defeated, they discard the aspectual information of the ESS entirely. In contrast, our principles still license the inference to subordination.

(36)   I threw a water balloon at Hans. He was (really) upset.

How can we make sense of these data, given the analysis developed in the previous section? And can (should?) we explain the distinct triggers of *Result* with exclamatives and emotives by a common principle? In what follows, we follow Altshuler (2021) in offering some possible avenues for answering these questions and relate them to our proposed analysis. The upshot will be that all ESSs with *Result* readings have a more complex event structure that does not fit into our present analysis—which means they are not counterexamples, but worthy of further investigation.

In order to make sense of the *Result* triggers in the two data sets above, we think it is worthwhile to consider the role that point of view may play in the interpretation of exclamatives and emotive adjectives. These classes of expressions have been argued to be perspectival expressions (Bylinina et al., 2014; Eckardt, 2015, 2021). The reason that we think that this is a promising avenue to pursue is that perspectival expressions arguably involve additional event structure to represent the perspectival anchor (Bylinina, 2017).[5] Let us assume that this correct and further assume that the additional event structure encodes a change-of-state component involved in the experiencer event on which there is some perspective. Then it would not be surprising that *Result* is triggered. Recall the proposed rule for ESSs (12), repeated below:

(12)   ESSs typically subordinate the state:
  a. $R(\alpha, \beta) \land ev(\alpha) \land st(\beta) > subord(R)$.
  b. $R(\alpha, \beta) \land ev(\alpha) \land st(\beta) \land \Diamond cause(e_\alpha, e_\beta) > subord(R)$.

Note that in the antecedent of this rule, the second argument of R, namely β, is stative. If we take $st(\beta)$ to be incompatible with β describing a change-of-state (even if it also describes a state), then the rule does not apply to such β. Thus if β describes a change-of-state, we would not expect R to subordinate. But, if it does not subordinate, the rule in (21b), repeated below, will ensure that *Result* is inferred.

(21) a. $R(\alpha, \beta) \land \Diamond cause(e_\beta, e_\alpha) > R = Explanation$.
    b. $R(\alpha, \beta) \land \Diamond cause(e_\alpha, e_\beta) > R = Result$.

Thus if exclamatives and emotive adjectives describe change-of-states (despite also describing states), the desired *Result* interpretation are compatible with and indeed predicted by our analysis.

This analysis is further supported by examples involving deverbals, that is, statives that are derived from eventives:

(37)   I threw a giant water balloon at Hans. . .
  a. . . . He was drenched.
  b. . . . His shirt was soaked.
  b. . . . He was not thrilled.

---

[5]See also work on the semantics of fiction by Altshuler and Maier (2018), who consider how perspectival constructions in an otherwise seemingly impersonal narrative could force the reader to accommodate a first-personal narrator to serve as the perspectival anchor. For more discussion of the way perspective relates to narrative progression, see Cumming 2021 and Anand and Toosarvandani 2021.

Notice that in all these examples, *Result* is the most natural interpretation. This is exactly what is expected given our assumption above, namely that $st(\beta)$ is incompatible with β describing a change-of-state (even if it also describes a state).

This analysis also motivates a more complex discourse structure in (30)-(36). In particular, we can treat the exclamative as its own discourse segment, which serves as the second argument of *Result*, with the initial eventive description serving as the initial argument. This complex discourse segment is then elaborated on by the stative description. This proposed discourse structure is provided below, for (33):

(38) (α) I threw a water balloon at Hans.
    (β) Splat!   ⎤-Result
    (γ) His shirt was wet.   ⎦-Elaboration

Whens segmented like this, the sequence α-β is not an ESS in the sense of the rule in (24), so we do not infer that they must cohere by a subordinating relation. Common sense seems to support that throwing a water balloon can cause *Splat!*, so we may infer *Result*. If this means that we ought to consider *Splat!* an eventive, then β-γ *is* an ESS, but since *Elaboration* is subordinating this interpretation is compatible with (24). Of course, this analysis will require a more detailed analysis of the content and discourse semantics of exclamatives that we do not provide here.[6]

While we have, thus far, not said anything about *Elaboration*, Altshuler (2016) argues that it plays an important role in narrative progression contexts. He considers the oft-cited example below, modified from Hinrichs (1986), which was argued by Lascarides and Asher (1993) to exemplify *Result*.

(39) Max switched off the light. It was pitch dark around him.

While *pitch dark* may be taken to be a perspectival expression, lending itself to the analysis just proposed, Altshuler (2016) instead argues that it merely exemplifies *Elaboration*. His idea is that the first sentence in the discourse entails that the room became darker (to some degree) in some salient location. The second sentence, then, elaborates on the degree to which it became darker. To wit, note that (40a) is degraded compared to (39). It can be improved with an exclamative, as in (40b).

(40) a. Jameson switched off the light. ?It was dark around him.
    b. Jameson switched off the light. Voilà! It was dark around him.

We take these data to indicate that the analysis we have developed in this paper is on the right track and that a fleshed out analysis of the role of perspectival expressions will vindicate our generalisation that ESSs are subordinating.

---

[6]From the off-the-shelf tools available, one may consider α and β to be cohering by the *metadiscursive* version of *Result*, i.e. that the event described in α causes the utterance of β (Asher and Lascarides, 2003). We are not entirely satisfied with this, as β also contributes a worldly consequence of α: the water balloon exploding or, at least, the speaker's reaction to this (as if the balloon were a dud, *Splat!* would be inappropriate).

6. Conclusion

This paper has explored the consequences of adopting Abusch (2014)'s hypothesis, repeated below, for the interpretation of a discourse.

(1) Abusch's Hypothesis: Pictures are stative depictions of the world.

We focused on the phenomenon of narrative progression and considered Abusch's proposal that aspectual differences between linguistic and pictorial narrative do not factor in providing a uniform analysis across media. Based on eventive-stative sequences in linguistic narrative, we argued that this proposal is empirically inadequate. Subsequently, we motivated a new answer within Segmented Discourse Representation Theory that attempts to preserve Abusch's insight that there can and should be a uniform analysis of narrative progression across media.

We proposed to exploit the (assumed) aspectual differences between pictorial and linguistic narratives. Adopting Abusch's hypothesis (i.e. that all pictures are stative) allowed us to describe a single pragmatic algorithm that makes the correct predictions for the data that Abusch's proposal did not and indeed could not account for. A key innovation of the analysis are the ideas that: (i) there are competing axioms that encode information about discourse structural properties of coherence relations (subordinating vs. coordinating) and (ii) the competition is sensitive to whether an argument of a coherence relation is stative. In the previous section, we saw data that seem to merit further investigation into (ii): exclamatives, deverbals and evaluative statives appear in ESSs that at first glance contradict our analysis, but there is reason to believe that these sequences exhibit more complex event structures that are, in fact, compatible with our proposal.

Finally, we note that our proposed algorithm only takes three coherence relations into consideration, namely *Result*, *Explanation* and *Background*. However, we saw that *Elaboration* also may play a key role in accounting for narrative progression. And it is certainly the case that *Narration* factors into any theory of narrative progression (Lascarides and Asher, 1993).[7] In future work, we plan to revisit this issue and integrate these and other relations to further develop the pragmatic algorithm proposed here. As it stands, this paper is a first step towards and a proof of concept for a general theory of narratives across media. If pictures are stative, it means that we can use the existing tools for discourse interpretation provided by SDRT to construct this theory.

References

Abusch, D. (2014). Temporal succession and aspectual type in visual narrative. In L. Crnič and U. Sauerland (Eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, Volume 1, pp. 9–29.

Altshuler, D. (2016). *Events, States and Times: An Essay on Narrative Discourse in English*. Berlin/Warsaw: de Gruyter. Open access, available at: https://www.degruyter.com/view/product/474350.

Altshuler, D. (2021). A puzzle about narrative progression and causal reasoning. In A. Stokke and E. Maier (Eds.), *The Language of Fiction*. Oxford: Oxford University Press.

---

[7] See Altshuler and Varasdi (2015); Cumming (2021); Anand and Toosarvandani (2021) for more recent discussion.

Altshuler, D. and E. Maier (2018). Death on the freeway: Imaginative resistance as narrator accommodation. In I. Frana, P. Menéndez-Benito, and R. Bhatt (Eds.), *Festschrift for Angelika Krazter*. to be available at: `https://scholarworks.umass.edu/ak_festsite`.

Altshuler, D. and R. Truswell (2021). *The Syntax-Discourse Interface. Extraction from Coordinate Structures*. Oxford: Oxford University Press.

Altshuler, D. and K. Varasdi (2015). A proof of definitional adequacy of RESULT and NARRATION. In S. D'Antonio, C.-R. Little, M. R. Moroney, and M. Wiegand (Eds.), *Proceedings of Semantics and Linguistics Theory 25*.

Anand, P. and M. Toosarvandani (2021). Narrative and point of view. In D. Altshuler (Ed.), *Linguistics meets Philosophy*. Cambridge, UK: Cambridge University Press.

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

Asher, N. and L. Vieu (2005). Subordinating and coordinating discourse relations. *Lingua 115*, 591–610.

Bylinina, L. (2017). Judge-Dependence in Degree Constructions. *Journal of Semantics 34*, 291–331.

Bylinina, L., E. McCready, and Y. Sudo (2014). The landscape of perspective shifting. http://web.mit.edu/ysudo/www/pdf/tubingen-handout.pdf.

Culicover, P. and R. Jackendoff (1997). Semantic subordination despite syntactic coordination. *Linguistic Inquiry 28*, 195–217.

Cumming, S. (2021). Narrative and point of view. In E. M. . A. Stokke (Ed.), *The Language of Fiction*. Oxford: OUP.

Davidson, D. (1967). The logical form of action sentences. In N. Rescher (Ed.), *The Logic of Decision and Action*, pp. 81–120. Univ. of Pittsburgh Press.

Deane, P. (1991). Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics 2*, 1–63.

Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics? *Linguistics and philosophy 9*(1), 37–61.

Eckardt, R. (2015). *The Semantics of Free Indirect Discourse*. Brill, Leiden, The Netherlands.

Eckardt, R. (2021). In search of the narrator. In E. M. . A. Stokke (Ed.), *The Language of Fiction*. Oxford: OUP.

Greenberg, G. (2013). Beyond resemblance. *Philosophical review 122*(2), 215–287.

Hinrichs, E. (1986). Temporal anaphora and discourses of English. *Linguistics and Philosophy 9*, 63–82.

Hobbs, J. (1985). On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University.

Hunter, J. and K. Thompson (2021). On the role of relations and structure in discourse interpretation. In D. Altshuler (Ed.), *Linguistics meets Philosophy*. Cambridge: Cambridge University Press.

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

Kehler, A. (2002). *Coherence, Reference and the Theory of Grammar*. CSLI Publications, Cambridge University Press.

Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In J. v. B. Renate Bartsch and P. van Emde Boas (Eds.), *Semantics and Contextual Expression*, pp. 75–115. Foris.

Lascarides, A. and N. Asher (1993). Temporal interpretation, discourse relations and common-sense entailment. *Linguistics and philosophy 16*(5), 437–493.

Lascarides, A. and N. Asher (2009). Agreement, disputes and commitments in dialogue. *Journal of Semantics 26*(2), 109–158.

Parsons, T. (1990). *Events in the Semantics of English: A study in subatomic semantics*. Cambridge, MA: MIT Press.

Polanyi, L. (1985). A theory of discourse structure and discourse coherence. In *Proceedings of the 21st Meeting of the Chicago Linguistics Society*, pp. 306–322. Chicago: The University of Chicago Press.

Schlöder, J. J. (2018). *Assertion and Rejection*. Ph. D. thesis, Institute for Logic, Language and Computation, University of Amsterdam.

Txurruka, I. (2003). The natural language conjunction *and*. *Linguistics and Philosophy 26*, 255–285.