Animation Motion in NarrativeML

Inderjeet Mani

Atelier Milford Paradise, Prachuap Kiri Khan, Thailand
inderjeet.mani@gmail.com

Abstract

This paper describes qualitative spatial representations relevant to cartoon motion incorporated into NarrativeML, an annotation scheme intended to capture some of the core aspects of narrative. These representations are motivated by linguistic distinctions drawn from cross-linguistic studies. Motion is modeled in terms of transitions in spatial configurations, using an expressive dynamic logic with the manner and path of motion being derived from a few basic primitives. The manner is elaborated to represent properties of motion that bear on character affect. Such representations can potentially be used to support cartoon narrative summarization and question-answering. The paper discusses annotation challenges, and the use of computer vision to help in annotation. Work is underway on annotating a cartoon corpus in terms of this scheme.

1998 ACM Subject Classification H.5.1 Multimedia Information Systems, I.2.4 Knowledge Representation Formalisms and Methods, I.2.7 Natural Language Processing, I.4.8 Scene Analysis: Motion

Keywords and phrases Cinematography, Motion, Qualitative Reasoning, Narrative, NarrativeML

Digital Object Identifier 10.4230/OASIcs.CMN.2016.3

1 Introduction

Motion is the essence of animated cartoons. Animators go to great lengths to create gestures and sequences of poses that create a vivid and appealing illusion of many different varieties of motion. What would the Road Runner cartoons be without the thrill of the chase, the characters' prolonged braking motions and sudden propulsions? Why are we so entertained by Wile E. Coyote's fantastic object-penetrating collisions and varieties of chasm plunges? One would expect that qualitative representations of the characters' spatiotemporal dynamics would be more relevant to narrative than their precise geometries or the equations describing their highly constrained, cartoon-physics trajectories. Ideally, these qualitative representations should reflect the narratologically-relevant cognitive abstractions used by the audience in describing movies, and at the same time, be computable. This paper describes qualitative spatial representations relevant to cartoon motion incorporated into NarrativeML [28], an annotation scheme intended to capture certain core aspects of narrative.

As the film theorist David Bordwell [3] explains, films offer the same rich stimuli for inferring motion that are presented in the real world. He quotes Paul Messaris [31]: "What distinguishes images (including motion pictures) from language and from other modes of communication is the fact that images reproduce many of the informational cues that people make use of in their perception of physical and social reality." These inferences about motion involve, as is well-known, optical flow [13], which tracks the changing positions of points in sequences of images impinging on the retina (see Section 4). Building on Bordwell's account, I suggest that language-mediated inferences about static and dynamic spatial relations are crucial for narrative. In such an analysis, the spatial concepts are best represented qualitatively, a proposal which may be novel to humanities (including film) narratologists.

Figure 1 (a) Bugs entering the painting. (b) Woozy motion.

In any medium, manner of movement can be relevant for inferring character properties, including affect and traits. However, cartoon motion seems very different from the motions characters undergo in narrative texts (e.g., the ghost stories studied by [16]). In addition to expressing a parody of motion in the real world, cartoon physics allows for all sorts of creative manners of motion. Consider, as an example, the manner of motion in the 2003 cartoon movie *Looney Tunes: Back in Action*(LTBIA)[1]. In Figure 1(a), Bugs Bunny is about to leap into the painting *The Persistence of Memory* to escape Elmer Fudd, but in Figure 1(b), his motion within the landscape of the painting has become sluggish, as has that of Daffy who has joined him, as their shapes experience Daliesque distortions in this embedded storyworld. Such woozy movements are narratologically significant, as they convey struggle as well as exhaustion.

A large corpus of cartoon movies annotated with systematic characterizations of motion and other narratologically-relevant information would be useful in examining similarities and differences across medium and genre. Such an effort could have implications for humanities narratology [29] by way of providing a precise conceptual framework to enhance narratological theories for media such as cartoons. Corpora annotated with such representations can potentially also be of practical use for training algorithms aimed at movie scene search and summarization.

In 'silent' cartoon movies like the Road Runner ones, the fact that we are speakers of natural language influences our narratological inferences, even when these are drawn from non-textual media. This suggests that annotating the narrative content of a movie should start with an ekphrasis consisting of brief descriptions in natural language. Using natural language descriptions as an additional input for the annotator, beyond the video, not only leverages information that is not directly present in the video, but in addition allows one to harness the rich conceptual resources that natural language provides. Earlier work by [27] has described how qualitative spatial representations can be used to formally represent and reason about well-known aspects of the semantics of spatial prepositions and motion verbs. The contribution of this paper is twofold: extending the representations in NarrativeML to incorporate motion, and the application to the narratives in animated cartoons.

NarrativeML is based on multiple layers of annotation, relying on tagging predicates and arguments in the sentences of the text using PropBank [33]. Events and their temporal relations are represented using TimeML [34], which in turn leverages the interval calculus [1]. The automatic application of TimeML to classical narratological analyses of text is discussed in [25]. NarrativeML also includes a partial temporal ordering of narrative events that share a common protagonist, called a Narrative Event Chain (NEC) [4]. Once incidental events are pruned away, the NEC answers the question as to what the protagonist did in the story. References to places and simple static relations between them are modeled using SpatialML

[1] `https://www.youtube.com/watch?v=97PLr9FK0sw`.

[26] and ISO-Space [36]. All these concepts form part of the fabula (or story). NarrativeML also represents the mapping to *sjuzhet* (or discourse), including the seven varieties of ordering described by [11] as well as narrative tempo and subordinated discourse[2].

In contrast to film narratology such as [2], NarrativeML takes the position that there is always a narrator, but she may or may not be present in the film. For focalization, the annotator of a movie will have to record from whose point of view the scene is being displayed, deciding whether it is the narrator, the 'camera', the audience, or a particular character. Here film presents a challenge. Genette's three-way characterization of focalization into omniscient, internal, and external as in [11], [12] is text-based and involves overlapping categories, as [8] among others has argued. In film, as [21] points out, there may be many different shades of focalization, based on camera angles, deep focus, shot length and scale, etc. A related question is what sort of theory of mind the narrator has with respect to the characters; in the case of a silent movie, gazing into minds may be realized by thought balloons or the focus of attention of the 'camera'. NarrativeML sidesteps the complexities here by allowing for focalization a fourth mixed category, called OTHER, while requiring that the annotator record the position of the viewer relative to figure and ground objects. Thus, above and beyond its role in motion, spatial representation is key to capturing narrative information related to focalization.

2 Spatial Representation

The spatial representations discussed here are motivated by linguistic analyses of prepositions and motion verbs across languages. Being qualitative and linguistically motivated, they are at an entirely different level of abstraction from the fine-grained ones used in animation systems. However, as I will argue, they are useful in representing static and dynamic spatial aspects of narrative.

2.1 Static Spatial Relations

The analysis of spatial prepositions and adpositions in language has come from a variety of theoretical frameworks, including AI and psycholinguistics, e.g., [30], descriptive linguistics, [17], formal semantics, [46], and cognitive linguistics, [23], [18]. Much of the analysis has focused on representing phrases like "the book on the table" and "the fruit in the bowl" in terms of *topological relations* between objects involving notions of coincidence, contact and containment. To formally represent such relations, ISO-Space, and thus NarrativeML, uses the Region Connection Calculus (RCC-8) [37]. In RCC-8, objects are conceived as non-empty, equi-dimensional regions. Based on a single primitive relation of connection between regions, RCC-8 defines the set of eight base relations shown in Figure 2(a). Thus, "the book on the table" may be represented by `EC(book, table)` and "the fruit in the bowl" by `IN(fruit, bowl)`, where `IN` is the disjunction of the base relations `TPP` and `NTPP`.

In addition to topological relations, languages distinguish spatial relations that reflect *orientations* of objects. Studies across languages [24] reveal that they use a basic inventory of three varieties of coordinate systems to describe orientation, that are unevenly distributed across languages. In the *intrinsic* frame, used in examples like "in the front of the picture" and "by the side of the boat", the linguistic relation R between a figure object (F) and

[2] Other aspects of NarrativeML, involving characters, their goals, plot structure, and audience responses are not discussed here.

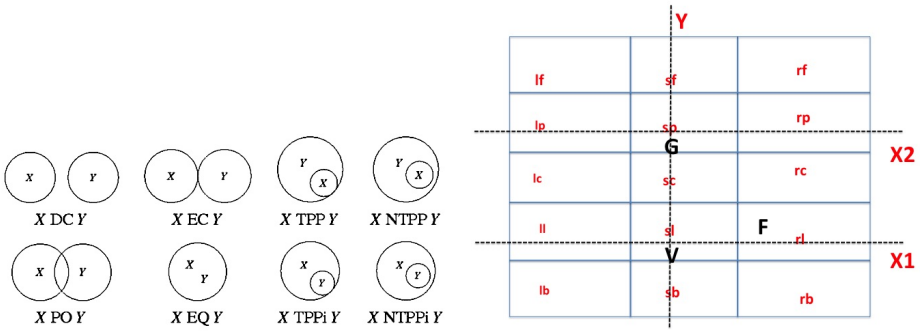

Figure 2 (a) *RCC-8* relations. (b) *Double Cross Calculus* example.

a ground object (G) is characterized in terms of particular facets of the ground object G, e.g., "front", "nose", "sides", etc., which are dependent on the object's affordances and are highly culture-dependent. In this frame, F lies "in a search domain extending from G on the basis of an angle or line projected from the center of G, through an anchor point A (usually the named facet 'R')" [24] (p. 42-3). The *absolute frame* of reference, e.g., "due north of St. Croix", involves a coordinate system where F is described in terms of fixed bearings (related to compass points and/or landscape markers) with respect to an origin on G. The *relative frame* involves a ternary relation, between F, G, and a third object, the viewer V, as in examples like "to the right of Bugs". Here a coordinate system is centered on V, with possibly another coordinate system centered on G arising from a geometric projection from V's coordinate system to G's, in turn providing intrinsic facets to G via V. Languages that have a relative frame always have an intrinsic frame as well, introducing ambiguity.

The ISO-Space representation, and as a result, NarrativeML, is neutral with respect to which qualitative representations should be used to capture orientation relations. Here we introduce three representations that will be used in the example under discussion.

A representation relevant to the *intrinsic frame* is the Dipole Calculus of [32], [7], which represents spatial relations based on oriented line segments called dipoles. Each dipole divides the plane into a left and right half, and the calculus accordingly specifies orientation relations between the start and end points of each dipole and the other. A start or end point on dipole B can be relatively to the left (l) or the right (r) of, or else start (s) or end (e) of, dipole A. Thus, in Figure 1(b), `llrr(Bugs, Daffy)`, meaning that the start and end of Daffy are to the left of Bugs and the start and end of Bugs are to the right of Daffy. This representation is compatible with "Daffy is to the left of Bugs" and "Bugs is on Daffy's right". When augmented with additional orientations: back (b), interior (i), and front (f), one gets a calculus with 69 base relations [32], which we will refer to as DC-69.

The *absolute frame* can be represented in the Cardinal Direction Calculus of [15], [39]. Here the minimum bounding rectangle of the ground region is made the central tile of a 9-element grid, and is labeled 'B', for bounding box. The figure region is then positioned on the grid, and the tiles it falls into are used to describe its orientation with respect to that central tile, yielding nine regions in all: B, S, SW, W, NW, N, NE, E, and SE. Thus, in Figure 1(b), with B over Bugs, we have `ESE(Daffy, Bugs)`. Given that the calculus has a base set of 511 relations, we will refer to it as CDC-511.

For the *relative frame* of reference, the Double Cross Calculus (DCC) of [10], [38], is relevant. Here we have a ternary relation between figure, ground, and viewer. As shown in Figure 2(b), the figure object F, viewer V, and ground G are construed as points, and a

line Y from V to G is extended to create a pair of half-planes, left (l) and right (r). A pair of lines, one (X1) perpendicular to the line Y and through V, and the other (X2) parallel through it and through G, creates three regions, forward (f), back (b), with a central region (c) in between. Consider applying it to F=Bugs in Figure 1(a). He is in the plane between the viewer and the ground G, the painting, so we have `rf(Bugs, PersistenceofMemory, Viewer)`. This is compatible with "Bugs is in front of the painting" and "Bugs is on the right in front of the painting". Likewise, in Figure 1(b), Daffy is to the right of Bugs from the viewer's point-of-view, so we have `rc(Daffy, Bugs, Viewer)`. Adding the relations of equality and inequality, we get a base set of 17 relations (DCC-17).

2.2 Motion

Having represented aspects of time and space, one needs to incorporate motion into NarrativeML. A fundamental cross-linguistic insight regarding motion comes from Leonard Talmy [40], [41], who points out that languages have two distinct strategies for expressing concepts of motion. In *manner-type languages* (English and other Germanic languages, also Slavic languages), the main verb expresses the manner or cause of motion, while path information is expressed elsewhere in the form of 'satellite' constituents[3]. In contrast, in *path-type languages* (Romance, Turkish, Semitic, and other languages), the verb expresses the path, whereas the manner is optionally expressed by adjuncts.

Adopting this classification, which has been extensively studied cross-linguistically along with its exceptions, [27] introduce a procedural semantics for motion in natural language, where motion is viewed in terms of transitions in spatial configurations. A distinction is made between *action-based predicates* (for manner-of-motion verbs like "bike", "drive", "fly", etc.) and *location-based predicates* (e.g., for path verbs like "arrive", "depart", etc.). Action-based predicates do not make reference to distinguished locations, but rather to the 'assignment' and 'reassignment' of locations of the object, through the action. The location-based predicates focus on points on a path, and thus they reference a distinguished location, and the location of the moving object is 'tested' to check its relation to this distinguished value.

The semantics for these predicates is expressed in Dynamic Interval Temporal Logic (DITL) from [35], a first-order dynamic logic (introduced by James Pustejovsky) where events are modeled as programs, and states refer to preconditions or post-conditions of these programs. This approach to modeling the semantics of motion, is explained in detail in [27]. The following programs, from [27] (p. 95-107), describe the basic constructs of motion needed.

Definition 1 shows how directed movement away from a source is represented in DITL[4]:

▶ **Definition 1** (Moving away).

```
DITL move_away(c, src) ≡ y:=src;
    (loc(c):=z, z ≠ y, dist(y, src) < dist(z, src); y:=z)+ /*
1.      Assign y to object location.
2.      Then reassign its location to z, which is further away
        from source than y.
3.      Iterate steps 1–2 one or more times.   */
```

[3] A satellite is "any constituent other than a noun-phrase or prepositional-phrase complement that is in a sister relation to the verb root" [40] (p. 102).

[4] In DITL, semicolon is a program sequencing operator and comma is a (higher-precedence) predicate conjunction operator.

One can now define non-primitive programs corresponding to motions that are lexicalized by motion verbs. Arriving is shown in Definition 2.

▶ **Definition 2** (Arriving as making contact at end of path).

```
reach(c, dest) ≡ (y:= loc(c); RCC-8DC(y, dest)?; move_toward(c, dest))+;
     (y:= loc(c); RCC-8EC(y, dest)?) /*
1.      Test if object is disconnected from the destination.
2.      If so, move towards the destination
3.      Iterate steps 1–2 one or more times
4.      Test if object touches the destination.  */
```

Manner of motion is not treated as a primitive, but arises as an elaboration of the components of the motion, namely figure, ground, event, path, and medium. This allows one to distinguish various manners of motion; for example, one can define sliding (Definition 3), which involves maintaining an extended connection with a surface, as well as bouncing (Definition 4), which involves alternating between an extended connection and disconnection.

▶ **Definition 3** (Sliding).

```
slide(c, surf) ≡ y:=loc(c),
    (loc(c):=z, z ≠ y, RCC-8EC(z, surf); y:=z)+
```

▶ **Definition 4** (Bouncing).

```
bounce(c, surf) ≡ y:=loc(c),
     (loc(c):=z, z ≠ y, RCC-8EC(z, surf); y:=z;
      loc(c):=z, z ≠ y, RCC-8DC(z, surf); y:=z)+
```

For representing affect associated with manners of motion, one has to introduce additional features into the framework. Here I build on the approach of [5], [45], who use natural language in input specifications to drive the motion of animated characters. I focus here on *Effort*, a concept taken from analysis of dance [22]. Effort is characterized (Table 1) in terms of four factors: Space, Weight, Time and Flow, with the left and right columns labeling the low and high ends respectively of a scale.

Thus, the woozy movement in Figure 1(b) is represented in NarrativeML as the event e, where `effort(e, f1)` is associated with the four factors, each on a five-point scale: space and weight as `space(f1, very_low)` & `weight(f1, very_high)`, with time and weight as `time(f1, very_low)` & `flow(f1, very_high)`. Bugs' and Daffy's flight across the landscape of the painting is increasingly tortured and slow, so in previous frames the flow value would have been freer.

Prolonged braking, a device essential to Road Runner and other cartoons, may be viewed as sliding with decreasing speed, as seen in Definition 5. A frazzled variant can be expressed via its Effort.

▶ **Definition 5** (Prolonged Braking).

```
slow-brake(c, surf) ≡ y:=loc(c);
    (loc(c):=z, z ≠ y, RCC-8EC(z, surf), speed(c, z) < speed(c, y);
     y:=z)+
```

Table 1 Effort in Laban's system, from [5].

Space: attention to the surroundings	
Indirect: flexible, meandering, wandering, multi-focus *waving away bugs, slashing through plant growth surveying a crowd of people, scanning a room for misplaced keys*	**Direct**: single focus, channeled, undeviating *pointing to a particular spot, threading a needle, describing the exact outline of an object*
Weight: attitude towards the impact of one's movement	
Light: buoyant, delicate, easily overcoming gravity, marked by decreasing pressure *dabbing paint on a canvas, pulling out a splinter, describing the movement of a feather*	**Strong**: powerful, having an impact, increasing pressure into the movement *punching, pushing a heavy object, wringing a towel, expressing a firmly held opinion*
Time: lack or sense of urgency	
Sustained: lingering, leisurely, indulging in time *stretching to yawn, stroking a pet*	**Sudden**: hurried, urgent *swatting a fly, lunging to catch a ball, grabbing a child from the path of danger, making a snap decision*
Flow: amount of control and bodily tension	
Free: uncontrolled, abandoned, unable to stop in the course of the movement *waving wildly, shaking off water, flinging a rock into a pond*	**Bound**: controlled, restrained *moving in slow motion, tai chi, fighting back tears, carefully carrying a cup of hot liquid*

3 Annotation Example

Sheep in the Island is a 2007 'silent' cartoon film from Korea that features a sheep stranded on a tropical island with a dragon duck[5]. It is a shipwreck narrative, with typical themes of dominance over nature and survival on a deserted island. Inspired by K-Pop culture, the film aims for universal appeal by limiting the presence of text and restricting the audio to non-linguistic verbal sounds and instrumental background music. It thus provides a simple test case for ekphrasis-based narrative annotation. A few sample frames relevant to the discussion below are shown in Figure 3.

The narrative is pre-segmented into sets of time intervals in the video, suggestive segment labels indicated with line comments (//). The time intervals are ordered chronologically, but are not contiguous. The input given to the annotator is shown here highlighted in yellow in Annotation 1. Its ekphrasis is shown alongside, along with the indices of events, entities, and times in NarrativeML[6].

▶ **Annotation 1** (SHEEP IN THE ISLAND).

```
1. // SETTING
2. 0:02-0:07_t1 island_x1 with rock_x2 and sand_x4 seen across sea_x3
3. // BOATS IN MOTION
4. 0:08-0:12_t2 gunboat_m1 approaches_e1 from right partly in front of island_x1
5. 0:13-0:17_t3 gunboat_m1 approaches_e2 seen from front looming large
6. 0:18-0:20_t4 gunboat_m1 approaches_e3 from left , seen from island_x1
7. 0:21-0:26_t5 larger boat_m2 approaches_e4 from right as gunboat_m1
                approaches_e5 from left
```

[5] https://mayhemandmuse.com/sheep-in-the-island-part-1/ and https://www.youtube.com/watch?v=YvR8LGOUpNA.

[6] This paper and the annotation environment use logical expressions rather than the underlying XML to which it is mapped. XML DTDs for NarrativeML are at http://tinyurl.com/inderjeetmani/home/NarrativeML.

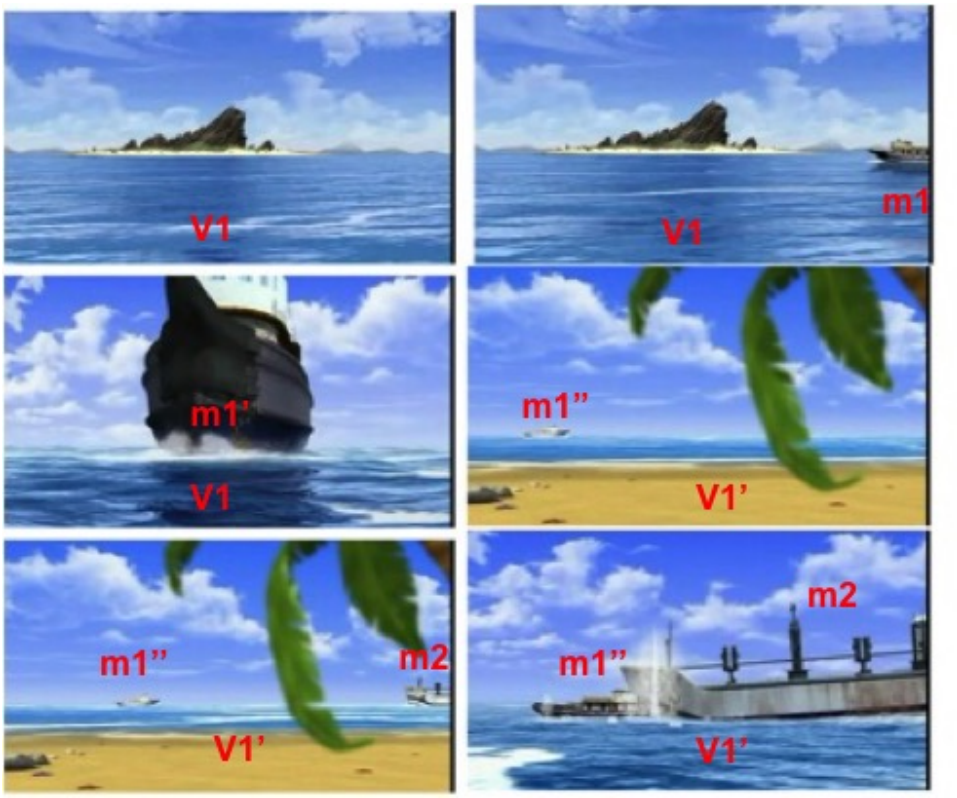

Figure 3 *Sheep in the Island* at 7, 11, 14, 19, 25, and 34 seconds.

```
8.  // A SHIPWRECK
9.  0:33-0:38t6 boatsm1,m2 crashe6
10. 0:38-0:41t7 boatsm1,m2 sinke7 as three boxesm3,m4,m5
              floate8 towards islandx1
11. 0:42-O:46t8 one boxm3 arrivese9 on islandx1
12. 0.47-1:05t9 boxm3 bounchese10 on sandx4
13. // ENTER THE SHEEP
14. 1:10-1:14t10 sheepc1 emergese11 from boxm3, seen from above
15. 1:15-1:17t11 sheepc1 jumpse12 and landse13 on sandx4
16. 1:20-1:22t12 sheepc1 approachese14
17. 1:23-1:26t13 sheepc1 turnse15 and walkse16 away
18. 1:27-1:28t14 sheepc1 turnse17 facing forward in head shot
19. // A HUNT INTERRUPTED
20. 1:30-1:32t15 sheepc1 observese18 frogc2 hoppinge19 on sandx4 in front
21. 1:33-1:35t16 sheepc1 pursuese20 frogc2
22. 1:36-1:41t17 sheepc1 catchese21 and holdse22 frogc2
23. 1:41-1:42t18 sheepc1 gets readye23 to devoure24 frogc2
24. :43-1:48t19 sheepc1 noticese25 a large boxm4 to its left
25. 1:49-1:56t20 sheepc1 slamse26 frogc2 about
26. 1:58-2:03t21 sheepc1 strollse27 around boxm4 to right edgez1,
              with stampz2 'DANGER' on front facez3
27. 2:05-2:06t22 boxm4 shakese28
28. // ENTER THE DRAGON
29. 2:07-2:11t23 dragon clawy4 emergese29 from boxm4,
              seen from above along with sheepc1
```

The individuation of events is based on the text, annotated in TimeML along with the time intervals[7]. The crucial thing in the BOATS IN MOTION segment is that in line 4, the gunboat m1 (with a front-protruding gun in the video) is seen in profile heading to the left parallel to the viewer V1 who is away from the island. This can also be seen visually in Figure 3 at 11 seconds.

Then, in line 5, the scene switches to a front shot of the same boat (Figure 3 at 14 seconds), the inference being that the viewer has changed orientation, not the boat[8]. In line 6, the gunboat is now seen from the island where the viewer now is, instead of from the sea, and it is now to the left of and parallel to the viewer (Figure 3 at 19 seconds). In line 7, the larger boat m2 (not the gunboat) approaches from the right, with the viewer still in the same position on the island (Figure 3 at 25 seconds), gearing up for a collision in the next segment A SHIPWRECK (line 9 ff., and Figure 3 at 34 seconds).

▶ **Annotation 2** (SETTING).

```
1. 0:02-0:07_t1 island_x1 with rock_x2 and sand_x4 seen across sea_x3
2. narrative(i1) & medium(i1, cartoon_animation) & narrative(i2)
    & medium(i2, text_annotation) & narrative_segment(i1, i3)
    & title(i3, 'SETTING')
3. & narrator(i1, N0) & narrator_type(N0, absent) & narrator(i2, N1)
4. & narrative_time(N0, =) & narrative_time(N1, =)
5. & narrative_order(N0, CHRONICLE) & narrative_order(N1, CHRONICLE)
6. & RCC-8EC (x2, x4) // rock_x2 is connected to sand_x4
7. & RCC-8NTPP (x4, x1) // sand_x4 is part of island_x1
8. & RCC-8NTPP (x2, x1) // rock_x2 is part of island_x1
9. & RCC-8EC (V1, x3) // Viewer_V1 is on sea_x3
10. & DCC-17sf(x1, x3, V1)
    // island_x1 is in far background with respect to Viewer_V1
```

Annotation 2 shows the NarrativeML annotation of the SETTING segment. Line 2 distinguishes the filmic narrative from the textual description. Line 3 indicates that the narrator of the description is in fact the annotator N1, differentiated from the filmic narrator N0, who is absent. Line 4 states that N1 narrates the scene descriptions as in a running commentary, so that the narrative time is simultaneous. The filmic narrator is also not using any devices to suggest retrospective or other temporal distance. Line 5 indicates that the events are narrated by the film as well as by the annotator in (i.e., CHRONICLE) order of occurrence. The RCC-8 relations in lines 6-9 capture coarse-grained topological relations in the SETTING, and the Double Cross Calculus (DCC-17) in line 10 is used to convey point of view, namely the relative frame where the viewer 'camera' is shooting across the sea to the island.

▶ **Annotation 3** (BOATS IN MOTION).

```
1. 0:08-0:12_t2 gunboat_m1 approaches_e1 from right
                partly in front of island_x1
2. IC-13EQUAL(e1, t2) & @(RCC-8DC(m1, x1), e1)
            // gunboat_m1 is disconnected from island_x1
3. & narrative_segment(i1, i4) & title(i4, 'BOATS IN MOTION')
```

[7] The BEFORE temporal relations indicating the chronological ordering of events in the fabula are left out for reasons of space.

[8] I use prime notation (V1', m1', etc.) in Figure 3 to remind the reader of an object's changed viewpoint.

```
4. & @(RCC-8EC(m1, x3), e1)          // gunboat_m1 floats on sea_x3
5. & face(m1, y1) & @(DC-69rrrl(y1, m1), e1) // left face_y1 of gunboat_m1
6. & @(DCC-17rc(y1, x1, V1), e1)
     // left face_y1 is between island_x1 and viewer
7. & @(DITLmove_away(y1, RB), e1)
    // RB = right boundary of viewing frame
8. 0:13-0:17_t3 gunboat_m1 approaches_e2 seen from front looming large
9. IC-13EQUAL(e2, t3) & @(RCC-8DC(m1, x1), e2) & effort(e2, f1)
10. & space(f1, very_high) & weight(f1, very_high)
     & time(f1, high) & flow(f1, high)
11. & @(RCC-8EC(m1, x3), e2) & edge(m1, y2) & @(DC-69sbsi(y2, m1), e2)
     // front edge_y2 of gunboat_m1
12. & @(DITLmove_toward(y2, V1), e2)
13. 0:18-0:20_t4 gunboat_m1 approaches_e3 from left, seen from island_x1
14. RCC-8EC(V1, x1) & IC-13EQUAL(e3, t4) // viewer is on island_x1
15. & @(RCC-8EC(m1, x3), e3)
16. & face(m1, y3) & @(DC-69lllr (y3, m1), e3)
      // right face_y3 of gunboat_m1
17. & @(DCC-17lf( y3, x1, V1), e3)
       // right face_y3 is to the left of viewer
18. & @(DITLmove_toward(y3, RB), e3)
19. 0:21-0:26_t5 larger boat_m2 approaches_e4 from right as gunboat_m1
           approaches_e5 from left
20. RCC-8EC(V1, x1) & IC-13EQUAL(e4, t5) & effort(e4, f2)
21. & space(f2, very_high) & weight(f2, high) & time(f2, neutral)
         & flow(f2, high)
22. & IC-13EQUAL(e5, t5) & effort(e5, f3)
23. & space(f3, high) & weight(f3, high)
         & time(f3, neutral) & flow(f3, neutral)
24. & @(RCC-8EC(m2, x3), e4) & @(RCC-8EC(m1, x3), e5)
         // boats float on sea_x3
25. & @(RCC-8DC(m2, x1), e4) & @(RCC-8DC(m1, x1), e5)
         // boats disconnected from island_x1
26. & face(m2, y4) & @(DC-69rrrl(y4, m2), e4)    // left face_y4 of boat_m2
27. & face(m1, y3) & @(DC-69lllr(y3, m1), e5)
         // right face_y3 of gunboat_m1
28. & @(DCC-17lf( y3, x1, V1), e5)
         // right face_y3 is to the left of viewer
29. & @(DCC-17rf( y4, x1, V1), e4)
         // left face_y4 is to the right of viewer
30. & @(DITLmove_away(y4, RB), e4) & @(DITLmove_toward(y3, RB), e5)
```

Annotation 3 turns to motion, which has until now not been represented in NarrativeML. In line 2, the @ predicate indicates that the separation of the gunboat from the island holds throughout e1. In line 5, the intrinsic left face y1 of the gunboat is characterized with an additional primitive spatial relation called `face`, using the Dipole Calculus (DC-69) to represent the left one, i.e., the gunboat dipole m1 is viewed as to the right and orthogonal to the left face dipole, i.e., `y1`↑`m1`→, yielding the relation `rrrl(y1, m1)`. This left face is moving away from the right boundary, as indicated by the $\texttt{move}_{away}$ predicate in line 7. In line 8, the scene changes to the front view of the gunboat, with its increased Effort, impelled as if by a sinister force, indicated in line 10. The gunboat's intrinsic front `edge` (another primitive) y2 is identified in line 11 using DC-69, where the two dipoles are represented as being on the same line. The DC-69 relation `sbsi(y2, m1)` expresses the fact that the start

of the gunboat m1 is at the start of its front edge and its end is behind its front edge, and the start of its front edge is at the start of the gunboat and its end is in the interior of the gunboat. The gunboat's front edge y2 is moving towards the viewer as indicated in line 12. Capturing the fact that the gunboat is speeding towards the viewer V1 while looming steadily larger is narratologically important, as actions with the viewer as target have the potential to increase suspense.

The movement to the right of the other boat is captured in the remaining lines. The Effort of the boats approaching each other is indicated in lines 21 (larger boat) and 23 (gunboat), with the larger boat with its greater apparent momentum indicated by increased Effort.

▶ **Annotation 4** (A SHIPWRECK).

```
1. 0:33-0:38_t6 boats_m1,m2 crash_e6
2. RCC-8EC(V1, x1) & IC-13EQUAL(e6, t6)
   & @(reach(m1, m2), e6) & narrative_segment(i1, i5)
   & title(i5, 'A SHIPWRECK')
3. & @(RCC-8EC(m1, x3), e6) & @(RCC-8EC(m2, x3), e6)
4. & @(RCC-8EC(m2, x3), e6) & @(RCC-8EC(m1, x3), e6)
5. & @(RCC-8DC(m2, x1), e6) & @(RCC-8DC(m1, x1), e6)
6. & @(RCC-8PO(y3, y4), e6)
      // right face of gunboat_m1 telescopes into left face of boat_m2
7. & effort(e6, f4) & space(f4, very_high) & weight(f4, very_high)
   & time(f4, very_high) & flow(f4, very_high)
8. 0:38-0:41_t7 boats_m1,m2 sink_e7 as three boxes_m3,m4,m5
            float_e8 towards island_x1
9. & face(x3, y5) & DC-69sfsi(y5, x3) // bottom of sea_x3
10. & RCC-8EC(V1, x1) & @( DITLmove_toward(m1, y5), e7)
    & @( DITLmove_toward(m2, y5), e7)
    & IC-13EQUAL(e7, t7) & IC-13EQUAL(e8, t7)
11. & @(RCC-8NTPP(m1, x3), e7) & @(RCC-8NTPP(m2, x3), e7)
      // boats submerged
12. & @(RCC-8EC(m3, x3), e8) & @(RCC-8EC(m4, x3), e8)
    & @(RCC-8EC(m5, x3), e8)
13. & @(DITLmove_toward(m3, x1), e8) & @(DITLmove_toward(m4, x1), e8)
    & @(DITLmove_toward(m5, x1), e8) // towards island_x1
14. 0:42-O:46_t8 one box_m3 arrives_e9 on island_x1
15. RCC-8EC(V1, x1) & IC-13EQUAL(e9, t8) & @(reach(m3, x1), e9)
16. 0.47-1:05_t9 box_m3 bounces_e10 on sand_x4
17. RCC-8EC(V1, x1) & IC-13EQUAL(e9, t9) & @(bounce(m3, x4), e10)
18. & NEC(m1, <e1,..,e7> & NEC(m2, <e4, e6, e7>)
19. & NEC(m3, <e8 ,..., e10)>) & NEC(m4, <e8>) & NEC(m4, <e8>)
20. & effort(e10, f5) & space(f5, low) & weight(f5, very_low)
   & time(f5, low) & flow(f5, low)
```

Annotation 4 begins with the boats crashing, which is seen as the right face of the gunboat telescoping into the left face of the larger boat (Figure 3 at 34 seconds). Line 7 indicates that the Effort is at the maximum for all its factors. In line 11, the boats are submerged below the sea, expressed in RCC-8. Line 12 has the three boxes floating on the sea, and in line 14 they move towards the island. The boxes emerge as by-products born of the crash, which is an early inflexion-point in the plot. In line 15, one box reaches the island, and in line 17, it bounces on the sand. Lines 18-19 indicate the NECs for the boats and the boxes. Line 20 characterize the effort involved in the bouncing of box m3, which is

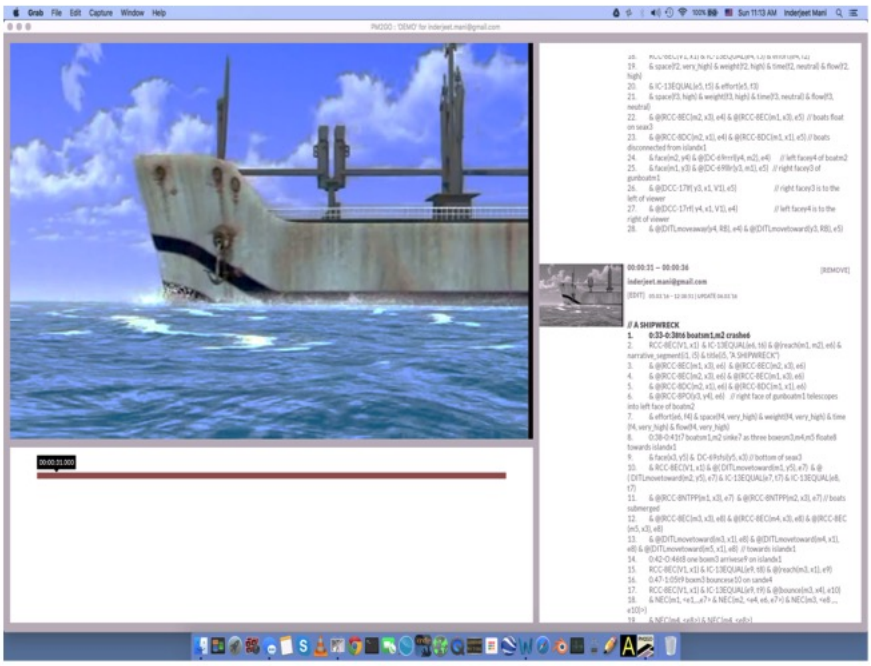

Figure 4 Annotating A SHIPWRECK at 0:33–0:38.

relatively unconstrained, propelled as the box is by the energy of the creature trapped inside. The self-propelled bouncing of the box foreshadows the emergence of new characters. Thus, although the entities in motion in the first three annotated segments (boats and boxes) do not involve the lifelike characters of the sheep and dragon duck, annotating their specific motions is relevant for plot structure as well as foreshadowing the arrival of those characters.

4 Annotation Challenges

Figure 4 shows the video annotation tool PM2GO being used to annotate segments of *Sheep in the Island*[9]. The video is shown on the left, with the player and interval selection below, and the annotations on the right: BOATS IN MOTION, above, and A SHIPWRECK below, using Annotations 3 and 4, respectively.

While individual movie ekphrases might be generated by crowd-sourcing, the annotations are too dense to be efficiently executed for large corpora without some level of automatic preprocessing. The good news here is that progress has been made on automatic labeling of semantic roles for PropBank, e.g., [14], SpatialML tagging [26] and Semantic Role Labeling (in the SemEval tasks) for figure-ground spatial relations [20]. For automatic TimeML tagging, there has been progress as well, though approaches seemed to have hit a ceiling of 70% F-measure on event-ordering across languages and tasks, e.g., TempEval [43], in part due to the paucity of annotated data[10]. Unfortunately, the annotation using PM2GO

[9] See `http://motionbank.org/sites/motionbank.org/files/pm2go_handbook_07_14.pdf`.
[10] Narrative texts auto-tagged with TimeML are available at `http://tinyurl.com/inderjeetmani/home/NarrativeML`.

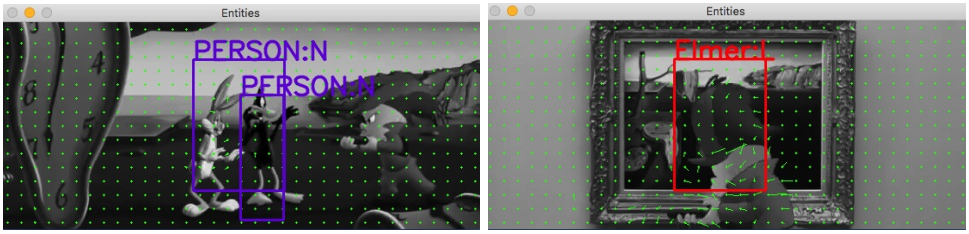

Figure 5 (a) Person detection. (b) Character and motion labeling.

does not use any automatic pre-processing. Integrating the TARSQI toolkit for TimeML tagging[11] and the SpatialML tagger[12] into an annotation pipeline is nontrivial since they are legacy software systems. Longer-term plans include re-implementing such capabilities on top of the far more modern Stanford CoreNLP toolkit[13] as well as migrating to a more narrative-friendly annotation workbench for video.

So far, the annotation of motion itself has not been automated. One possibility here is to leverage the field of computer vision, which has been advancing rapidly. It seems reasonable to populate some of the ekphrases and their annotations with suggestions from video processing. Figure 5 shows some results from applying computer vision tools from OpenCV[14] to *Looney Tunes: Back in Action*. In Figure 5(a), Bugs and Daffy have been classified as people using a Histogram of Oriented Gradients (HOG) [6] pre-trained on images of people; note that Elmer has been missed. Figure 5(b) shows that Elmer has been detected as an object and labeled correctly, using the Haar classifier cascade of [44], trained on labeled images from a corpus of Bugs Bunny cartoons. The system has also correctly identified Elmer's direction of movement (left) using an optical flow detector [9]. In addition to improving the accuracy of such computer vision methods with more training data, it should be possible to extend them to automatically label the type of motion, as in [42].

While NarrativeML has been used to annotate numerous examples, it has not as such been applied to text corpora in the large, let alone to ekphrases for movies, so important questions of annotation reliability and efficiency remain open. These latter questions are the focus of current research, applied to a corpus of cartoon movies. To simplify the task, the pre-selected set of frames to be annotated is restricted to relatively short time intervals, with the guidelines focused on creation of the ekphrasis and its NarrativeML for that set.

5 Conclusion

In terms of expressiveness, these additions to NarrativeML (constituting version 0.2) allow for the annotation of relevant narrative information in cartoon movies, at a level of abstraction guided by natural language and representing key semantic distinctions related to space and motion. The annotation scheme is thus attractive for representing spatial relations, focalization and motion in cartoons, and could potentially be used for humanities narratology and practical applications as described in Section 1. The scheme might also be embedded in authoring environments for animation.

[11] http://www.timeml.org/tarsqi/toolkit/download.html
[12] http://www.timeml.org/tarsqi/toolkit/download.html
[13] http://nlp.stanford.edu/software/corenlp.shtml
[14] http://opencv.org

Of course, there is still much that is missing that would shed light on narrative. For the intrinsic frame, where object shape is important, the dipole calculus is not that suitable. For focalization, there needs to be a characterization of relevant shot types, as discussed in [19], as well as the varieties of shot transition or cut. The varying distance, focus, orientation, and area of interest of the 'camera' are also crucial for film narrative. In addition, for the cartoon genre, character shape, as well as more elaborate motion manners and their velocities may be revealing of character affect. Recording this sort of information in narrative corpora could be very valuable. Nevertheless, reasoning with such qualitative representations is not always tractable, and maximal tractable subsets of calculus relations, when found, often require discarding key relations. Combining representations and adding dimensions only add to the complexity. Finally, there are numerous annotation challenges discussed in Section 4, some of which can be addressed by computer vision.

References

1 James Allen. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, 26(11):832–843, 1983.

2 David Bordwell. *Narrative in the Fiction Film*.Madison: University of Wisconsin Press, 1985.

3 David Bordwell. *Common Sense + Film Theory = Common-Sense Film Theory?* `http://www.davidbordwell.net/essays/commonsense.php`.

4 Nathanael Chambers. *Inducing Event Schemas and their Participants from Unlabeled Text*. Ph.D. Dissertation, Department of Computer Science, Stanford University, 2011.

5 Diane Chi. *A Motion Control Scheme for Animating Expressive Figure Arm Movements*. PhD Thesis. University of Pennsylvania, 1999.

6 Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CPVR)*, 2005, pages 886–893.

7 Frank Dylla and Reinhard Moratz. Empirical complexity issues of practical spatial reasoning about relative position. In *Workshop on Spatial and Temporal Reasoning at ECAI 2004*, Valencia, Spain, August 2004.

8 William F. Edmiston. *Hindsight and Insight: Focalization in Four Eighteenth-Century French Novels*. University Park, PA: Penn State University Press, 1991.

9 Gunnar Farnebäck. Two-frame Motion Estimation Based on Polynomial Expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis*, pages 363–370, 2003.

10 Christian Freksa. Using orientation information for qualitative spatial reasoning. In A. U. Frank, I. Campari, and U. Formentini (eds.), *Theories and methods of spatiotemporal reasoning in geographic space*, Springer, Berlin, pages 162–178, 1992.

11 Gerard Genette. *Narrative Discourse* (trans. Jane Lewin). Ithaca: Cornell University Press, 1980.

12 Gerard Genette. *Narrative Discourse Revisited* (trans. Jane Lewin). Ithaca: Cornell University Press, 1988.

13 James J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, 1950.

14 Daniel Gildea and Daniel Jurafsky. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3), pages 245–288, 2002.

15 R. Goyal and M. J. Egenhofer. Consistent queries over cardinal directions across different levels of detail. In *Proceedings of the 11th International Workshop on Database and Expert Systems Applications*, 2000.

16 David Herman. Spatial Cognition in Natural-Language Narratives. In M. Mateas and P. Sengers (eds.), *Working notes of the Narrative Intelligence Symposium*, pages 21–25. AAAI Fall Symposium Series. Menlo Park, CA: AAAI Press, 1999.

17 A. Herskovits. *Language and Spatial Cognition: an interdisciplinary study of the prepositions in English*. Cambridge University Press, 1986.

18 Ray Jackendoff. *Semantic Structures*. Cambridge, MA: MIT Press, 1990.

19 Manfred Jahn. *Narratology: A Guide to the Theory of Narrative*. English Department, University of Cologne, 2003. `http://www.uni-koeln.de/~ame02/ppp.htm`.

20 Parisa Kordjamshidi, M. Van Otterlo, and M. F. Moens. Spatial role labeling: Towards extraction of spatial relations from natural language. *ACM Transactions on Speech and Language Processing (TSLP)*, 8 (3), 4, 2011.

21 Markus Kuhn and Johann N. Schmidt. Narration in Film. In Peter Huhn et al. (eds.), *The Living Handbook of Narratology*, paragraph 28. Hamburg: Hamburg University Press, 2014. `http://www.lhn.uni-hamburg.de/article/narration-film-revised-version-uploaded-22-april-2014`.

22 Rudolf Laban and F. C. Lawrence. *Effort: Economy in Body Movement*. Plays, Inc., 1974.

23 G. Lakoff. *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press, 1987.

24 S. C. Levinson. *Space in Language and Cognition*. Cambridge University Press, 2003.

25 Inderjeet Mani. *The Imagined Moment*. Lincoln: University of Nebraska Press, 2010.

26 Inderjeet Mani, Christine Doran, David Harris, Justin Hitzeman, Robert Quimby, Justin Richer, Ben Wellner, Scott Mardis, and Seamus Clancy. SpatialML: annotation scheme, resources, and evaluation. *Language Resources and Evaluation*, 44(3):263–280, 2010.

27 Inderjeet Mani and James Pustejovsky. *Interpreting Motion: Grounded Representations for Spatial Language*. New York: Oxford University Press, 2012.

28 Inderjeet Mani. *Computational Modeling of Narrative*. Synthesis Lectures on Language Technologies, Morgan & Claypool, 2013.

29 Jan Christoph Meister. *Computing Action. A Narratological Approach*. Berlin: de Gruyter, 2003.

30 George A. Miller and Philip N. Johnson-Laird. *Language and Perception*. Belknap Press of Harvard University Press, 1976.

31 Paul Messaris. *Visual Literacy: Image, Mind, and Reality*. Boulder: Westview Press, page 165, 1994.

32 Reinhard Moratz, Jochen Renz, and DiedrichWolter. Qualitative spatial reasoning about line segments. In W. Horn (ed.), *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI)*. Berlin, Germany, IOS Press 2000.

33 Martha Palmer, Dan Gildea, and Paul Kingsbury. The Proposition Bank: a corpus annotated with semantic roles. *Computational Linguistics*, 31(1):71–105, 2005.

34 James Pustejovsky, Bob Ingria, Roser Sauri, Jose Castano, Jessica Littman, Rob Gaizauskas, Andrea Setzer, Graham Katz and Inderjeet Mani. The specification language TimeML. In Inderjeet Mani, James Pustejovsky and Robert Gaizauskas (eds.), *The Language of Time: A Reader*, New York: Oxford University Press, pages 49–562, 2005.

35 James Pustejovsky and Jessica L. Moszkowicz. The Qualitative Spatial Dynamics of Motion in Language. In M. Bhatt, H. Guesgen, S. Woelfl, and S. Hazarika (eds.), *Qualitative Spatial and Temporal Reasoning: Emerging Applications, Trends and Future Directions*. Journal of Spatial Cognition and Computation, 11(1): 15–44, 2011.

36 James Pustejovsky, Jessica L. Moszkowicz, and Marc Verhagen. The current status of ISO-Space. Joint ISA-7 Workshop on Interoperable Semantic Annotation SRSL-3, *Workshop on Semantic Representation for Spoken Language*, I2MRT Workshop on Multimodal Resources and Tools, 2012.

37 D. A. Randell, Z. Cui. and A. G. Cohn. A Spatial Logic on Regions and Connection. In *Proceedings of 3rd Int. Conf. on Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, pages 165–176, 1992.

38 Alexander Scivos and Bernhard Nebel. Double-Crossing: Decidability and Computational Complexity of a Qualitative Calculus for Navigation. In *Proceedings COSIT-2001*, Springer-Verlag, 2001.

39 Spiros Skiadopoulos and Manolis Koubarakis. On the consistency of cardinal direction constraints. *Artificial Intelligence* 163, pages 91–135, 2005.

40 Leonard Talmy. *Toward a Cognitive Semantics*. MIT Press, 2000.

41 Leonard Talmy. Main Verb Properties and Equipollent Framing. In Guo JianSheng et al. (eds.), *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*. Lawrence Erlbaum Associates, 2009.

42 Subhashini Venugopalan, Huijuan Xu, Jeff Donahue, Marcus Rohrbach, Raymond Mooney and Kate Saenko. Translating Videos to Natural Language Using Deep Recurrent Neural Networks. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015)*, Denver, Colorado, June 2015, pages 149–1504.

43 Marc Verhagen, Roser Sauri, Tommaso Caselli and James Pustejovsky. SemEval-2010 Task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2)*, Uppsala, 2010, pages 57–62.

44 Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CPVR)*, pages 511–518, 2001.

45 Liwei Zhao, Monica Costa, and Norman I. Badler. Interpreting Movement Manner. In *Computer Animation 2000 (CA'00)*, Philadelphia, Pennsylvania, 2000, pages 98–103.

46 Joost Zwarts and Yoad Winter. Vector space semantics: A model-theoretic analysis of locative prepositions. *Journal of Logic, Language and Information* 9(2):171–213, 2000.