

# Towards Generating Narratives from a Preverbal Message: A DRT-based approach

Katja Jasinskaja  
Universität Heidelberg/IMS, Universität Stuttgart

Antje Roßdeutscher  
IMS, Universität Stuttgart

## 1 Background

Speakers of different languages tell the same story differently. The empirical basis for this finding is a multilingual corpus of spoken narratives which all retell the same animation film. (See Carroll et al. (2008) and references therein.) The fragment of the film that we are concentrating on shows a clay man who looks for water in a world made of paper. Twice he is in danger of being overrun by a large piece of paper blown by the wind, whereas the third time he is actually knocked down by one piece. He finally finds a small puddle of water, but while he is trying to collect the water from the ground, the wet paper breaks in and he falls into the hole. Typical solutions of an English and a German speaker are given in (1) and (2).

- (1) a. and you see him land on a new plane [...]  
b. and eh he wakes up  
c. and eh just as he is opening his eyes from the fall [...]  
d. he looks in front of him  
e. and there's this piece of paper coming straight for him  
f. and he's scared  
g. and he jumps up onto his knees  
h. and the piece of paper misses him
- (2) a. als n'ächstes sieht man das m'ännchen fallen [...]  
b. und ∅ geht n' paar schritte  
c. wird ∅ dann von so nem wehenden blatt umgeworfen  
d. und ∅ h'ört auf einmal wieder dieses tropfen  
e. und ∅ sieht dann auch n' 'äh n st'ück papier  
f. wo also so ne kleine pf'ütze schon is  
g. und er kniet sich dann davor

First, the solutions differ in that the English speaker establishes a causal chain in (1), whereas this is not so in (2), e.g. doing some paces in (2b) does not naturally

lead to be knocked down, cf. (2c). It is the use of sentence-internal *dann* ‘then’ that establishes narrative connection between the two event descriptions, see also (2g). Second, the German speaker skips talking about all the dangerous situations except the one where the protagonist gets directly affected, i.e. knocked over by the piece of paper, cf. line (2c), whereas the English speaker also describes in detail the first situation where the man manages to avoid the collision. If German speakers tried to do the same, we would have discourses like (3), but we do not find them in the corpus, although (3) is coherent. It is very untypical for German speakers to fill the subject position with an object other than the protagonist as in (3c) and (3e).

- (3) a. und er wird dann wach  
 b. und öffnet dann die Augen  
 c. und da kommt gerade ein Papier auf ihn zu  
 d. und er springt auf  
 e. und das Papier fliegt vorbei

This illustrates the generalization that German speakers select the protagonist as a global topic for the whole narrative, which then systematically surfaces as the grammatical subject of the clauses, whereas the events in which the protagonist is not or only marginally involved (i.e. without dramatic consequences for himself) are simply not narrated. Thus the typical German narrative can be grosso modo characterized as following the instruction: “Select events for description that are changes of properties of the protagonist; connect those descriptions by anaphoric temporal adverbials.” In contrast, the majority of English speakers produce solutions that could be characterized as following the instruction: “Introduce some fictitious witness time  $t_{now}$  of a narrator witnessing the story; fill the period  $t_{now}$  by producing continuous event descriptions, either ‘shot by shot’ or causally connected.”

Why do the English and the German speakers assess the narration task differently? First of all, one should take into account that it is remarkably difficult to produce a coherent narrative spontaneously. Therefore experienced story-tellers develop narration strategies that make their task easier. What counts as an “easy option” in turn depends on the grammar of the language in question, i.e. it is crucial which grammatical features require obligatory marking:

- (4) **Leading Hypothesis.** If it is obligatory to mark a grammatical feature sentence by sentence, the speaker will choose its value globally.

This means that the speaker will set the value of the obligatory feature as a default for all the sentences in the discourse, although it can be overridden locally for the sake of coherence.

In verb-second languages the preverbal position must be filled by exactly one constituent and that constituent may, but need not be the subject of the sentence. A German speaker must decide not only which salient discourse referent to select for subject but also where to place it. A global decision is advantageous here because

it reduces the set of decisions to be made for each particular sentence to either placing the globally selected subject or a temporal adverb in the sentence-initial position. Choosing the subject is ranked higher in the decision process than event selection. Whether or not an event is selected for description depends on whether or not that event is suitable for being described with the protagonist as subject. If the selected events turn out not to be causally connected the temporal adverbial guarantees establishing the story line, cf. (2g).

Speakers of a strict SVO language have no choice as to where to place the subject, therefore a global selection of the subject has no strategic advantage. However, in contrast to German speakers, who neither have the obligation nor the possibility to choose aspect, English speakers have to make that choice in each sentence. Some of them opt globally against progressive and at the same time against the introduction of the narrator and his witness-time  $t_{now}$  to be filled by a causal chain. Their narration solutions are similar to those of the German speakers. The majority of English speakers, however, make their global choice for the progressive and the deictic narrator-centered perspective on the story events. For example, the speaker of (1) starts and ends episodes in the progressive (not documented in the lines). Of course, a causal chain cannot be established unless the speaker locally switches to perfective aspect. E.g. in (1f), (1g) and (1h) the global default decision for the progressive is overridden for the sake of providing causal antecedents for their successors.

If event selection follows the causal chain this is unsuitable for if not incompatible with any further restriction on subject selection. As a consequence English speakers simply select the salient agent of the event for subject and change the topic if the salient agent changes. To sum up: (4) leads to different **hierarchies of decisions** necessary for describing or non-describing a particular event.

The decision hierarchy rules out discourses like (3) because the events in (3c) and (3e), where the protagonist is neither the agent nor is strongly affected by the action, are either filtered out by global event selection criteria and will not be verbalized, or if the speaker locally undoes that global decision and verbalizes them for the sake of coherence, the piece of paper will be demoted to a non-subject position, the subject position being reserved for the protagonist. Thus wordings like *er sieht ein Papier auf sich zukommen* (he sees a paper coming) and *er weicht dem Papier aus* (he avoids the paper) would be generated for (3c) and (3e) instead. The passive voice in (2c) shows a further typical demotion operation.

## 2 Goals

Our main goal is to reconstruct the empirical findings presented above in a formal theory a central part of which is the hierarchy of decisions informally described above.

We are not aiming at developing generation components for computational applications, nor modeling low level psychological production processes. Rather, we

want to model the influence of cross-linguistic variation on the strategic decisions of speakers in the process of discourse production. At the same time this study presents an attempt at applying Discourse Representation Theory (DRT) in the direction of generation. Originally, DRT was developed as a theory of discourse interpretation (Kamp and Reyle, 1993), but more recent declarative formulations of DRT (e.g. Kamp and Roßdeutscher, 2005) should, in principle, be applicable in both directions. This paper explores the second possibility.

### 3 Analysis

We assume levels of knowledge representation that are shared by all speakers. One contains referents for objects that can be seen in the film, as well as time points and intervals of the film time, and DRS-conditions for propositions that can be judged true on the basis of the picture and the sound track. At another level that representation is enriched with all sorts of common-sensical inferences—the “pragmatic” interpretation of the film—in particular causal relations between states and changes of state, inferences about the normal consequences of some events (e.g. if this sheet of paper continues moving and the man does not move, it will crash on him); and finally, ascriptions of perception and mental states—beliefs, emotions, intentions—to the protagonist. A fragment of this representation is shown in Fig. 1. The DRS describes the episode where the protagonist ( $m$ ) wakes up, sees a sheet of paper ( $sheet_1$ ) flying towards him, and avoids it by sitting up, cf. (1).

The Davidsonian notion of eventuality widely adopted in formal semantics, including DRT, is inspired by the goal of explaining a number of linguistic phenomena, such as adverbial modification, tense and aspect. Thus what counts as a state or an event depends ultimately on what relevant distinctions are made by the grammar and the lexicon of a particular language. It is unclear that such a notion is suitable for representing non-linguistic knowledge (interpretation of a film), which we assume to be language-neutral and shared by all participants of the study regardless of their first language. For our purposes we identify states as time-dependent propositions, i.e. a pair of a (definite) proposition and a time interval during which it holds, e.g.  $s_1, s_2$  in Fig. 1. A special kind of “state” is a gradual change of some property over a period of time, cf.  $s_6, s_{47}$ . Events are momentaneous changes from a proposition  $p$  to  $\neg p$ , identified by a pair of states whose first elements are  $p$  and  $\neg p$  respectively, and the second elements are time intervals such that the end point of one interval is the starting point of the other, cf.  $e_1$  in Fig. 1. The relation of immediate cause  $\Rightarrow$  holds between a state and an event if the state is one of the necessary preconditions for the event to happen, and overlaps with the event in time, e.g. the protagonist  $m$  looking in the direction of the sheet of paper  $sheet_1$  ( $s_{11}$ ), as well as his eyes being open ( $s_2$ ) are preconditions for him getting to see  $sheet_1$  ( $s_2 \Rightarrow e_3; s_{11} \Rightarrow e_3$ ). An event immediately causes its poststate, e.g.  $e_1 \Rightarrow s_2$ . The language for describing spacial objects and relations is largely borrowed from Kamp and Roßdeutscher (2005).

$m, head, eyes, spine, legs, sheet_1, w_2, t_1 \dots t_{21}, t^{1:25} \dots t^{2:27}, s_1 \dots s_{57}, e_1 \dots e_{18}$ $s_1 : (\neg open(eyes), \langle t_1^{1:28}, t_1^{1:31} \rangle)$ $s_2 : (open(eyes), \langle t_1^{1:31}, t_1^{1:37} \rangle)$ $e_1 : (s_1, s_2), \quad e_1 \Rightarrow s_2$ $t_1 \leq t^{1:29}, \quad t^{1:37} \leq t_2 \quad (\text{the whole movement})$ $s_6 : (move(sheet_1), \langle t_1, t_2 \rangle) \quad (\text{the whole movement})$ $s_7 : (move(sheet_1), \langle t_2^{1:31}, t^{1:34} \rangle); s_7 \subset s_6 \quad (\text{part 1 visible to } m)$ $s_8 : (PATHSURF(sheet_1, w_2), \langle t_1, t_2 \rangle)$ $s_{10} : (OBS(m) \parallel \langle m, sheet_1 \rangle), \langle t_3, t^{1:34} \rangle, \quad t_1 \leq t_3 \leq t_2^{1:31}$ $s_{11} : (ALIGN(OBS(m), \langle m, sheet_1 \rangle), \langle t_3, t^{1:34} \rangle)$ $s_{14} : (\neg see(m, s_7), \langle t_1^{1:25}, t_2^{1:31} \rangle)$ $s_{15} : (see(m, s_7), \langle t_2^{1:31}, t^{1:34} \rangle)$ $e_3 : (s_{14}, s_{15}); \quad s_2 \Rightarrow e_3; \quad s_{11} \Rightarrow e_3; \quad s_7 \Rightarrow e_3; \quad e_3 \Rightarrow s_{15}$ $s_{16} : (INCSURF(head, w_2), \langle t_2^{1:31}, t_2^{1:33} \rangle)$ $t_6 \leq t_2^{1:32}, t^{1:34} \leq t_7 \leq t^{1:37}, t_{20} \approx t^{1:34}, t_{20} \leq t_{21}$ $s_{22} : (\neg believe(m, [s_{29} : (APPLYFORCE(sheet_1, m), \langle t_{20}, t_{21} \rangle)]), \langle t^{1:25}, t_6 \rangle)$ $s_{23} : (believe(m, [s_{29} : (APPLYFORCE(sheet_1, m), \langle t_{20}, t_{21} \rangle)]), \langle t_6, t_7 \rangle)$ $e_5 : (s_{22}, s_{23}); \quad s_8 \Rightarrow \dots \Rightarrow e_5; \quad s_{15} \Rightarrow \dots \Rightarrow e_5; \quad s_{16} \Rightarrow \dots \Rightarrow e_5; \quad e_5 \Rightarrow s_{23}$ $t_{12} < t_{13}$ $s_{32} : (\neg scared(m), \langle t^{1:25}, t_{12} \rangle)$ $s_{33} : (scared(m), \langle t_{12}, t_{13} \rangle)$ $e_8 : (s_{32}, s_{33}); \quad s_{23} \Rightarrow e_8; \quad e_8 \Rightarrow s_{33}$ $t_6 \leq t_{14}; \quad t_{14} < t_{15}; \quad t_{14} \leq t_1^{1:33}; \quad t^{1:34} \leq t_{15}$ $s_{36} : (\neg intend(m, CAUSE(m, \neg APPLYFORCE(sheet_1, m))), \langle t^{1:25}, t_{14} \rangle)$ $s_{37} : (intend(m, CAUSE(m, \neg APPLYFORCE(sheet_1, m))), \langle t_{14}, t_{15} \rangle)$ $e_{10} : (s_{36}, s_{37}); \quad s_{23} \Rightarrow e_{10}; \quad e_{10} \Rightarrow s_{37}$ $t_{14} \leq t_{18}; \quad t_{18} < t_{19}; \quad t_{18} \leq t_1^{1:33}; \quad t^{1:34} \leq t_{19}$ $s_{40} : (\neg intend(m, CAUSE(m, ALIGN(spine, VERT))), \langle t^{1:25}, t_{18} \rangle)$ $s_{41} : (intend(m, CAUSE(m, ALIGN(spine, VERT))), \langle t_{18}, t_{19} \rangle)$ $e_{12} : (s_{40}, s_{41}); \quad s_{37} \Rightarrow \dots \Rightarrow e_{12}; \quad e_{12} \Rightarrow s_{41}$ $s_{46} : (\neg [\angle(spine, HOR) \text{ changes from } 0^\circ \text{ to } 90^\circ], \langle t^{1:25}, t_1^{1:33} \rangle)$ $s_{47} : (\angle(spine, HOR) \text{ changes from } 0^\circ \text{ to } 90^\circ, \langle t_1^{1:33}, t^{1:34} \rangle)$ $e_{15} : (s_{46}, s_{47}) \quad s_{41} \Rightarrow e_{15}; \quad e_{15} \Rightarrow s_{47}$ $s_{50} : (\neg ALIGN(spine, VERT), \langle t^{1:25}, t^{1:34} \rangle)$ $s_{51} : (ALIGN(spine, VERT), \langle t^{1:34}, t_1^{1:55} \rangle)$ $e_{17} : (s_{50}, s_{51}) \quad s_{41} \Rightarrow e_{17}; \quad s_{47} \Rightarrow e_{17}; \quad e_{17} \Rightarrow s_{51}$ $s_{53} : (\neg INCSURF(head, w_2), \langle t_2^{1:33}, t^{2:27} \rangle)$ $e_{18} : (s_{16}, s_{53}) \quad s_{47} \Rightarrow e_{18}; \quad e_{18} \Rightarrow s_{53}$ $s_{55} : (legs \parallel HOR, \langle t^{1:25}, t_1^{1:44} \rangle)$ $s_{56} : (\neg APPLYFORCE(sheet_1, m), \langle t^{1:25}, t^{2:27} \rangle)$ $s_{57} : (\neg APPLYFORCE(sheet_1, m), \langle t_{20}, t_{21} \rangle)$ $s_{53} \Rightarrow s_{57}$
--

Figure 1: A fragment of the DRS representing the interpretation of the film

This representation is input to the event and predicate selection procedure. The core of the story is always formed by causally connected changes of state, i.e. events. States, including dynamic processes extended in time, are talked about only if they constitute a pre- or poststate of some relevant event, or if they cause some relevant event. The speaker follows the temporal axis event by event applying the language specific decision hierarchy such as (5) for German or (6) for English.

$$(5) \quad G: \left\{ \begin{array}{l} \{ \langle \text{Subj}(e)=m \rangle, \langle \boxed{t_{dann}} \mid \vee \boxed{\text{Subj}}, \boxed{\text{Subj}} \mid \vee \boxed{t_{dann}} \rangle \} \prec \\ \{ \langle \text{select}(e), \neg \text{select}(e) \rangle, \langle \text{predicate}(e,m) \rangle \} \end{array} \right\}$$

$$(6) \quad E_{majority}: \left\{ \begin{array}{l} \{ \langle \text{select}(e), \neg \text{select}(e) \rangle, \langle e \in t_{now}, e \notin t_{then} \rangle \} \prec \\ \{ \langle \text{Subj}(e)=\text{agent}(e) \rangle, \langle \text{predicate}(e) \rangle \} \end{array} \right\}$$

(5) says that the selection of the protagonist ( $m$ ) as subject and of the word order options is prior to event and predicate selection, which is ultimately responsible e.g. for filtering out events where  $m$  has low proto-agent properties (Dowty, 1991) and does not make a prototypical subject. In the episode represented in the DRS in Fig. 1, the man is experiencer or subject of a mental state in the events  $e_3, e_5, e_8$  so these events are good candidates for filtering out. Opening his eyes referred to by  $e_1$  may or may not be regarded as an intentional action. The events  $e_{10}$  through  $e_{18}$  refer to different aspects of the man sitting up, from forming the intention to prevent collision with  $sheet_1$  to the change of location of the man's head. These events are causally related and as a whole form a complex eventuality controlled and carried out by  $m$ . However, (a) this action is originally caused by the movement  $s_7$  of the sheet of paper, and that together with the events surrounding it have been filtered out as not properly involving the protagonist, and (b) the man sitting up also has no essential consequences for the rest of the story (later he stands up properly and the action continues). Therefore the mention of this eventuality is also easy to skip. As a result, the German speakers normally do not narrate the episode with  $sheet_1$  at all, cf. (2).

In contrast, for a typical English speaker event selection is prior to and independent of the choice of subject, cf. (6). The speaker has an option to choose the temporal frame covering the episode in Fig. 1 for their witness time  $t_{now}$ , and if they do so, the causally connected events within that frame are selected for narration. This is why the English speakers, such as (1), are more likely than the Germans to recount this episode. Which of these events are put in the foreground, and which are implied or presupposed is decided later at the stage of predicate selection. In sum, the result of applying the decision hierarchies in (5) or (6) is a model for *the preverbal message* in the sense of Levelt (1999). Thus we predict that this preverbal message is language-specific.

Next, if the event is selected for narration, rules of lexical selection are applied to the preverbal message and specify which verb will be chosen for the event description. For instance, if  $e_5$  ( $m$  realizes that he is in danger of being hit by  $sheet_1$ ) or any of its consequences  $e_8$  ( $m$  is scared),  $e_{10}$  ( $m$  wants to avoid  $sheet_1$ ),

$e_{12} \oplus e_{15} \oplus e_{17}$  (*m* sits up) is selected for narration, the cause states  $s_6, s_7$  (*sheet*<sub>1</sub> moves),  $s_8$  (the path of *sheet*<sub>1</sub>'s movement is  $w_2$ ),  $s_{16}$  (*m*'s head is located on that path) are also selected. This combination of state descriptions licenses the choice of *the sheet of paper / das Papier* as subject and *move/fly/roll towards the man / auf das Männchen zu fliegen* as predicate (we ignore the mode of motion). Since the beginning and the end of this movement are not shown in the film and are unspecified in the corresponding DRS, we only find definite temporal locations that are embedded in the maximal movement state  $s_7$ , e.g.  $\langle t_2^{1:31}, t_2^{1:34} \rangle \subset s_7$ . This licenses the use of the progressive form in English. The event  $e_3$  marks the point when the man starts to see *sheet*<sub>1</sub> moving. Since neither German nor English have a concise way of referring to this event, such as an inchoative form of *see / sehen*, if the  $e_3$  is selected for narration this is done via reference to its poststate *see the sheet of paper flying towards him*. The sheet of paper *sheet*<sub>1</sub> never hits the man in the whole scene ( $s_{56} : (\neg \text{APPLYFORCE}(\textit{sheet}_1, m), \langle t_2^{1:25}, t_2^{2:27} \rangle)$ ). States that never change, and especially those identified by negative propositions, are normally not narrated. However, if there is an expectation or belief that the proposition should have obtained at some point then negated predicates can be produced. The protagonist's belief that *sheet*<sub>1</sub> might hit him ( $s_{23}$ ) in the time interval  $\langle t_{20}, t_{21} \rangle$  and the fact that it doesn't ( $s_{57}$ ) licenses the production of sentences like *The sheet of paper does not hit him* or *The sheet of paper misses him*. Moreover,  $s_{57}$  is brought about by *m*'s action originating from his intention to do something so that *sheet*<sub>1</sub> does not hit him ( $e_{10}$ ). Thus  $e_{10} \Rightarrow \dots \Rightarrow s_{57}$  licenses the use of such predicates as *avoid the sheet of paper*. Finally, the fact that the intention to prevent collision with *sheet*<sub>1</sub> ( $e_{10}$ ) gives rise to another intention  $e_{12}$  to sit up, i.e. align the spine with the absolute vertical while the legs remain resting on the ground ( $s_{55}$ ), which is carried out in  $e_{17}$ , licenses the production of sentences like *he sits/gets/jumps up (in order) to avoid the sheet of paper*. In sum, these examples are supposed to show how the DRT-based representation of the preverbal knowledge passing through the filter of the event selection procedure gives rise to a variety of expressions for the same content in one language or two different languages.

## 4 Conclusion

This paper presents a case study in applying the DRT formalism to modeling cross-linguistic differences in discourse planning in the process of narrative production. At this point we can say that the DRT component for representing spacial objects and relations by Kamp and Roßdeutscher (2005) in combination with the DRT treatment of intentionality and mental attitudes and the proposed notion of state, event, and immediate cause, are suitable for representing language-independent non-verbal knowledge in such a way that it reflects all the relevant distinctions that can make their way into different language-specific solutions of the narrative task. The episode on which we tested this application is representative in the sense that it is a case where two participants (the man and the sheet of paper) interact on roughly

“equal terms” and their relative proto-agent status flips from event to event. The fact that this episode is difficult for the German and relatively easy for the English narrators is predicted by differences between the German and the English decision hierarchies in narrative planning.

## References

- Carroll, M., Roßdeutscher, A., von Stutterheim, C., and Lambert, M. (2008). Subordination in narratives and macrostructural planning: Taking a comparative point of view. In Fabricius-Hansen, C. and Ramm, W., editors, *Subordination versus coordination in sentence and text from a cross-linguistic perspective*, Amsterdam. Benjamins.
- Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language*, 67:547–619.
- Kamp, H. and Reyle, U. (1993). *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Publishers.
- Kamp, H. and Roßdeutscher, A. (2005). A part of the conceptual background theory, concerned with time, space and motion. Technical report, IMS, Stuttgart University.
- Levelt, W. J. (1999). Producing spoken language: a blueprint of the speaker. In Brown, Collin, M. and Hagoort, P., editors, *The Neurocognition of Language*, pages 83–121. Oxford University Press.