Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2015-009 March 30, 2015

A Suite of Techniques for Describing Activity in Terms of Events

Gary C. Borchardt

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

A Suite of Techniques for Describing Activity in Terms of Events

Gary C. Borchardt

MIT Computer Science and Artificial Intelligence Laboratory

1. Introduction

Despite the ease with which humans can describe and reason about unfolding activity in terms of events, enabling machines to exhibit these abilities has proved to be extremely difficult. This report investigates the potential of using language-encoded human knowledge of what typically happens during the temporal unfolding of various types of events as a basis for constructing a software system that can recognize event occurrences from sensor data and perform a range of related reasoning tasks concerning those event occurrences.

Two central insights are explored in this work:

- ***Language can provide a window on the workings of the human mind.*** This view, while not going so far as to claim that language *is* the machinery of the mind, nevertheless asserts that important parts of the mind's machinery can be observed and articulated through the medium of language. By this view, we look to the vocabulary and expressions of language to provide primitives for representing general-purpose human knowledge related to event recognition and reasoning about events, and we look to human introspection and interview to extract this knowledge for use by machines. Low-level scene information is also encoded using these language-motivated primitives, and event recognition is accomplished in software by matching models of typical event occurrences to encoded scene information.

- ***Concurrently addressing multiple, related capabilities can speed development and lead to an efficient, comprehensive solution.*** This view suggests that the task of event recognition and various tasks related to reasoning about events should not be studied separately. By examining event recognition along with supporting capabilities such as prediction and partial recognition, plus tasks dependent on event recognition such as summarization, explanation and question answering, it is more likely that we will be able to gather sufficient constraint to converge quickly on a set of related mechanisms that exhibit satisfactory performance on all of the targeted capabilities.

Building on these insights, a set of techniques has been implemented as part of the IMPACT reasoning system, supporting the tasks of event recognition, summarization of event sequences, explanation of recognized events, explanation of non-recognized events, prediction of event completions, and question answering regarding events and scene information. The techniques operate on sequences of timestamped, three-dimensional scene positions and contacts for humans, body parts, and objects, provided by a Microsoft Kinect sensor plus associated software. Given this setup, a satisfactory initial level of performance has been achieved on all of the addressed tasks, executing on a conventional laptop computer in faster than real time when compared to the observed activity. Regarding event recognition, on a set of 64 recorded, 10-second activity sequences, the system correctly recognizes 83% of noted event occurrences for a selection of event types, including identification of participants, their roles, and times of occurrence for the events. Considering the complete sets of event occurrences proposed by the system for the 64 recorded sequences, 81% of these proposed event occurrences were deemed to be reasonable accounts of the associated activity, whether or not these event instances might normally be noted by humans. Regarding summarization, for the 64 recorded sequences, the system eliminates an average of 27% of its proposed event instances, appropriately deeming these events as redundant descriptions of the observed activity. Comparable performance has been achieved on the explanation, prediction and question answering tasks. As such, the techniques provide a reference implementation of how these capabilities might be targeted, together, using a set of similar mechanisms operating on human-supplied knowledge encoded in a language-motivated representation. Moving forward, this reference implementation can be extended by exploiting parallelism, quantitative encoding and machine learning where appropriate, and it can be assessed as a potential model of human cognition using a range of experimental procedures.

The following subsections of this section list motivations for major design choices taken in construction of the implemented system. Following this, Section 2 describes formation of the input to the system, Section 3 describes each of the developed techniques plus further potential techniques, and Section 4 provides a concluding discussion of the effort. Appendix A presents the corpus of 64 recorded activity sequences used for development and testing of the system, and Appendix B presents 102 event models developed for use by the system.

1.1 Why Multiple Techniques?

The described work targets six capabilities, ranging from event recognition to prediction and question answering. There are several motivations for addressing these capabilities together, rather than separately. First, the capabilities interact to a significant degree. Summarization can streamline the presentation of results from event recognition and explanation. Event recognition can be a subtask carried out during explanation. Prediction of event completions can occur in the normal course of recognizing event occurrences, and question answering can serve to guide the application of event recognition. Second, given that the human mind is a resource-limited system, it is not unreasonable to assume that shared mechanisms are responsible for the accomplishment of these related tasks in humans, and thus by addressing the tasks together, using similar implementations, we increase our likelihood of producing a faithful model of human cognition. Third, each of these capabilities is a complex, multi-level processing problem with many, many potential approaches. By insisting that a single, coherent approach support all of these capabilities at once, we can draw additional constraint to help us achieve a solution for each of the capabilities considered individually as well.

1.2 Why Three-Dimensional Input?

The described work takes as input a sequence of timestamped, three-dimensional positions and contacts for human "skeletons"—including body parts such as hands, shoulders and feet—plus associated physical objects. One might question whether this input suitably mirrors information available to humans when they observe and process unfolding scenes. In response, it can be seen that input such as this is appropriate, by observing that humans do indeed form very robust, three-dimensional understandings of their surroundings. While stereo vision, head movement, and scene activity can help humans learn to form these three-dimensional understandings, the acquired capability is so robust that we can infer such positions and contacts in our surroundings using only one eye, without moving our heads, and without any activity occurring in the observed scene. Given the presence of this robust capability in humans, it is difficult to imagine that having access to detailed three-dimensional position and contact information does not contribute in a significant way to our own human ability to recognize events and reason about those event occurrences, and thus we should not be deterred from creating machine implementations of relatively higher-level event processing using such input.

Computer vision research has for some time investigated techniques for three-dimensional identification and tracking of humans and objects in a scene, yet the recent development of sensors like the Kinect has led to significant advances in performance for these tasks (e.g., [Shotton *et al.*, 2012; Song and Xiao, 2014]). While the mechanism of the Kinect and related sensors is decidedly not cognitively-motivated—in particular, relying on active projection of infrared patterns onto the observed scene rather than passive observation of the scene—nevertheless, the output of these sensors—the three-dimensional information they produce—does fit with human cognition and can serve as an appropriate substitute while passive computer vision systems remain in development for producing robust three-dimensional information about an observed scene.

1.3 Why Language-Based Representations?

The techniques described in this report operate on information encoded using a language-oriented representation called "transition space" [Borchardt, 1992; Borchardt, 1994; Borchardt *et al.*, 2014], motivated by several considerations in the cognitive psychology literature (e.g., as documented in [Miller and Johnson-Laird, 1976]). In turn, transition space specifications are encoded using a language called Möbius that depicts the syntax and semantics of simple English expressions [Borchardt, 2014]. Information at all levels of processing is represented in a language-oriented way. There are several reasons for this design choice.

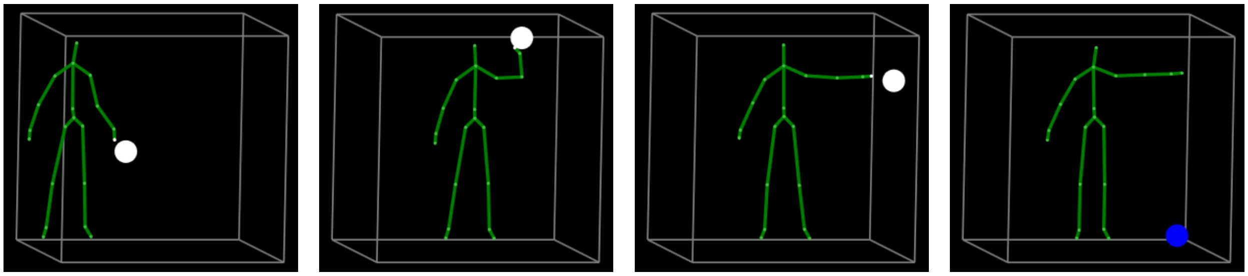
First, while it is certainly true that language can express the occurrence of events with statements such as “The human dropped the object.”, language can also go much deeper to express lower-level information associated with these occurrences. For example, we can describe individual attribute changes by asserting that the human’s control of the object has disappeared or the object’s elevation has decreased between two points in time, we can describe momentary states by asserting that the object has a particular speed or heading at a particular point in time, and, at a very low level, we can express pairwise comparisons of values by asserting that the vertical speed of the object or the elevation of the object is greater at one time than it is at another time. This ability of language to shed light on detailed human knowledge regarding the unfolding of events has important consequences. When we ask people to reflect on what happens during particular types of events—say, “dropping an object”—in terms of underlying changes, states and comparisons, we get readily formed and relatively consistent accounts, and by virtue of the fact that these accounts are rendered in language and are thus understandable to others, we can discuss the extracted accounts and come to a rough, consensual agreement on the details of what typically happens during these types of events. If humans have such detailed knowledge available to them, and this knowledge can even be brought into rough agreement across individuals, it is hard to imagine that this knowledge is not utilized by humans during their own recognition and reasoning about events, and so we should not feel reluctant to make use of language-centered, detailed, human accounts of what happens during various types of events as a basis for implementing machine recognition and reasoning about events.

Second, while language-oriented representations and operations may be important to human recognition and reasoning about events, even if these might be of less than central importance, they still provide constraint on an immensely complex problem. Rather than approaching event recognition and reasoning about events as capabilities to be learned from a blank slate, requiring large numbers of training examples and careful structuring to avoid overfitting to particular contexts, if we commit our implementation to use language-oriented attributes like “speed”, “heading”, “distance”, “contact”, “being inside”, “being above”, and so forth, we can then exploit the general-purpose utility of these attributes and converge more quickly on a robust capability.

A third motivation for using language-oriented representations and operations also deserves mention. If we are to address event recognition not in isolation, but rather in the context of other, related operations like summarization, explanation, prediction and question answering, then since many of these operations are intimately connected to language processing, if we target event recognition using a language-based approach, we can hope to gain considerable leverage in accomplishing these related operations as well. This is the case in the work reported here, as it is fairly straightforward to present the results of event recognition in language, present summaries and explanations in language, and relate input questions, encoded in language, to elements of represented structures.

2. Language-Oriented Scene Modeling from Kinect-Generated Input

For the techniques described here, input to the IMPACT system consists of sequences of timestamped, three-dimensional positions of tracked human “skeletons” and accompanying physical objects, plus contact information between tracked components. Figure 1 illustrates such a sequence, presented as representative frames within a 10-second course of activity. In this sequence, a man faces the sensor, holding a ball downwards in his left hand. He takes a couple of steps to his left, faces the sensor again, holds the ball up, holds it out to his left, and drops the ball, which falls, bounces a few times and comes to rest.



2013-10-16T23:40:44.600 2013-10-16T23:40:48.600 2013-10-16T23:40:50.100 2013-10-16T23:40:54.300

Figure 1: Sample input sequence.

Input sequences such as this are generated using a Microsoft Kinect for Windows sensor plus added software. Of itself, the Microsoft Kinect can generate tracked, three-dimensional “skeleton” positions for 1 or 2 humans in indoor settings, within a range of approximately 1 to 4 meters from the sensor, including body parts such as hands, feet, elbows, knees, shoulders and heads. In addition, the Kinect produces a depth map of its sensed field of view, plus a color image. The skeleton information, depth map and color image are updated at up to 30 frames per second.

For the work reported here, the Kinect's information on tracked human skeletons has been utilized, and additional software has been created to track physical objects approximately the size of a basketball and to compute potential contacts between hands, feet, and tracked physical objects. The frame rate for generated position and contact information has been set to 10 frames per second, to take advantage of inherent smoothing offered by lowered frame rates, and to speed the processing of the implemented recognition and reasoning operations, which do not appear to require higher frame rates.

Tracking of physical objects is accomplished by real-time post-processing of the depth map information provided by the Kinect sensor, in a sequence of steps. This calculation identifies areas of the depth map that stand out in sufficient relief with respect to their surroundings, exhibit a roughly circular shape and an expected size of roughly 10–30cm in horizontal and vertical dimensions, and which exist for a sufficient number of initial frames or continue the last-calculated trajectory of a previously-identified object. Contact between identified objects, human hands and human feet is calculated as potential contact based on positional proximity, accompanied by calculations of persistence and smoothing.

The above calculations provide the system with enough information to form low-level, language-based specifications of the observed activity, suitable for encoding in the transition space representation. Within the transition space representation, time-oriented information is specified using a hierarchy of five types of quantities:

objects are entities of importance in the scene, represented as parsable strings—e.g., "Human 73", "Human 73 Right Foot", "Object 51" and time "2013-10-16T21:50:17.700",

attributes are language-motivated properties, relationships, and other functions of one or two participants—e.g., "position" or "speed" of an object, or "contact" between two objects,

states are instantaneous values of attributes—e.g., the speed of an object at a particular time, or whether or not contact exists between two objects at a particular time,

changes are comparisons of attribute values between two time points—e.g., an "increase" in the distance between two objects, a "change" in an object's heading, or contact "disappearing" between two objects, and

events are collections of changes and states brought into focus for a particular analysis—e.g., "Human 1817 picks up Object 444 from 2013-10-16T21:34:15.600 to 2013-10-16T21:34:17.700.".

In turn, assertions in the transition space representation are encoded using the Möbius language, which specifies simple English constructions in a parsed form that includes basic syntactic and semantic information. As an example, a statement of a transition space "state", encoded in Möbius, is as follows:

```
equal(
  subject:attribute
    speed(article: the, of:tangible_object "Object 104",
      at:time "2013-10-16T23:40:51.300"),
  object:value "2.665 m/s").
```

For convenience, a simple language generation mechanism can be applied to Möbius expressions to render them in a form that is more readable, as, for example, the following rendering of the above Möbius expression:

> The speed of Object 104 at 2013-10-16T23:40:51.300 is 2.665 m/s.

The characterization of scene activity initially computed from the Kinect sensor's output provides transition space "states" for four attributes: an object being an instance of a type of object, an object being a part of another object, the position of an object, and contact between two objects. From this characterization, the IMPACT system calculates states for a broader set of language-motivated attributes listed in Figure 2. All attributes have a "null" value in their range. In addition, attributes are characterized as being either boolean in nature (having a single non-"null" value), qualitative (having an unordered set of non-"null" values), or quantitative (having an ordered set of non-"null" values).

<object> being an instance of <object>	(boolean)
<object> being a part of <object>	(boolean)
the position of <tangible object>	(qualitative)
the contact between <tangible object> and <tangible object>	(boolean)
the elapsed time from <time> to <time>	(quantitative)
the vertical position of <tangible object>	(quantitative)
the horizontal position of <tangible object>	(qualitative)
the vertical orientation of <tangible object>	(qualitative)
the distance between <tangible object> and <tangible object>	(quantitative)
the vertical distance between <tangible object> and <tangible object>	(quantitative)
the horizontal distance between <tangible object> and <tangible object>	(quantitative)
the speed of <tangible object>	(quantitative)
the vertical speed of <tangible object>	(quantitative)
the horizontal speed of <tangible object>	(quantitative)
the heading of <tangible object>	(qualitative)
the vertical heading of <tangible object>	(qualitative)
the horizontal heading of <tangible object>	(qualitative)
the control of <tangible object> by <human>	(boolean)

Figure 2: Attributes used within the application.

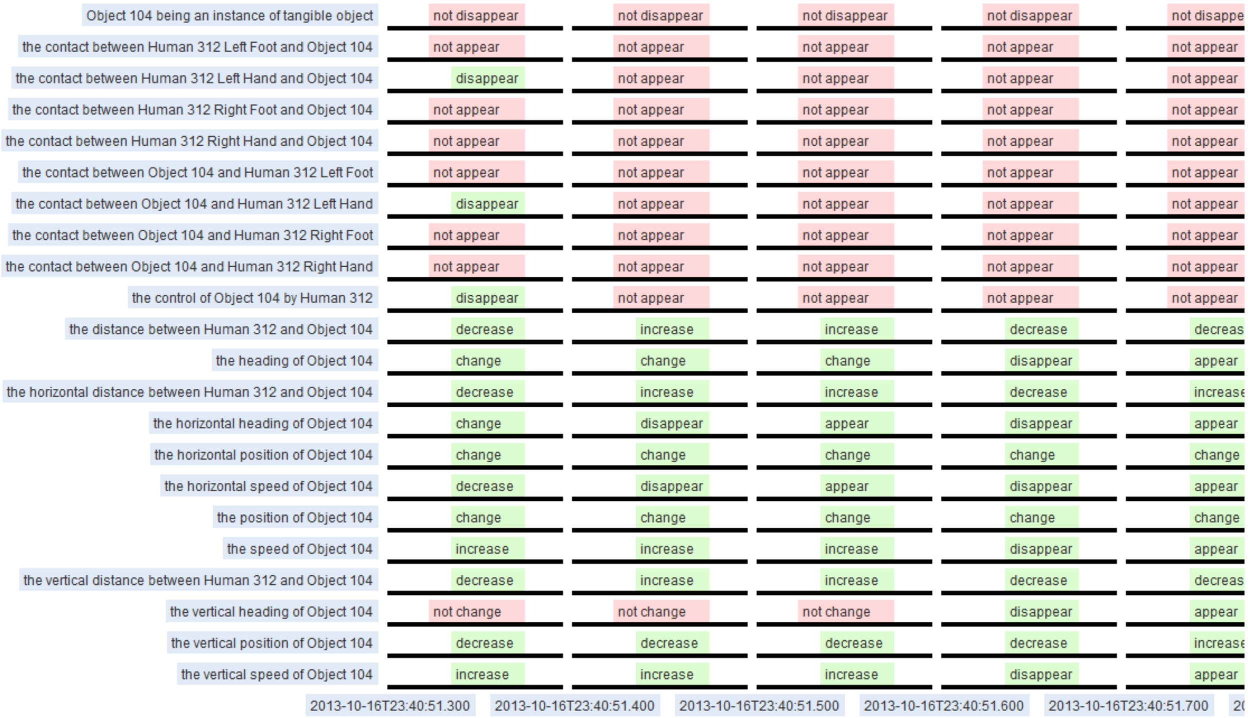
Once the system has computed states for attributes in the extended set of attributes, it continues by computing language-motivated changes between successive frames of the sequence. Figure 3 lists ten varieties of change described within the transition space representation. These varieties of change cover the range of possibilities that arise when an attribute of one or two specified objects is asserted to either equal or not equal the “null” value at an earlier time and again at a later time, plus possibly exhibit a relationship between its values at the earlier and later times, where one value equals, does not equal, exceeds, or does not exceed the other value. In this set, the varieties “change” and “not change” are specializations of “not disappear” suitable for qualitative or quantitative attributes, and “increase”, “not increase”, “decrease” and “not decrease” are specializations of “not disappear” suitable for quantitative attributes.

appear	not appear	
disappear	not disappear	*boolean attributes*
change	not change	*qualitative attributes*
increase	not increase	
decrease	not decrease	*quantitative attributes*

Figure 3: Ten varieties of change.

Assembled together, the encoded scene information can be depicted as in Figure 4, regarding the activity sequence illustrated in Figure 1. This diagram lists a subset of the total set of instantiated attributes along its vertical axis, time points along its horizontal axis, and changes along each row, between the time points. Definite changes (“appear”, “disappear”, “change”, “increase” and “decrease”) appear in green, while non-changes and possible changes (“not appear”, “not disappear”, “not change”, “not increase” and “not decrease”) are listed in red. Calculated state information is not indicated in these diagrams. The excerpt in Figure 4 begins with the human releasing the object held in his left hand and continues with the object falling and beginning to bounce.

Once the system has computed the information illustrated in Figure 4, including both changes and states in its set of language-motivated attributes of participating objects, it is then ready to perform event recognition and related reasoning operations regarding events.



	2013-10-16T23:40:51.300	2013-10-16T23:40:51.400	2013-10-16T23:40:51.500	2013-10-16T23:40:51.600	2013-10-16T23:40:51.700	20
Object 104 being an instance of tangible object	not disappear	not disappear	not disappear	not disappear	not disappe	
the contact between Human 312 Left Foot and Object 104	not appear	not appear	not appear	not appear	not appear	
the contact between Human 312 Left Hand and Object 104	disappear	not appear	not appear	not appear	not appear	
the contact between Human 312 Right Foot and Object 104	not appear	not appear	not appear	not appear	not appear	
the contact between Human 312 Right Hand and Object 104	not appear	not appear	not appear	not appear	not appear	
the contact between Object 104 and Human 312 Left Foot	not appear	not appear	not appear	not appear	not appear	
the contact between Object 104 and Human 312 Left Hand	disappear	not appear	not appear	not appear	not appear	
the contact between Object 104 and Human 312 Right Foot	not appear	not appear	not appear	not appear	not appear	
the contact between Object 104 and Human 312 Right Hand	not appear	not appear	not appear	not appear	not appear	
the control of Object 104 by Human 312	disappear	not appear	not appear	not appear	not appear	
the distance between Human 312 and Object 104	decrease	increase	increase	decrease	decreas	
the heading of Object 104	change	change	change	disappear	appear	
the horizontal distance between Human 312 and Object 104	decrease	increase	increase	decrease	increase	
the horizontal heading of Object 104	change	disappear	appear	disappear	appear	
the horizontal position of Object 104	change	change	change	change	change	
the horizontal speed of Object 104	decrease	disappear	appear	disappear	appear	
the position of Object 104	change	change	change	change	change	
the speed of Object 104	increase	increase	increase	disappear	appear	
the vertical distance between Human 312 and Object 104	decrease	increase	increase	decrease	decreas	
the vertical heading of Object 104	not change	not change	not change	disappear	appear	
the vertical position of Object 104	decrease	decrease	decrease	decrease	increase	
the vertical speed of Object 104	increase	increase	increase	disappear	appear	

Figure 4: Excerpt of encoded scene information for the sequence illustrated in Figure 1.

3. Techniques for Describing Activity in Terms of Events

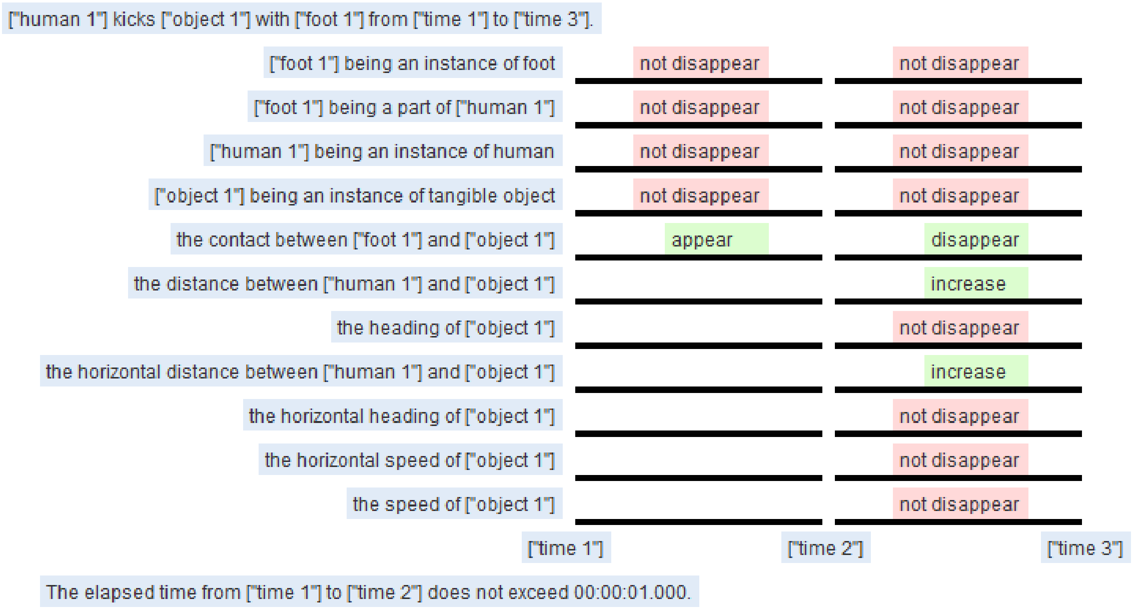
The techniques described in the following subsections make use of three components of information and/or processing: (1) a record of encoded scene information, produced by the mechanisms described in Section 2, (2) a set of event models that provide reference knowledge about what typically happens during the unfolding of particular types of events, and (3) a matcher that uses the event models to identify instances of those types of events within the scene information. Each technique adds a high-level algorithm that coordinates the use of these three components.

An iterative process was used to incrementally refine the above three components during development and fine-tuning of the techniques described below. In some cases, this refinement process has resulted in new attributes being included for calculation within the encoded scene information, or existing attributes being calculated in slightly different ways. In other cases, event models for new types of events have been added to the library of event models, or modifications have been made to existing event models. In still other cases, modifications have been made to the matcher, enabling it to search more broadly or narrowly for event instances. By manipulating these three components in tandem as new input examples have been considered during development, it has been possible to converge on a set of techniques that both provide satisfactory performance at their targeted tasks and that do so using data and knowledge that can be articulated in language and that, through inspection, match human intuition about data and knowledge that should be appropriate in these contexts.

3.1 Event Recognition

Appendix B lists 102 event models created for use with the techniques described here. The events that have been modeled involve 0 or 1 object and 0 to 2 humans, and the models have been created in a symbolic editor that generates transition space assertions encoded in the Möbius language. All of the event models appearing in Appendix B are used with the event recognition, prediction and question answering capabilities. Subsets of these event models are used with the explanation and summarization capabilities, or for evaluation of the event recognition capability, as indicated in Appendix B.

Figure 5 illustrates an event model for an event of a human kicking an object with one of his or her feet. The model is presented in a form that renders the underlying Möbius expressions in simple English, but allows variables, enclosed in brackets, to appear in a form that more closely resembles their Möbius encoding. Changes in each concerned attribute are listed between time points, with unconstrained activity indicated by the omission of change specifications for particular attributes between particular pairs of time points. Additional elements of state information—typically constraints on attribute values and temporal durations—appear below the grid portion of the diagram.



["human 1"] kicks ["object 1"] with ["foot 1"] from ["time 1"] to ["time 3"].

["foot 1"] being an instance of foot	not disappear	not disappear
["foot 1"] being a part of ["human 1"]	not disappear	not disappear
["human 1"] being an instance of human	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear
the contact between ["foot 1"] and ["object 1"]	appear	disappear
the distance between ["human 1"] and ["object 1"]		increase
the heading of ["object 1"]		not disappear
the horizontal distance between ["human 1"] and ["object 1"]		increase
the horizontal heading of ["object 1"]		not disappear
the horizontal speed of ["object 1"]		not disappear
the speed of ["object 1"]		not disappear

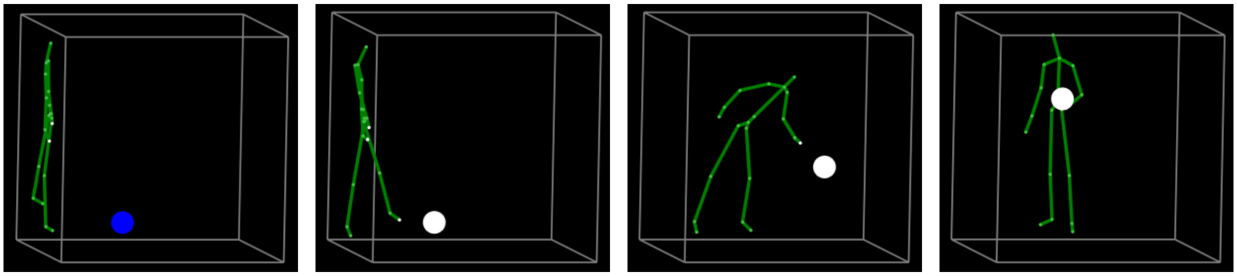
["time 1"] ["time 2"] ["time 3"]

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:00:01.000.

Figure 5: Event model for a human kicking an object with a particular foot.

In the event model depicted in Figure 5, contact first appears between the human's foot and the object. Motion of the object is unconstrained during this time, as are changes in distance between the human and the object, allowing the kicking to follow from either a stationary or moving human or object. In the second interval of the event model, contact disappears between the foot and the object, and the object moves while the distance between the human and the object increases. A duration constraint specifies that the first interval of time may not exceed 1 second. With this constraint, the model will not match circumstances in which, say, the human's foot comes into contact with the object, then rests awhile, then pushes the ball away.

Event models are matched to encoded scene information that has been created in the manner outlined in Section 2. As an example of matching involving the above-described event model for "kicking an object", Figure 6 depicts one of the 64 recorded sequences used in this effort. In this recorded sequence, a human starts by bending to look at a ball on the ground, kicks the ball slightly with his left foot, steps, bends over and picks up the ball, and then turns, holding the ball.

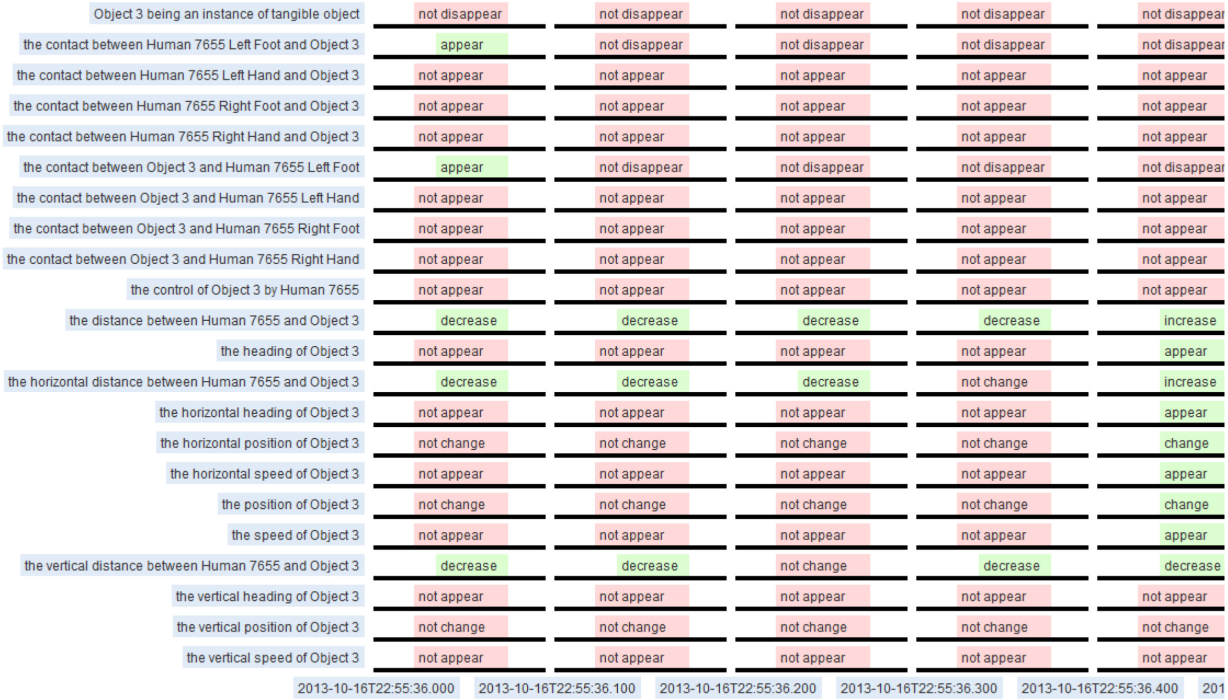


2013-10-16T22:55:32.400 2013-10-16T22:55:36.400 2013-10-16T22:55:39.300 2013-10-16T22:55:42.300

Figure 6: A recorded input sequence.

Figure 7 presents an excerpt of the encoded scene information generated for the input example illustrated in Figure 6. As before, only change information is indicated in this type of diagram, although there is also a considerable amount of state information incorporated within the underlying, encoded scene information. The incorporated state information includes, for example, a set of assertions that the ball has particular, numerical speeds at particular instants in time.

In the excerpt of activity illustrated in Figure 7, Human 7655's left foot comes into contact with Object 3, and then a few frames later, Object 3 starts moving.



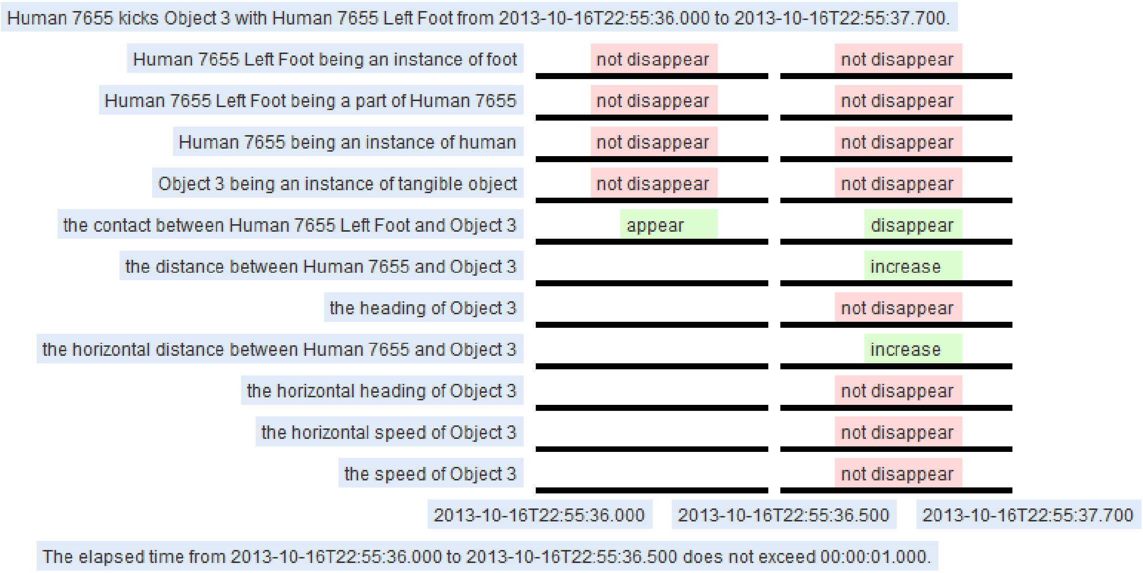
	2013-10-16T22:55:36.000	2013-10-16T22:55:36.100	2013-10-16T22:55:36.200	2013-10-16T22:55:36.300	2013-10-16T22:55:36.400	201
Object 3 being an instance of tangible object	not disappear	not disappear	not disappear	not disappear	not disappear	
the contact between Human 7655 Left Foot and Object 3	appear	not disappear	not disappear	not disappear	not disappear	
the contact between Human 7655 Left Hand and Object 3	not appear	not appear	not appear	not appear	not appear	
the contact between Human 7655 Right Foot and Object 3	not appear	not appear	not appear	not appear	not appear	
the contact between Human 7655 Right Hand and Object 3	not appear	not appear	not appear	not appear	not appear	
the contact between Object 3 and Human 7655 Left Foot	appear	not disappear	not disappear	not disappear	not disappear	
the contact between Object 3 and Human 7655 Left Hand	not appear	not appear	not appear	not appear	not appear	
the contact between Object 3 and Human 7655 Right Foot	not appear	not appear	not appear	not appear	not appear	
the contact between Object 3 and Human 7655 Right Hand	not appear	not appear	not appear	not appear	not appear	
the control of Object 3 by Human 7655	not appear	not appear	not appear	not appear	not appear	
the distance between Human 7655 and Object 3	decrease	decrease	decrease	decrease	increase	
the heading of Object 3	not appear	not appear	not appear	not appear	appear	
the horizontal distance between Human 7655 and Object 3	decrease	decrease	decrease	not change	increase	
the horizontal heading of Object 3	not appear	not appear	not appear	not appear	appear	
the horizontal position of Object 3	not change	not change	not change	not change	change	
the horizontal speed of Object 3	not appear	not appear	not appear	not appear	appear	
the position of Object 3	not change	not change	not change	not change	change	
the speed of Object 3	not appear	not appear	not appear	not appear	appear	
the vertical distance between Human 7655 and Object 3	decrease	decrease	not change	decrease	decrease	
the vertical heading of Object 3	not appear	not appear	not appear	not appear	not appear	
the vertical position of Object 3	not change	not change	not change	not change	not change	
the vertical speed of Object 3	not appear	not appear	not appear	not appear	not appear	

Figure 7: Excerpt of encoded scene information for the activity sequence depicted in Figure 6.

To perform event recognition, the implemented system utilizes a technique called "core–periphery matching". This technique consists of a goal-directed, hierarchical matching of an event model to scene information in which a match for the entire event is formed by seeking and combining matches for individual intervals of the model and matches for individual intervals of the model are formed by seeking and combining matches for individual states and changes in those intervals. This process allows for "stretching" of the event model's time intervals as necessary to fit the speed or slowness of unfolding activity in the observed scene.

A detailed description of core–periphery matching appears in [Borchardt *et al.*, 2014], which describes a recent effort involving recognition of vehicle events from vehicle track data. Core–periphery matching of an event model to encoded scene information proceeds by first identifying potential variable bindings and temporal extents for the event model's first transition—its first interval of changes and states—then continues by extending these matches to accommodate the matches for subsequent transitions. Each transition is matched by considering a first change in that transition, then another change, and so on, until state constraints for that transition are also considered. At the lowest level, an individual change in a transition is matched by identifying potential variable bindings and temporal extents for that change, with each potential temporal extent having a "core" and a "periphery" span of time in the scene. Within the "core" interval of the temporal extent for a change match, any subinterval is guaranteed to exhibit the desired change (e.g., a sought-after "increase" in an event model will match any subsequence of a sequence of several “increase” changes in the scene). Within the surrounding “periphery” interval of the temporal extent for a change match, as long as some portion of the core interval is also included, the desired change is guaranteed to be exhibited (e.g., a leading sequence of “not change” specifications in the scene can extend the match of a sought-after “increase” in the event model, as long at least one “increase” from the core interval of matched scene changes is also included from the temporal extent of the change match). Identifying “cores” plus “peripheries” in the matching of individual changes to scene information allows the matching process to flexibly combine change matches into transition matches and then combine transition matches into event matches.

For the recorded example whose scene information appears in Figure 7, core-periphery matching produces a match for the “kicking” event model of Figure 5, as depicted in Figure 8.



Human 7655 kicks Object 3 with Human 7655 Left Foot from 2013-10-16T22:55:36.000 to 2013-10-16T22:55:37.700.

Human 7655 Left Foot being an instance of foot	not disappear	not disappear
Human 7655 Left Foot being a part of Human 7655	not disappear	not disappear
Human 7655 being an instance of human	not disappear	not disappear
Object 3 being an instance of tangible object	not disappear	not disappear
the contact between Human 7655 Left Foot and Object 3	appear	disappear
the distance between Human 7655 and Object 3		increase
the heading of Object 3		not disappear
the horizontal distance between Human 7655 and Object 3		increase
the horizontal heading of Object 3		not disappear
the horizontal speed of Object 3		not disappear
the speed of Object 3		not disappear

2013-10-16T22:55:36.000 2013-10-16T22:55:36.500 2013-10-16T22:55:37.700

The elapsed time from 2013-10-16T22:55:36.000 to 2013-10-16T22:55:36.500 does not exceed 00:00:01.000.

Figure 8: Matched event instance of “kicking an object”.

Depending on the circumstances exhibited in a scene, core-periphery matching of an event model may result in longer or shorter total durations for identified event instances, subject to duration constraints specified for individual intervals within the event model. For the above “kicking” event model, for example, a quick progression from appearance to disappearance of contact, combined with a short movement of the kicked object, would yield a much shorter identified instance of “kicking”.

The 64 recorded activity sequences used for development and testing of the implemented techniques are described in Appendix A. Using the iterative refinement method described above, it was possible to create a set of intuitive, language-based event models, depicted in Appendix B, and match these event models to the recorded sequences using core–periphery matching to achieve satisfactory performance in the recognition of event instances. Across the 64 input sequences, manually assessing the quality of the system's event recognition for a set of 20 significant event types indicated in Appendix B, it was determined that 83% of event instances that should have been recognized were correctly recognized by the system, including identification of participants, their roles, and approximate time bounds for the occurrences. Considering the full sets of event instances proposed by the system in connection with the 64 input sequences, 81% of these system-postulated events were deemed to be appropriate in the context of the depicted activity, whether or not these event instances might normally be noted by humans.

3.2 Summarization

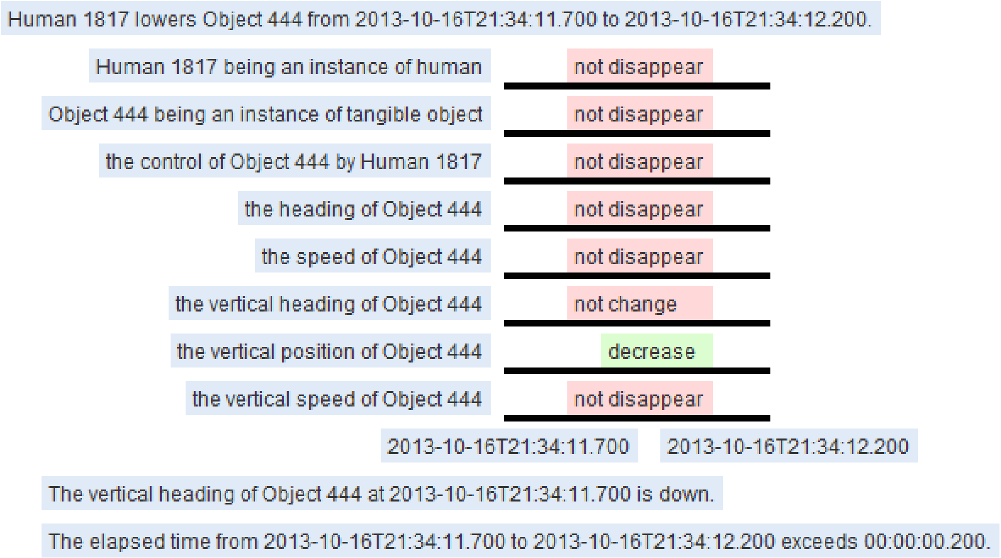
One particularly important aspect of summarization is the minimization of redundant information. In reporting event occurrences, an effective summary will omit mention of events that can be directly inferred from the occurrence of other events that are mentioned in the summary.

One way to identify redundant events in an event summary is to consider each event's relative coverage of scene information. In this context, scene information can be considered at a very detailed level of granularity. The scene modeling mechanism described in Section 2 produces a set of states (e.g., instantaneous positions, speeds, contacts, etc.) and changes (e.g., an increase in vertical position between two time points); however, these types of assertions can be further decomposed into lower-level assertions: collections of pairwise comparisons between quantities, where each comparison specifies that one quantity equals, does not equal, exceeds, or does not exceed a second quantity, and each of the two indicated quantities is either a timestamped attribute value or a constant, reference value. Comparisons of this sort are referred to in [Borchardt, 2014] as "Level 1 Relative Attribute Value Expressions", or "Level 1 RAVEs". Some examples of Level 1 RAVEs are as follows:

> The position of Object 383 at 2014-07-29T18:10:16.200 equals (2.133, 1.189, 1.395).
>
> The contact between Human 2027 Left Hand and Object 383 at 2014-07-29T18:10:16.200 does not equal null.
>
> The speed of Object 383 at 2014-07-29T18:10:16.300 exceeds the speed of Object 383 at 2014-07-29T18:10:16.200.

Each event instance produced by the event recognition process has associated with it a set of Level 1 RAVEs within the scene information—those Level 1 RAVEs in the scene that have been used to support matches of states and changes in the corresponding event model as that event instance was recognized. A useful summarization strategy is, then, as follows: if all of the scene Level 1 RAVEs associated with, or "covered by", one recognized event instance are also covered by other event instances in the event summary, then the first event instance can be considered redundant and can be excluded from the event summary—its mention communicates no additional knowledge about what happened during the unfolding activity.

As an example, Figure 9 illustrates an instance of "lowering an object" recognized in one of the recorded sequences. This event instance is portrayed as changes and states associated with mapped time points from the event model used in the match and corresponds to a much finer-granularity set of changes and states—and ultimately Level 1 RAVEs—within the encoded scene information.



Human 1817 lowers Object 444 from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:12.200.

Human 1817 being an instance of human — not disappear

Object 444 being an instance of tangible object — not disappear

the control of Object 444 by Human 1817 — not disappear

the heading of Object 444 — not disappear

the speed of Object 444 — not disappear

the vertical heading of Object 444 — not change

the vertical position of Object 444 — decrease

the vertical speed of Object 444 — not disappear

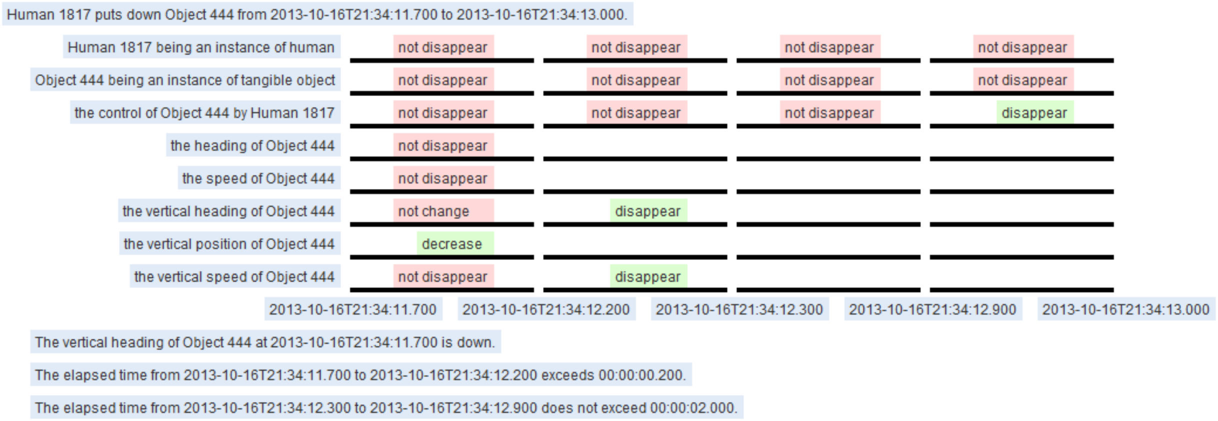
2013-10-16T21:34:11.700 2013-10-16T21:34:12.200

The vertical heading of Object 444 at 2013-10-16T21:34:11.700 is down.

The elapsed time from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:12.200 exceeds 00:00:00.200.

Figure 9: Recognized event instance of "lowering an object".

Figure 10 illustrates a second recognized event instance, for "putting down an object". This event instance also corresponds to a much finer-granularity set of statements within the encoded scene information. Since all of the scene information covered by the event instance of "lowering" illustrated in Figure 9 is also covered by the event instance of "putting down" illustrated in Figure 10, the instance of "lowering" can reasonably be excluded from a summary of the observed activity, so long as the instance of "putting down" is included, or other event instances are retained that also cover this scene information. This strategy matches human intuition: it would be redundant to state both that a person has put down an object and that the person has lowered the object during the initial part of the "putting down" event.



Human 1817 puts down Object 444 from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:13.000.

Human 1817 being an instance of human	not disappear	not disappear	not disappear	not disappear
Object 444 being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of Object 444 by Human 1817	not disappear	not disappear	not disappear	disappear
the heading of Object 444	not disappear			
the speed of Object 444	not disappear			
the vertical heading of Object 444	not change	disappear		
the vertical position of Object 444	decrease			
the vertical speed of Object 444	not disappear	disappear		

2013-10-16T21:34:11.700 2013-10-16T21:34:12.200 2013-10-16T21:34:12.300 2013-10-16T21:34:12.900 2013-10-16T21:34:13.000

The vertical heading of Object 444 at 2013-10-16T21:34:11.700 is down.

The elapsed time from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:12.200 exceeds 00:00:00.200.

The elapsed time from 2013-10-16T21:34:12.300 to 2013-10-16T21:34:12.900 does not exceed 00:00:02.000.

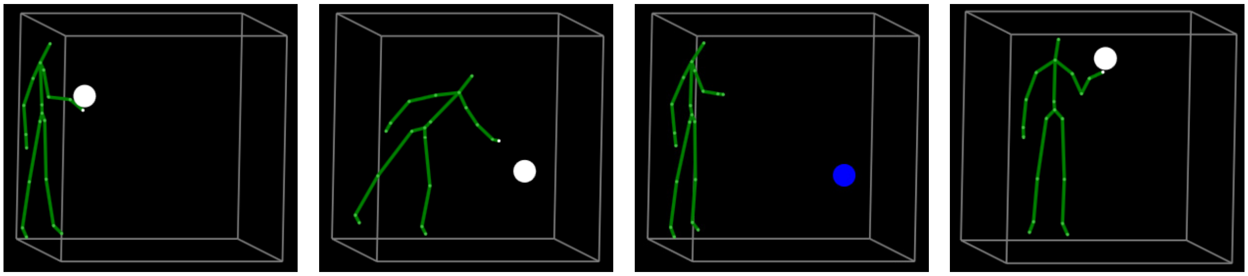
Figure 10: Recognized event instance of "putting down an object".

Preliminary work utilizing this approach to summarization is described in [Borchardt *et al.*, 2014]. In that work, a fixed summarization strategy was employed: particular types of states and changes within the scene information were designated as being significant for purposes of calculating coverage of scene information for summarization. In the work described here, a flexible strategy has been employed, where different elements of scene information can be marked as "significant" for calculating coverage, depending on the goals at hand. Considering the components of the underlying representation in relative, increasing order of complexity—starting with time points and objects, then progressing to attributes, states, changes, and events—it would appear to be the case that "masking" these elements from the summarization process—marking them as already covered—has the effect of constraining the resultant summary in a progression of coarse to fine ways. For example, masking time points outside a particular time range can be used to generate a summary of activity within that time range. Objects, humans and parts of humans of lesser concern can be masked from the summarization process, so as to focus attention on important objects, humans and parts of humans. Particular attributes can be masked—e.g., those assessed on individual objects such as "speed" and "heading". Particular types of states and changes can be masked—e.g., numerical values for speed and distance, non-presence of contact, or increases and decreases in distance. Particular types of events can be masked—e.g., events involving humans but no objects. In all of these cases, masking of particular items of lesser importance from the summarization process will have the effect of focusing summarization toward finding a suitable set of recognized events that cover, and hence describe, the remaining "unmasked" scene information.

For the work described here, the IMPACT system implements the following general-purpose masking strategy when summarizing recorded examples. This strategy was determined by experimentation; however, a range of related strategies might also be expected to yield suitable results:

- no masking of **time points**: summarize the entire temporal extent of the recorded example,
- no masking of **objects**: all humans, parts of humans, and physical objects are considered
- masking of the following **attributes**: being an instance of a type of object, being a part of an object, position and horizontal position, contact between objects, elapsed time between time points, speed and horizontal speed, and heading and horizontal heading,
- masking of **state information** that involves comparisons to constant, reference values,
- for the unmasked attribute varieties, masking all remaining **state and change information** *except*:
  - for vertical position, vertical speed, distance, horizontal distance and vertical distance: value comparisons between time points, where one value is asserted to exceed or not exceed the other,
  - for vertical orientation: information on whether attribute values exist (do not equal "null") or do not exist (equal "null") at particular time points, and value comparisons between time points, where one value is asserted to equal or not equal the other,
  - for vertical heading: value comparisons between time points, where one value is asserted to not equal the other,
  - for control of an object by a human: instances of the change varieties "appear", "disappear" and "not disappear", and
- masking of **event types**, limiting the set used to the 49 events indicated in Appendix B as event types employed for summarization.

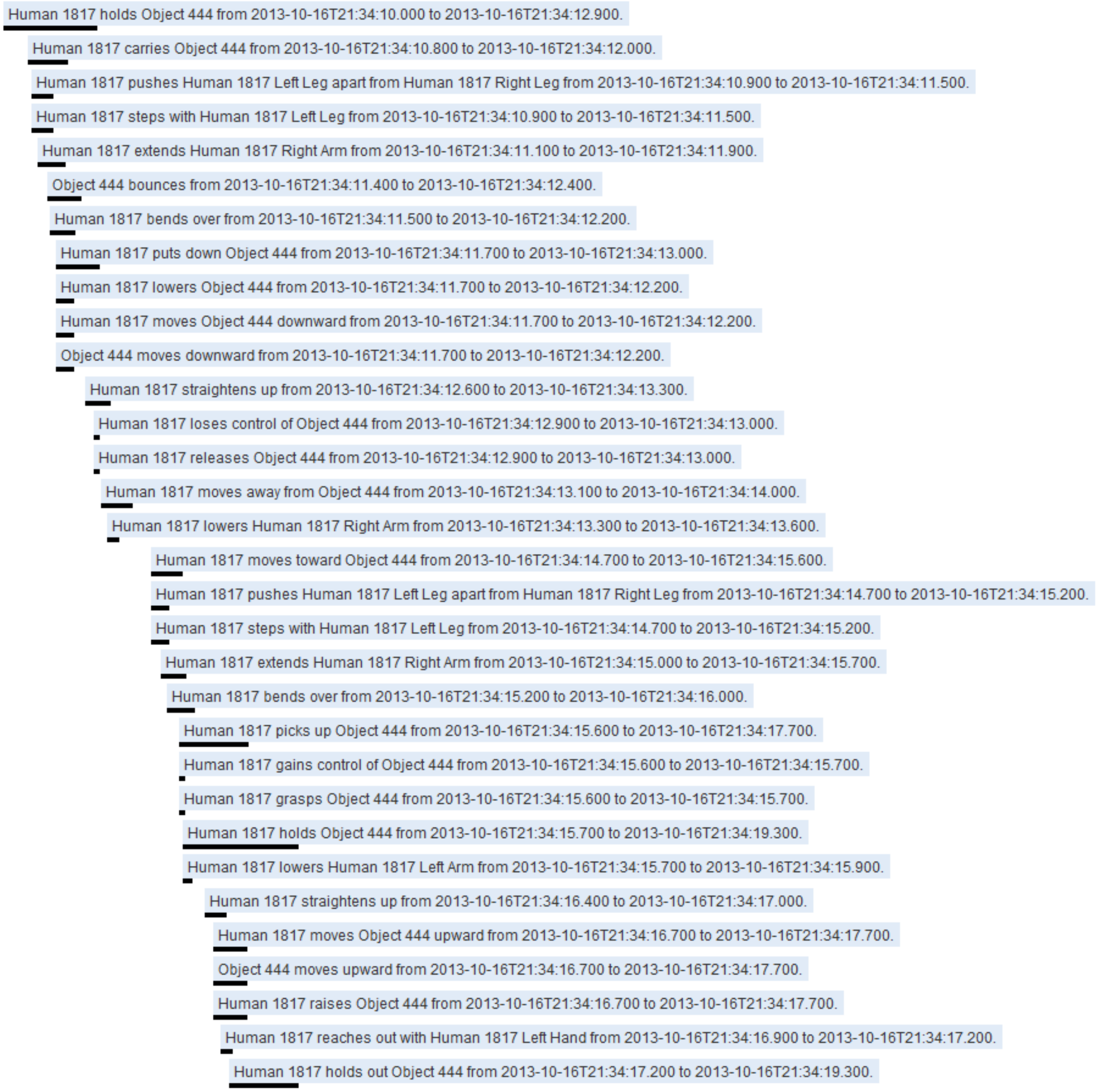
As an example of the application of IMPACT's summarization strategies, Figure 11 depicts one of the 64 recorded activity sequences, in which a man begins by stepping, reaching and placing a ball on a surface, stands back, reaches out and picks up the ball, and holds the ball at shoulder level.



2013-10-16T21:34:09.800 2013-10-16T21:34:12.200 2013-10-16T21:34:14.600 2013-10-16T21:34:19.500

Figure 11: A recorded input sequence.

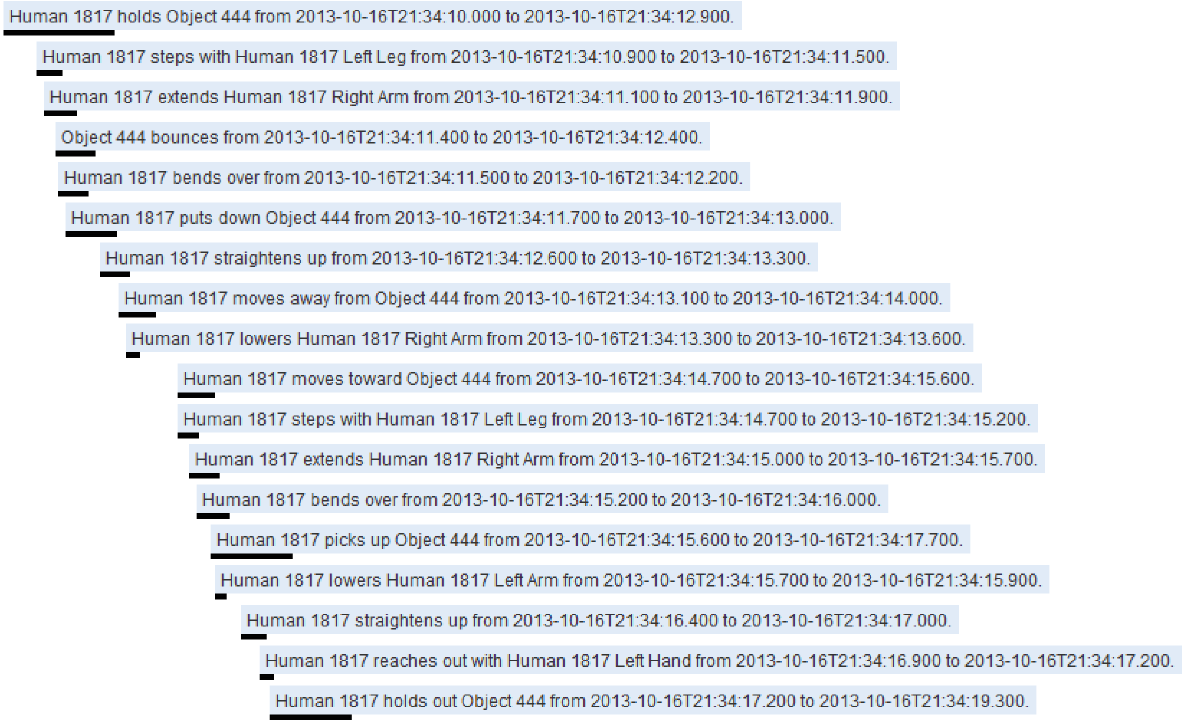
Figure 12 lists an initial, unsummarized set of 32 event instances recognized by IMPACT for this recorded example. This set is limited only by the exclusion of event types not employed for summarization; for efficiency, when summarization is desired, masked event types are not subjected to event recognition in the first place. In the listing of events in Figure 12, each event is positioned horizontally to indicate its approximate ordering in the temporal sequence. A black bar beneath each event indicates its relative starting and ending time points.



Human 1817 holds Object 444 from 2013-10-16T21:34:10.000 to 2013-10-16T21:34:12.900.
Human 1817 carries Object 444 from 2013-10-16T21:34:10.800 to 2013-10-16T21:34:12.000.
Human 1817 pushes Human 1817 Left Leg apart from Human 1817 Right Leg from 2013-10-16T21:34:10.900 to 2013-10-16T21:34:11.500.
Human 1817 steps with Human 1817 Left Leg from 2013-10-16T21:34:10.900 to 2013-10-16T21:34:11.500.
Human 1817 extends Human 1817 Right Arm from 2013-10-16T21:34:11.100 to 2013-10-16T21:34:11.900.
Object 444 bounces from 2013-10-16T21:34:11.400 to 2013-10-16T21:34:12.400.
Human 1817 bends over from 2013-10-16T21:34:11.500 to 2013-10-16T21:34:12.200.
Human 1817 puts down Object 444 from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:13.000.
Human 1817 lowers Object 444 from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:12.200.
Human 1817 moves Object 444 downward from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:12.200.
Object 444 moves downward from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:12.200.
Human 1817 straightens up from 2013-10-16T21:34:12.600 to 2013-10-16T21:34:13.300.
Human 1817 loses control of Object 444 from 2013-10-16T21:34:12.900 to 2013-10-16T21:34:13.000.
Human 1817 releases Object 444 from 2013-10-16T21:34:12.900 to 2013-10-16T21:34:13.000.
Human 1817 moves away from Object 444 from 2013-10-16T21:34:13.100 to 2013-10-16T21:34:14.000.
Human 1817 lowers Human 1817 Right Arm from 2013-10-16T21:34:13.300 to 2013-10-16T21:34:13.600.
Human 1817 moves toward Object 444 from 2013-10-16T21:34:14.700 to 2013-10-16T21:34:15.600.
Human 1817 pushes Human 1817 Left Leg apart from Human 1817 Right Leg from 2013-10-16T21:34:14.700 to 2013-10-16T21:34:15.200.
Human 1817 steps with Human 1817 Left Leg from 2013-10-16T21:34:14.700 to 2013-10-16T21:34:15.200.
Human 1817 extends Human 1817 Right Arm from 2013-10-16T21:34:15.000 to 2013-10-16T21:34:15.700.
Human 1817 bends over from 2013-10-16T21:34:15.200 to 2013-10-16T21:34:16.000.
Human 1817 picks up Object 444 from 2013-10-16T21:34:15.600 to 2013-10-16T21:34:17.700.
Human 1817 gains control of Object 444 from 2013-10-16T21:34:15.600 to 2013-10-16T21:34:15.700.
Human 1817 grasps Object 444 from 2013-10-16T21:34:15.600 to 2013-10-16T21:34:15.700.
Human 1817 holds Object 444 from 2013-10-16T21:34:15.700 to 2013-10-16T21:34:19.300.
Human 1817 lowers Human 1817 Left Arm from 2013-10-16T21:34:15.700 to 2013-10-16T21:34:15.900.
Human 1817 straightens up from 2013-10-16T21:34:16.400 to 2013-10-16T21:34:17.000.
Human 1817 moves Object 444 upward from 2013-10-16T21:34:16.700 to 2013-10-16T21:34:17.700.
Object 444 moves upward from 2013-10-16T21:34:16.700 to 2013-10-16T21:34:17.700.
Human 1817 raises Object 444 from 2013-10-16T21:34:16.700 to 2013-10-16T21:34:17.700.
Human 1817 reaches out with Human 1817 Left Hand from 2013-10-16T21:34:16.900 to 2013-10-16T21:34:17.200.
Human 1817 holds out Object 444 from 2013-10-16T21:34:17.200 to 2013-10-16T21:34:19.300.

Figure 12: An initial set of recognized event instances for the recorded sequence.

To form a summary of this event sequence, the system first applies the masking strategies listed above, then engages in an iterative process by which events are removed from the list if they fail to uniquely cover any unmasked scene information. During this iterative process, smaller event instances—instances that cover fewer scene Level 1 RAVEs—are considered for exclusion before larger event instances are considered. This process continues until no additional event instances can be omitted from the summary without leaving some unmasked, scene Level 1 RAVEs uncovered by event instances remaining in the summary. When this process is applied to the initial set of 32 recognized events listed in Figure 12, a reduced set of 18 event instances is produced, as listed in Figure 13.



Human 1817 holds Object 444 from 2013-10-16T21:34:10.000 to 2013-10-16T21:34:12.900.
Human 1817 steps with Human 1817 Left Leg from 2013-10-16T21:34:10.900 to 2013-10-16T21:34:11.500.
Human 1817 extends Human 1817 Right Arm from 2013-10-16T21:34:11.100 to 2013-10-16T21:34:11.900.
Object 444 bounces from 2013-10-16T21:34:11.400 to 2013-10-16T21:34:12.400.
Human 1817 bends over from 2013-10-16T21:34:11.500 to 2013-10-16T21:34:12.200.
Human 1817 puts down Object 444 from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:13.000.
Human 1817 straightens up from 2013-10-16T21:34:12.600 to 2013-10-16T21:34:13.300.
Human 1817 moves away from Object 444 from 2013-10-16T21:34:13.100 to 2013-10-16T21:34:14.000.
Human 1817 lowers Human 1817 Right Arm from 2013-10-16T21:34:13.300 to 2013-10-16T21:34:13.600.
Human 1817 moves toward Object 444 from 2013-10-16T21:34:14.700 to 2013-10-16T21:34:15.600.
Human 1817 steps with Human 1817 Left Leg from 2013-10-16T21:34:14.700 to 2013-10-16T21:34:15.200.
Human 1817 extends Human 1817 Right Arm from 2013-10-16T21:34:15.000 to 2013-10-16T21:34:15.700.
Human 1817 bends over from 2013-10-16T21:34:15.200 to 2013-10-16T21:34:16.000.
Human 1817 picks up Object 444 from 2013-10-16T21:34:15.600 to 2013-10-16T21:34:17.700.
Human 1817 lowers Human 1817 Left Arm from 2013-10-16T21:34:15.700 to 2013-10-16T21:34:15.900.
Human 1817 straightens up from 2013-10-16T21:34:16.400 to 2013-10-16T21:34:17.000.
Human 1817 reaches out with Human 1817 Left Hand from 2013-10-16T21:34:16.900 to 2013-10-16T21:34:17.200.
Human 1817 holds out Object 444 from 2013-10-16T21:34:17.200 to 2013-10-16T21:34:19.300.

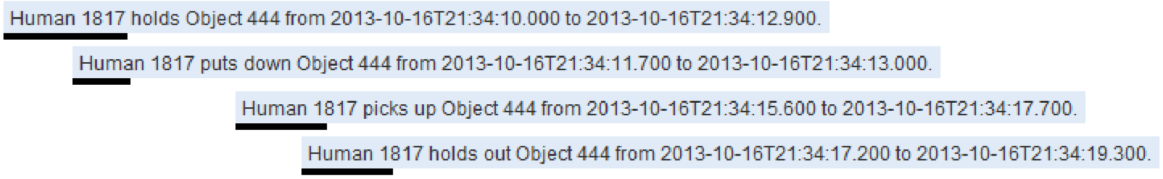
Figure 13: A general-purpose summary of the events listed in Figure 12.

Using this general-purpose summarization strategy on the 64 recorded examples listed in Appendix A, the number of event instances in the generated summary was found to be, overall, about 27% smaller than the original number of event instances recognized. This reduction ranged from approximately 11% for recorded sequences #001 through #008—concerning lower-level activity not involving movable objects—to 43% for recorded sequences #011 through #018—concerning higher-level activity that does involve movable objects.

The degree of reduction produced by summarization also depends on the strategy employed. Treating summarization as a flexible goal, one might imagine invoking specific summarization strategies that are tuned to specific needs. If we are only interested in events involving transportation or transfer of objects by humans, for example, we might alternatively use a more specialized masking strategy such as the following:

- no masking of **time points**,
- no masking of **objects**,
- masking of *all* **attributes** except control of an object by a human,
- masking of **state information** that involves comparisons to constant, reference values,
- for control of an object by a human, masking all remaining **state and change information** except instances of the change varieties “appear”, “disappear” and “not disappear”, and
- masking of **event types**, limiting the set used to the 49 summarization events indicated in Appendix B.

Using this strategy on the same recorded example considered in Figures 11, 12 and 13, the event summary depicted in Figure 14 is produced, for an 87% reduction in the number of events listed.



```
Human 1817 holds Object 444 from 2013-10-16T21:34:10.000 to 2013-10-16T21:34:12.900.
    Human 1817 puts down Object 444 from 2013-10-16T21:34:11.700 to 2013-10-16T21:34:13.000.
              Human 1817 picks up Object 444 from 2013-10-16T21:34:15.600 to 2013-10-16T21:34:17.700.
                  Human 1817 holds out Object 444 from 2013-10-16T21:34:17.200 to 2013-10-16T21:34:19.300.
```

Figure 14: Special-purpose event summary for the recorded sequence concerned in Figures 11, 12 and 13.

3.3 Explanation

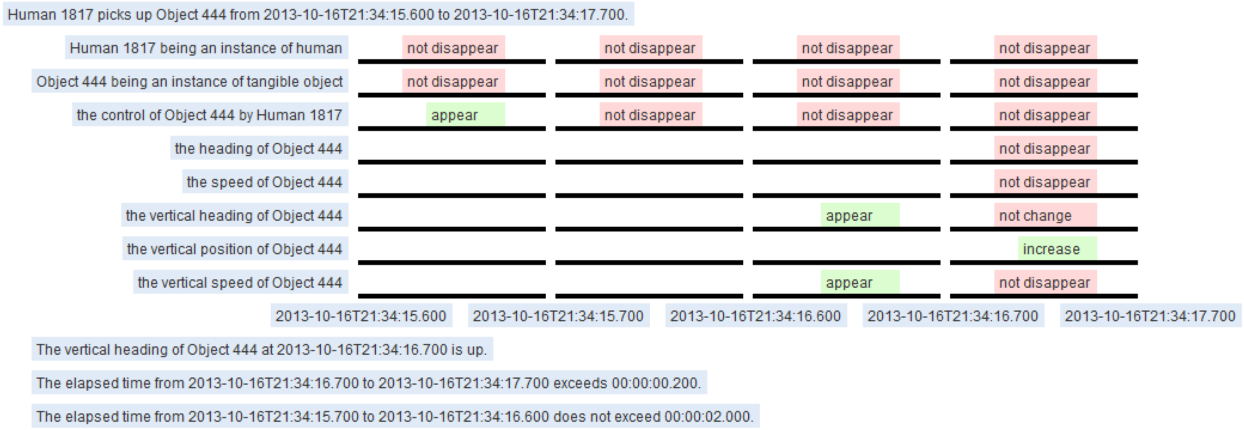
Two kinds of explanation have been explored here. A first variety of explanation provides justification for recognized events. This justification can include the listing of other, simpler events that have also been recognized, and it can also include the listing of scene states and changes that help support the conclusion that a particular event has occurred. The second variety of explanation provides justification for non-recognition of particular types of events. When presented with a query that asks why a particular type of event was not recognized to occur, this type of explanation can list states and changes which would have been expected to occur in the scene, had the event in question taken place, but were found to be missing. Explanation of non-recognized events relies on partial matching of event models to scene information.

3.3.1 Explanation of Recognized Events

When a system determines that a particular event has occurred in a given situation, it can be very informative for the system to be able to explain why it has come to that conclusion. A straightforward way to do this, given a language-based representation, would be simply to list the scene information covered by the event instance to be explained, at the level of states and changes. For example, for the recorded sequence appearing above in Figure 11, Section 3.2, the following event instance is recognized:

```
Human 1817 picks up Object 444 from 2013-10-16T21:34:15.600 to
    2013-10-16T21:34:17.700.
```

A crude explanation of why the system has recognized a "picking up" event within the recorded activity would be to list the states and changes of the matched event model, as illustrated in Figure 15.



Human 1817 picks up Object 444 from 2013-10-16T21:34:15.600 to 2013-10-16T21:34:17.700.

Human 1817 being an instance of human	not disappear	not disappear	not disappear	not disappear
Object 444 being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of Object 444 by Human 1817	appear	not disappear	not disappear	not disappear
the heading of Object 444				not disappear
the speed of Object 444				not disappear
the vertical heading of Object 444			appear	not change
the vertical position of Object 444				increase
the vertical speed of Object 444			appear	not disappear

2013-10-16T21:34:15.600 2013-10-16T21:34:15.700 2013-10-16T21:34:16.600 2013-10-16T21:34:16.700 2013-10-16T21:34:17.700

The vertical heading of Object 444 at 2013-10-16T21:34:16.700 is up.

The elapsed time from 2013-10-16T21:34:16.700 to 2013-10-16T21:34:17.700 exceeds 00:00:00.200.

The elapsed time from 2013-10-16T21:34:15.700 to 2013-10-16T21:34:16.600 does not exceed 00:00:02.000.

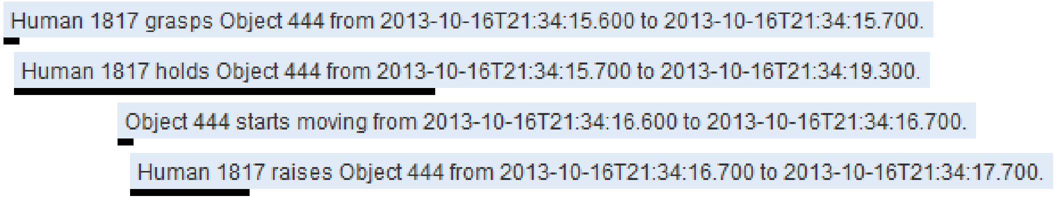
Figure 15: Rough explanation of event recognition formed by listing states and changes.

A better approach would be to use event recognition and summarization to try to produce a more compact description of the covered scene information. This approach was taken in the work described here. After the IMPACT system has recognized an event instance, if a request has been made for an explanation of why that instance was recognized, the system starts by recognizing all instances of a larger "explanation" set of 95 event types as indicated in Appendix

B, then attempts to summarize the targeted event's covered scene information by performing the following tasks:

- It "masks" all scene Level 1 RAVEs not covered by the event instance to be explained, leaving unmasked the subset of scene information specifically associated with this event instance.
- It assesses, for each recognized event instance in the "explanation" set, what proportion of its covered scene Level 1 RAVEs also fall within the coverage of the event instance to be explained. This information is used to prioritize event instances for exclusion or inclusion in the event summary.
- It initializes an event summary by including all event instances it has recognized in the "explanation" set, but specifically excluding the event instance to be explained.
- It masks attributes of lesser importance to summarization, plus state information involving comparisons to reference values, as in the general summarization strategy described in Section 3.2.
- Next, it iteratively removes event instances from the event summary, considering first those event instances that have the lowest percentage of their scene Level 1 RAVE coverage overlapping with the scene coverage of the event to be explained. As with the summarization procedure described in Section 3.2, event instances are excluded from the event summary if they do not uniquely cover any unmasked Level 1 RAVEs in the encoded scene information.
- Finally, after iterative removal of event instances has subsided, there may be some scene states and changes associated with the event to be explained that are not covered by any explaining event instances. These states and changes are rendered in English and included in the event summary.

For the above example of a recognized instance of "picking up", when this procedure is carried out by the IMPACT system, the explanation in Figure 16 is produced:



Human 1817 grasps Object 444 from 2013-10-16T21:34:15.600 to 2013-10-16T21:34:15.700.

Human 1817 holds Object 444 from 2013-10-16T21:34:15.700 to 2013-10-16T21:34:19.300.

Object 444 starts moving from 2013-10-16T21:34:16.600 to 2013-10-16T21:34:16.700.

Human 1817 raises Object 444 from 2013-10-16T21:34:16.700 to 2013-10-16T21:34:17.700.

Figure 16: Improved explanation of event recognition for an instance of "picking up".

Using this approach, explanation of recognized event instances will always succeed. At worst, if it is not possible to summarize covered scene information in terms of other recognized events, then the full set of associated scene states and changes will be listed as an explanation of the recognition. Where other, often smaller, recognized event instances do cover portions of the scene information associated with the event to be explained, then a more compact summary explanation is produced.

3.3.2 Explanation of Non-Recognized Events

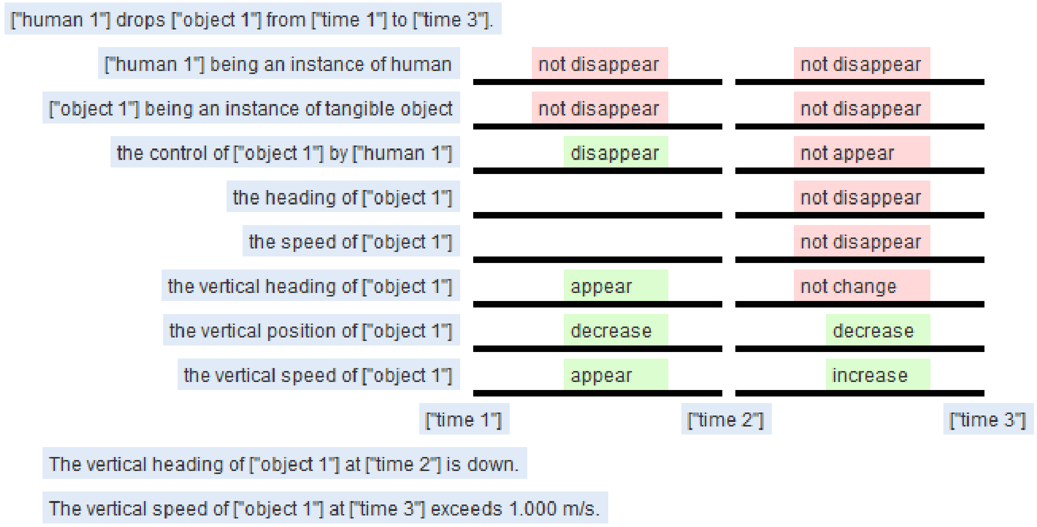
To generate an explanation of why a particular segment of activity was not recognized as an occurrence of a specified type of event, a different kind of procedure must be employed. In this case, there is no recognized event instance and associated scene information from which to produce an explanation.

The matcher used to associate event models with scene information in the work described here has been modified to operate in an alternate mode in which partial matches are sought: matches where, say, 80% or more of the component states and changes in an event model are matched to scene information. Using this mode of operation, when presented with a request to explain the absence of a particular event recognition, the system attempts to form partial matches using successively weakening thresholds for percentage of event model states and changes that must be matched to scene information. As the matcher progresses chronologically through the time intervals in the targeted event model, it enforces its minimum threshold at each time interval—requiring a minimum percentage of states and changes matched within that time interval—and it also imposes the threshold cumulatively at the end of the event model match in order to bring into consideration the matching of supplementary assertions in the event models. If the matcher is able to produce a partial match of the targeted event, then, as an explanation of why that event was not recognized, the system lists those states and changes of the associated event model that were unmatched in the partial match.

As an example, in the recorded sequence of activity discussed above in Sections 3.2 and 3.3.1, a man places a ball on a surface. For this recorded sequence, a request may be submitted to the system, asking it why the following partially instantiated event was not recognized:

```
Human 1817 drops Object 444 from [>=("2013-10-16T21:34:10.000")] to
   [<("2013-10-16T21:34:15.000")]?
```

The associated event model for an event of “dropping an object” is illustrated in Figure 17.



["human 1"] drops ["object 1"] from ["time 1"] to ["time 3"].

["human 1"] being an instance of human	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	disappear	not appear
the heading of ["object 1"]		not disappear
the speed of ["object 1"]		not disappear
the vertical heading of ["object 1"]	appear	not change
the vertical position of ["object 1"]	decrease	decrease
the vertical speed of ["object 1"]	appear	increase

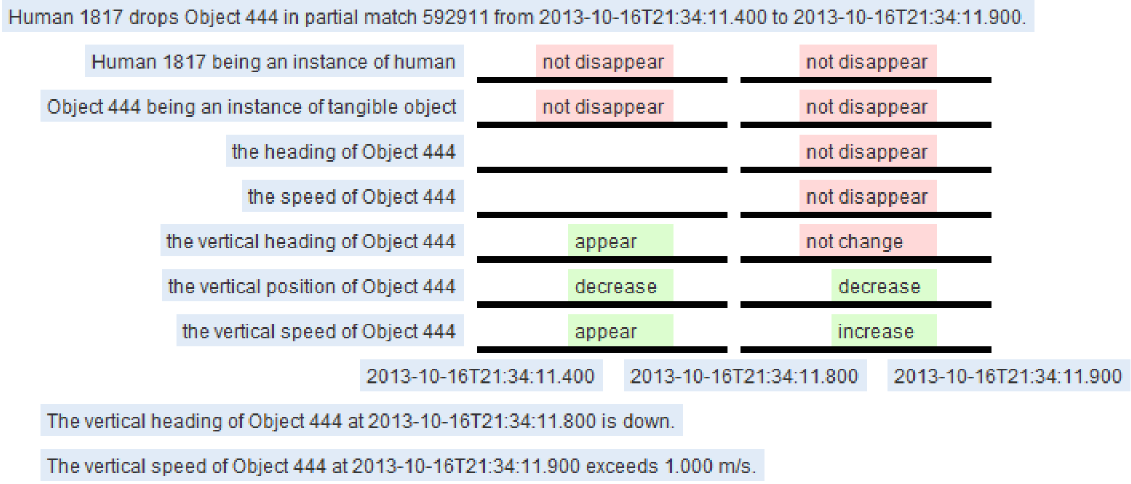
["time 1"] ["time 2"] ["time 3"]

The vertical heading of ["object 1"] at ["time 2"] is down.

The vertical speed of ["object 1"] at ["time 3"] exceeds 1.000 m/s.

Figure 17: Event model for "dropping an object".

Using the matcher in the manner described above, a partial match is obtained for the requested dropping event, at a threshold level of 85%. This partial match is illustrated in Figure 18.



Human 1817 drops Object 444 in partial match 592911 from 2013-10-16T21:34:11.400 to 2013-10-16T21:34:11.900.

Human 1817 being an instance of human	not disappear	not disappear
Object 444 being an instance of tangible object	not disappear	not disappear
the heading of Object 444		not disappear
the speed of Object 444		not disappear
the vertical heading of Object 444	appear	not change
the vertical position of Object 444	decrease	decrease
the vertical speed of Object 444	appear	increase

2013-10-16T21:34:11.400 2013-10-16T21:34:11.800 2013-10-16T21:34:11.900

The vertical heading of Object 444 at 2013-10-16T21:34:11.800 is down.

The vertical speed of Object 444 at 2013-10-16T21:34:11.900 exceeds 1.000 m/s.

Figure 18: Partial match for the requested "dropping" event.

In this partial match, there are two changes of the event model in Figure 17 that have not been matched, concerning control of the object by the human. As a response to the request for an explanation of why a full "dropping" event was not recognized, the system lists these changes, rendered in English. The event was not recognized because the following circumstances were not observed:

```
The control of Object 444 by Human 1817 disappears between
    2013-10-16T21:34:11.400 and 2013-10-16T21:34:11.800.

The control of Object 444 by Human 1817 does not appear between
    2013-10-16T21:34:11.800 and 2013-10-16T21:34:11.900.
```

In this instance, while it was the case that the object did move downward and with increasing speed, as in a “dropping” event, it was not the case that the human relinquished control and continued with an absence of control of the object during that same time.

Explanation of non-recognition of events will succeed only if sufficient scene information exists to match a tested threshold percentage of a targeted event model's states and changes. Within the implemented system, progressive weakening of the threshold for partial matching was stopped at the level of 60% of an event model’s states and changes matched. In general, matching with lower-percentage thresholds was frequently found to produce multiple partial matches for a query, resulting in significantly reduced explanatory power for any one partial match.

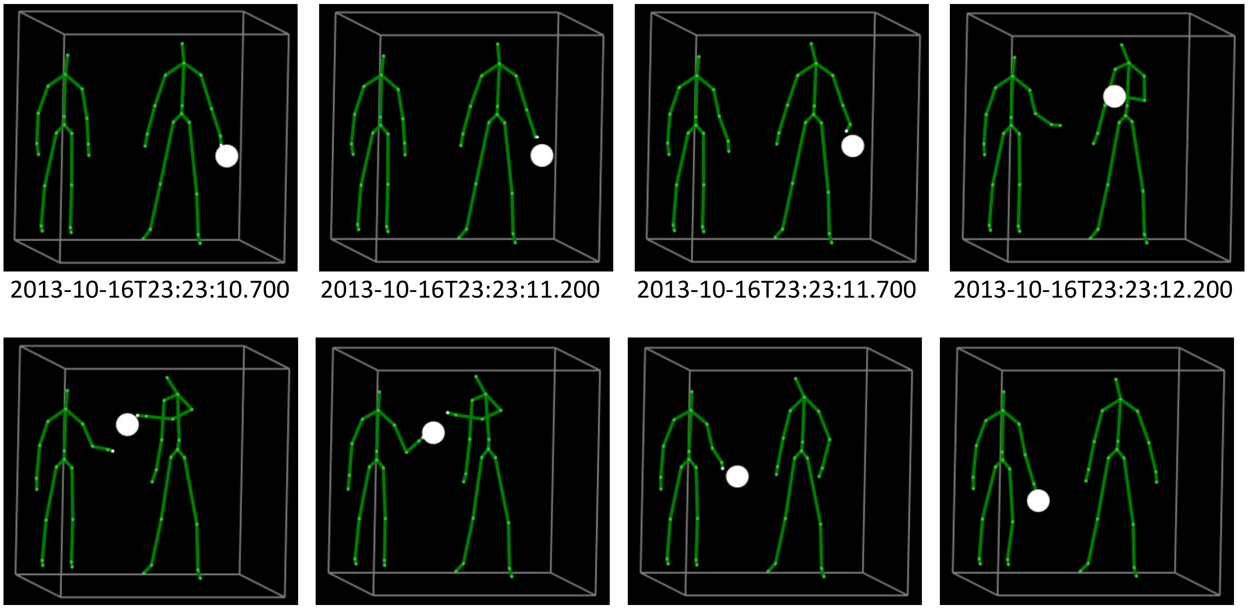
Explanation of non-recognition of events produced in this manner can provide useful information in several respects. In some cases, these explanations can identify anomalous input that, on first glance, might appear to portray an instance of some type of event, but on closer examination, proves not to be an instance of that event. In other cases, these explanations can be used to improve system performance—modifying the way particular attributes are calculated in the scene information, for example, or modifying event models used for matching.

3.4 Prediction

It is natural to think of event recognition as occurring in tandem with the flow of input: forming hypotheses on the basis of initial information up to some point in time in the observed scene, then refining and verifying those hypotheses as information at subsequent times becomes available. Core-periphery matching of event models works in this manner by initially identifying matches for states and changes in the first interval—the first transition—of an event model, then extending that interval's matches and each subsequent interval's matches as necessary to accommodate states and changes sought within the following interval of the event model.

When events are recognized in this sequential manner, it can be expected that it would be possible to predict the completion of a particular type of event, given an increment of scene information that portrays an initial portion of an occurrence of that type of event. The matcher used for event recognition in this work has been modified to operate alternatively in such an incremental manner. Given a segment of scene information continuing up to a specified, ending time point, the matcher operating in this incremental manner will return a number of unfinished matches, where scene information has been associated with initial portions of various event models, but the final portions of those event models remain to be matched by yet-to-be-received information. If the system is then given a request to predict the completion of a particular type of event, it can respond by listing the unfinished matches it has for that type of event.

Figure 19 portrays a portion of the activity in one of the recorded sequences listed in Appendix A. In this portion, a human gives an object to a second human.



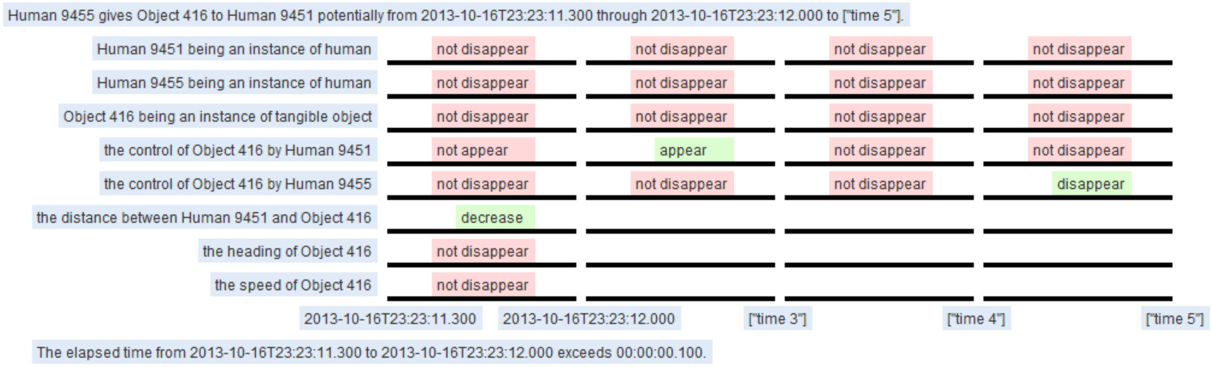
2013-10-16T23:23:10.700 2013-10-16T23:23:11.200 2013-10-16T23:23:11.700 2013-10-16T23:23:12.200

2013-10-16T23:23:12.700 2013-10-16T23:23:13.200 2013-10-16T23:23:13.700 2013-10-16T23:23:14.200

Figure 19: A portion of a recorded sequence involving a "giving" event.

If the matcher is run in its incremental mode of operation with an ending time point of "2013-10-16T23:23:12.000", the matcher will generate three unfinished matches for the event model of "giving an object to a human". The most complete of these matches is listed in Figure 20.

Here, Human 9455 is seen as potentially giving Object 416 to Human 9451, with the first transition in the event having completed and the remaining transitions awaiting completion. Within the first, matched transition, Human 9455 is seen as maintaining control of Object 416 while Object 416 moves closer to Human 9451. However, the system has not yet observed Human 9451 coming into control of Object 416.



Human 9455 gives Object 416 to Human 9451 potentially from 2013-10-16T23:23:11.300 through 2013-10-16T23:23:12.000 to ["time 5"].

Human 9451 being an instance of human	not disappear	not disappear	not disappear	not disappear
Human 9455 being an instance of human	not disappear	not disappear	not disappear	not disappear
Object 416 being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of Object 416 by Human 9451	not appear	appear	not disappear	not disappear
the control of Object 416 by Human 9455	not disappear	not disappear	not disappear	disappear
the distance between Human 9451 and Object 416	decrease			
the heading of Object 416	not disappear			
the speed of Object 416	not disappear			

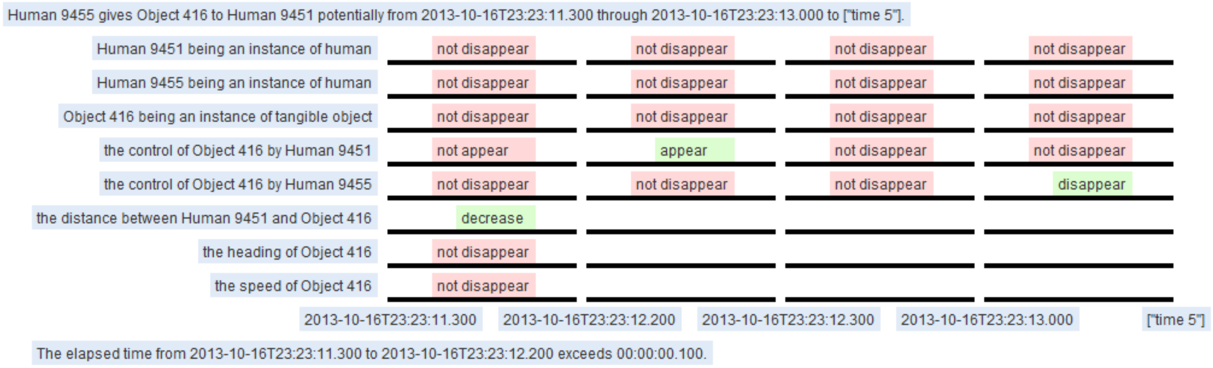
2013-10-16T23:23:11.300 2013-10-16T23:23:12.000 ["time 3"] ["time 4"] ["time 5"]

The elapsed time from 2013-10-16T23:23:11.300 to 2013-10-16T23:23:12.000 exceeds 00:00:00.100.

Figure 20: An unfinished match for Human 9455 giving Object 416 to Human 9451, given input data through time "2013-10-16T23:23:12.000".

The remaining two unfinished matches generated by the matcher leave additional portions of the “give” event model unmatched. In one match, the variable `["time 2"]` appears instead of time point "2013-10-16T23:23:12.000" in Figure 12, addressing the possibility that Object 416 will continue to get closer to Human 9451 as new input data are processed. In the other unfinished match, the variable `["time 1"]` appears instead of time point "2013-10-16T23:23:11.300", addressing the possibility that a later, distinct decrease in distance will begin a “giving” event between these participants.

For this example, if incremental matching is run up to the time point "2013-10-16T23:23:13.000", three different unfinished matches of the “giving” event model are generated. The most complete of these matches is illustrated in Figure 21. Here, the first three transitions of a “giving” event are specified as having completed, with remaining, yet-unobserved activity being disappearance of control of Object 416 by Human 9455.



Human 9455 gives Object 416 to Human 9451 potentially from 2013-10-16T23:23:11.300 through 2013-10-16T23:23:13.000 to ["time 5"].

Human 9451 being an instance of human	not disappear	not disappear	not disappear	not disappear
Human 9455 being an instance of human	not disappear	not disappear	not disappear	not disappear
Object 416 being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of Object 416 by Human 9451	not appear	appear	not disappear	not disappear
the control of Object 416 by Human 9455	not disappear	not disappear	not disappear	disappear
the distance between Human 9451 and Object 416	decrease			
the heading of Object 416	not disappear			
the speed of Object 416	not disappear			

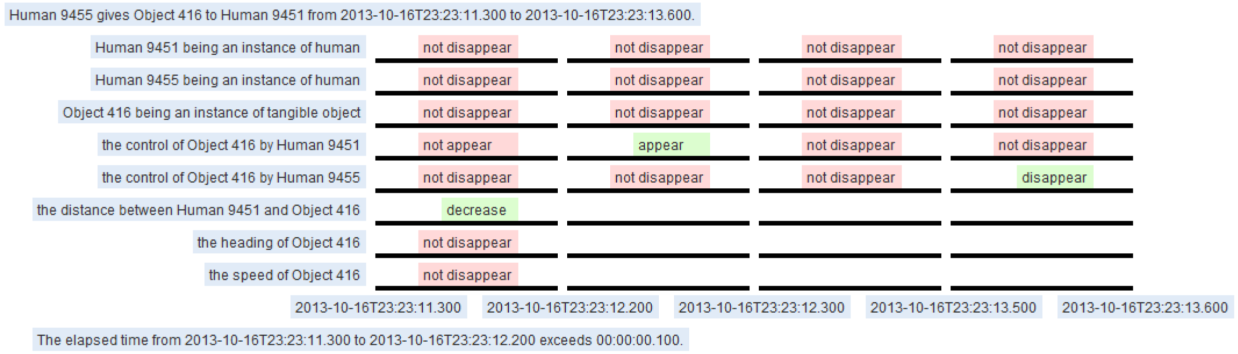
2013-10-16T23:23:11.300 2013-10-16T23:23:12.200 2013-10-16T23:23:12.300 2013-10-16T23:23:13.000 ["time 5"]

The elapsed time from 2013-10-16T23:23:11.300 to 2013-10-16T23:23:12.200 exceeds 00:00:00.100.

Figure 21: An unfinished match for Human 9455 giving Object 416 to Human 9451, given input data through time "2013-10-16T23:23:13.000".

The other two unfinished matches returned by the system indicate completion of one and two transitions in the event model, respectively, with endpoints of these transitions as yet unspecified. Taken together, these 3 unfinished matches motivate a prediction that, given appropriate, subsequent input scene information, it is potentially the case that Human 9451 is in the process of giving Object 416 to Human 9451.

Following the prediction of “giving” through to completion, if the matcher is run with a later endpoint, say "2013-10-16T23:23:14.000", a full match of the “giving” event model will be generated, as illustrated in Figure 22.



Human 9455 gives Object 416 to Human 9451 from 2013-10-16T23:23:11.300 to 2013-10-16T23:23:13.600.

Human 9451 being an instance of human	not disappear	not disappear	not disappear	not disappear
Human 9455 being an instance of human	not disappear	not disappear	not disappear	not disappear
Object 416 being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of Object 416 by Human 9451	not appear	appear	not disappear	not disappear
the control of Object 416 by Human 9455	not disappear	not disappear	not disappear	disappear
the distance between Human 9451 and Object 416	decrease			
the heading of Object 416	not disappear			
the speed of Object 416	not disappear			

2013-10-16T23:23:11.300 2013-10-16T23:23:12.200 2013-10-16T23:23:12.300 2013-10-16T23:23:13.500 2013-10-16T23:23:13.600

The elapsed time from 2013-10-16T23:23:11.300 to 2013-10-16T23:23:12.200 exceeds 00:00:00.100.

Figure 22: Completed match for Human 9455 giving Object 416 to Human 9451.

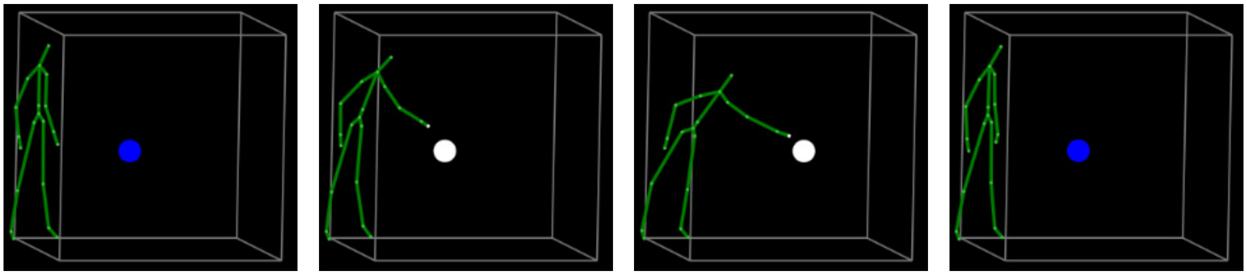
Operating the core-periphery matcher in an incremental mode forces it to reveal its matches under construction. Since it is not possible for the matcher to complete the matching of any event model without first forming unfinished, incomplete matches of this sort, then it is the case that for every event instance eventually recognized by the system, it is possible to query the matcher as it is processing input information after the start of that event instance but before the end and to receive such predictions of completion of the event of interest.

3.5 Question Answering

An inherent byproduct of using language as a basis for representing scene information and event occurrences is that it is relatively straightforward to answer a range of questions, from high-level questions about events and their participants to low-level questions about underlying states and changes. Humans also possess this ability, of course, and the fact that humans can so easily introspect about states, changes and events in an observed scene would seem to suggest that humans’ internal representation of this information, if not overtly language-based, is at least easily associated with language. Furthermore, if such a language-based or language-related encoding of information is present in humans, it would seem likely that this encoding contributes in at least some way to human reasoning about activity.

In a previous application involving event recognition concerning vehicle activities [Borchardt, *et al.*, 2014], we constructed a question-answering interface that allows users to submit their questions in unrestricted English. In that application, user questions, which may involve a variety of syntactic forms and vocabulary elements, are first converted into a canonical, language-based form by the START information access system. START's interpretation of user questions is based on representation in terms of nested ternary expressions, coupled with matching on the basis of natural language annotations [Katz, 1990; Katz, 1997]. In the previous application concerning vehicle activities, START's interpretation of user questions results in the re-expression of those questions as Möbius requests, which are then answered by the IMPACT system.

In the work described here, the ability to ask questions in unrestricted English has not been added, and it is the case that user questions must be pre-encoded in Möbius prior to answering. A complete back end for question answering has been constructed, however, using the IMPACT system, and this has enabled IMPACT to answer a broad range of question varieties. Figure 23 presents a recorded sequence of activity that serves as an illustration of IMPACT's support for question answering. In this recorded sequence, a man starts by reaching to touch a ball resting on a surface, then retracts his hand, reaches again to push the ball, retracts his hand again, and reaches again to pull the ball back.



2013-10-16T23:53:08.400 2013-10-16T23:53:10.300 2013-10-16T23:53:14.000 2013-10-16T23:53:18.300

Figure 23: A recorded sequence of activity.

The following are examples of questions that can be answered by the implemented system. Each question is listed first in natural English, then in a form, also employed in the above paragraphs, that presents the Möbius encoding of that question in a way that combines generated English with symbolic variables enclosed in brackets. IMPACT's responses are listed third, in readable English, using a similar generation algorithm applied to the underlying Möbius encodings.

- questions about scene states and changes:

    ("Is Human 578's left hand in contact with Object 233 at 23:53:10.000?")

    The contact between Human 578 Left Hand and Object 233 exists at 2013-10-16T23:53:10.000?

  The contact between Human 578 Left Hand and Object 233 exists at 2013-10-16T23:53:10.000.

    ("What is the speed of Human 578's left hand at 23:53:15.700?")

    The speed of Human 578 Left Hand at 2013-10-16T23:53:15.700 is [what]?

  1.849 m/s

    ("What happens to the distance between Human 578 and Object 233 between 23:53:15.700 and 23:53:15.800?")

    The distance between Human 578 and Object 233 [@=(increase)] between 2013-10-16T23:53:15.700 and 2013-10-16T23:53:15.800?

  The distance between Human 578 and Object 233 decreases between 2013-10-16T23:53:15.700 and 2013-10-16T23:53:15.800.

- questions about entities—including times—involved in scene states and changes:

    ("When does contact appear between Human 578's left hand and Object 233?")

    The contact between Human 578 Left Hand and Object 233 appears between [] and []?

  The contact between Human 578 Left Hand and Object 233 appears between 2013-10-16T23:53:09.600 and 2013-10-16T23:53:09.700.
  The contact between Human 578 Left Hand and Object 233 appears between 2013-10-16T23:53:12.200 and 2013-10-16T23:53:12.300.
  The contact between Human 578 Left Hand and Object 233 appears between 2013-10-16T23:53:16.200 and 2013-10-16T23:53:16.300.

(“What humans are present in the scene?”)

[what] is an instance of human at []?

Human 578

(“What items come into contact with Object 233?"

The contact between [what] and Object 233 exists at []?

Human 578 Left Hand

- questions that perform restricted searches over scene states and changes:

(“When is the horizontal speed of Object 233 greater than 1.0 m/s?”)

The horizontal speed of Object 233 at [what] is [>=("1.000 m/s"), <>(null)]?

2013-10-16T23:53:13.300
2013-10-16T23:53:13.400
2013-10-16T23:53:17.200

(“How does the horizontal speed of Object 233 change over 23:53:13.100 to 23:53:13.500?”)

The horizontal speed of Object 233 [@=(increase)] between [>=("2013-10-16T23:53:13.100")] and [<=("2013-10-16T23:53:13.500")]?

The horizontal speed of Object 233 increases between 2013-10-16T23:53:13.100 and 2013-10-16T23:53:13.200.
The horizontal speed of Object 233 increases between 2013-10-16T23:53:13.200 and 2013-10-16T23:53:13.300.
The horizontal speed of Object 233 increases between 2013-10-16T23:53:13.300 and 2013-10-16T23:53:13.400.
The horizontal speed of Object 233 decreases between 2013-10-16T23:53:13.400 and 2013-10-16T23:53:13.500.

- questions about available event models:

  ("What events can be recognized?")

  [what] is registered for a recognition model?

  (…a listing of 102 event types…)

- questions about component states, changes and assertions of event models:

  ("What happens when a human grasps an object?")

  [what] is a component of [=(grasp(subject:human [some, "human 1"], object:tangible_object [some, "object 1"], from:time [some, "time 1"], to:time [some, "time 2"]).)]?

  ["human 1"] does not cease to be an instance of human between ["time 1"] and ["time 2"].
  ["object 1"] does not cease to be an instance of tangible object between ["time 1"] and ["time 2"].
  The control of ["object 1"] by ["human 1"] appears between ["time 1"] and ["time 2"].
  The speed of ["object 1"] at ["time 1"] does not exceed 0.800 m/s.

- questions about recognized event instances:

  ("Does Human 578 push Object 233?")

  Human 578 pushes Object 233 from [] to []?

  Human 578 pushes Object 233 from 2013-10-16T23:53:13.100 to 2013-10-16T23:53:13.500.

  ("What occurrences are there of extending an arm?")

  [] extends [] from [] to []?

  Human 578 extends Human 578 Left Arm from 2013-10-16T23:53:09.400 to 2013-10-16T23:53:10.100.
  Human 578 extends Human 578 Left Arm from 2013-10-16T23:53:12.000 to 2013-10-16T23:53:12.400.
  Human 578 extends Human 578 Left Arm from 2013-10-16T23:53:15.600 to 2013-10-16T23:53:16.000.

- questions about component states, changes and assertions of recognized event instances:

  ("What happens when Human 578 releases Object 233 between 23:53:17.900 and 23:53:18.000?")

  [what] is a component of [=(release(subject:human "Human 578", object:tangible_object "Object 233", from:time "2013-10-16T23:53:17.900", to:time "2013-10-16T23:53:18.000").)]?

  Human 578 does not cease to be an instance of human between 2013-10-16T23:53:17.900 and 2013-10-16T23:53:18.000.
  Object 233 does not cease to be an instance of tangible object between 2013-10-16T23:53:17.900 and 2013-10-16T23:53:18.000.
  The control of Object 233 by Human 578 disappears between 2013-10-16T23:53:17.900 and 2013-10-16T23:53:18.000.
  The speed of Object 233 at 2013-10-16T23:53:18.000 does not exceed 0.800 m/s.

- questions about entities—including times—involved in recognized event instances:

  ("When does Human 578 begin pulling Object 233?")

  Human 578 pulls Object 233 from [what] to []?

  2013-10-16T23:53:16.900

  ("What object does Human 578 pull?")

  Human 578 pulls [what] from [] to []?

  Object 233

Questions such as these are answered by performing searches over language-encoded scene information, knowledge of what happens during various types of events, recognized instances of events, and component states and changes associated with recognized event instances. Any Möbius question that refers to this range of encoded information will elicit a suitable response, consistent with the information existing within the system.

3.6 Other Potential Capabilities

If it is possible to achieve reasonable performance on event recognition, summarization, explanation, prediction and question answering using human-supplied knowledge expressed in a language-motivated representation, then it can be asked what other capabilities might be supported by this general approach. One way to explore the possibilities is to consider a conceptual model of human reasoning about events that motivates the above-described techniques. This model, illustrated in Figure 24, assumes three broad channels through which interactions with the world are carried out: perception, language, and motor control. Within the intersection of all three channels—that is, influenced by all three—a realm of reasoning about events is hypothesized, where the underlying information and operations can to an extent be articulated in language and could thus be assumed, at least partially, to be supported by language-based representations and reasoning. [Borchardt, 1994] presents a model of this realm of reasoning in which the underlying representations of timestamped information are associated along three broad dimensions: (1) causality, which is aligned with the temporal ordering of timestamped information, (2) abstraction, which concerns principally non-information-preserving transformations of timestamped information—largely compression of information, and (3) analogy, which concerns principally information-preserving transformations of timestamped information.

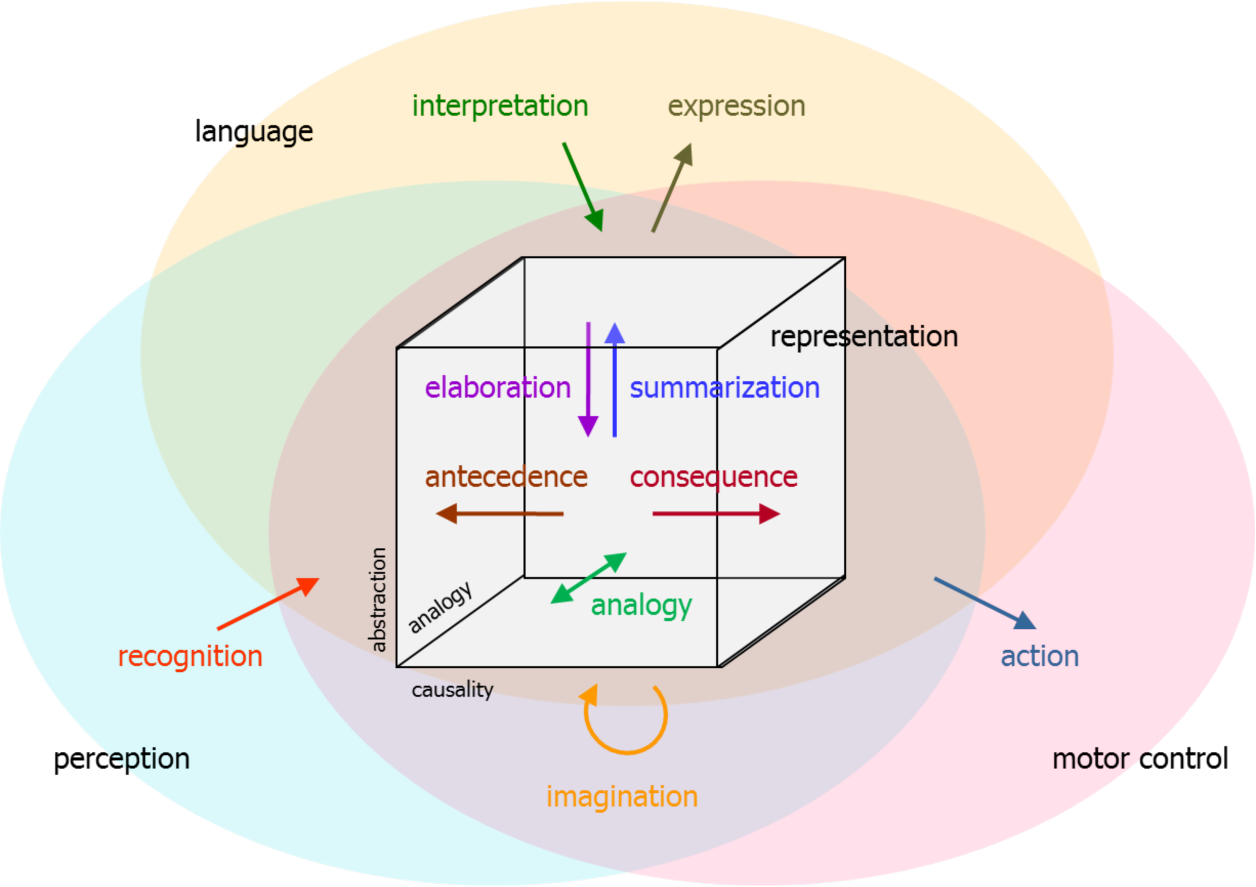


Figure 24: A conceptual model of capabilities for reasoning about events.

Specific varieties of reasoning about events and other timestamped information appear as labeled arrows in Figure 24. Several of these have been addressed in the sections above. Event recognition takes perceptual input and casts this in language-oriented states and changes, then continues by abstracting that level of description to the higher level of event occurrences. Summarization of event sequences is also an abstraction operation, reducing the number of events that are included in a description. Explanation of recognized events and explanation of non-recognized events are both elaboration operations that add detail to a description. Prediction of event completions is one variety of reasoning about consequence. Question answering concerns interpretation of language input, only partially addressed here, followed by extraction of encoded, timestamped information at all levels of abstraction, followed by expression of the encoded information in language.

There are also a number of other varieties of reasoning, as depicted in this model, that may be amenable to the general approach outlined here—that is, representation of information in language-oriented ways, plus reasoning accomplished by matching and transformation operations. Examples of these other varieties of reasoning are:

- ***advanced summarization,*** in which not only is redundant information removed from a generated description of activity, but as well, information deemed overly detailed is restated at higher levels of abstraction,
- ***envisioning***, in which described events are elaborated into states and changes in order to assess their compatibility and ordering with respect to other events, and feasibility within the context of other timestamped information available to the reasoner,
- ***interpretation of metaphors***, similar to envisioning, in which metaphorical references are first elaborated as if literal, then mapped in terms of attributes and object types, to other domains, where they are similarly assessed for compatibility, ordering, and feasibility,
- ***analogical reasoning***, in which larger clusters of event models from one domain are mapped in terms of attributes and object types to another domain for purposes of carrying out other reasoning operations,
- ***prediction through event chaining***, a form of reasoning about consequence, in which event models are placed in sequence by identifying overlaps where the end of one event's unfolding substantially matches the beginning of another event's unfolding,
- ***explanation through event chaining***, in which event models are chained in the direction of antecedence, from observed or reported events to hypothesized preceding events that could have led to their occurrence, and

- ***advanced imagination*** and ***planning***, which may be modeled as involving larger loops of multiple operations: envisioning events; testing for compatibility, ordering and feasibility; predicting consequences and explaining in terms of antecedents, and recognizing event occurrences from assembled state and change information.

In [Borchardt, 1994], several of these varieties of reasoning are explored in the context of understanding verbal descriptions of activity, where the input is a collection of reported event instances, arriving through the language channel of interaction with the world. Following from the work described here, it would appear to be the case that the same kinds of reasoning operations could also take place, using similar representations and mechanisms, on the basis of input derived from observations of a changing scene—inputs received through the perceptual channel of interaction with the world.

4. Conclusions

The work presented in this report amounts to a proof-of-concept investigation into whether it may be possible to simultaneously address two goals: (1) to replicate, in some manner of performance, the interrelated capabilities of event recognition, summarization, explanation, prediction, and question answering, and (2) to do so using language-encoded, human-supplied knowledge about what typically happens during various types of events. Both goals have been accomplished by this investigation. On the set of 64 recorded activity sequences, the system recognizes 83% of noted event occurrences and proposes event occurrences deemed appropriate 81% of the time, summarization matches human intuition and reduces the number of events included in a description by approximately 27%, and explanation, prediction and question answering have produced results they are expected to produce, uniformly in the targeted contexts. The supporting event models, listed in Appendix B, do, to a reasonable extent, match human intuition regarding the temporal unfolding of the targeted types of events. Furthermore, the approach explored in this work yields itself to relatively quick development and is efficient in its execution—operating in real-time on a standard laptop computer and sharing resources and techniques across the related functional capabilities.

There are two principal areas of opportunity for extending this work. First is to improve the performance of the implemented system. New attributes can be added: speed and heading relative to body orientation, support, attachment, containment, and so forth. New event models can be added: shaking hands, exchanging two objects, turning and stepping relative to body orientation, and so forth. Existing event models can be refined by applying them to a larger corpus of recorded activity sequences. Machine learning can be applied to fine-tune the event models. Additional, related capabilities such as those described in Section 3.6 can be implemented.

The second area of opportunity is to assess and refine the approach taken here through reference to experimental results in cognitive science. Part of this has to do with psychological experiments: do subjects predict that certain types of events can co-occur, or be in conflict with one another, or lead to the occurrence of other events in the same ways predicted by software matching between event models of the sort developed here? Another part is linguistic: do subjects find as semantically related or unrelated, as redundant or non-redundant various combinations of statements concerning event occurrences and the occurrence of underlying states and changes? This sort of investigation might exploit variants of the "but" test described in [Bendix, 1966]. Finally, examination of models of neural circuitry can shed light on whether it may be plausible to view human recognition of events and reasoning about events as one of matching models of typical temporal unfolding of events to encoded scene information. For example, there is some evidence that particular circuits in the human brain may be responsible for "change detection" regarding attributes like object direction and possibly speed [Riggall and Postle, 2012]. Of particular interest would be whether change detection circuits might exist for a range of significant attributes, and whether other circuits might assemble clusters of temporally-organized, perceived attribute changes and states into affirmative recognitions of event instances.

Using results from cognitive science to guide the development of machine capabilities for recognition and reasoning about events has important, potential advantages. It is entirely possible that recognition of events and reasoning about events are so intimately connected with human cognition that the only way to replicate these capabilities robustly may be to pursue a cognitively-inspired approach. Also, even if there exist other mechanisms, not cognitively motivated, by which recognition and reasoning about events might be accomplished, it is likely that natural mechanisms are relatively more efficient due to the longevity of their existence and refinement in nature.

A particularly significant way in which language may shed light on mechanisms for reasoning about events involves language-oriented attributes. It is often argued that unguided statistical learning of machine capabilities can lead to overfitting of even reasonably large datasets by exploiting incidental, context-specific features—features that do not carry over to the broader range of related contexts not covered by the training data set. In contrast, it may be the case that human nature and cognitive development have provided us with important, relatively context-independent attributes for use in reasoning about scenes, and that, given our facility with language, we have come to identify these attributes with names. It is interesting to note that the named attributes we have to describe simple physical interactions do have a substantial degree of context independence: “position”, “speed”, “heading”, “elevation”, “distance”, being “inside”, “contact”, “support”, “pressure”, “attachment”, and so forth, are all largely viewpoint-independent, scale independent, rotation-independent, and illumination independent. It is possible, then, that these attributes have come into use out of necessity and utility, and are now identifiable to humans through the language references we have associated with them. Constructing machine capabilities based on the use of such attributes thus has the potential to leverage significant aspects of human evolution and development.

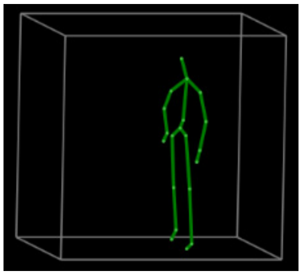
A. Input Examples

The following 16 activity sequences were recorded using a Kinect for Windows sensor plus supporting software for tracking mid-sized portable objects and calculating contacts between hands, feet and objects. Four instances of each activity sequence were recorded, yielding 64 recorded examples in all. Each sequence lasts 10 seconds, with the three-dimensional position information and contact information recorded at 10 frames per second. Within the sequences, 1 or 2 human participants appear, plus 0 or 1 object. For the object, an inflatable beach ball was used, approximately 25 cm in diameter.

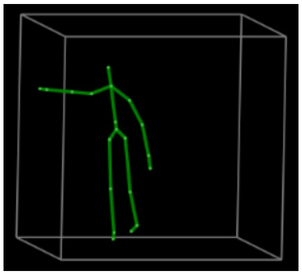
In the descriptions that follow, each sequence’s activity is summarized in English, and representative frames of the activity are portrayed, drawn from one of the four recorded instances for that sequence. By default, body parts are depicted in green and objects are depicted in blue. However, when there is contact between hands, feet and objects, these elements are depicted in white. As the recorded sequences were created for purposes of development and experimentation, they include not only events addressed by the research described here, but also other types of events not yet addressed by this work.

Sequence 001

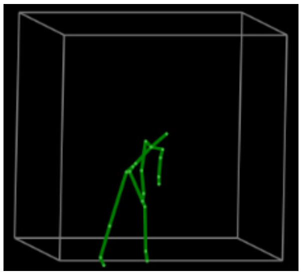
A man stands at right, facing left, takes a step to the left, extends and lowers his right arm, turns to face the sensor, reaches down with his right arm, and straightens up.



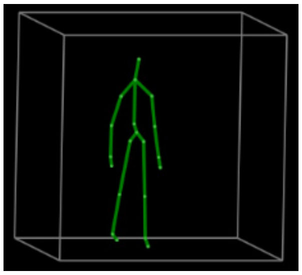
2013-05-09T11:31:46.000



2013-05-09T11:31:48.400



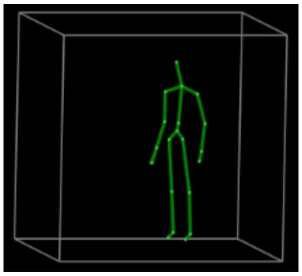
2013-05-09T11:31:51.400

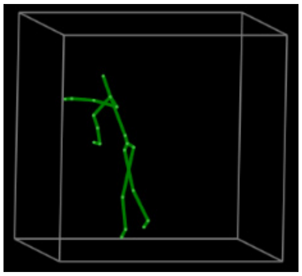


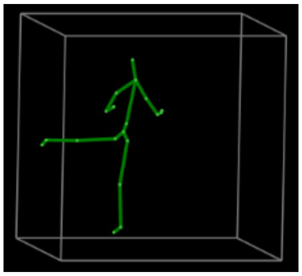
2013-05-09T11:31:55.900

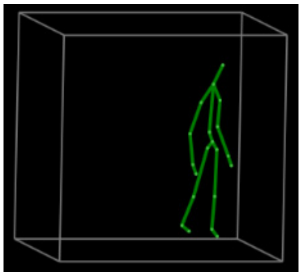
Sequence 002

A man stands at right, facing left, takes a step to the left, punches with his left arm, kicks with his right leg, then turns and walks to the right.


2013-05-09T11:44:02.000

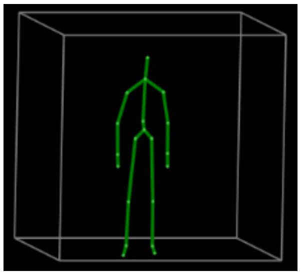

2013-05-09T11:44:04.000

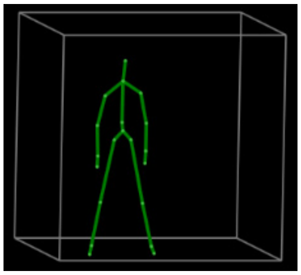

2013-05-09T11:44:05.400

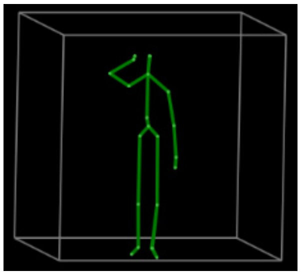

2013-05-09T11:44:11.900

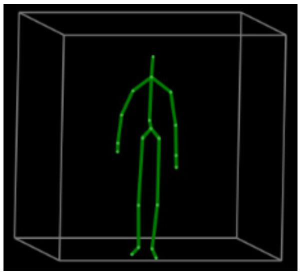
Sequence 003

A man stands facing the sensor, leans to shake his right foot, then his left foot, widens his stance, narrows his stance, and salutes.


2013-05-09T11:54:56.000

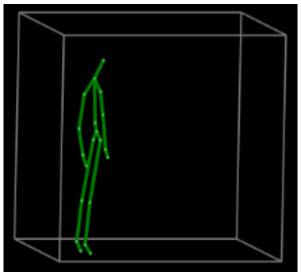

2013-05-09T11:55:02.600

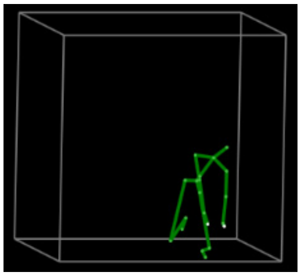

2013-05-09T11:55:04.800

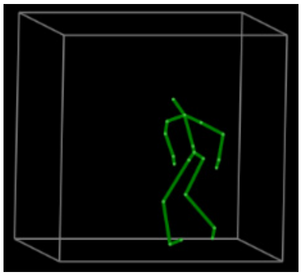

2013-05-09T11:55:05.900

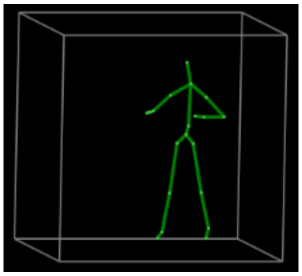
Sequence 004

A man stands at left, walks right, stumbles, stands up, bends over, straightens up, stretches his back, twists to the left, and untwists to face the sensor.


2013-05-09T12:07:09.000

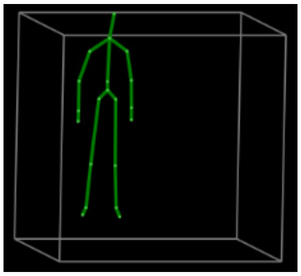

2013-05-09T12:07:11.500

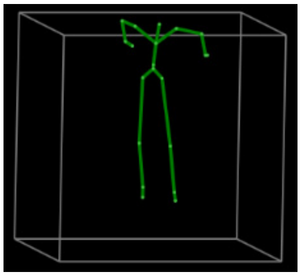

2013-05-09T12:07:12.900

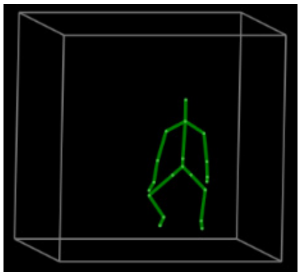

2013-05-09T12:07:18.900

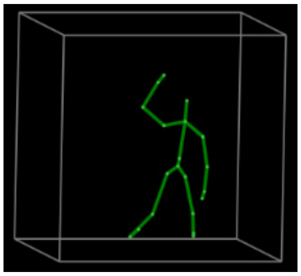
Sequence 005

A man stands at left, facing the sensor, jumps sideways and lands to the right, facing the sensor, squats, lowers both knees to the ground, raises his right arm, and waves.


2013-05-09T12:20:43.000

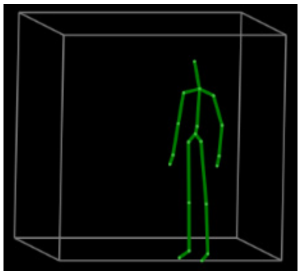

2013-05-09T12:20:45.000

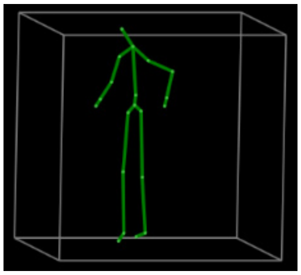

2013-05-09T12:20:47.800

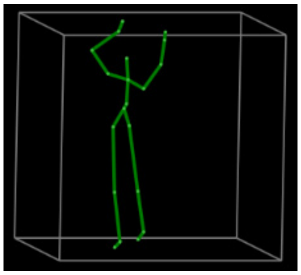

2013-05-09T12:20:52.900

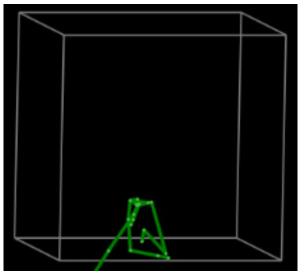
Sequence 006

A man stands at right, facing left, takes a step to the left, jumps up and down twice, raises both arms, and crumples to the ground.


2013-05-09T12:25:40.000

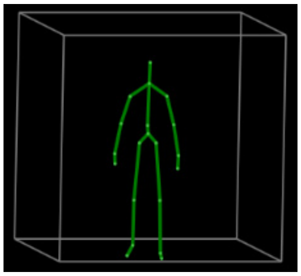

2013-05-09T12:25:44.000

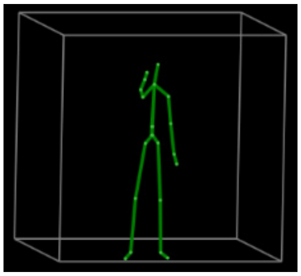

2013-05-09T12:25:46.600

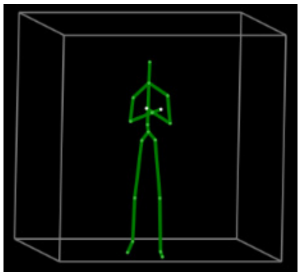

2013-05-09T12:25:49.900

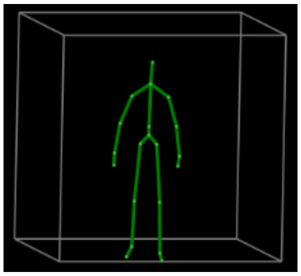
Sequence 007

A man stands facing the sensor, reaches toward the sensor with his right arm, lowers his right arm, crosses his arms, lowers his arms, twists his shoulders to the left, and straightens out his shoulders.


2013-05-09T12:31:41.000

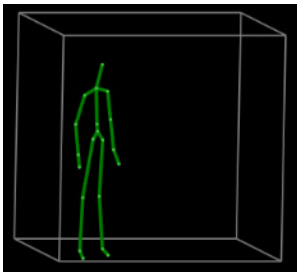

2013-05-09T12:31:43.400

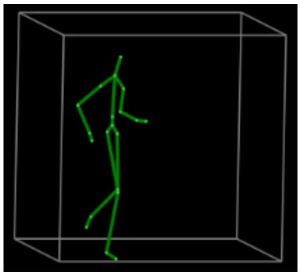

2013-05-09T12:31:47.600

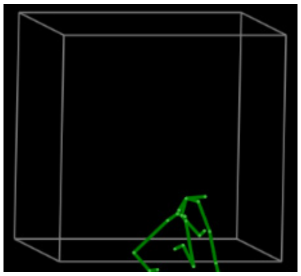

2013-05-09T12:31:50.900

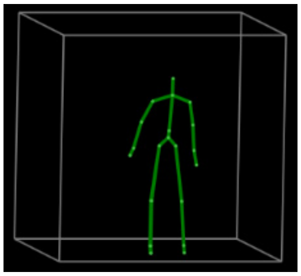
Sequence 008

A man stands at left, facing right, runs in place, crumples to the ground, stands up, and takes a bow.


2013-05-09T12:39:03.000

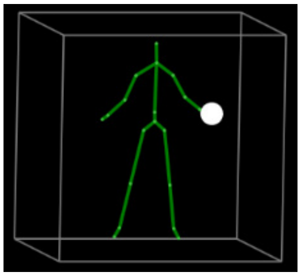

2013-05-09T12:39:04.500

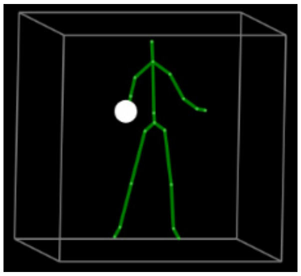

2013-05-09T12:39:07.300

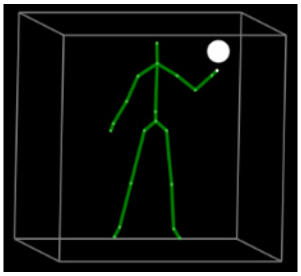

2013-05-09T12:39:12.900

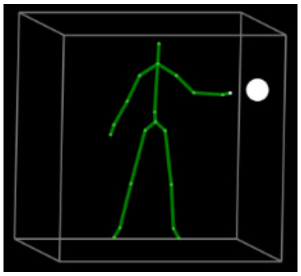
Sequence 011

A man stands facing the sensor, holding a ball in his left hand, tosses the ball to his right hand and back to his left hand, waves the ball back and forth, then up and down, then back and forth.


2013-10-16T21:19:09.700

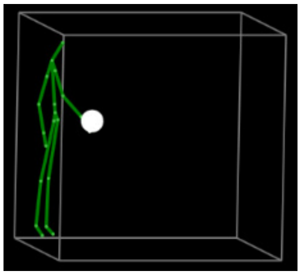

2013-10-16T21:19:12.100

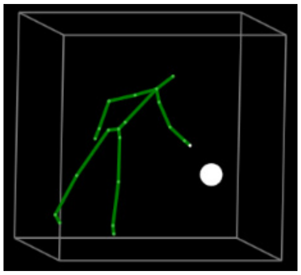

2013-10-16T21:19:16.300

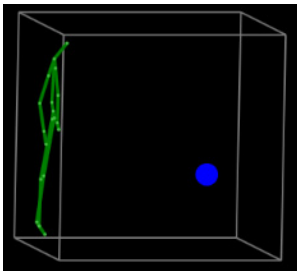

2013-10-16T21:19:19.400

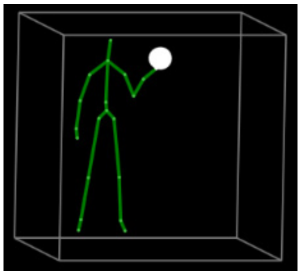
Sequence 012

A man stands at left, facing right, holding a ball in his left hand, steps and reaches to the right, places the ball on an unrecorded bench, straightens up and backs off to the left, then reaches forward and picks up the ball with his left hand, faces the sensor and holds the ball up at shoulder level.


2013-10-16T21:32:22.800

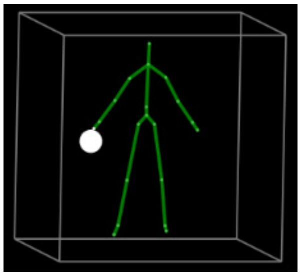

2013-10-16T21:32:25.600

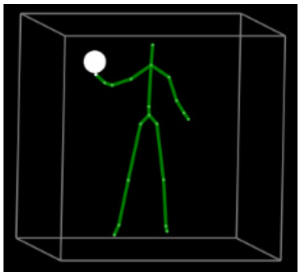

2013-10-16T21:32:27.500

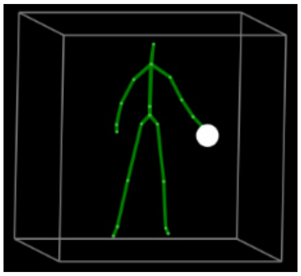

2013-10-16T21:32:32.500

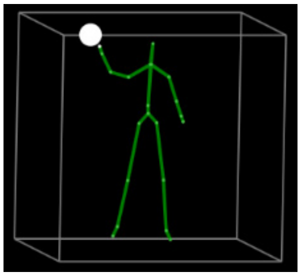
Sequence 013

A man stands facing the sensor, holding a ball in his right hand, rapidly exchanges the ball several times between his hands, shakes the ball up and down with his right hand, rapidly exchanges the ball again several times between hands, shakes the ball up and down with his left hand, transfers the ball to his right hand, and holds the ball up at head level.


2013-10-16T21:50:17.700

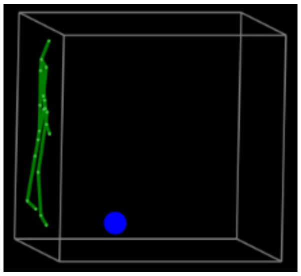

2013-10-16T21:50:22.000

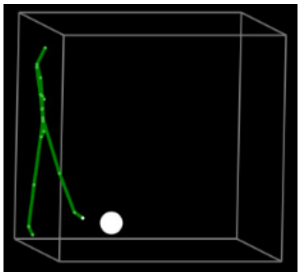

2013-10-16T21:50:25.700

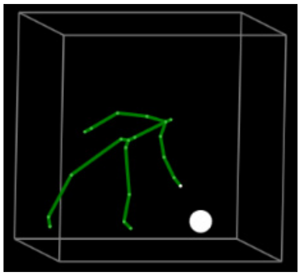

2013-10-16T21:50:27.400

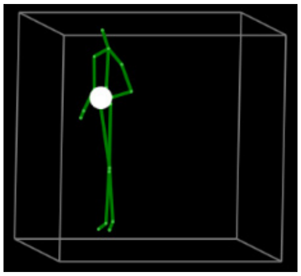
Sequence 014

A man stands at left, facing right, bends slightly to look at a ball on the ground, kicks the ball gently to the right with his left foot, steps to the right, bends over and picks up the ball with his left hand, and turns to the left, holding the ball.


2013-10-16T22:48:12.700

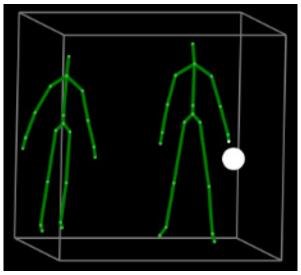

2013-10-16T22:48:16.600

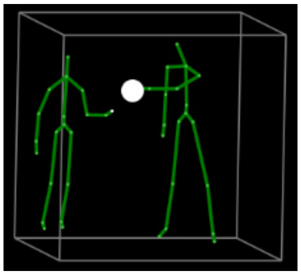

2013-10-16T22:48:19.000

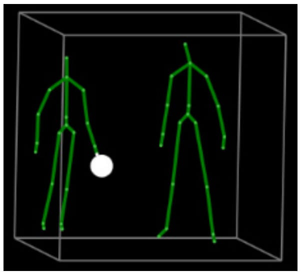

2013-10-16T22:48:22.600

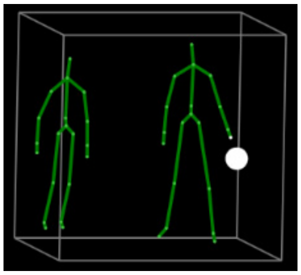
Sequence 015

A man stands at right, facing the sensor, holding a ball in his left hand, hands the ball to a woman standing at left, facing the sensor, who takes the ball with her left hand and holds it briefly, then hands it back to the man, who takes the ball with his left hand and holds the ball.


2013-10-16T23:18:04.600

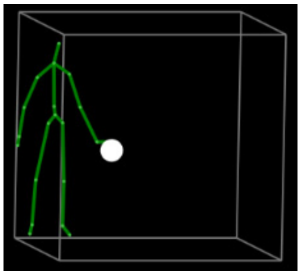

2013-10-16T23:18:07.100

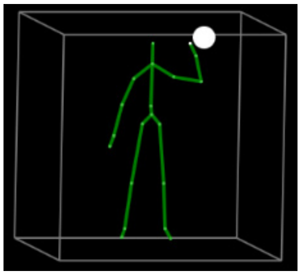

2013-10-16T23:18:08.700

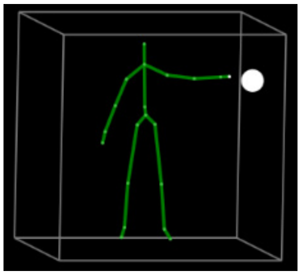

2013-10-16T23:18:14.300

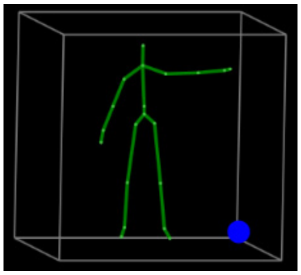
Sequence 016

A man stands at left, facing the sensor, holding a ball in his left hand, takes a step to the right, faces the sensor, holds the ball up at head height, holds the ball out to the right, then drops the ball, which falls to the ground and bounces a few times before coming to rest.


2013-10-16T23:39:59.000

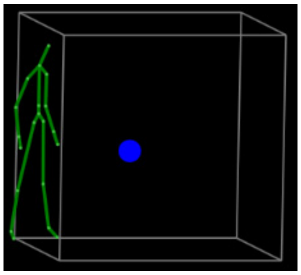

2013-10-16T23:40:02.800

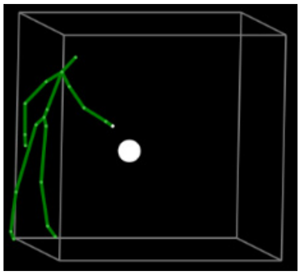

2013-10-16T23:40:04.400

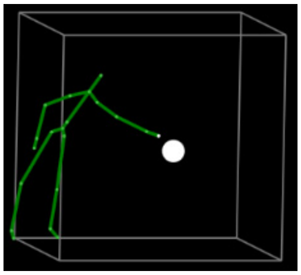

2013-10-16T23:40:08.700

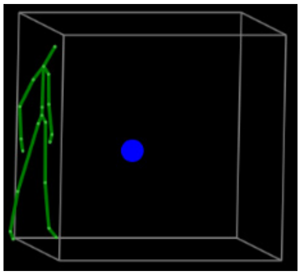
Sequence 017

A man stands at left, facing right, reaches to briefly touch a ball resting on an unrecorded table, retracts his hand, reaches again to push the ball to the right, retracts his hand, and reaches a third time to pull the ball back to the left.


2013-10-16T23:53:08.400

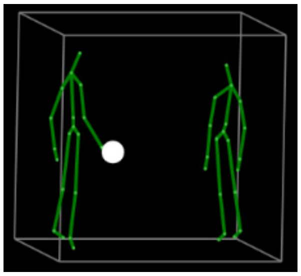

2013-10-16T23:53:10.300

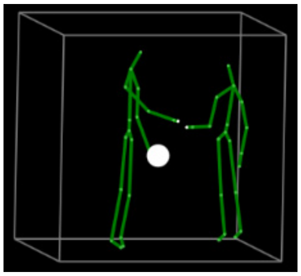

2013-10-16T23:53:14.000

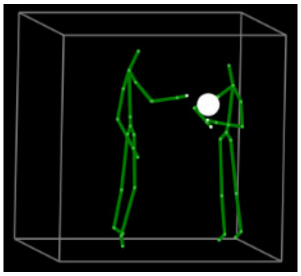

2013-10-16T23:53:18.300

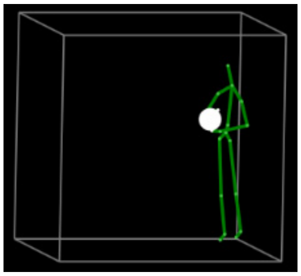
Sequence 018

A man stands at left, facing right with a ball in his left hand, takes a step to the right and shakes right hands with a woman at right, facing left, then gives the ball to the woman, who receives the ball with both hands and holds the ball as the man turns and walks away to the left.


2014-07-29T18:07:57.100


2014-07-29T18:07:59.500

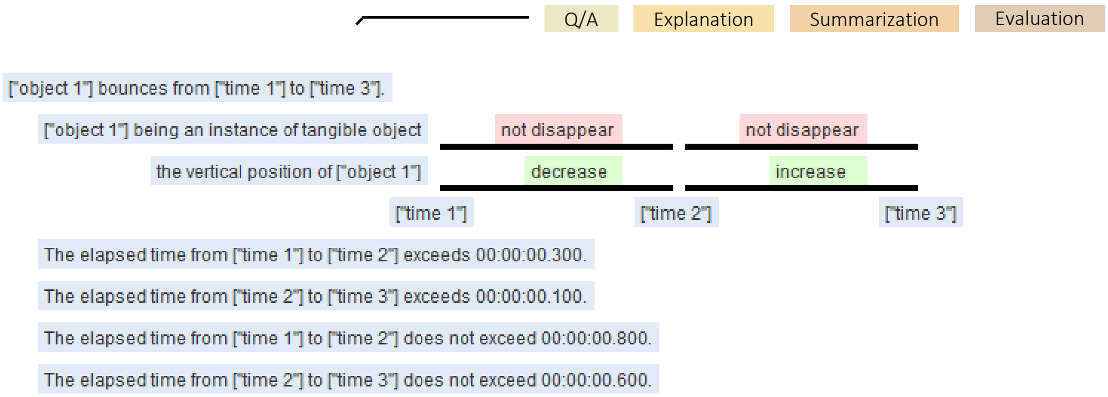

2014-07-29T18:08:01.600


2014-07-29T18:08:06.800

B. Event Models

Following are event models for 102 event types addressed in the research described here. The models are organized into clusters based on the number of participating humans and objects. Portrayal of the models uses a format in which the underlying Möbius encoding of information is rendered in a form of readable English, with specifications of Möbius variables appearing in brackets. Also included with each model is an indication of whether or not that model has been used in support of the question answering, explanation, and summarization capabilities, and in the evaluation of event recognition.

0 Humans, 1 Object



["object 1"] bounces from ["time 1"] to ["time 3"].

["object 1"] being an instance of tangible object

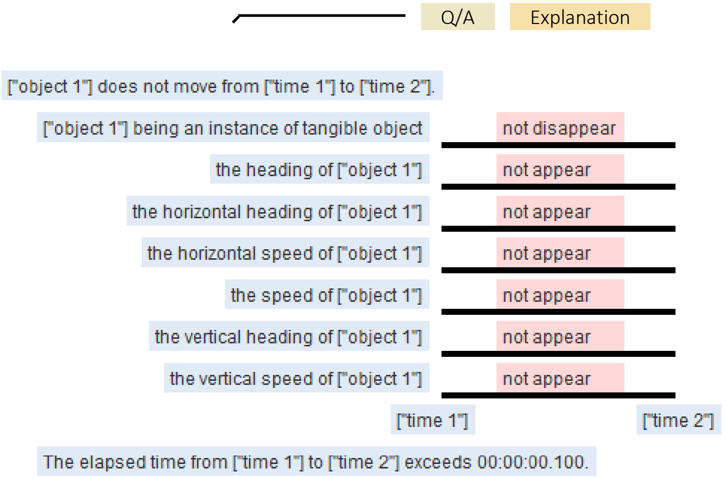
the vertical position of ["object 1"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.300.

The elapsed time from ["time 2"] to ["time 3"] exceeds 00:00:00.100.

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:00:00.800.

The elapsed time from ["time 2"] to ["time 3"] does not exceed 00:00:00.600.



["object 1"] does not move from ["time 1"] to ["time 2"].

["object 1"] being an instance of tangible object

the heading of ["object 1"]

the horizontal heading of ["object 1"]

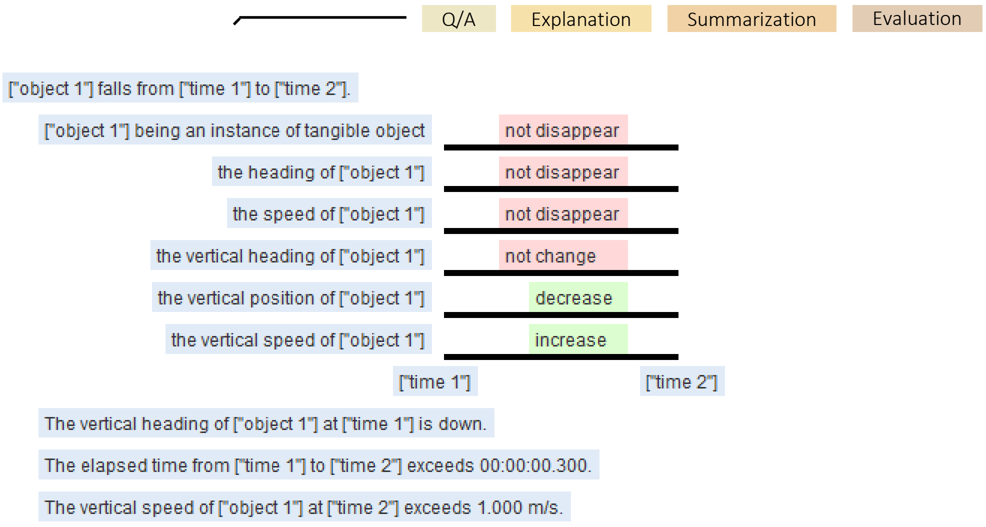
the horizontal speed of ["object 1"]

the speed of ["object 1"]

the vertical heading of ["object 1"]

the vertical speed of ["object 1"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



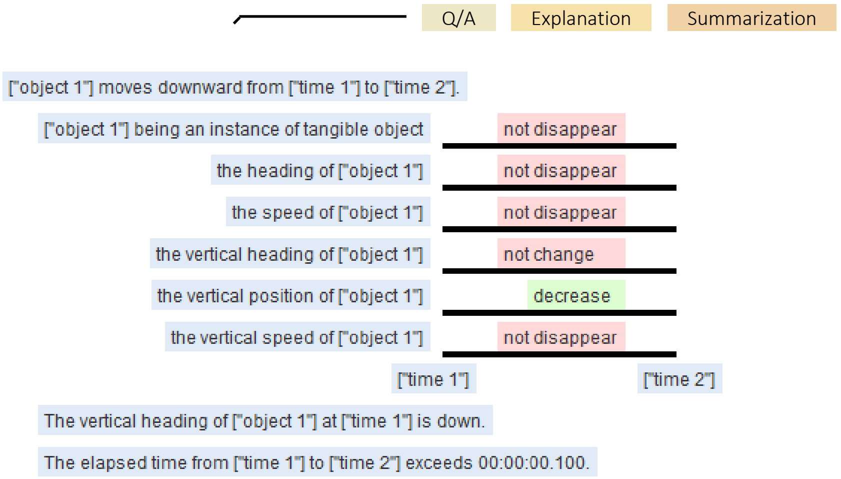
["object 1"] falls from ["time 1"] to ["time 2"].

["object 1"] being an instance of tangible object — not disappear
the heading of ["object 1"] — not disappear
the speed of ["object 1"] — not disappear
the vertical heading of ["object 1"] — not change
the vertical position of ["object 1"] — decrease
the vertical speed of ["object 1"] — increase
["time 1"] ["time 2"]

The vertical heading of ["object 1"] at ["time 1"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.300.

The vertical speed of ["object 1"] at ["time 2"] exceeds 1.000 m/s.

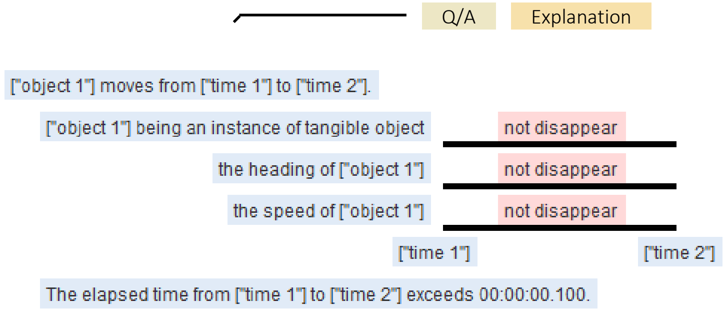


["object 1"] moves downward from ["time 1"] to ["time 2"].

["object 1"] being an instance of tangible object — not disappear
the heading of ["object 1"] — not disappear
the speed of ["object 1"] — not disappear
the vertical heading of ["object 1"] — not change
the vertical position of ["object 1"] — decrease
the vertical speed of ["object 1"] — not disappear
["time 1"] ["time 2"]

The vertical heading of ["object 1"] at ["time 1"] is down.

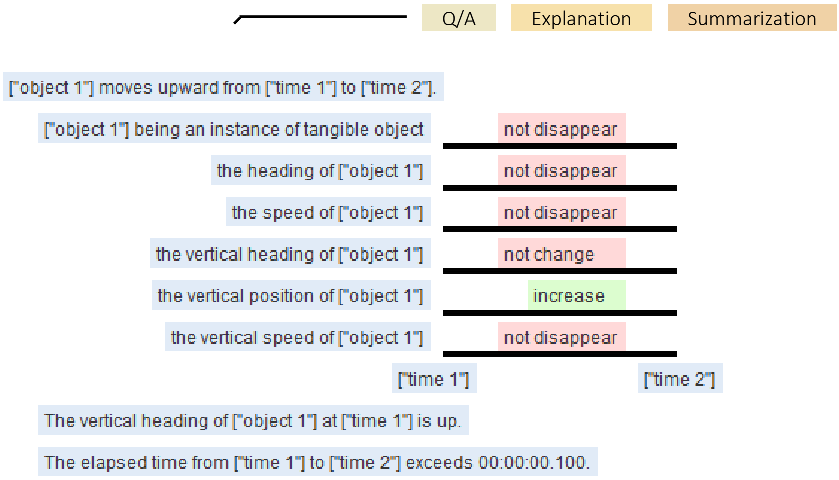
The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



["object 1"] moves from ["time 1"] to ["time 2"].

["object 1"] being an instance of tangible object — not disappear
the heading of ["object 1"] — not disappear
the speed of ["object 1"] — not disappear
["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.


Q/A Explanation Summarization

["object 1"] moves upward from ["time 1"] to ["time 2"].

["object 1"] being an instance of tangible object — not disappear

the heading of ["object 1"] — not disappear

the speed of ["object 1"] — not disappear

the vertical heading of ["object 1"] — not change

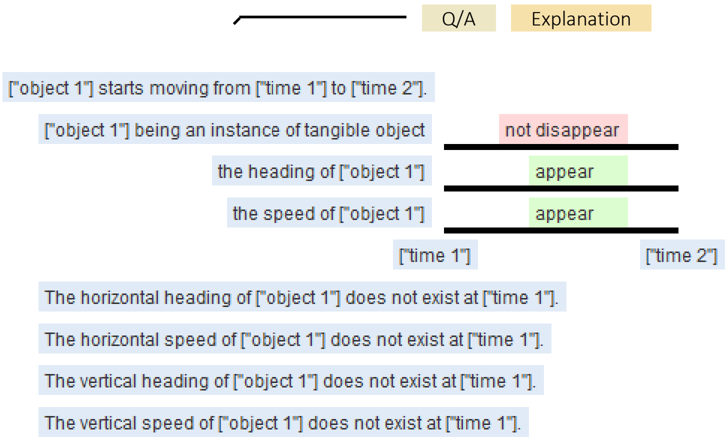
the vertical position of ["object 1"] — increase

the vertical speed of ["object 1"] — not disappear

["time 1"] ["time 2"]

The vertical heading of ["object 1"] at ["time 1"] is up.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.


Q/A Explanation

["object 1"] starts moving from ["time 1"] to ["time 2"].

["object 1"] being an instance of tangible object — not disappear

the heading of ["object 1"] — appear

the speed of ["object 1"] — appear

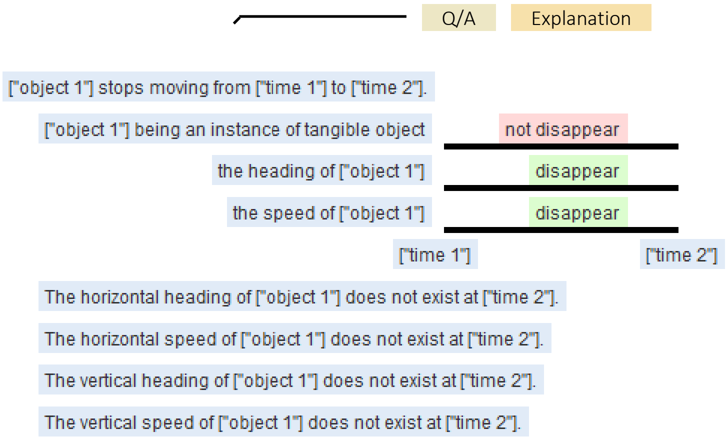
["time 1"] ["time 2"]

The horizontal heading of ["object 1"] does not exist at ["time 1"].

The horizontal speed of ["object 1"] does not exist at ["time 1"].

The vertical heading of ["object 1"] does not exist at ["time 1"].

The vertical speed of ["object 1"] does not exist at ["time 1"].


Q/A Explanation

["object 1"] stops moving from ["time 1"] to ["time 2"].

["object 1"] being an instance of tangible object — not disappear

the heading of ["object 1"] — disappear

the speed of ["object 1"] — disappear

["time 1"] ["time 2"]

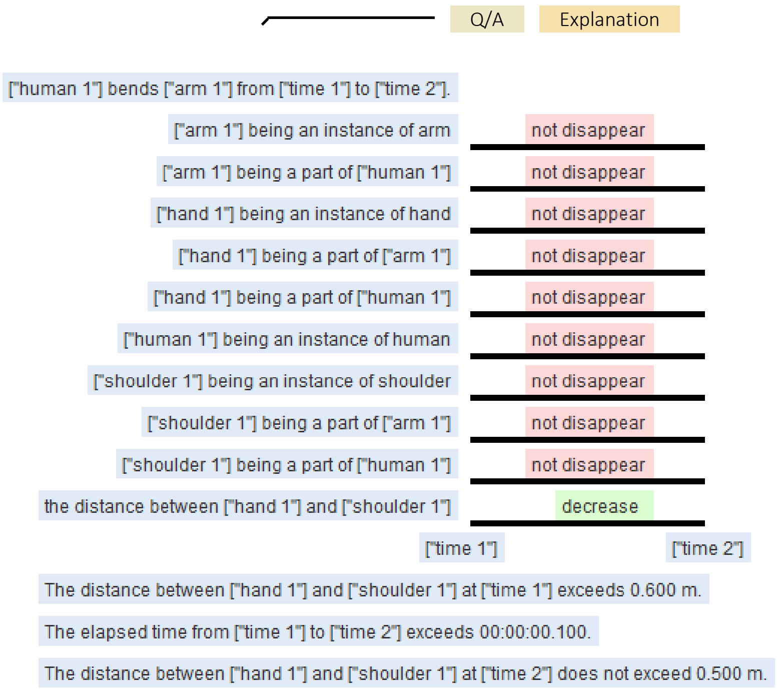
The horizontal heading of ["object 1"] does not exist at ["time 2"].

The horizontal speed of ["object 1"] does not exist at ["time 2"].

The vertical heading of ["object 1"] does not exist at ["time 2"].

The vertical speed of ["object 1"] does not exist at ["time 2"].

1 Human, 0 Objects



Q/A Explanation

["human 1"] bends ["arm 1"] from ["time 1"] to ["time 2"].

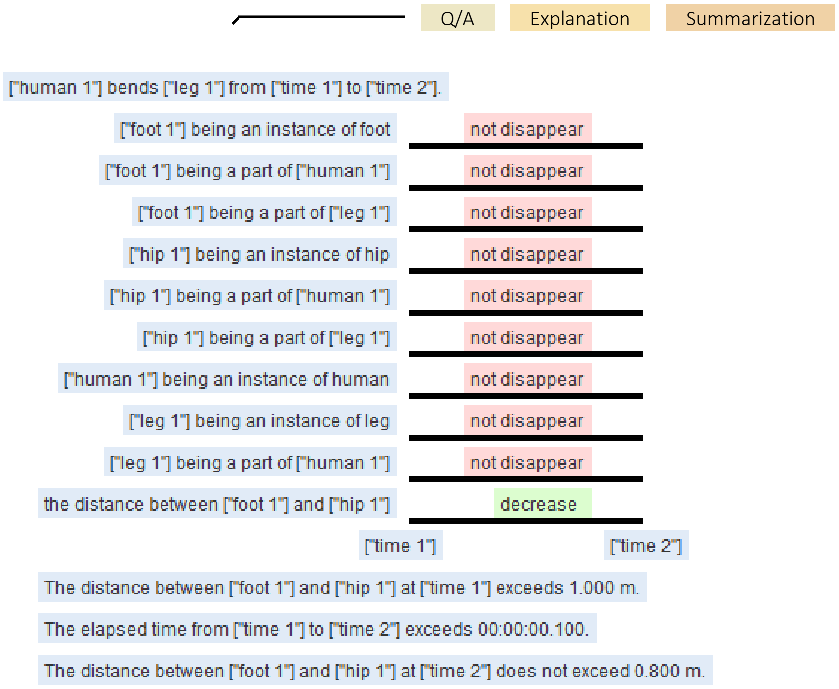
["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the distance between ["hand 1"] and ["shoulder 1"]	decrease

["time 1"] ["time 2"]

The distance between ["hand 1"] and ["shoulder 1"] at ["time 1"] exceeds 0.600 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

The distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] does not exceed 0.500 m.



Q/A Explanation Summarization

["human 1"] bends ["leg 1"] from ["time 1"] to ["time 2"].

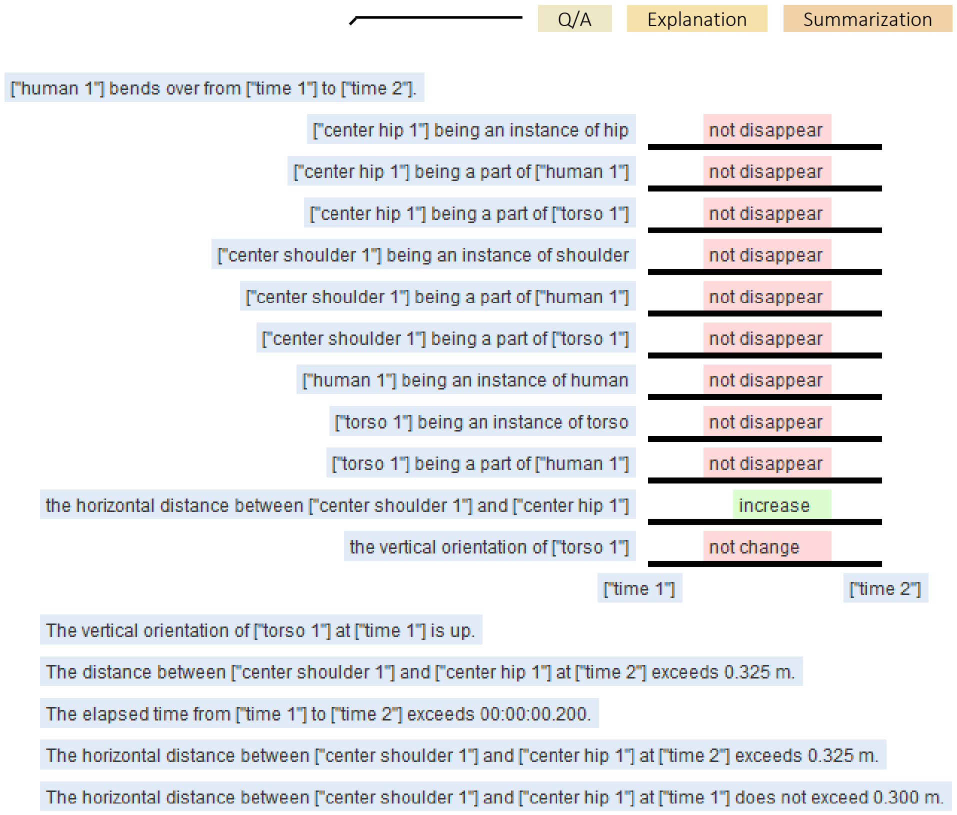
["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["foot 1"] being a part of ["leg 1"]	not disappear
["hip 1"] being an instance of hip	not disappear
["hip 1"] being a part of ["human 1"]	not disappear
["hip 1"] being a part of ["leg 1"]	not disappear
["human 1"] being an instance of human	not disappear
["leg 1"] being an instance of leg	not disappear
["leg 1"] being a part of ["human 1"]	not disappear
the distance between ["foot 1"] and ["hip 1"]	decrease

["time 1"] ["time 2"]

The distance between ["foot 1"] and ["hip 1"] at ["time 1"] exceeds 1.000 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

The distance between ["foot 1"] and ["hip 1"] at ["time 2"] does not exceed 0.800 m.



Q/A Explanation Summarization

["human 1"] bends over from ["time 1"] to ["time 2"].

["center hip 1"] being an instance of hip	not disappear
["center hip 1"] being a part of ["human 1"]	not disappear
["center hip 1"] being a part of ["torso 1"]	not disappear
["center shoulder 1"] being an instance of shoulder	not disappear
["center shoulder 1"] being a part of ["human 1"]	not disappear
["center shoulder 1"] being a part of ["torso 1"]	not disappear
["human 1"] being an instance of human	not disappear
["torso 1"] being an instance of torso	not disappear
["torso 1"] being a part of ["human 1"]	not disappear
the horizontal distance between ["center shoulder 1"] and ["center hip 1"]	increase
the vertical orientation of ["torso 1"]	not change
	["time 1"] ["time 2"]

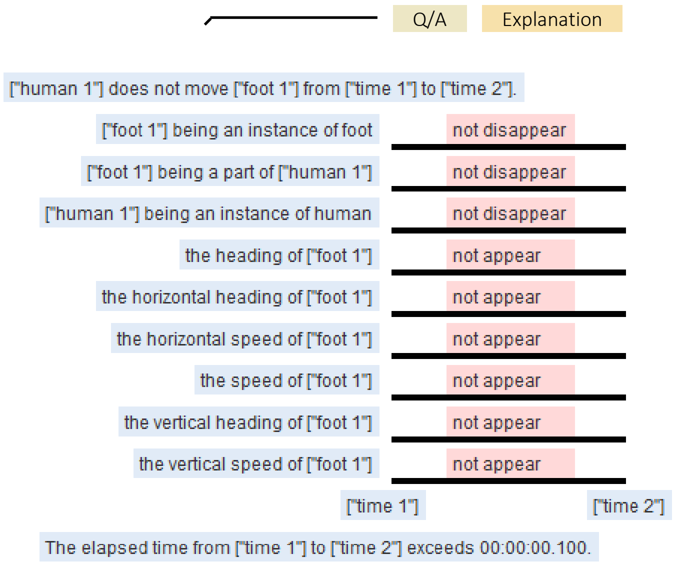
The vertical orientation of ["torso 1"] at ["time 1"] is up.

The distance between ["center shoulder 1"] and ["center hip 1"] at ["time 2"] exceeds 0.325 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.200.

The horizontal distance between ["center shoulder 1"] and ["center hip 1"] at ["time 2"] exceeds 0.325 m.

The horizontal distance between ["center shoulder 1"] and ["center hip 1"] at ["time 1"] does not exceed 0.300 m.

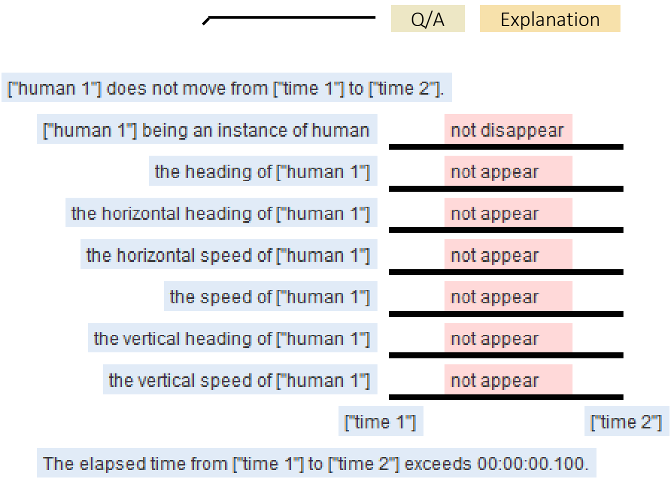


Q/A Explanation

["human 1"] does not move ["foot 1"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["foot 1"]	not appear
the horizontal heading of ["foot 1"]	not appear
the horizontal speed of ["foot 1"]	not appear
the speed of ["foot 1"]	not appear
the vertical heading of ["foot 1"]	not appear
the vertical speed of ["foot 1"]	not appear
	["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A

Explanation

["human 1"] does not move from ["time 1"] to ["time 2"].

["human 1"] being an instance of human — not disappear

the heading of ["human 1"] — not appear

the horizontal heading of ["human 1"] — not appear

the horizontal speed of ["human 1"] — not appear

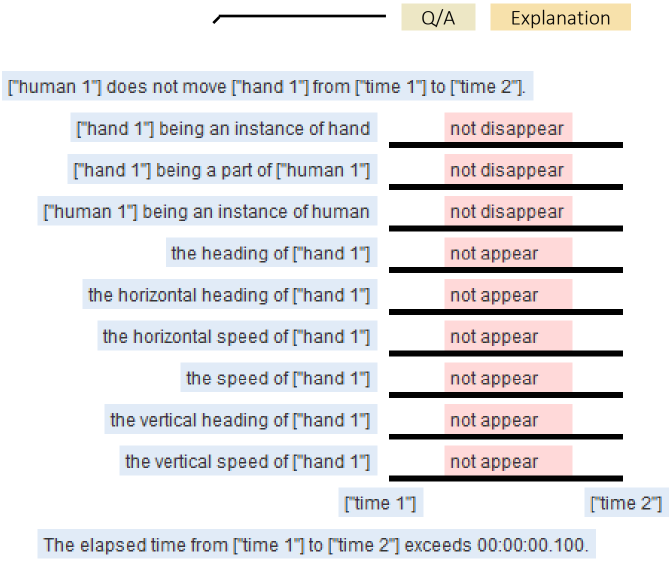
the speed of ["human 1"] — not appear

the vertical heading of ["human 1"] — not appear

the vertical speed of ["human 1"] — not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A

Explanation

["human 1"] does not move ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand — not disappear

["hand 1"] being a part of ["human 1"] — not disappear

["human 1"] being an instance of human — not disappear

the heading of ["hand 1"] — not appear

the horizontal heading of ["hand 1"] — not appear

the horizontal speed of ["hand 1"] — not appear

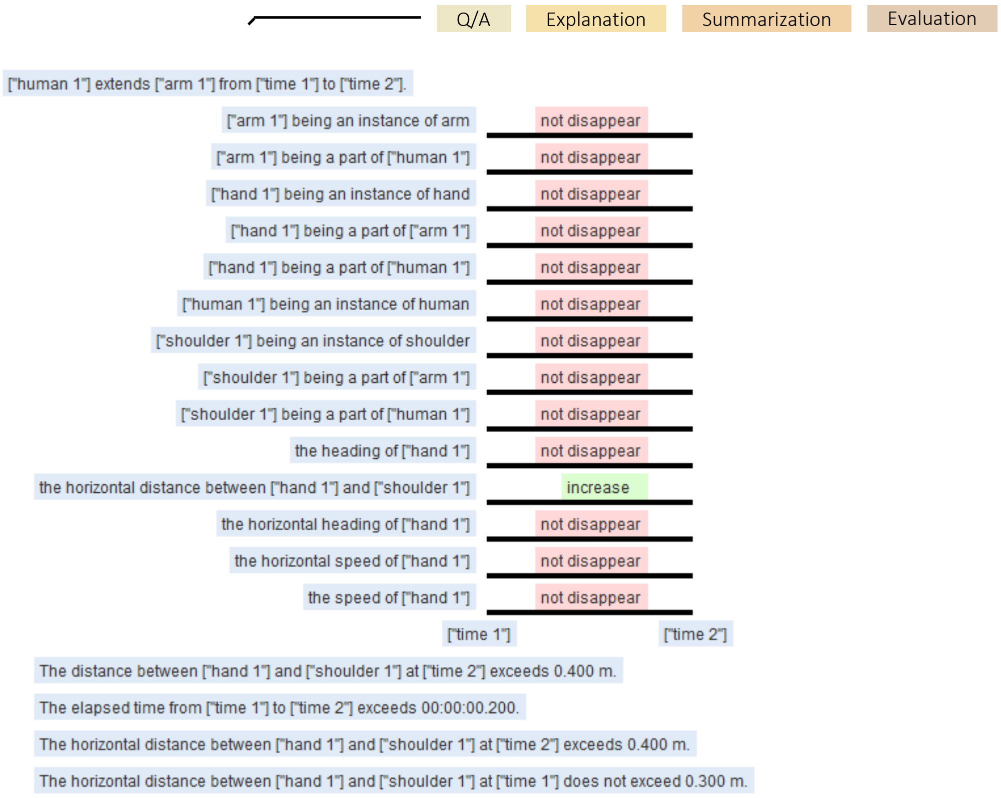
the speed of ["hand 1"] — not appear

the vertical heading of ["hand 1"] — not appear

the vertical speed of ["hand 1"] — not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



["human 1"] extends ["arm 1"] from ["time 1"] to ["time 2"].

["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the heading of ["hand 1"]	not disappear
the horizontal distance between ["hand 1"] and ["shoulder 1"]	increase
the horizontal heading of ["hand 1"]	not disappear
the horizontal speed of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear

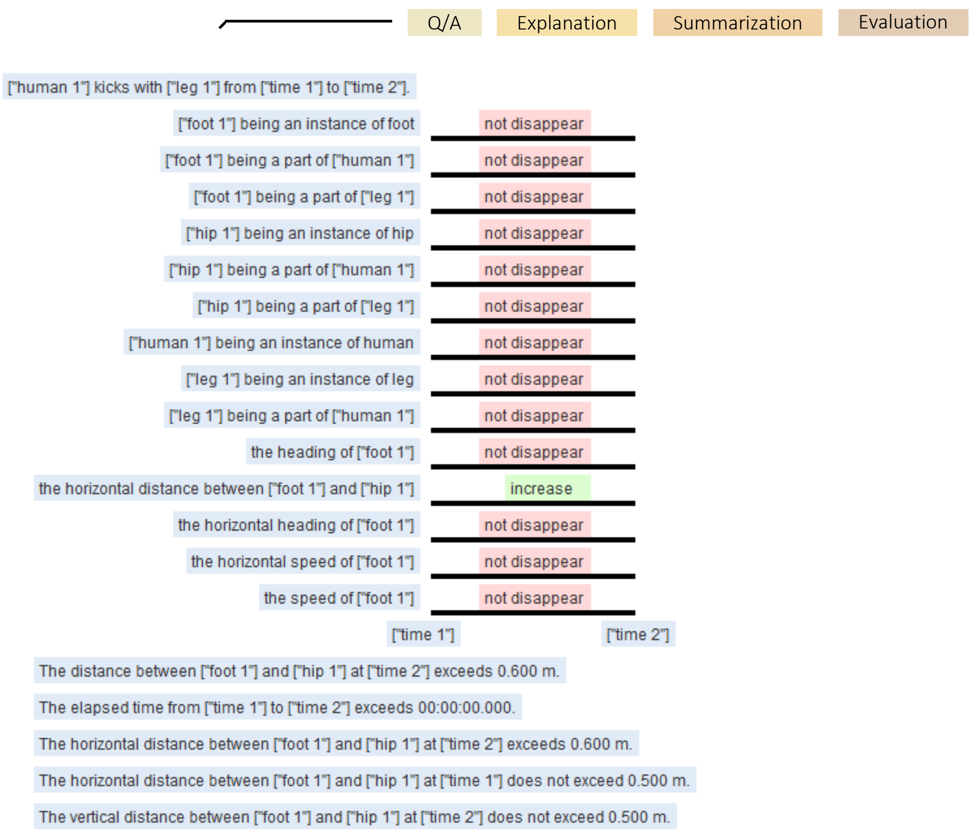
["time 1"] ["time 2"]

The distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.400 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.200.

The horizontal distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.400 m.

The horizontal distance between ["hand 1"] and ["shoulder 1"] at ["time 1"] does not exceed 0.300 m.



["human 1"] kicks with ["leg 1"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["foot 1"] being a part of ["leg 1"]	not disappear
["hip 1"] being an instance of hip	not disappear
["hip 1"] being a part of ["human 1"]	not disappear
["hip 1"] being a part of ["leg 1"]	not disappear
["human 1"] being an instance of human	not disappear
["leg 1"] being an instance of leg	not disappear
["leg 1"] being a part of ["human 1"]	not disappear
the heading of ["foot 1"]	not disappear
the horizontal distance between ["foot 1"] and ["hip 1"]	increase
the horizontal heading of ["foot 1"]	not disappear
the horizontal speed of ["foot 1"]	not disappear
the speed of ["foot 1"]	not disappear

["time 1"] ["time 2"]

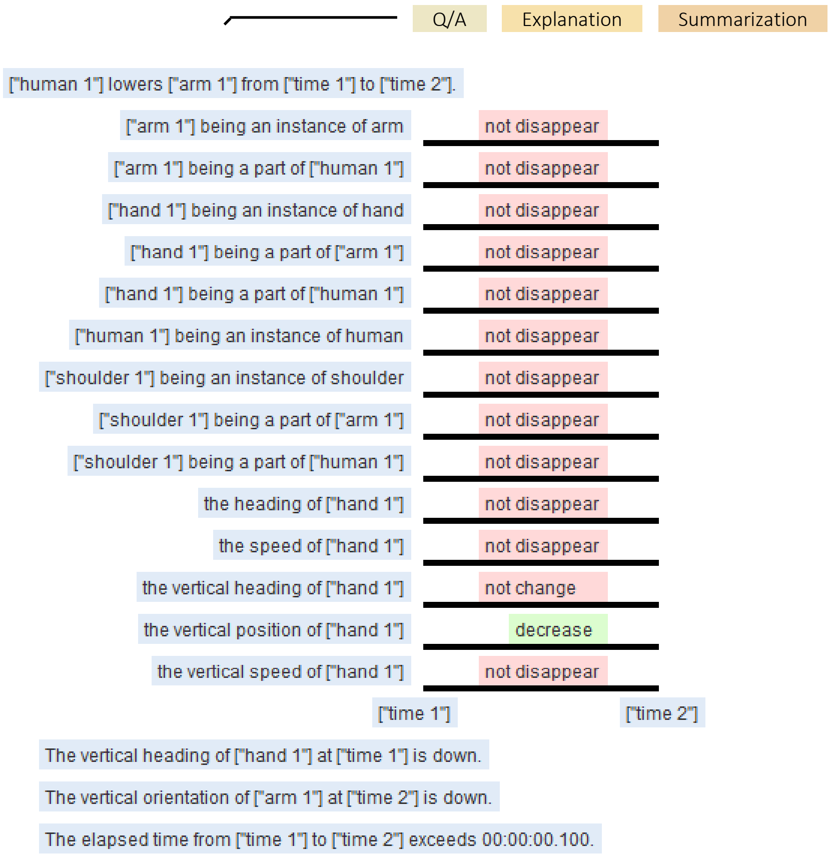
The distance between ["foot 1"] and ["hip 1"] at ["time 2"] exceeds 0.600 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.000.

The horizontal distance between ["foot 1"] and ["hip 1"] at ["time 2"] exceeds 0.600 m.

The horizontal distance between ["foot 1"] and ["hip 1"] at ["time 1"] does not exceed 0.500 m.

The vertical distance between ["foot 1"] and ["hip 1"] at ["time 2"] does not exceed 0.500 m.



["human 1"] lowers ["arm 1"] from ["time 1"] to ["time 2"].

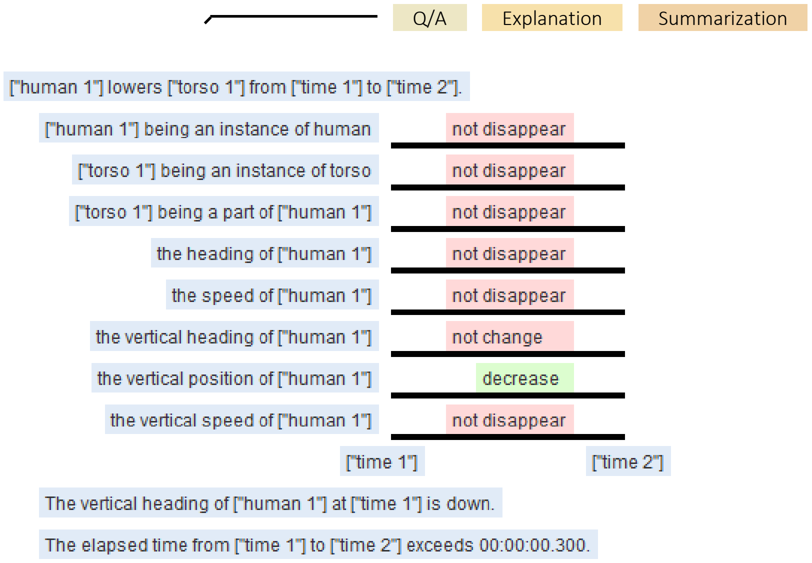
["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the heading of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear
the vertical heading of ["hand 1"]	not change
the vertical position of ["hand 1"]	decrease
the vertical speed of ["hand 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["hand 1"] at ["time 1"] is down.

The vertical orientation of ["arm 1"] at ["time 2"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



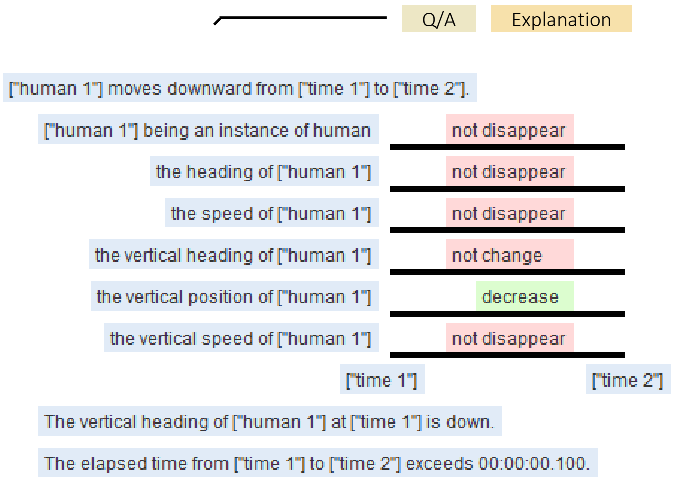
["human 1"] lowers ["torso 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["torso 1"] being an instance of torso	not disappear
["torso 1"] being a part of ["human 1"]	not disappear
the heading of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear
the vertical heading of ["human 1"]	not change
the vertical position of ["human 1"]	decrease
the vertical speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["human 1"] at ["time 1"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.300.



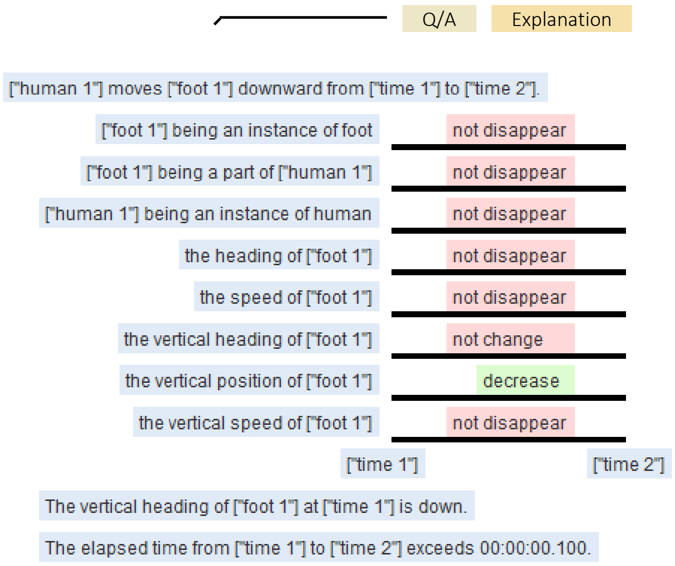
["human 1"] moves downward from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
the heading of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear
the vertical heading of ["human 1"]	not change
the vertical position of ["human 1"]	decrease
the vertical speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["human 1"] at ["time 1"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



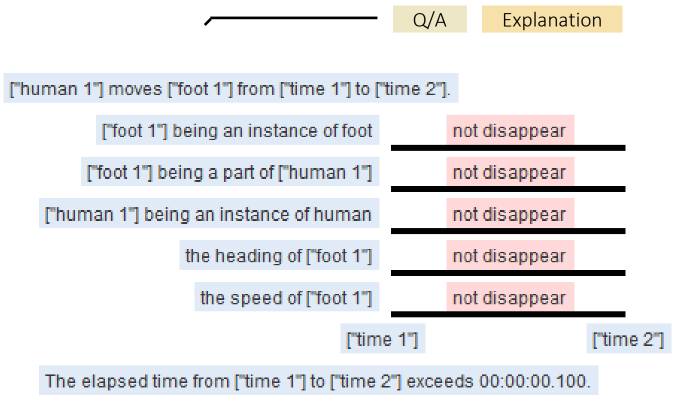
["human 1"] moves ["foot 1"] downward from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["foot 1"]	not disappear
the speed of ["foot 1"]	not disappear
the vertical heading of ["foot 1"]	not change
the vertical position of ["foot 1"]	decrease
the vertical speed of ["foot 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["foot 1"] at ["time 1"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

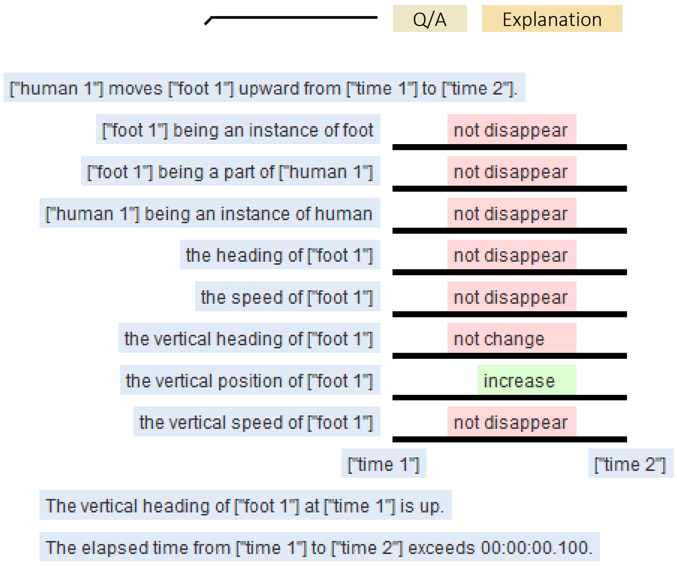


["human 1"] moves ["foot 1"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["foot 1"]	not disappear
the speed of ["foot 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A Explanation

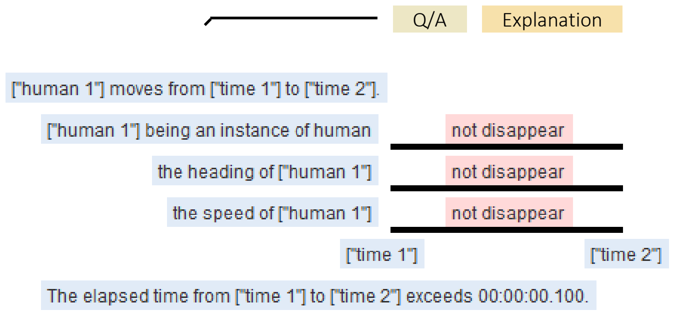
["human 1"] moves ["foot 1"] upward from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["foot 1"]	not disappear
the speed of ["foot 1"]	not disappear
the vertical heading of ["foot 1"]	not change
the vertical position of ["foot 1"]	increase
the vertical speed of ["foot 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["foot 1"] at ["time 1"] is up.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



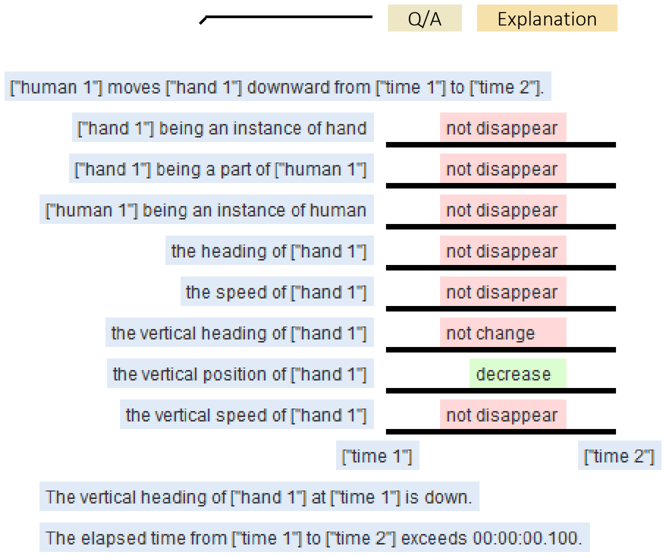
Q/A Explanation

["human 1"] moves from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
the heading of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A Explanation

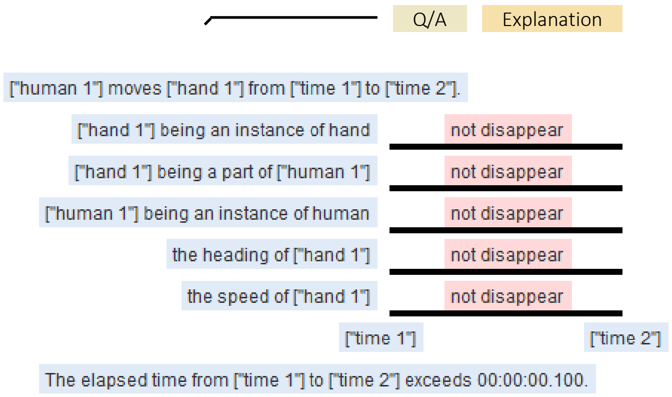
["human 1"] moves ["hand 1"] downward from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear
the vertical heading of ["hand 1"]	not change
the vertical position of ["hand 1"]	decrease
the vertical speed of ["hand 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["hand 1"] at ["time 1"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

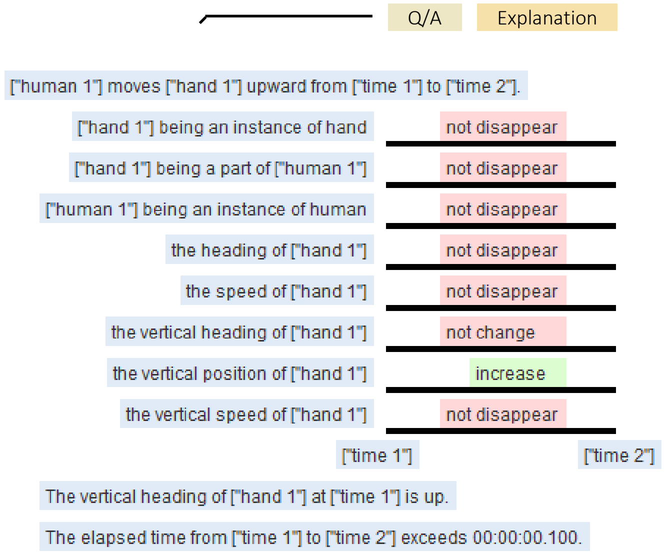


["human 1"] moves ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



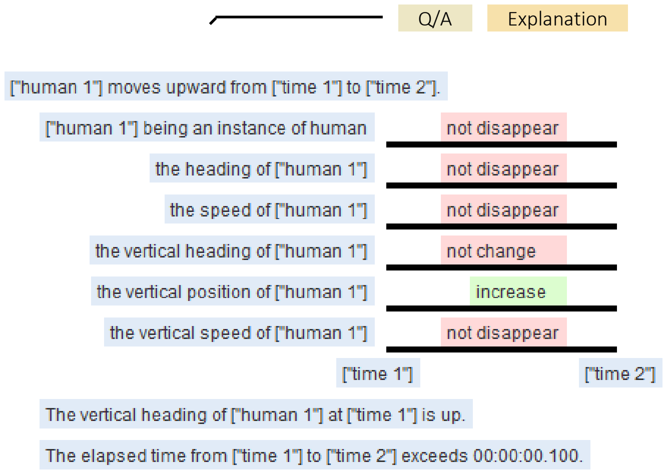
["human 1"] moves ["hand 1"] upward from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear
the vertical heading of ["hand 1"]	not change
the vertical position of ["hand 1"]	increase
the vertical speed of ["hand 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["hand 1"] at ["time 1"] is up.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



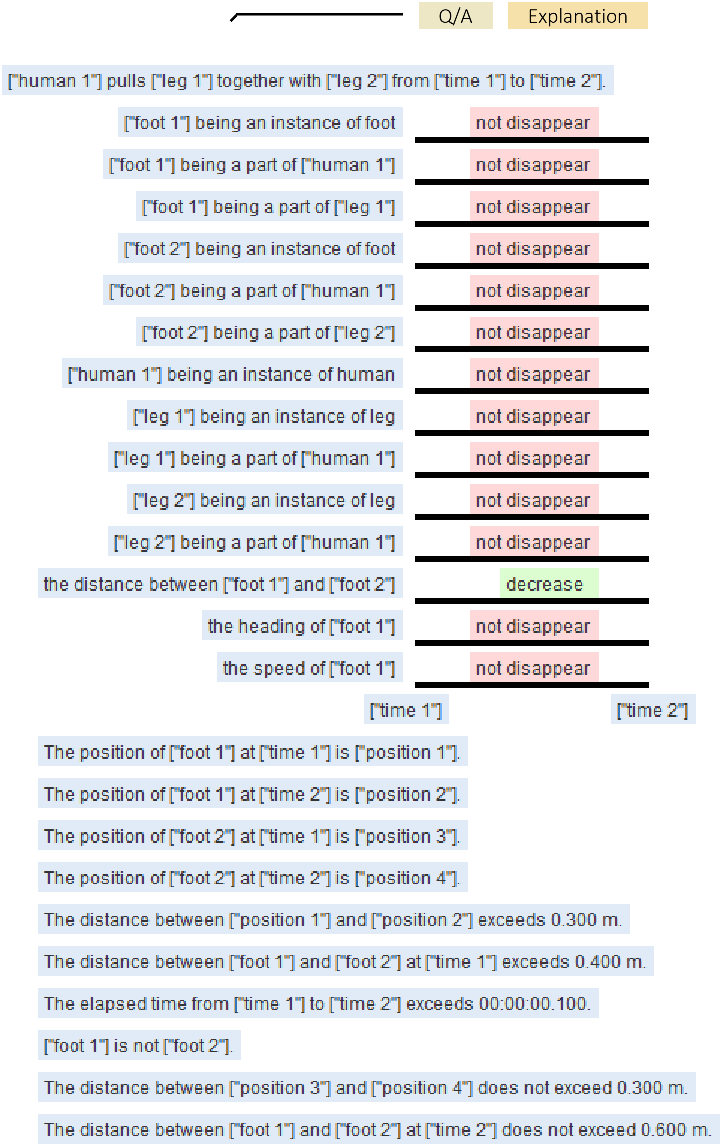
["human 1"] moves upward from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
the heading of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear
the vertical heading of ["human 1"]	not change
the vertical position of ["human 1"]	increase
the vertical speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["human 1"] at ["time 1"] is up.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A

Explanation

["human 1"] pulls ["leg 1"] together with ["leg 2"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["foot 1"] being a part of ["leg 1"]	not disappear
["foot 2"] being an instance of foot	not disappear
["foot 2"] being a part of ["human 1"]	not disappear
["foot 2"] being a part of ["leg 2"]	not disappear
["human 1"] being an instance of human	not disappear
["leg 1"] being an instance of leg	not disappear
["leg 1"] being a part of ["human 1"]	not disappear
["leg 2"] being an instance of leg	not disappear
["leg 2"] being a part of ["human 1"]	not disappear
the distance between ["foot 1"] and ["foot 2"]	decrease
the heading of ["foot 1"]	not disappear
the speed of ["foot 1"]	not disappear

["time 1"] ["time 2"]

The position of ["foot 1"] at ["time 1"] is ["position 1"].

The position of ["foot 1"] at ["time 2"] is ["position 2"].

The position of ["foot 2"] at ["time 1"] is ["position 3"].

The position of ["foot 2"] at ["time 2"] is ["position 4"].

The distance between ["position 1"] and ["position 2"] exceeds 0.300 m.

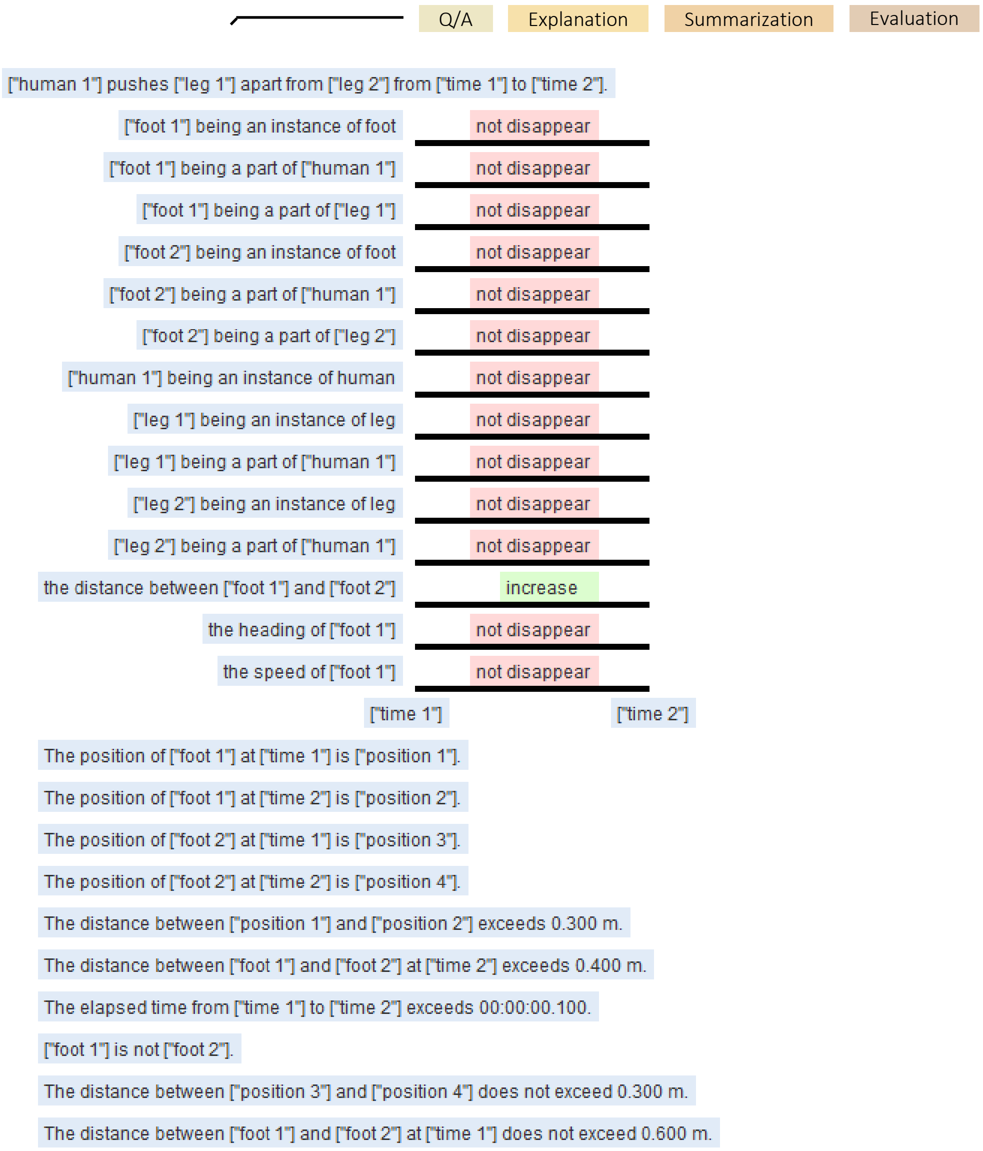
The distance between ["foot 1"] and ["foot 2"] at ["time 1"] exceeds 0.400 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

["foot 1"] is not ["foot 2"].

The distance between ["position 3"] and ["position 4"] does not exceed 0.300 m.

The distance between ["foot 1"] and ["foot 2"] at ["time 2"] does not exceed 0.600 m.



["human 1"] pushes ["leg 1"] apart from ["leg 2"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["foot 1"] being a part of ["leg 1"]	not disappear
["foot 2"] being an instance of foot	not disappear
["foot 2"] being a part of ["human 1"]	not disappear
["foot 2"] being a part of ["leg 2"]	not disappear
["human 1"] being an instance of human	not disappear
["leg 1"] being an instance of leg	not disappear
["leg 1"] being a part of ["human 1"]	not disappear
["leg 2"] being an instance of leg	not disappear
["leg 2"] being a part of ["human 1"]	not disappear
the distance between ["foot 1"] and ["foot 2"]	increase
the heading of ["foot 1"]	not disappear
the speed of ["foot 1"]	not disappear

["time 1"] ["time 2"]

The position of ["foot 1"] at ["time 1"] is ["position 1"].

The position of ["foot 1"] at ["time 2"] is ["position 2"].

The position of ["foot 2"] at ["time 1"] is ["position 3"].

The position of ["foot 2"] at ["time 2"] is ["position 4"].

The distance between ["position 1"] and ["position 2"] exceeds 0.300 m.

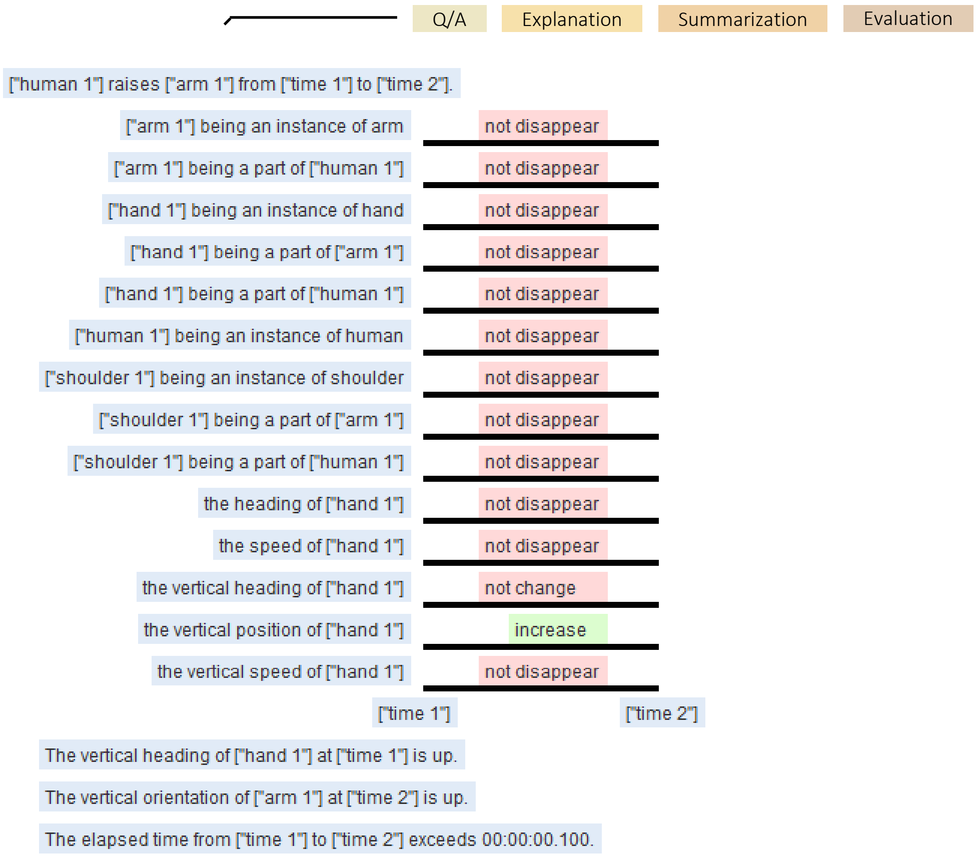
The distance between ["foot 1"] and ["foot 2"] at ["time 2"] exceeds 0.400 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

["foot 1"] is not ["foot 2"].

The distance between ["position 3"] and ["position 4"] does not exceed 0.300 m.

The distance between ["foot 1"] and ["foot 2"] at ["time 1"] does not exceed 0.600 m.



["human 1"] raises ["arm 1"] from ["time 1"] to ["time 2"].

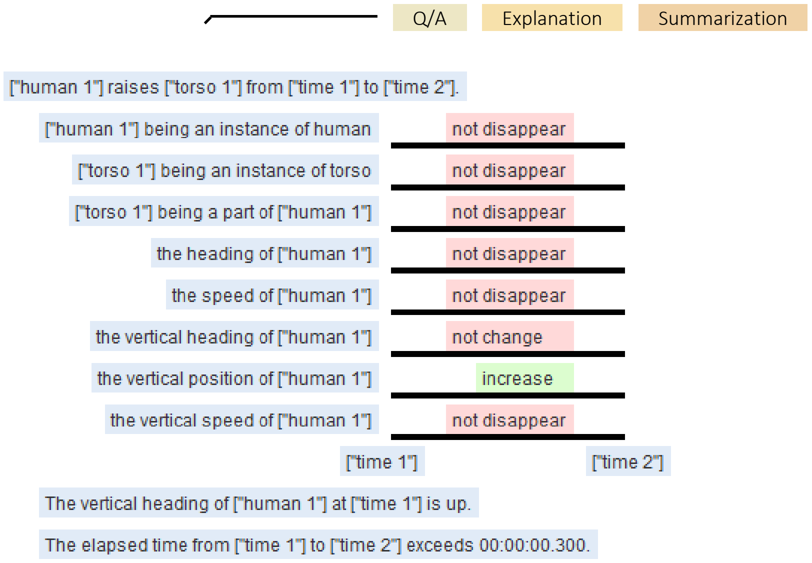
["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the heading of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear
the vertical heading of ["hand 1"]	not change
the vertical position of ["hand 1"]	increase
the vertical speed of ["hand 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["hand 1"] at ["time 1"] is up.

The vertical orientation of ["arm 1"] at ["time 2"] is up.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



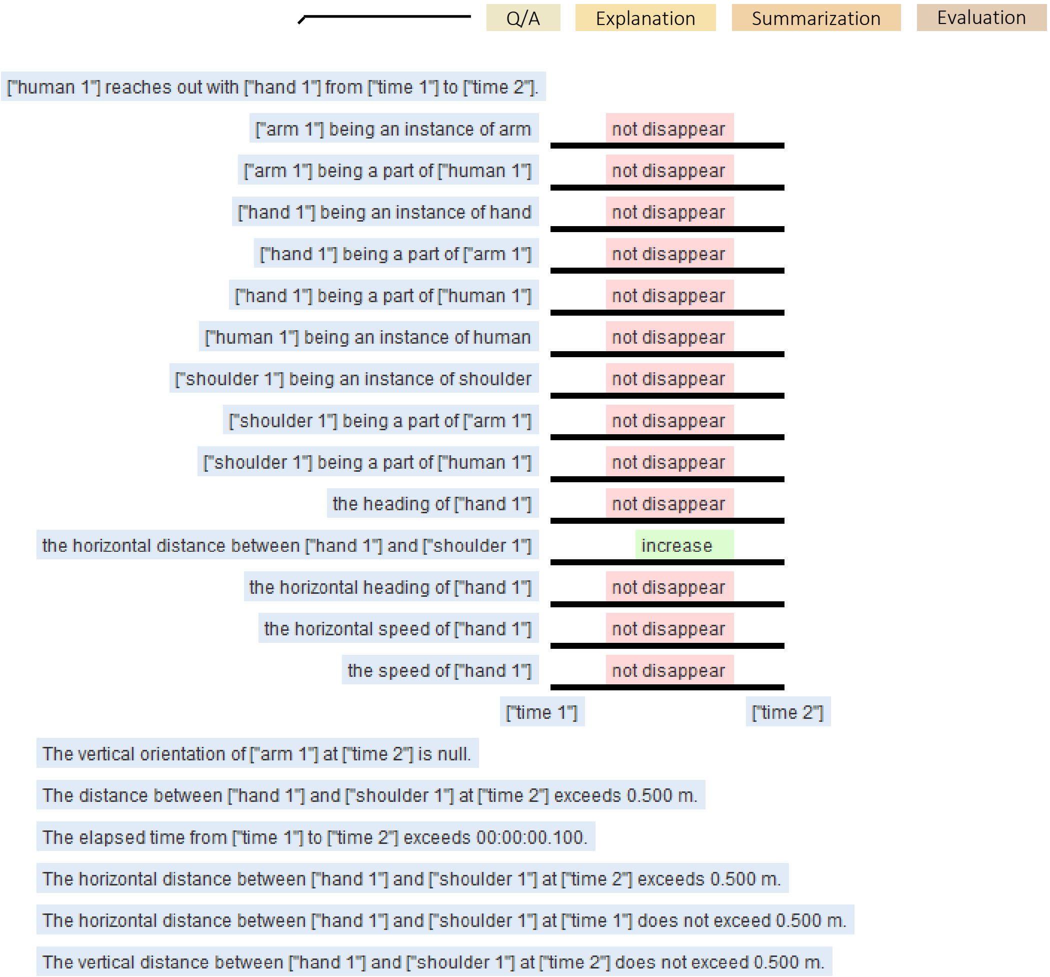
["human 1"] raises ["torso 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["torso 1"] being an instance of torso	not disappear
["torso 1"] being a part of ["human 1"]	not disappear
the heading of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear
the vertical heading of ["human 1"]	not change
the vertical position of ["human 1"]	increase
the vertical speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["human 1"] at ["time 1"] is up.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.300.



["human 1"] reaches out with ["hand 1"] from ["time 1"] to ["time 2"].

["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the heading of ["hand 1"]	not disappear
the horizontal distance between ["hand 1"] and ["shoulder 1"]	increase
the horizontal heading of ["hand 1"]	not disappear
the horizontal speed of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear
	["time 1"] ["time 2"]

The vertical orientation of ["arm 1"] at ["time 2"] is null.

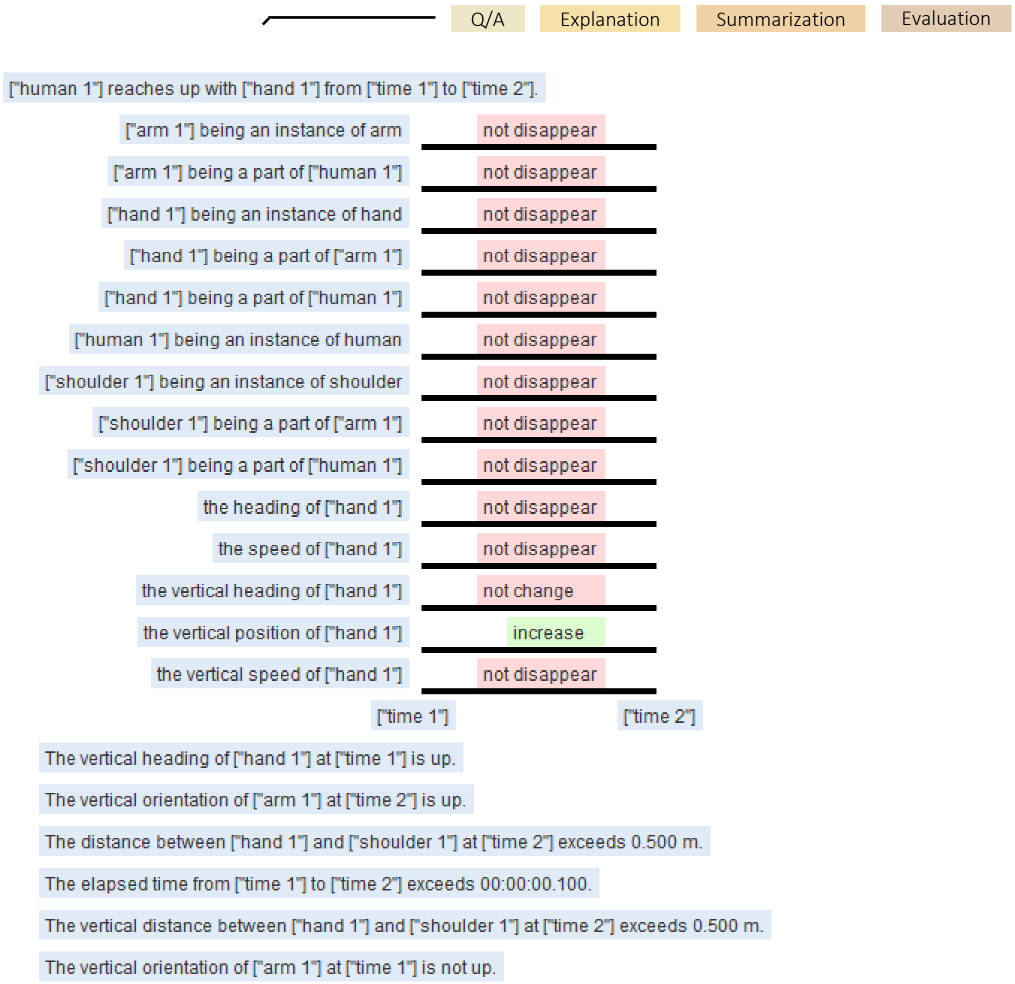
The distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.500 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

The horizontal distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.500 m.

The horizontal distance between ["hand 1"] and ["shoulder 1"] at ["time 1"] does not exceed 0.500 m.

The vertical distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] does not exceed 0.500 m.



["human 1"] reaches up with ["hand 1"] from ["time 1"] to ["time 2"].

["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the heading of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear
the vertical heading of ["hand 1"]	not change
the vertical position of ["hand 1"]	increase
the vertical speed of ["hand 1"]	not disappear
	["time 1"] ["time 2"]

The vertical heading of ["hand 1"] at ["time 1"] is up.

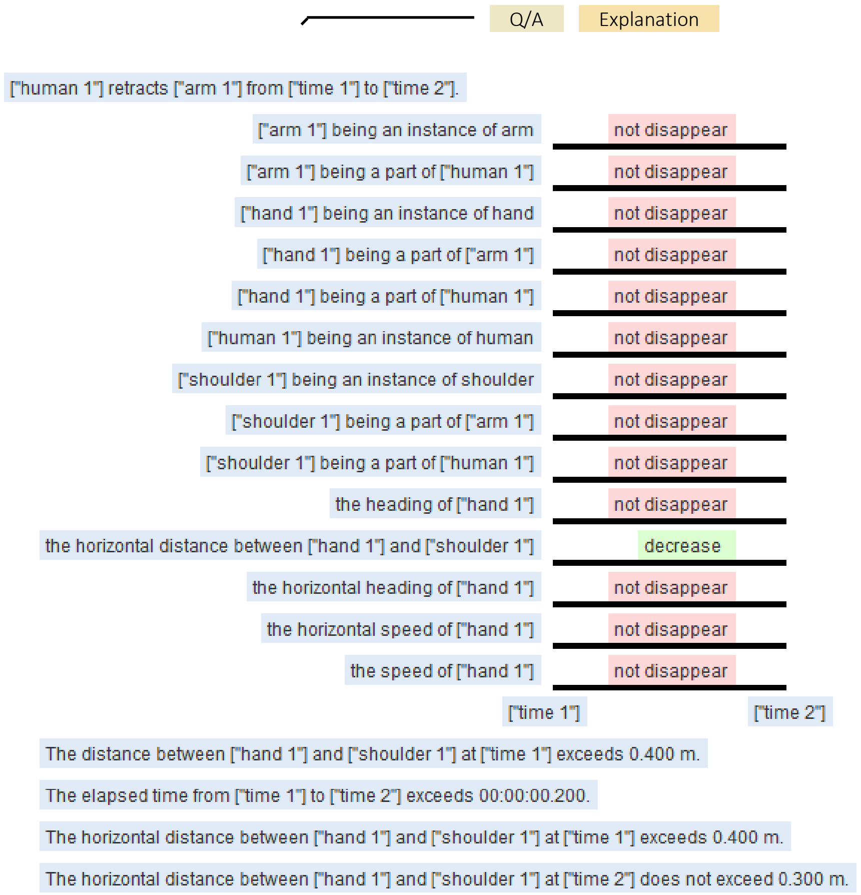
The vertical orientation of ["arm 1"] at ["time 2"] is up.

The distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.500 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

The vertical distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.500 m.

The vertical orientation of ["arm 1"] at ["time 1"] is not up.



["human 1"] retracts ["arm 1"] from ["time 1"] to ["time 2"].

["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the heading of ["hand 1"]	not disappear
the horizontal distance between ["hand 1"] and ["shoulder 1"]	decrease
the horizontal heading of ["hand 1"]	not disappear
the horizontal speed of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear

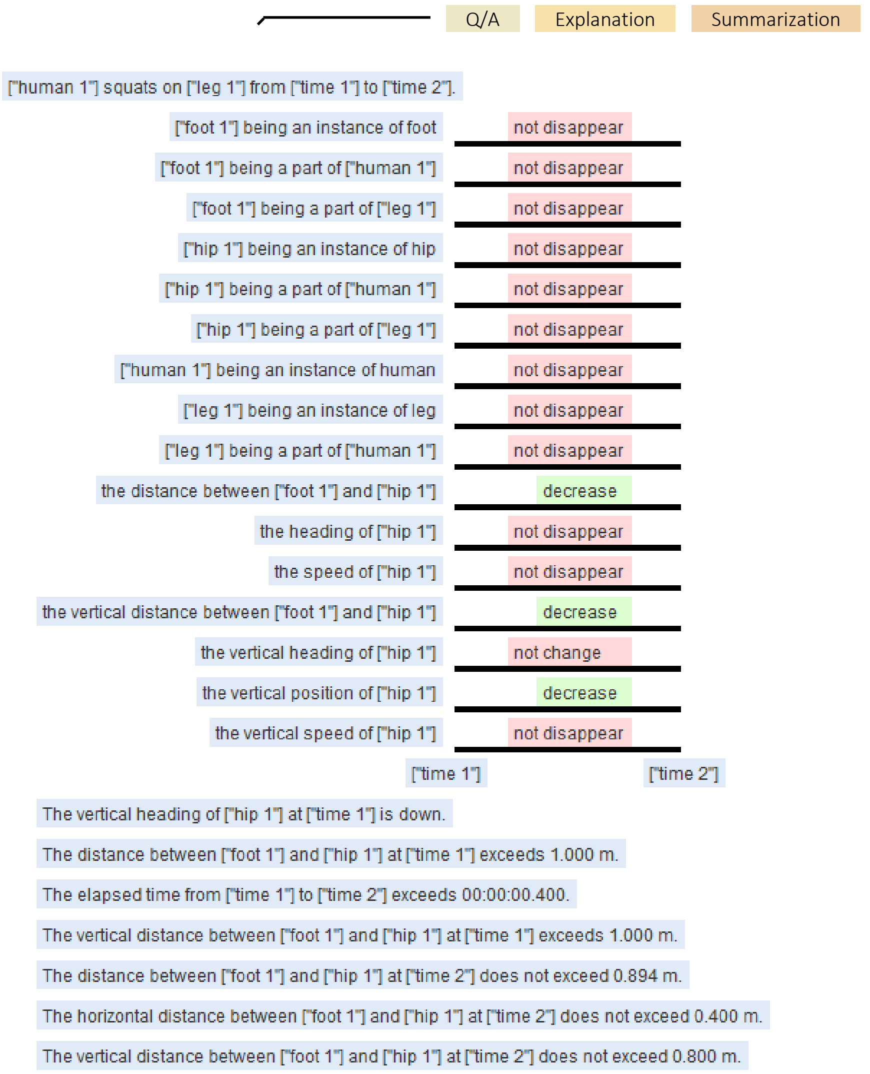
["time 1"] ["time 2"]

The distance between ["hand 1"] and ["shoulder 1"] at ["time 1"] exceeds 0.400 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.200.

The horizontal distance between ["hand 1"] and ["shoulder 1"] at ["time 1"] exceeds 0.400 m.

The horizontal distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] does not exceed 0.300 m.



Q/A Explanation Summarization

["human 1"] squats on ["leg 1"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["foot 1"] being a part of ["leg 1"]	not disappear
["hip 1"] being an instance of hip	not disappear
["hip 1"] being a part of ["human 1"]	not disappear
["hip 1"] being a part of ["leg 1"]	not disappear
["human 1"] being an instance of human	not disappear
["leg 1"] being an instance of leg	not disappear
["leg 1"] being a part of ["human 1"]	not disappear
the distance between ["foot 1"] and ["hip 1"]	decrease
the heading of ["hip 1"]	not disappear
the speed of ["hip 1"]	not disappear
the vertical distance between ["foot 1"] and ["hip 1"]	decrease
the vertical heading of ["hip 1"]	not change
the vertical position of ["hip 1"]	decrease
the vertical speed of ["hip 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["hip 1"] at ["time 1"] is down.

The distance between ["foot 1"] and ["hip 1"] at ["time 1"] exceeds 1.000 m.

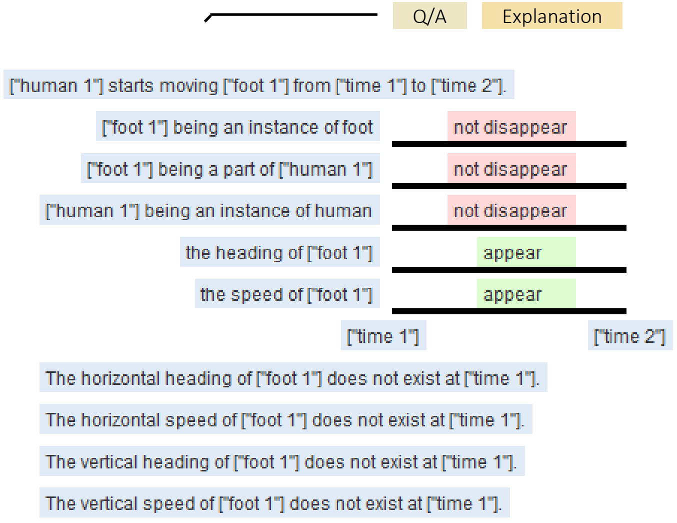
The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The vertical distance between ["foot 1"] and ["hip 1"] at ["time 1"] exceeds 1.000 m.

The distance between ["foot 1"] and ["hip 1"] at ["time 2"] does not exceed 0.894 m.

The horizontal distance between ["foot 1"] and ["hip 1"] at ["time 2"] does not exceed 0.400 m.

The vertical distance between ["foot 1"] and ["hip 1"] at ["time 2"] does not exceed 0.800 m.



Q/A Explanation

["human 1"] starts moving ["foot 1"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["foot 1"]	appear
the speed of ["foot 1"]	appear

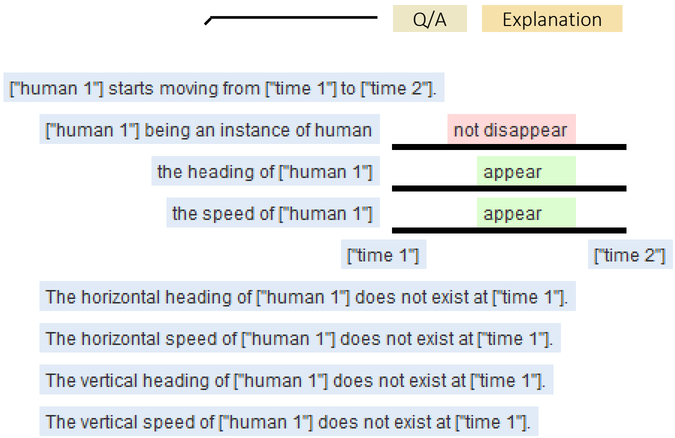
["time 1"] ["time 2"]

The horizontal heading of ["foot 1"] does not exist at ["time 1"].

The horizontal speed of ["foot 1"] does not exist at ["time 1"].

The vertical heading of ["foot 1"] does not exist at ["time 1"].

The vertical speed of ["foot 1"] does not exist at ["time 1"].



["human 1"] starts moving from ["time 1"] to ["time 2"].

["human 1"] being an instance of human — not disappear

the heading of ["human 1"] — appear

the speed of ["human 1"] — appear

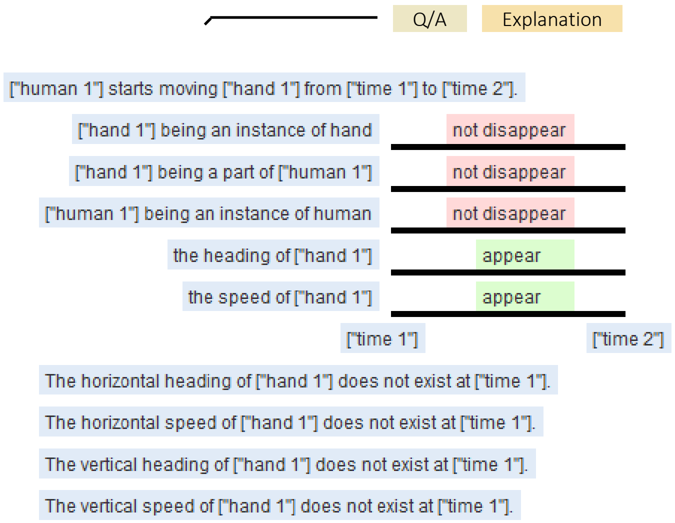
["time 1"] ["time 2"]

The horizontal heading of ["human 1"] does not exist at ["time 1"].

The horizontal speed of ["human 1"] does not exist at ["time 1"].

The vertical heading of ["human 1"] does not exist at ["time 1"].

The vertical speed of ["human 1"] does not exist at ["time 1"].



["human 1"] starts moving ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand — not disappear

["hand 1"] being a part of ["human 1"] — not disappear

["human 1"] being an instance of human — not disappear

the heading of ["hand 1"] — appear

the speed of ["hand 1"] — appear

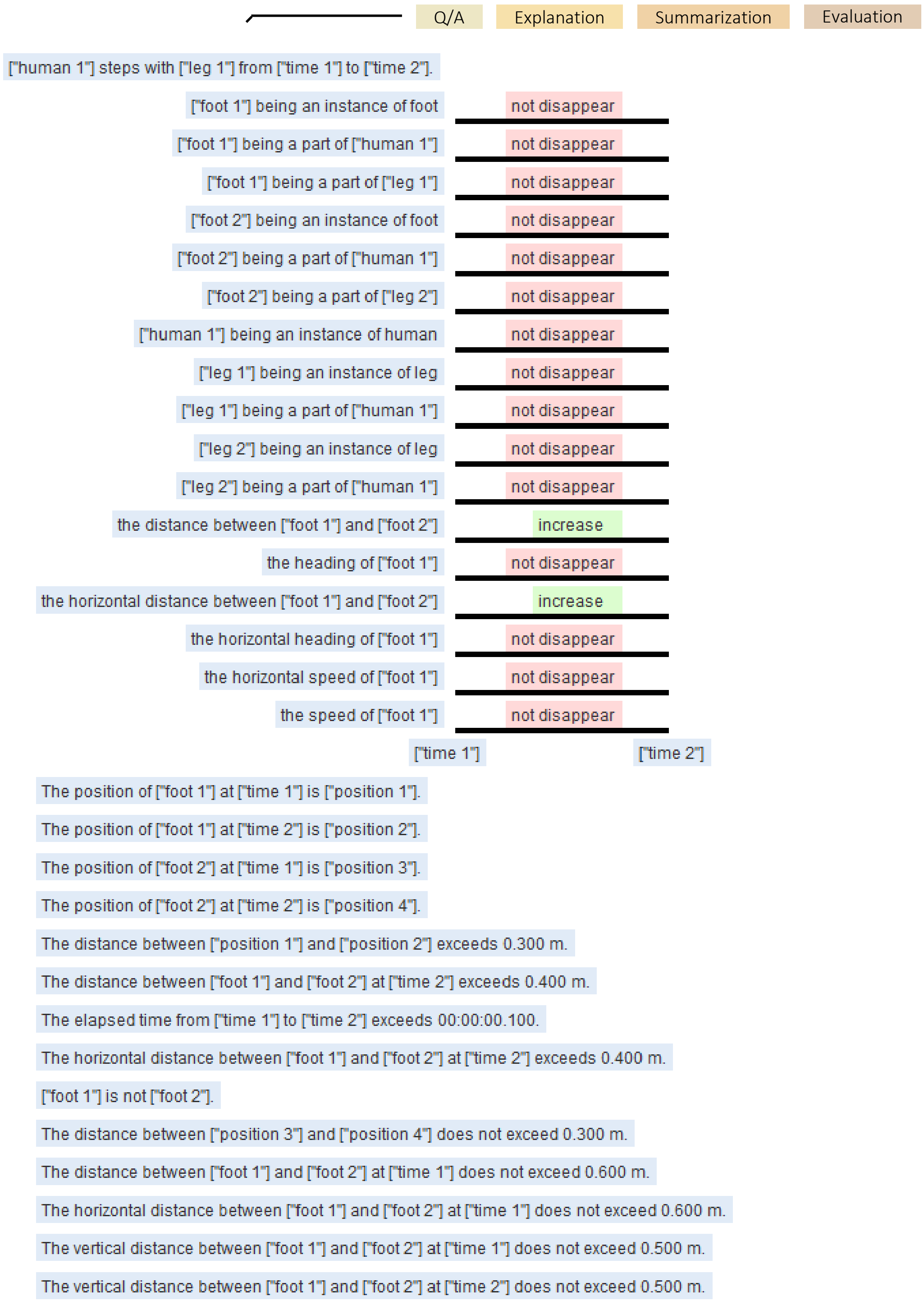
["time 1"] ["time 2"]

The horizontal heading of ["hand 1"] does not exist at ["time 1"].

The horizontal speed of ["hand 1"] does not exist at ["time 1"].

The vertical heading of ["hand 1"] does not exist at ["time 1"].

The vertical speed of ["hand 1"] does not exist at ["time 1"].



["human 1"] steps with ["leg 1"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["foot 1"] being a part of ["leg 1"]	not disappear
["foot 2"] being an instance of foot	not disappear
["foot 2"] being a part of ["human 1"]	not disappear
["foot 2"] being a part of ["leg 2"]	not disappear
["human 1"] being an instance of human	not disappear
["leg 1"] being an instance of leg	not disappear
["leg 1"] being a part of ["human 1"]	not disappear
["leg 2"] being an instance of leg	not disappear
["leg 2"] being a part of ["human 1"]	not disappear
the distance between ["foot 1"] and ["foot 2"]	increase
the heading of ["foot 1"]	not disappear
the horizontal distance between ["foot 1"] and ["foot 2"]	increase
the horizontal heading of ["foot 1"]	not disappear
the horizontal speed of ["foot 1"]	not disappear
the speed of ["foot 1"]	not disappear

["time 1"] ["time 2"]

The position of ["foot 1"] at ["time 1"] is ["position 1"].

The position of ["foot 1"] at ["time 2"] is ["position 2"].

The position of ["foot 2"] at ["time 1"] is ["position 3"].

The position of ["foot 2"] at ["time 2"] is ["position 4"].

The distance between ["position 1"] and ["position 2"] exceeds 0.300 m.

The distance between ["foot 1"] and ["foot 2"] at ["time 2"] exceeds 0.400 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

The horizontal distance between ["foot 1"] and ["foot 2"] at ["time 2"] exceeds 0.400 m.

["foot 1"] is not ["foot 2"].

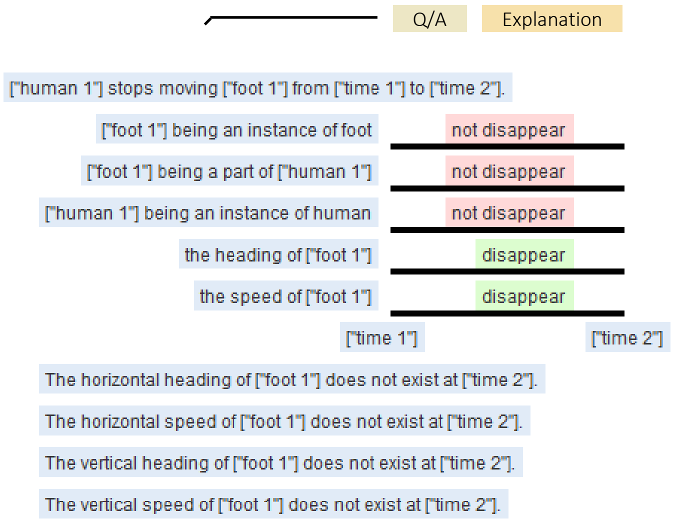
The distance between ["position 3"] and ["position 4"] does not exceed 0.300 m.

The distance between ["foot 1"] and ["foot 2"] at ["time 1"] does not exceed 0.600 m.

The horizontal distance between ["foot 1"] and ["foot 2"] at ["time 1"] does not exceed 0.600 m.

The vertical distance between ["foot 1"] and ["foot 2"] at ["time 1"] does not exceed 0.500 m.

The vertical distance between ["foot 1"] and ["foot 2"] at ["time 2"] does not exceed 0.500 m.



["human 1"] stops moving ["foot 1"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot	not disappear
["foot 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["foot 1"]	disappear
the speed of ["foot 1"]	disappear

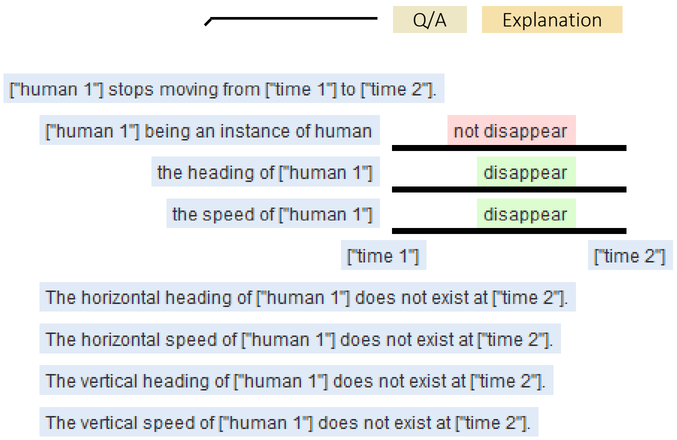
["time 1"] ["time 2"]

The horizontal heading of ["foot 1"] does not exist at ["time 2"].

The horizontal speed of ["foot 1"] does not exist at ["time 2"].

The vertical heading of ["foot 1"] does not exist at ["time 2"].

The vertical speed of ["foot 1"] does not exist at ["time 2"].



["human 1"] stops moving from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
the heading of ["human 1"]	disappear
the speed of ["human 1"]	disappear

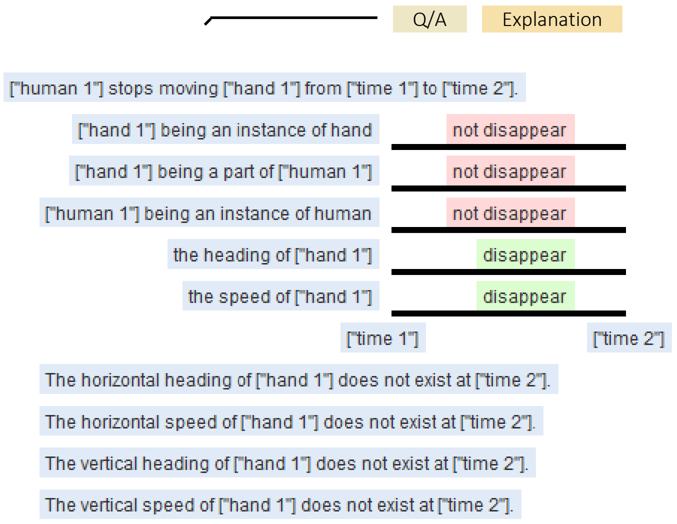
["time 1"] ["time 2"]

The horizontal heading of ["human 1"] does not exist at ["time 2"].

The horizontal speed of ["human 1"] does not exist at ["time 2"].

The vertical heading of ["human 1"] does not exist at ["time 2"].

The vertical speed of ["human 1"] does not exist at ["time 2"].



["human 1"] stops moving ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
the heading of ["hand 1"]	disappear
the speed of ["hand 1"]	disappear

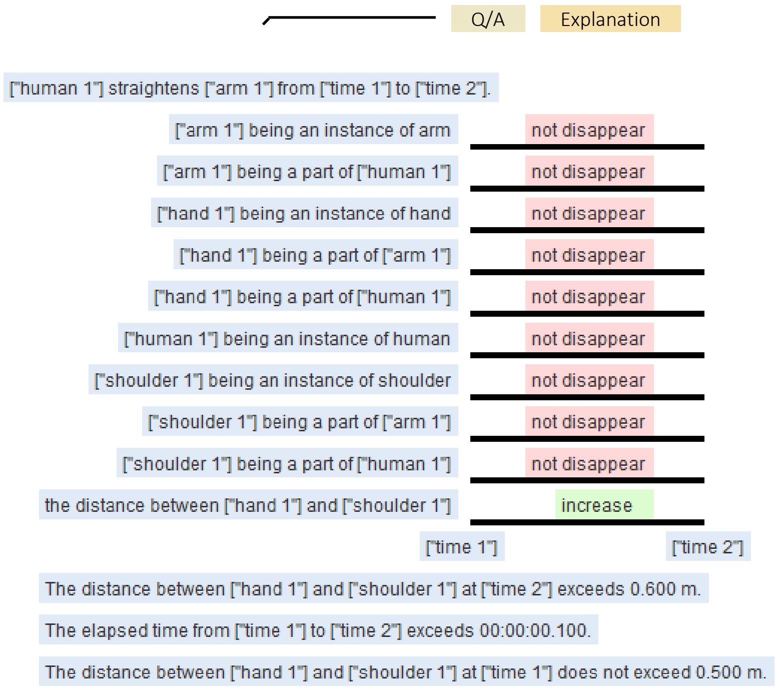
["time 1"] ["time 2"]

The horizontal heading of ["hand 1"] does not exist at ["time 2"].

The horizontal speed of ["hand 1"] does not exist at ["time 2"].

The vertical heading of ["hand 1"] does not exist at ["time 2"].

The vertical speed of ["hand 1"] does not exist at ["time 2"].



Q/A Explanation

["human 1"] straightens ["arm 1"] from ["time 1"] to ["time 2"].

["arm 1"] being an instance of arm — not disappear

["arm 1"] being a part of ["human 1"] — not disappear

["hand 1"] being an instance of hand — not disappear

["hand 1"] being a part of ["arm 1"] — not disappear

["hand 1"] being a part of ["human 1"] — not disappear

["human 1"] being an instance of human — not disappear

["shoulder 1"] being an instance of shoulder — not disappear

["shoulder 1"] being a part of ["arm 1"] — not disappear

["shoulder 1"] being a part of ["human 1"] — not disappear

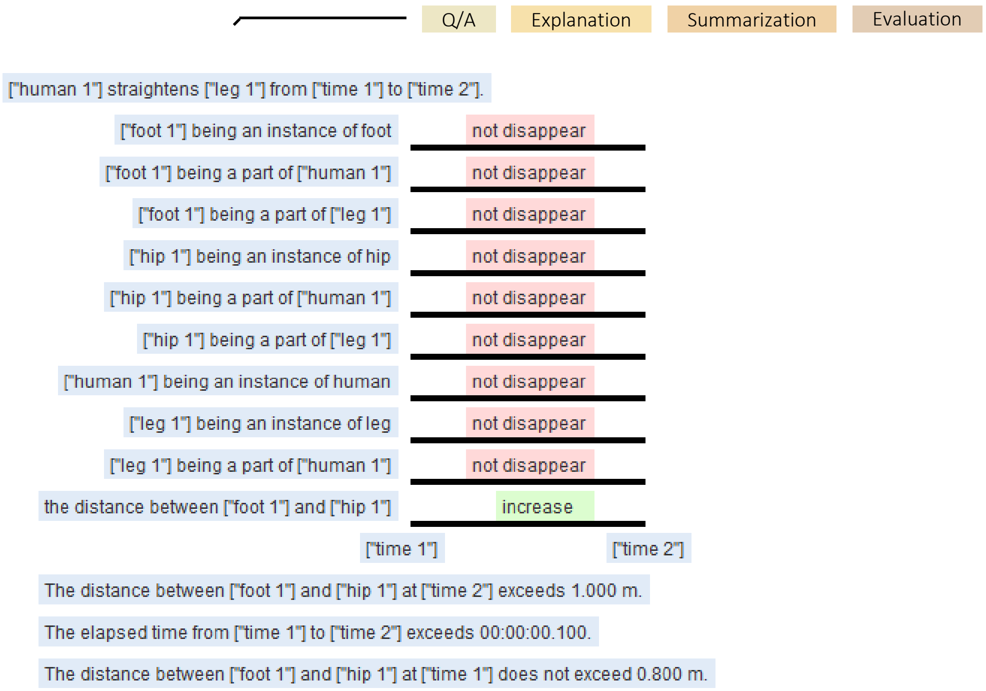
the distance between ["hand 1"] and ["shoulder 1"] — increase

["time 1"] ["time 2"]

The distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.600 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

The distance between ["hand 1"] and ["shoulder 1"] at ["time 1"] does not exceed 0.500 m.



Q/A Explanation Summarization Evaluation

["human 1"] straightens ["leg 1"] from ["time 1"] to ["time 2"].

["foot 1"] being an instance of foot — not disappear

["foot 1"] being a part of ["human 1"] — not disappear

["foot 1"] being a part of ["leg 1"] — not disappear

["hip 1"] being an instance of hip — not disappear

["hip 1"] being a part of ["human 1"] — not disappear

["hip 1"] being a part of ["leg 1"] — not disappear

["human 1"] being an instance of human — not disappear

["leg 1"] being an instance of leg — not disappear

["leg 1"] being a part of ["human 1"] — not disappear

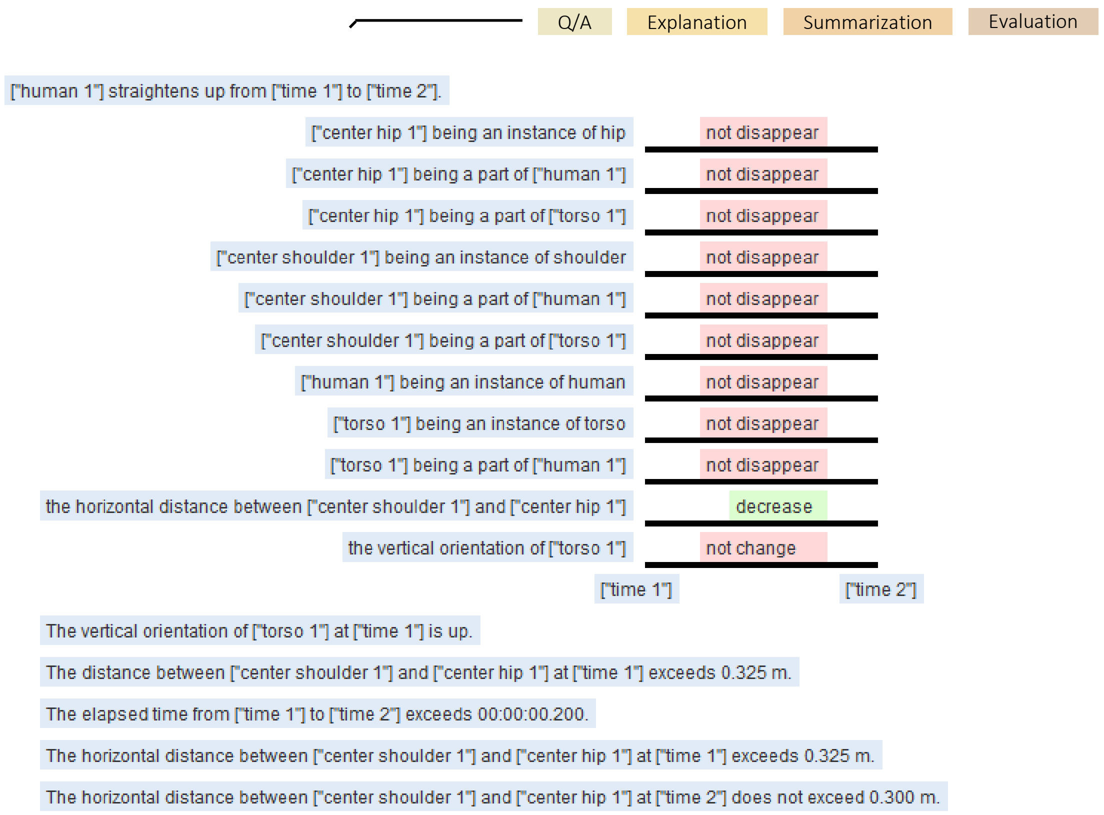
the distance between ["foot 1"] and ["hip 1"] — increase

["time 1"] ["time 2"]

The distance between ["foot 1"] and ["hip 1"] at ["time 2"] exceeds 1.000 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

The distance between ["foot 1"] and ["hip 1"] at ["time 1"] does not exceed 0.800 m.



Q/A Explanation Summarization Evaluation

["human 1"] straightens up from ["time 1"] to ["time 2"].

["center hip 1"] being an instance of hip	not disappear
["center hip 1"] being a part of ["human 1"]	not disappear
["center hip 1"] being a part of ["torso 1"]	not disappear
["center shoulder 1"] being an instance of shoulder	not disappear
["center shoulder 1"] being a part of ["human 1"]	not disappear
["center shoulder 1"] being a part of ["torso 1"]	not disappear
["human 1"] being an instance of human	not disappear
["torso 1"] being an instance of torso	not disappear
["torso 1"] being a part of ["human 1"]	not disappear
the horizontal distance between ["center shoulder 1"] and ["center hip 1"]	decrease
the vertical orientation of ["torso 1"]	not change
	["time 1"] ["time 2"]

The vertical orientation of ["torso 1"] at ["time 1"] is up.

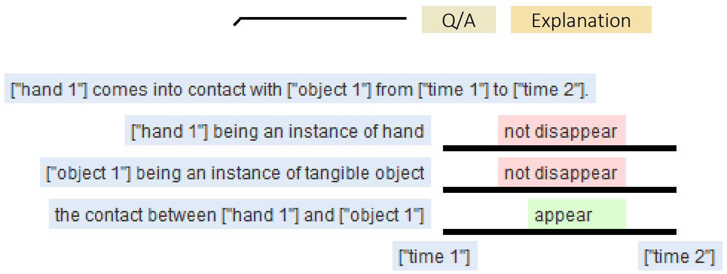
The distance between ["center shoulder 1"] and ["center hip 1"] at ["time 1"] exceeds 0.325 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.200.

The horizontal distance between ["center shoulder 1"] and ["center hip 1"] at ["time 1"] exceeds 0.325 m.

The horizontal distance between ["center shoulder 1"] and ["center hip 1"] at ["time 2"] does not exceed 0.300 m.

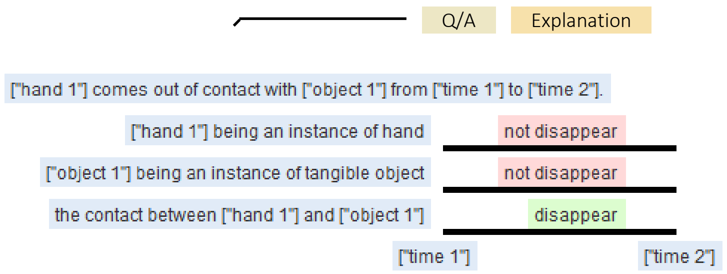
1 Human, 1 Object



Q/A Explanation

["hand 1"] comes into contact with ["object 1"] from ["time 1"] to ["time 2"].

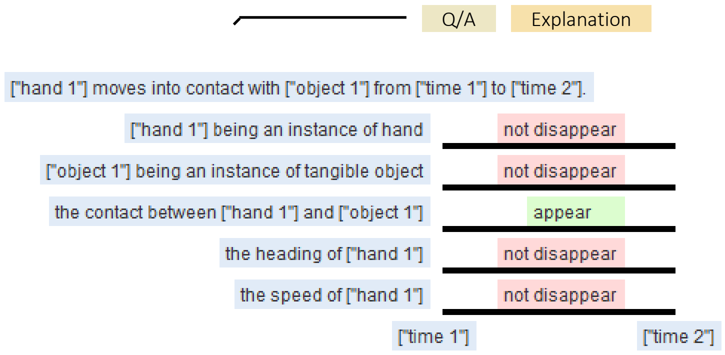
["hand 1"] being an instance of hand	not disappear
["object 1"] being an instance of tangible object	not disappear
the contact between ["hand 1"] and ["object 1"]	appear
	["time 1"] ["time 2"]



Q/A Explanation

["hand 1"] comes out of contact with ["object 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["object 1"] being an instance of tangible object	not disappear
the contact between ["hand 1"] and ["object 1"]	disappear
	["time 1"] ["time 2"]



Q/A Explanation

["hand 1"] moves into contact with ["object 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand — not disappear

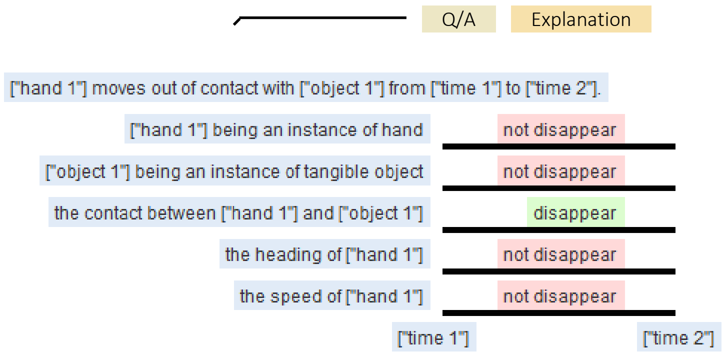
["object 1"] being an instance of tangible object — not disappear

the contact between ["hand 1"] and ["object 1"] — appear

the heading of ["hand 1"] — not disappear

the speed of ["hand 1"] — not disappear

["time 1"] ["time 2"]



Q/A Explanation

["hand 1"] moves out of contact with ["object 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand — not disappear

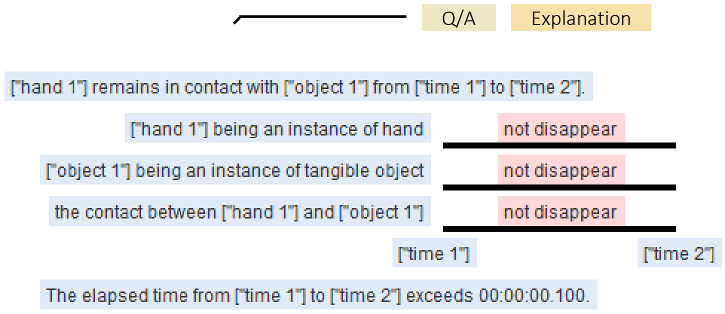
["object 1"] being an instance of tangible object — not disappear

the contact between ["hand 1"] and ["object 1"] — disappear

the heading of ["hand 1"] — not disappear

the speed of ["hand 1"] — not disappear

["time 1"] ["time 2"]



Q/A Explanation

["hand 1"] remains in contact with ["object 1"] from ["time 1"] to ["time 2"].

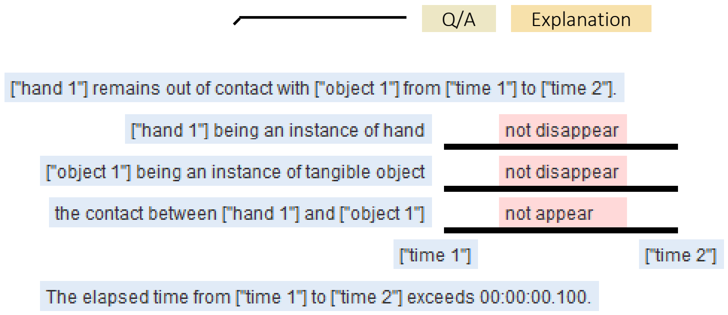
["hand 1"] being an instance of hand — not disappear

["object 1"] being an instance of tangible object — not disappear

the contact between ["hand 1"] and ["object 1"] — not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A Explanation

["hand 1"] remains out of contact with ["object 1"] from ["time 1"] to ["time 2"].

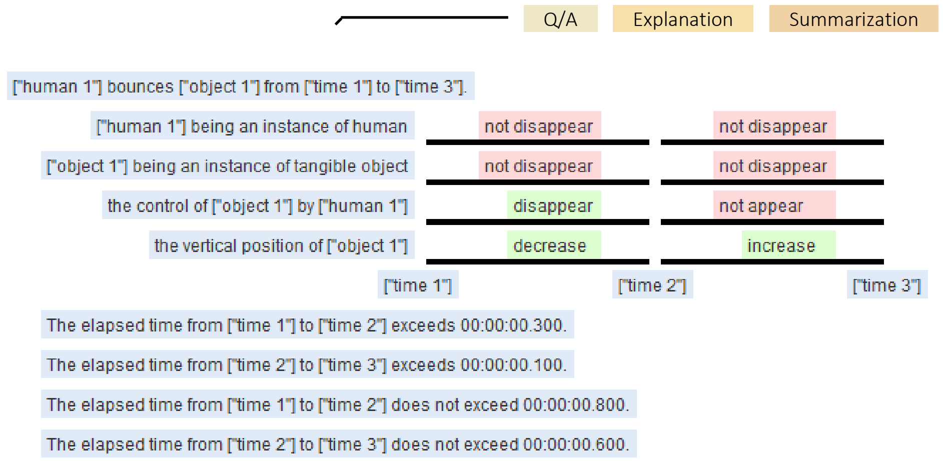
["hand 1"] being an instance of hand — not disappear

["object 1"] being an instance of tangible object — not disappear

the contact between ["hand 1"] and ["object 1"] — not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



["human 1"] bounces ["object 1"] from ["time 1"] to ["time 3"].

	["time 1"] – ["time 2"]	["time 2"] – ["time 3"]
["human 1"] being an instance of human	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	disappear	not appear
the vertical position of ["object 1"]	decrease	increase

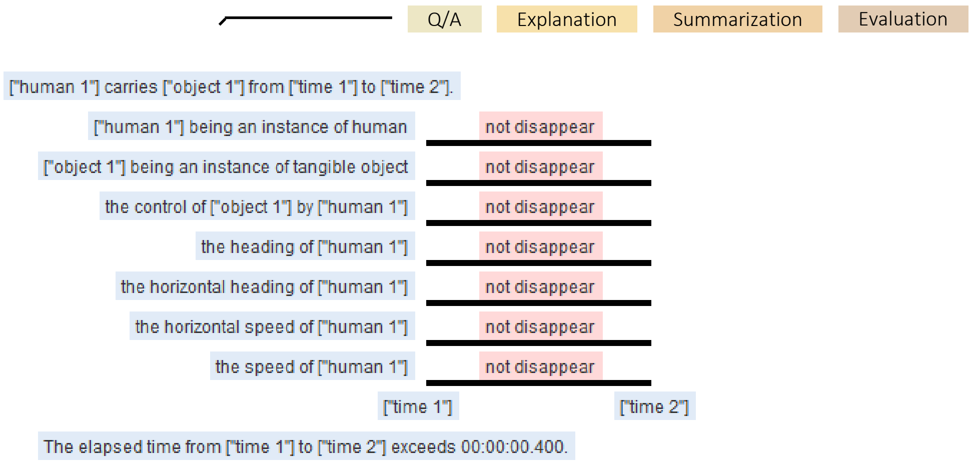
["time 1"] ["time 2"] ["time 3"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.300.

The elapsed time from ["time 2"] to ["time 3"] exceeds 00:00:00.100.

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:00:00.800.

The elapsed time from ["time 2"] to ["time 3"] does not exceed 00:00:00.600.

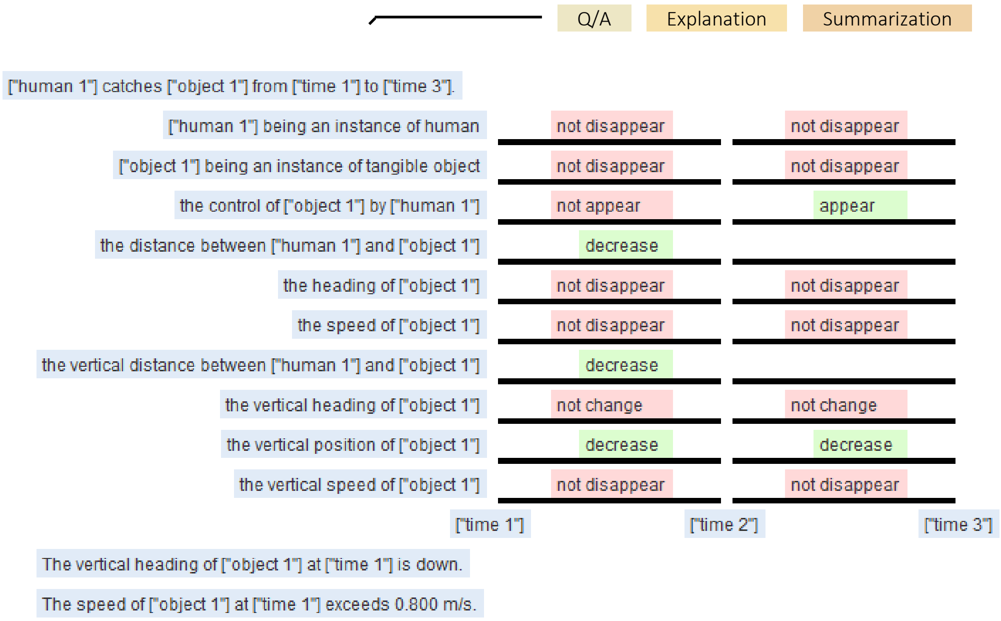


["human 1"] carries ["object 1"] from ["time 1"] to ["time 2"].

	["time 1"] – ["time 2"]
["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the heading of ["human 1"]	not disappear
the horizontal heading of ["human 1"]	not disappear
the horizontal speed of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.



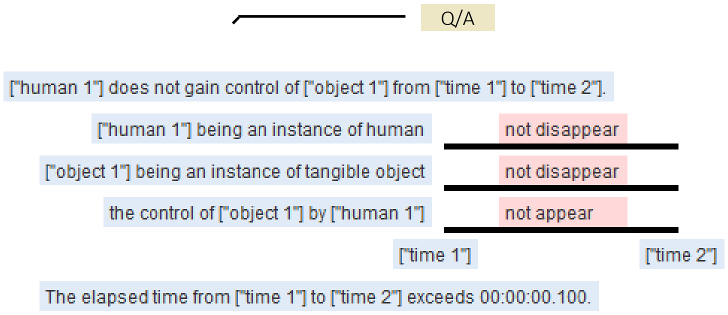
["human 1"] catches ["object 1"] from ["time 1"] to ["time 3"].

	["time 1"] – ["time 2"]	["time 2"] – ["time 3"]
["human 1"] being an instance of human	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	not appear	appear
the distance between ["human 1"] and ["object 1"]	decrease	
the heading of ["object 1"]	not disappear	not disappear
the speed of ["object 1"]	not disappear	not disappear
the vertical distance between ["human 1"] and ["object 1"]	decrease	
the vertical heading of ["object 1"]	not change	not change
the vertical position of ["object 1"]	decrease	decrease
the vertical speed of ["object 1"]	not disappear	not disappear

["time 1"] ["time 2"] ["time 3"]

The vertical heading of ["object 1"] at ["time 1"] is down.

The speed of ["object 1"] at ["time 1"] exceeds 0.800 m/s.



Q/A

["human 1"] does not gain control of ["object 1"] from ["time 1"] to ["time 2"].

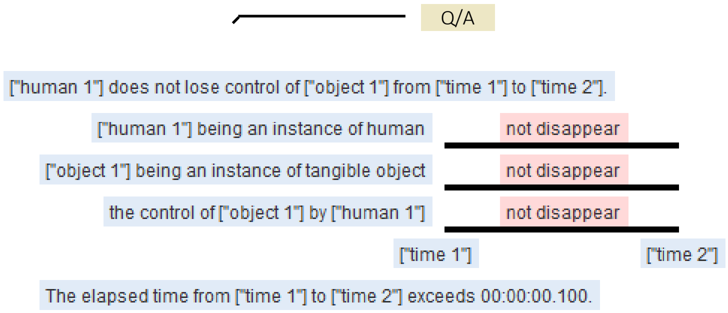
["human 1"] being an instance of human — not disappear

["object 1"] being an instance of tangible object — not disappear

the control of ["object 1"] by ["human 1"] — not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A

["human 1"] does not lose control of ["object 1"] from ["time 1"] to ["time 2"].

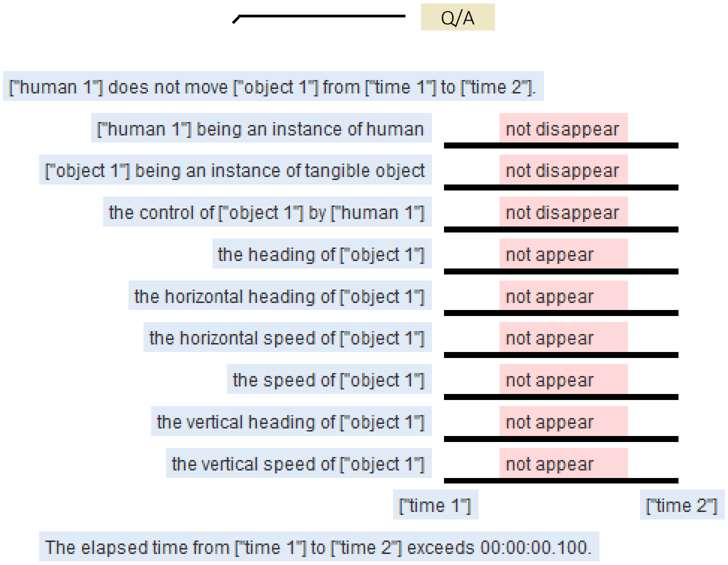
["human 1"] being an instance of human — not disappear

["object 1"] being an instance of tangible object — not disappear

the control of ["object 1"] by ["human 1"] — not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A

["human 1"] does not move ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human — not disappear

["object 1"] being an instance of tangible object — not disappear

the control of ["object 1"] by ["human 1"] — not disappear

the heading of ["object 1"] — not appear

the horizontal heading of ["object 1"] — not appear

the horizontal speed of ["object 1"] — not appear

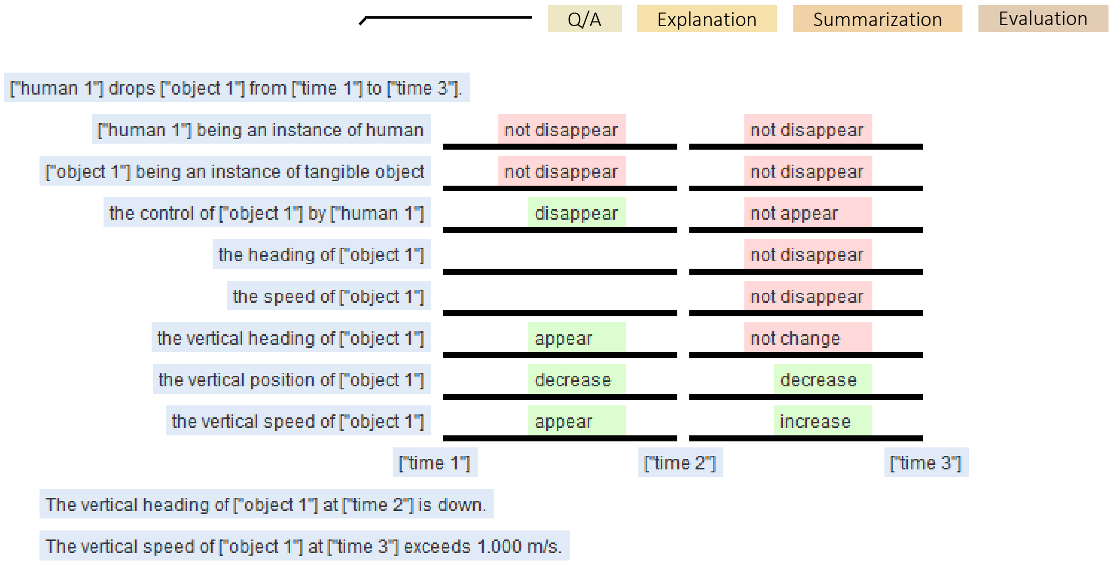
the speed of ["object 1"] — not appear

the vertical heading of ["object 1"] — not appear

the vertical speed of ["object 1"] — not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A | Explanation | Summarization | Evaluation

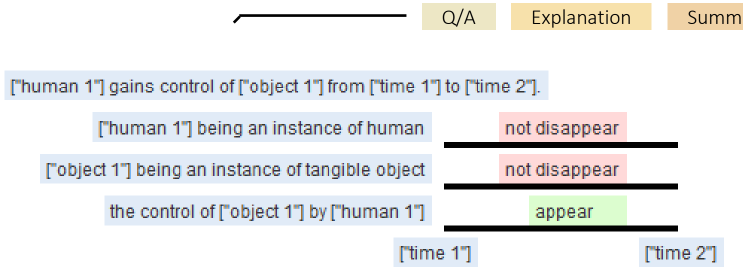
["human 1"] drops ["object 1"] from ["time 1"] to ["time 3"].

["human 1"] being an instance of human	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	disappear	not appear
the heading of ["object 1"]		not disappear
the speed of ["object 1"]		not disappear
the vertical heading of ["object 1"]	appear	not change
the vertical position of ["object 1"]	decrease	decrease
the vertical speed of ["object 1"]	appear	increase

["time 1"] ["time 2"] ["time 3"]

The vertical heading of ["object 1"] at ["time 2"] is down.

The vertical speed of ["object 1"] at ["time 3"] exceeds 1.000 m/s.

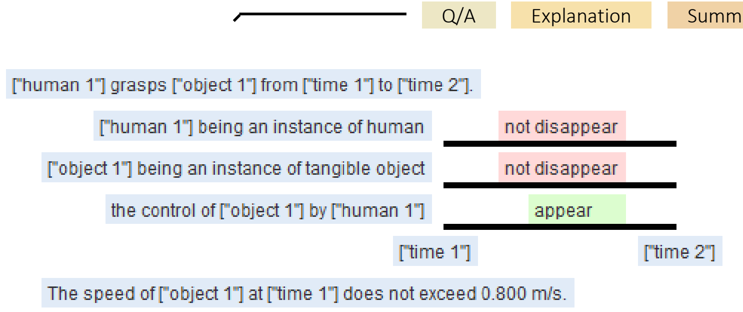


Q/A | Explanation | Summarization

["human 1"] gains control of ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	appear

["time 1"] ["time 2"]



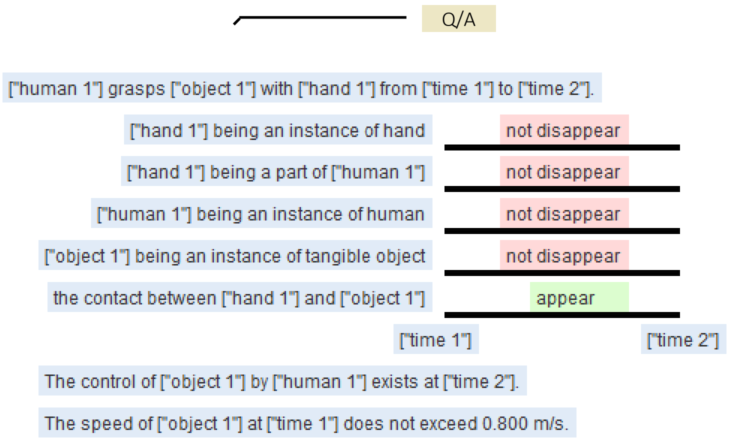
Q/A | Explanation | Summarization

["human 1"] grasps ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	appear

["time 1"] ["time 2"]

The speed of ["object 1"] at ["time 1"] does not exceed 0.800 m/s.



Q/A

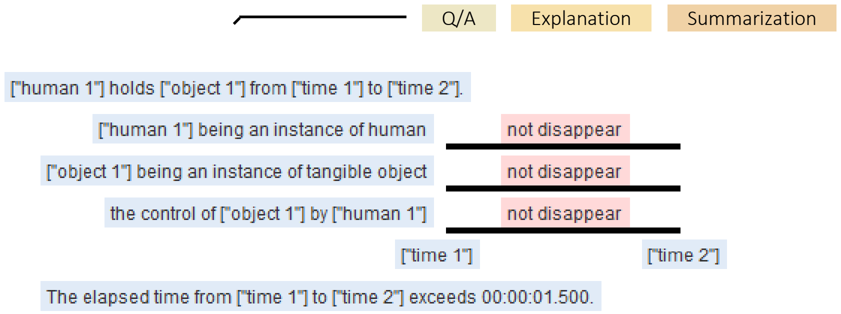
["human 1"] grasps ["object 1"] with ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the contact between ["hand 1"] and ["object 1"]	appear

["time 1"] ["time 2"]

The control of ["object 1"] by ["human 1"] exists at ["time 2"].

The speed of ["object 1"] at ["time 1"] does not exceed 0.800 m/s.



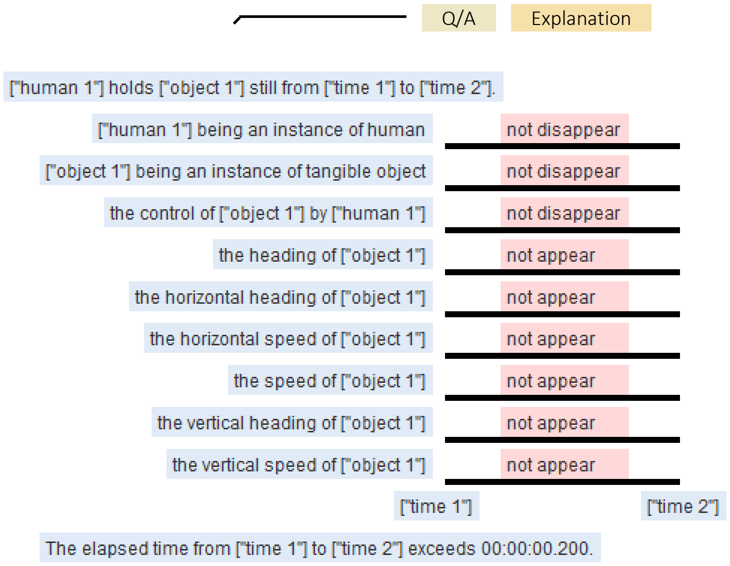
Q/A Explanation Summarization

["human 1"] holds ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:01.500.



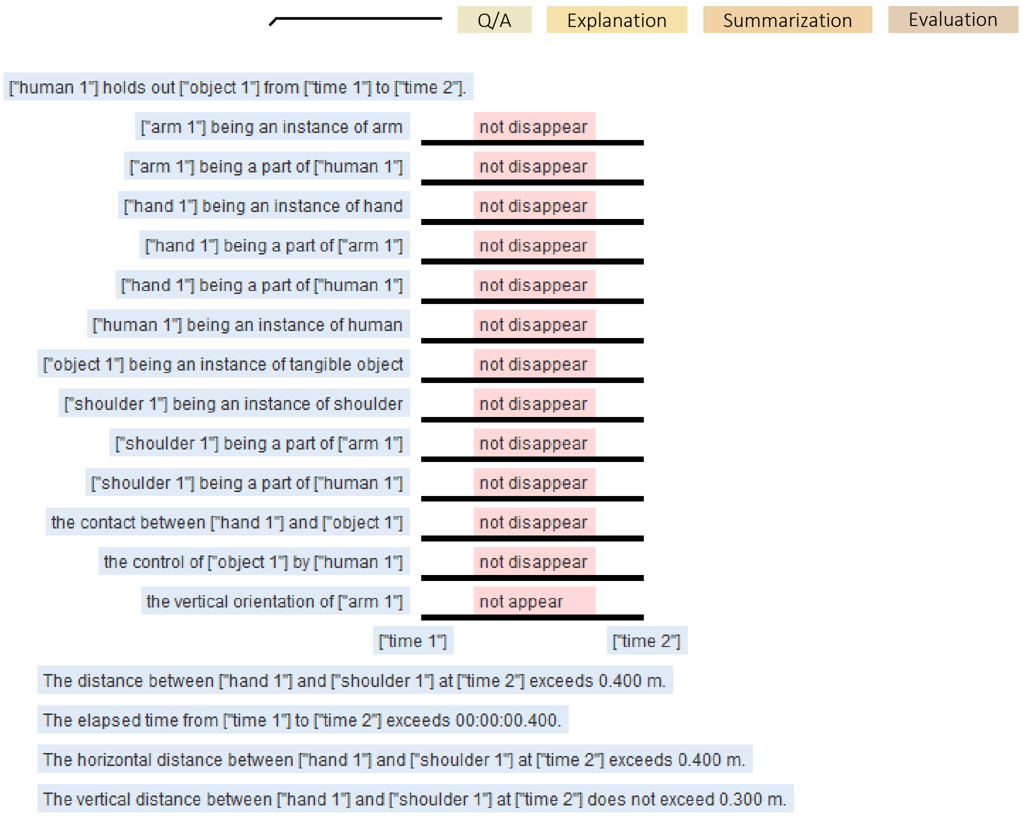
Q/A Explanation

["human 1"] holds ["object 1"] still from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the heading of ["object 1"]	not appear
the horizontal heading of ["object 1"]	not appear
the horizontal speed of ["object 1"]	not appear
the speed of ["object 1"]	not appear
the vertical heading of ["object 1"]	not appear
the vertical speed of ["object 1"]	not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.200.



["human 1"] holds out ["object 1"] from ["time 1"] to ["time 2"].

["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the contact between ["hand 1"] and ["object 1"]	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the vertical orientation of ["arm 1"]	not appear

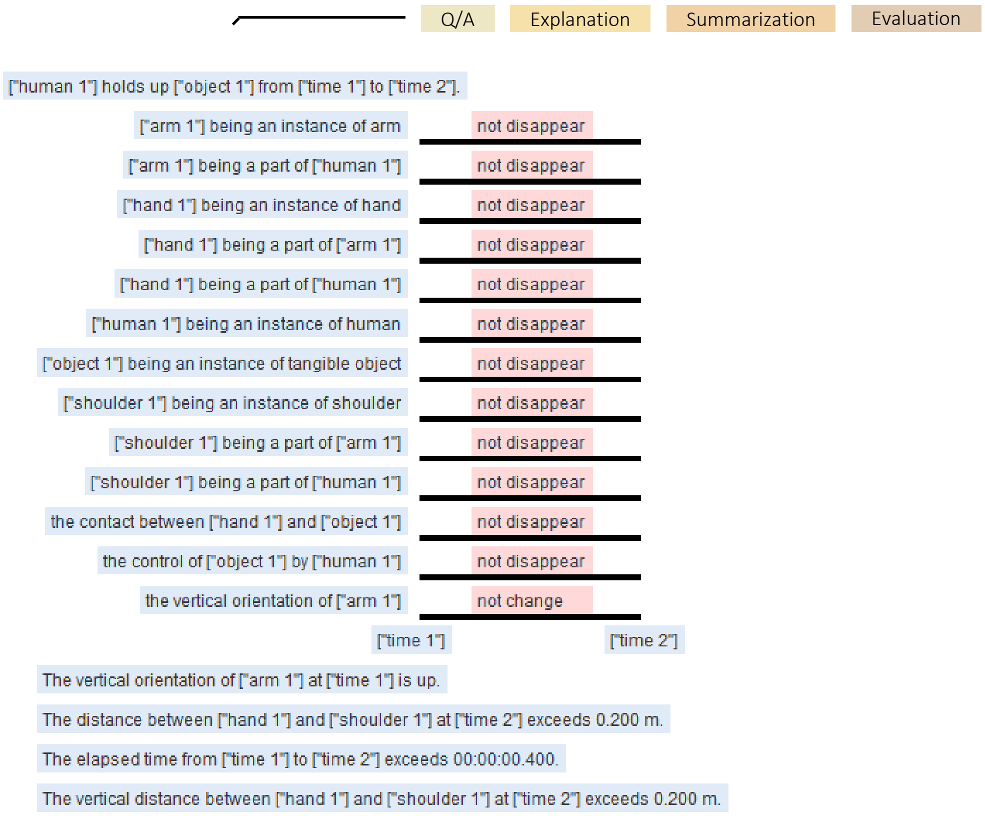
["time 1"] ["time 2"]

The distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.400 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The horizontal distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.400 m.

The vertical distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] does not exceed 0.300 m.



["human 1"] holds up ["object 1"] from ["time 1"] to ["time 2"].

["arm 1"] being an instance of arm	not disappear
["arm 1"] being a part of ["human 1"]	not disappear
["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["arm 1"]	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
["shoulder 1"] being an instance of shoulder	not disappear
["shoulder 1"] being a part of ["arm 1"]	not disappear
["shoulder 1"] being a part of ["human 1"]	not disappear
the contact between ["hand 1"] and ["object 1"]	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the vertical orientation of ["arm 1"]	not change

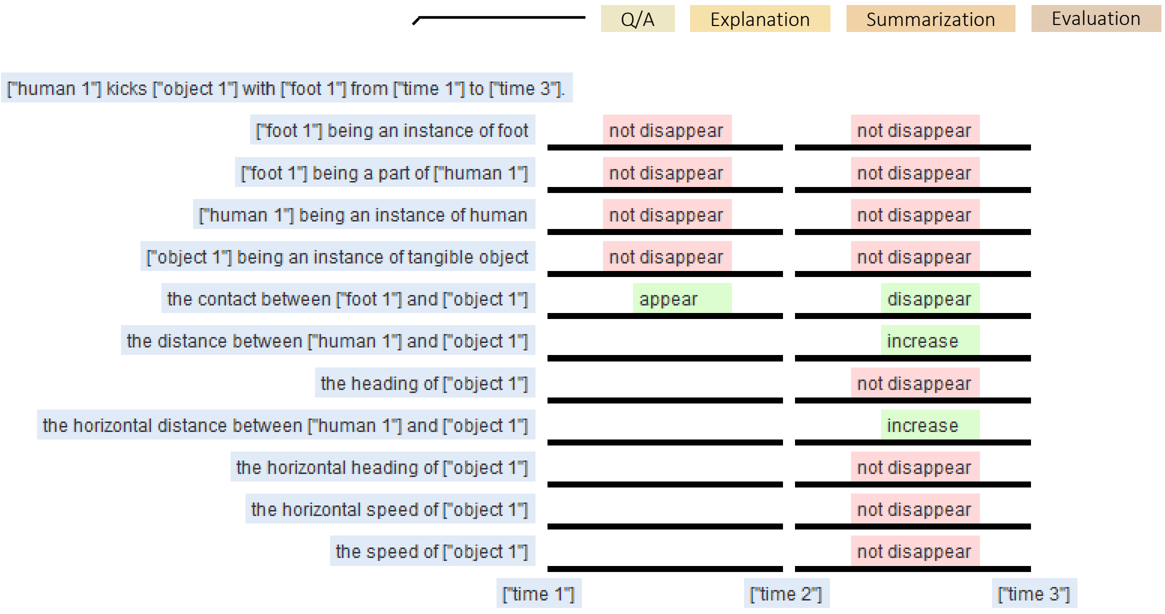
["time 1"] ["time 2"]

The vertical orientation of ["arm 1"] at ["time 1"] is up.

The distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.200 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The vertical distance between ["hand 1"] and ["shoulder 1"] at ["time 2"] exceeds 0.200 m.



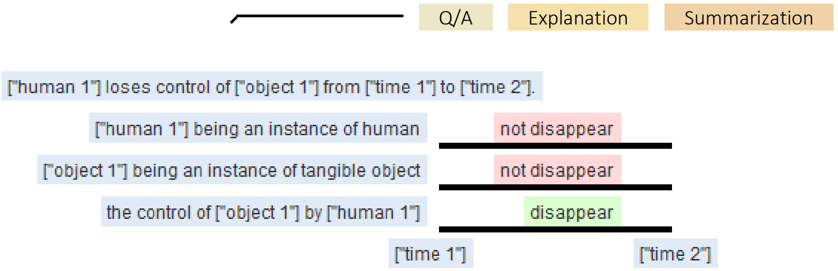
Q/A Explanation Summarization Evaluation

["human 1"] kicks ["object 1"] with ["foot 1"] from ["time 1"] to ["time 3"].

["foot 1"] being an instance of foot	not disappear	not disappear
["foot 1"] being a part of ["human 1"]	not disappear	not disappear
["human 1"] being an instance of human	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear
the contact between ["foot 1"] and ["object 1"]	appear	disappear
the distance between ["human 1"] and ["object 1"]		increase
the heading of ["object 1"]		not disappear
the horizontal distance between ["human 1"] and ["object 1"]		increase
the horizontal heading of ["object 1"]		not disappear
the horizontal speed of ["object 1"]		not disappear
the speed of ["object 1"]		not disappear

["time 1"] ["time 2"] ["time 3"]

The elapsed time from ["time 1"] to ["time 2"] does not exceed 00:00:01.000.

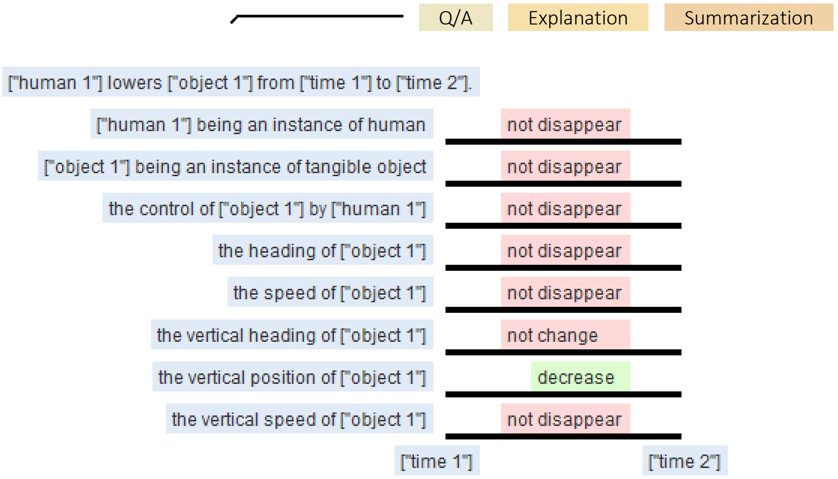


Q/A Explanation Summarization

["human 1"] loses control of ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	disappear

["time 1"] ["time 2"]



Q/A Explanation Summarization

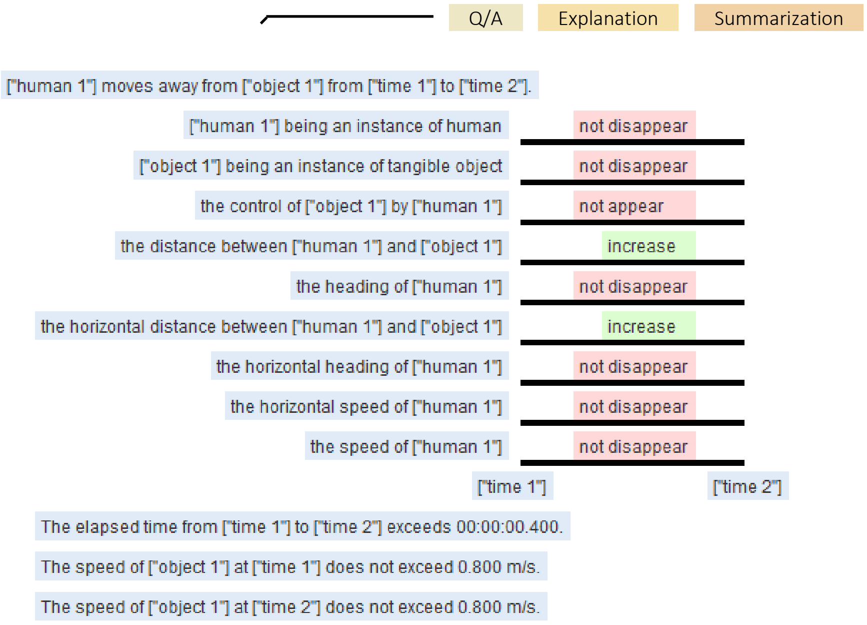
["human 1"] lowers ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the heading of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear
the vertical heading of ["object 1"]	not change
the vertical position of ["object 1"]	decrease
the vertical speed of ["object 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["object 1"] at ["time 1"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.200.



Q/A Explanation Summarization

["human 1"] moves away from ["object 1"] from ["time 1"] to ["time 2"].

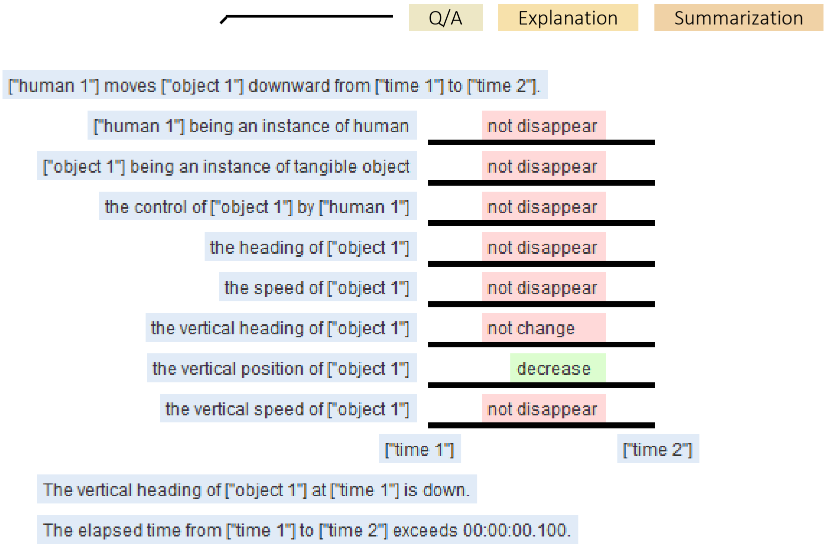
["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not appear
the distance between ["human 1"] and ["object 1"]	increase
the heading of ["human 1"]	not disappear
the horizontal distance between ["human 1"] and ["object 1"]	increase
the horizontal heading of ["human 1"]	not disappear
the horizontal speed of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The speed of ["object 1"] at ["time 1"] does not exceed 0.800 m/s.

The speed of ["object 1"] at ["time 2"] does not exceed 0.800 m/s.



Q/A Explanation Summarization

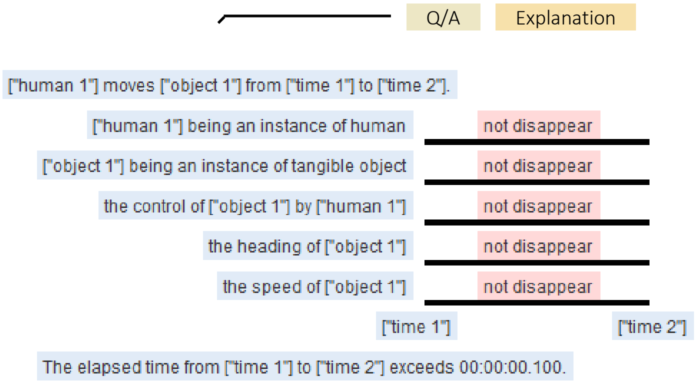
["human 1"] moves ["object 1"] downward from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the heading of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear
the vertical heading of ["object 1"]	not change
the vertical position of ["object 1"]	decrease
the vertical speed of ["object 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["object 1"] at ["time 1"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



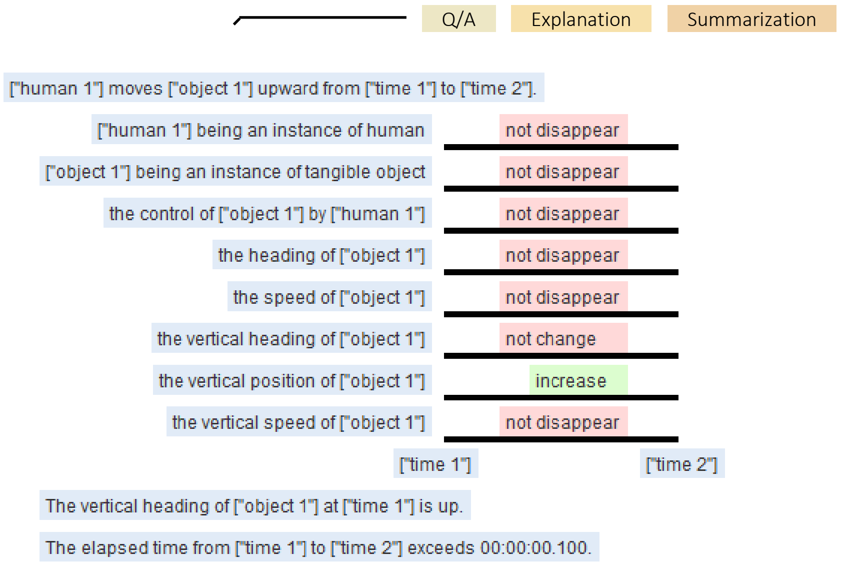
Q/A Explanation

["human 1"] moves ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the heading of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



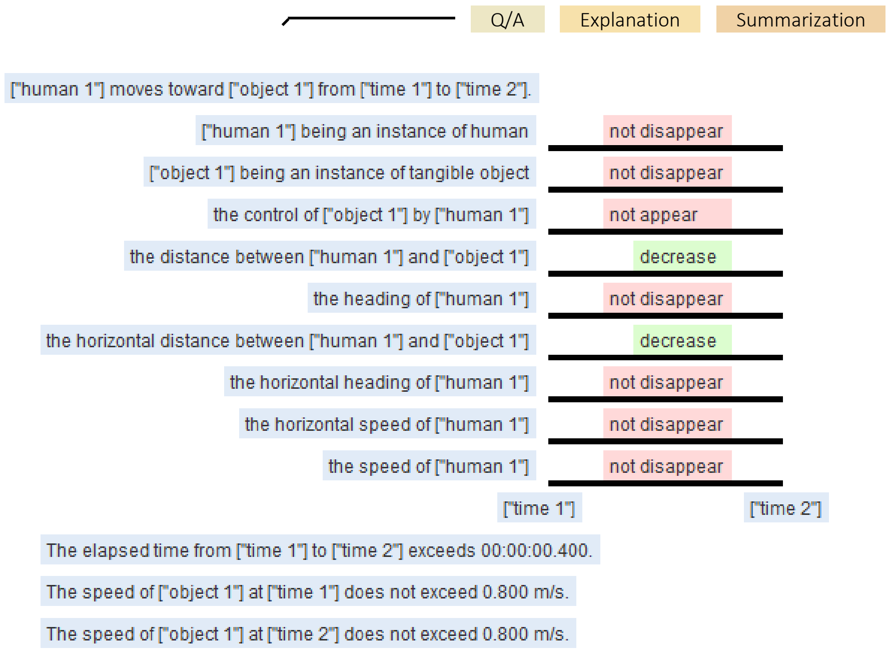
["human 1"] moves ["object 1"] upward from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the heading of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear
the vertical heading of ["object 1"]	not change
the vertical position of ["object 1"]	increase
the vertical speed of ["object 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["object 1"] at ["time 1"] is up.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



["human 1"] moves toward ["object 1"] from ["time 1"] to ["time 2"].

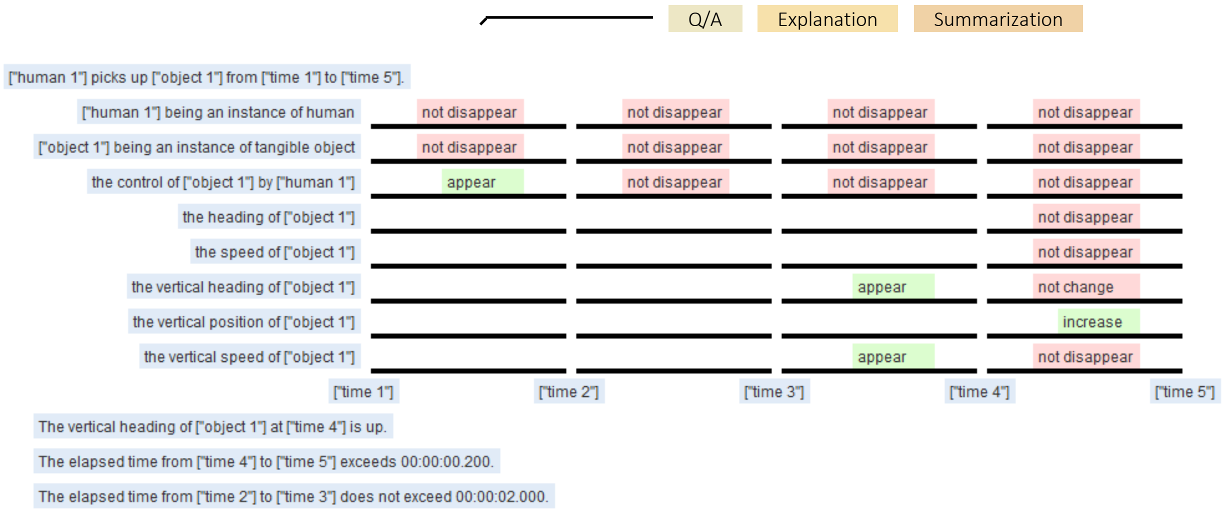
["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not appear
the distance between ["human 1"] and ["object 1"]	decrease
the heading of ["human 1"]	not disappear
the horizontal distance between ["human 1"] and ["object 1"]	decrease
the horizontal heading of ["human 1"]	not disappear
the horizontal speed of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The speed of ["object 1"] at ["time 1"] does not exceed 0.800 m/s.

The speed of ["object 1"] at ["time 2"] does not exceed 0.800 m/s.



["human 1"] picks up ["object 1"] from ["time 1"] to ["time 5"].

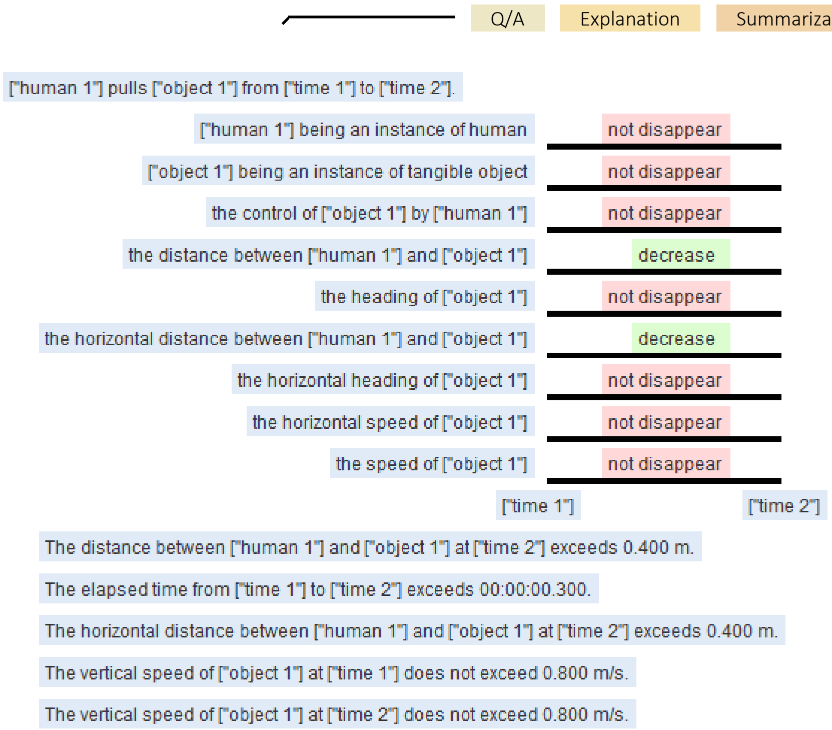
["human 1"] being an instance of human	not disappear	not disappear	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	appear	not disappear	not disappear	not disappear
the heading of ["object 1"]				not disappear
the speed of ["object 1"]				not disappear
the vertical heading of ["object 1"]			appear	not change
the vertical position of ["object 1"]				increase
the vertical speed of ["object 1"]			appear	not disappear

["time 1"] ["time 2"] ["time 3"] ["time 4"] ["time 5"]

The vertical heading of ["object 1"] at ["time 4"] is up.

The elapsed time from ["time 4"] to ["time 5"] exceeds 00:00:00.200.

The elapsed time from ["time 2"] to ["time 3"] does not exceed 00:00:02.000.



["human 1"] pulls ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the distance between ["human 1"] and ["object 1"]	decrease
the heading of ["object 1"]	not disappear
the horizontal distance between ["human 1"] and ["object 1"]	decrease
the horizontal heading of ["object 1"]	not disappear
the horizontal speed of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear

["time 1"] ["time 2"]

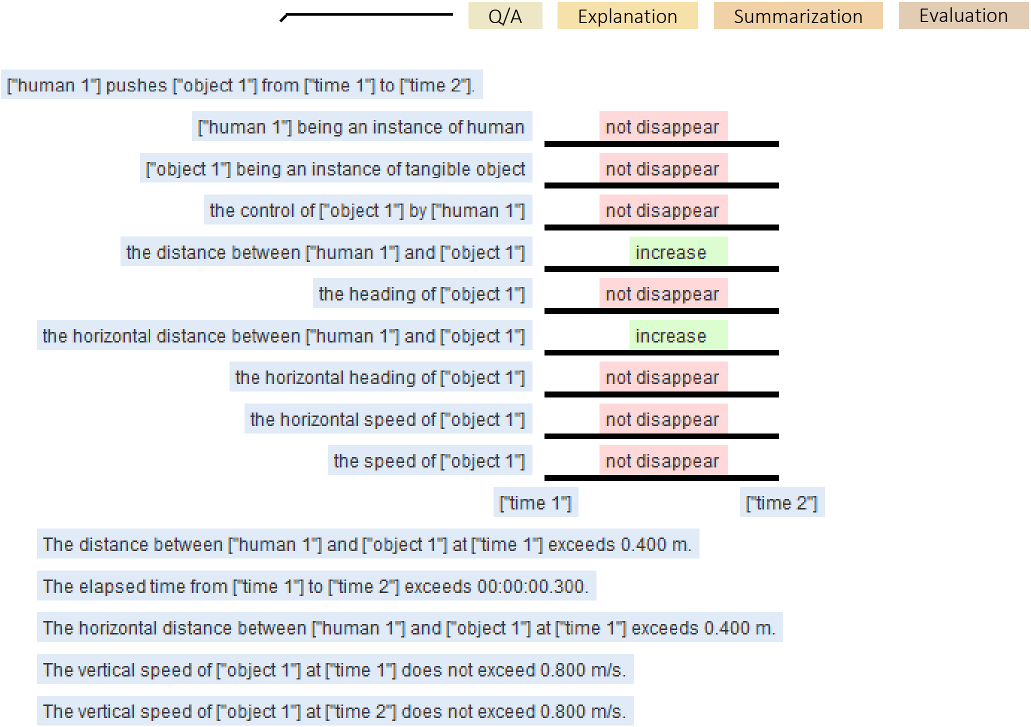
The distance between ["human 1"] and ["object 1"] at ["time 2"] exceeds 0.400 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.300.

The horizontal distance between ["human 1"] and ["object 1"] at ["time 2"] exceeds 0.400 m.

The vertical speed of ["object 1"] at ["time 1"] does not exceed 0.800 m/s.

The vertical speed of ["object 1"] at ["time 2"] does not exceed 0.800 m/s.



["human 1"] pushes ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the distance between ["human 1"] and ["object 1"]	increase
the heading of ["object 1"]	not disappear
the horizontal distance between ["human 1"] and ["object 1"]	increase
the horizontal heading of ["object 1"]	not disappear
the horizontal speed of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear

["time 1"] ["time 2"]

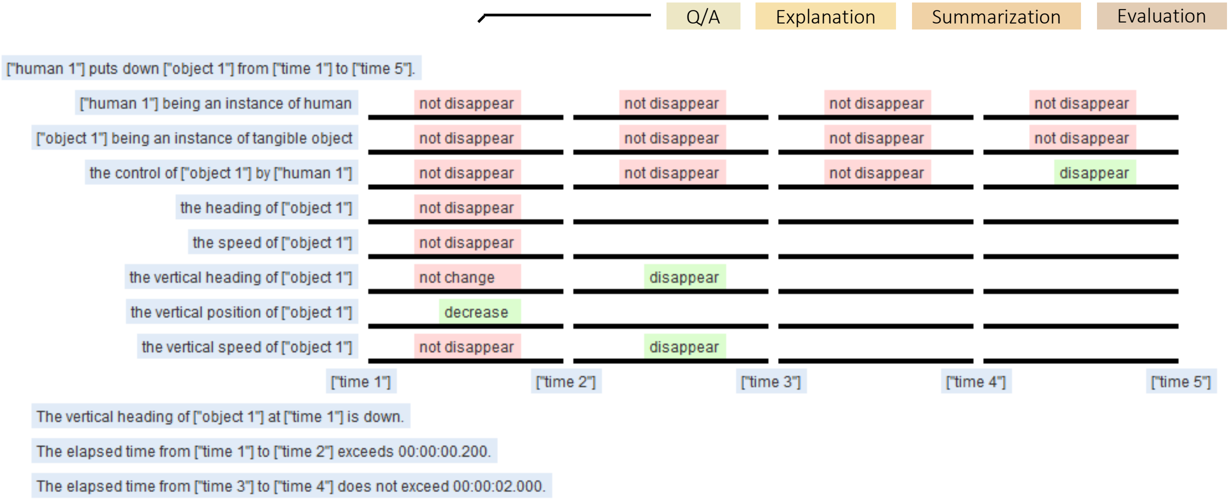
The distance between ["human 1"] and ["object 1"] at ["time 1"] exceeds 0.400 m.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.300.

The horizontal distance between ["human 1"] and ["object 1"] at ["time 1"] exceeds 0.400 m.

The vertical speed of ["object 1"] at ["time 1"] does not exceed 0.800 m/s.

The vertical speed of ["object 1"] at ["time 2"] does not exceed 0.800 m/s.



["human 1"] puts down ["object 1"] from ["time 1"] to ["time 5"].

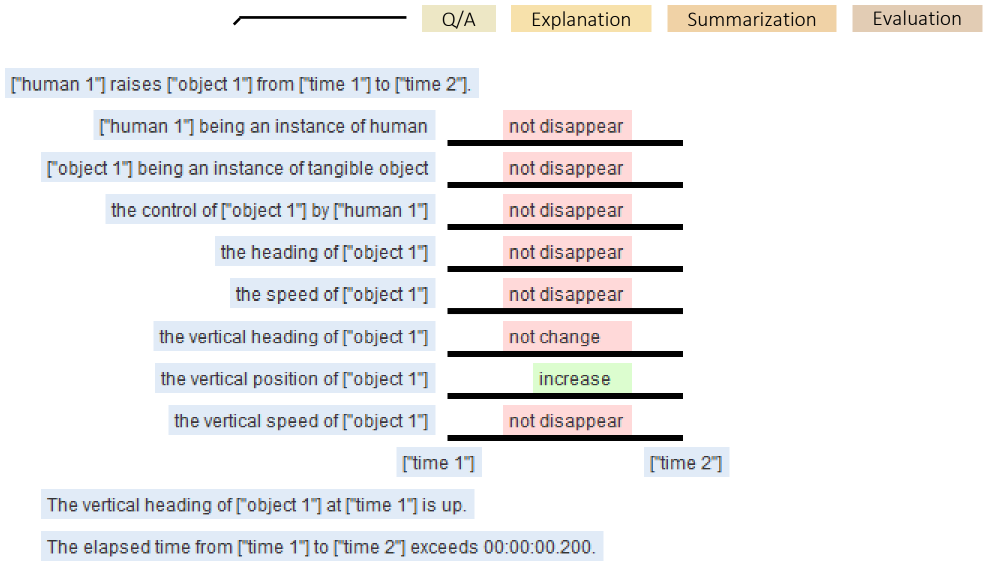
["human 1"] being an instance of human	not disappear	not disappear	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	not disappear	not disappear	not disappear	disappear
the heading of ["object 1"]	not disappear			
the speed of ["object 1"]	not disappear			
the vertical heading of ["object 1"]	not change	disappear		
the vertical position of ["object 1"]	decrease			
the vertical speed of ["object 1"]	not disappear	disappear		

["time 1"] ["time 2"] ["time 3"] ["time 4"] ["time 5"]

The vertical heading of ["object 1"] at ["time 1"] is down.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.200.

The elapsed time from ["time 3"] to ["time 4"] does not exceed 00:00:02.000.



Q/A Explanation Summarization Evaluation

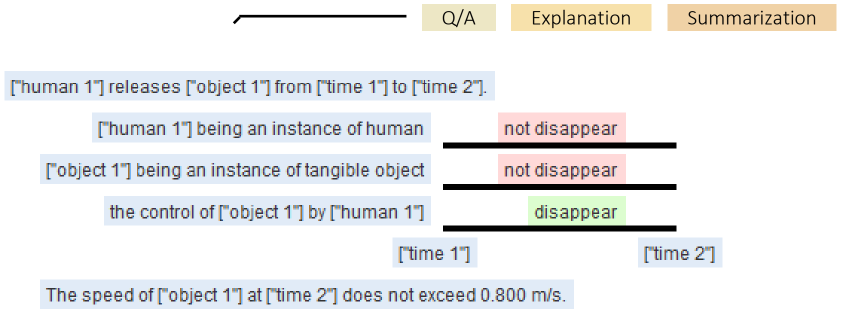
["human 1"] raises ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not disappear
the heading of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear
the vertical heading of ["object 1"]	not change
the vertical position of ["object 1"]	increase
the vertical speed of ["object 1"]	not disappear

["time 1"] ["time 2"]

The vertical heading of ["object 1"] at ["time 1"] is up.

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.200.



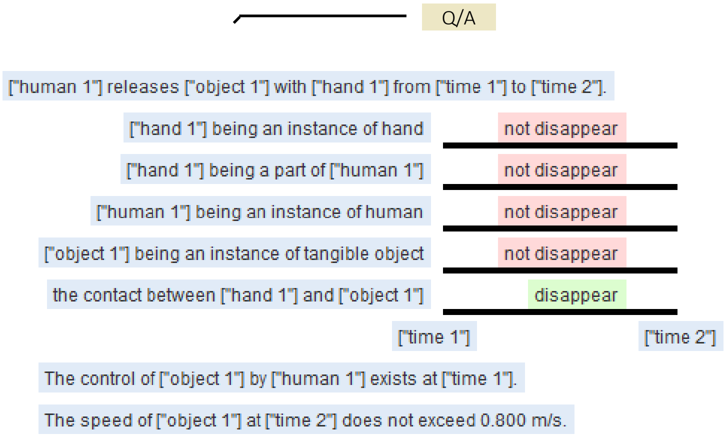
Q/A Explanation Summarization

["human 1"] releases ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	disappear

["time 1"] ["time 2"]

The speed of ["object 1"] at ["time 2"] does not exceed 0.800 m/s.



Q/A

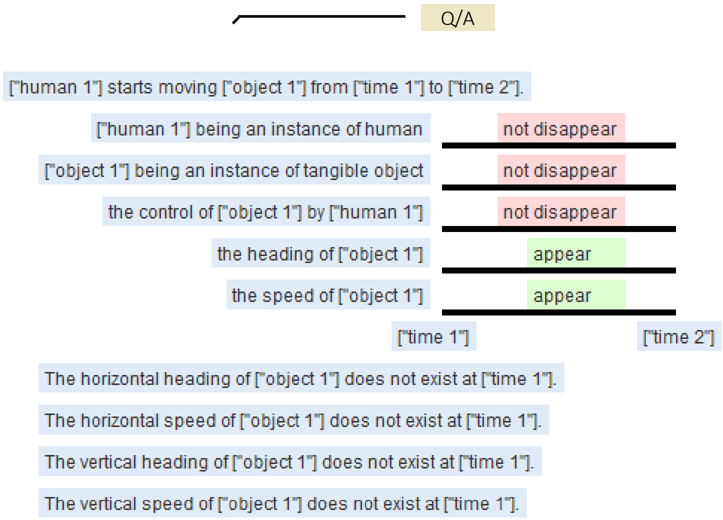
["human 1"] releases ["object 1"] with ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["hand 1"] being a part of ["human 1"]	not disappear
["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the contact between ["hand 1"] and ["object 1"]	disappear

["time 1"] ["time 2"]

The control of ["object 1"] by ["human 1"] exists at ["time 1"].

The speed of ["object 1"] at ["time 2"] does not exceed 0.800 m/s.



Q/A

["human 1"] starts moving ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human — not disappear

["object 1"] being an instance of tangible object — not disappear

the control of ["object 1"] by ["human 1"] — not disappear

the heading of ["object 1"] — appear

the speed of ["object 1"] — appear

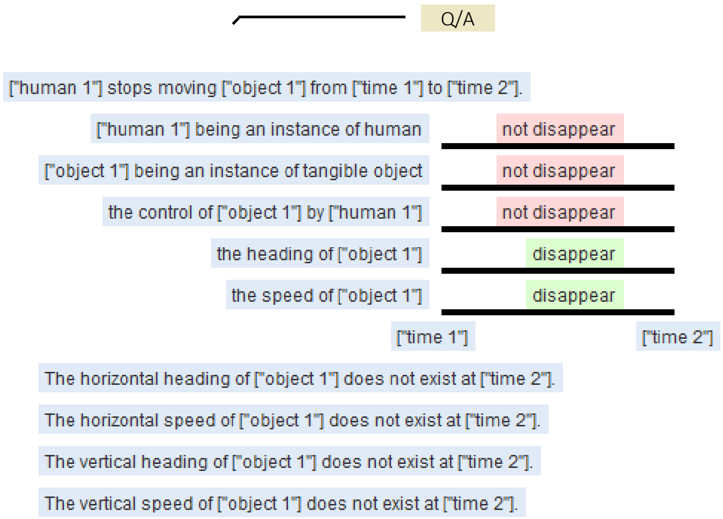
["time 1"] ["time 2"]

The horizontal heading of ["object 1"] does not exist at ["time 1"].

The horizontal speed of ["object 1"] does not exist at ["time 1"].

The vertical heading of ["object 1"] does not exist at ["time 1"].

The vertical speed of ["object 1"] does not exist at ["time 1"].



Q/A

["human 1"] stops moving ["object 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human — not disappear

["object 1"] being an instance of tangible object — not disappear

the control of ["object 1"] by ["human 1"] — not disappear

the heading of ["object 1"] — disappear

the speed of ["object 1"] — disappear

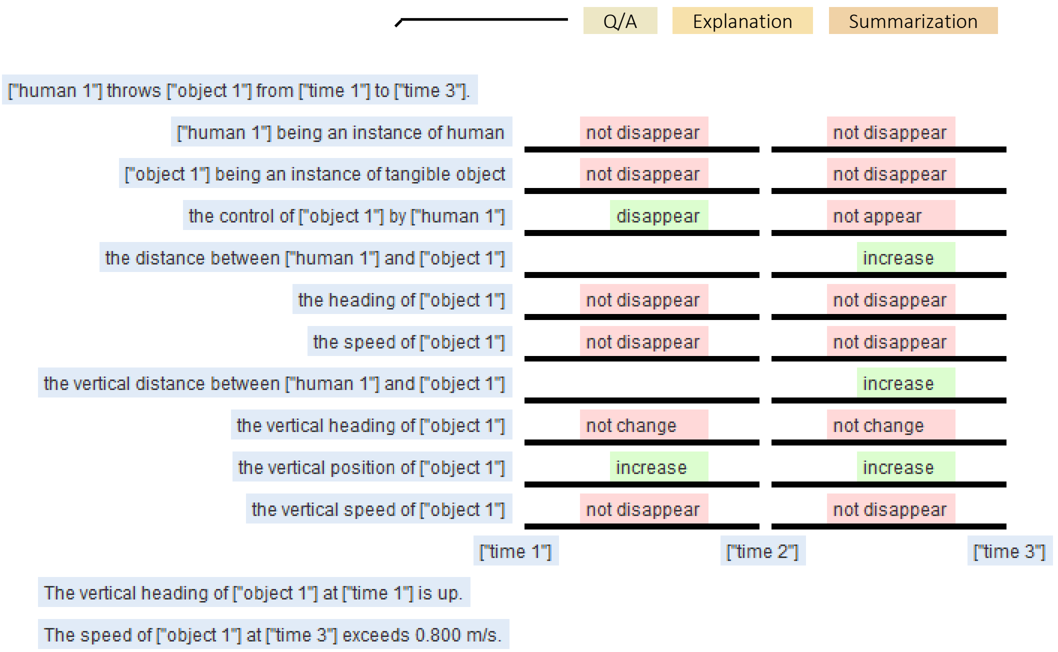
["time 1"] ["time 2"]

The horizontal heading of ["object 1"] does not exist at ["time 2"].

The horizontal speed of ["object 1"] does not exist at ["time 2"].

The vertical heading of ["object 1"] does not exist at ["time 2"].

The vertical speed of ["object 1"] does not exist at ["time 2"].



Q/A Explanation Summarization

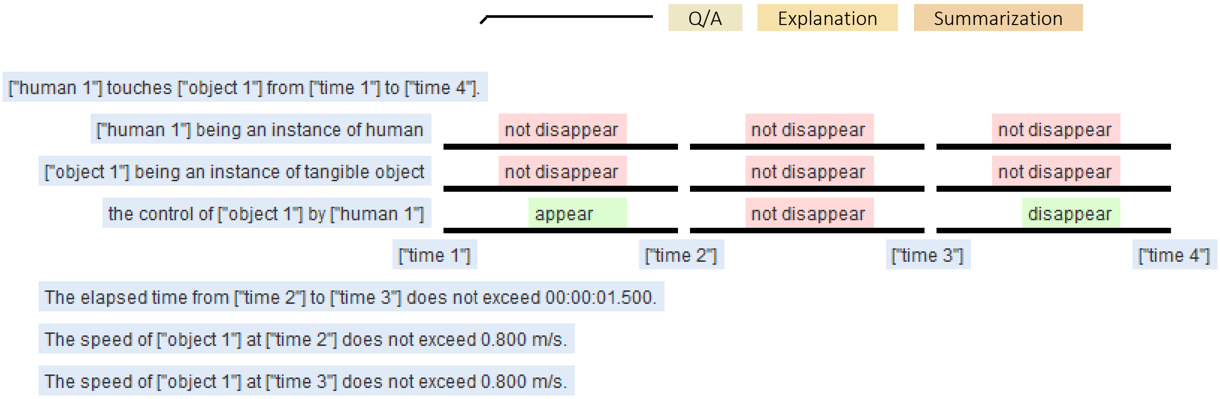
["human 1"] throws ["object 1"] from ["time 1"] to ["time 3"].

["human 1"] being an instance of human	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	disappear	not appear
the distance between ["human 1"] and ["object 1"]		increase
the heading of ["object 1"]	not disappear	not disappear
the speed of ["object 1"]	not disappear	not disappear
the vertical distance between ["human 1"] and ["object 1"]		increase
the vertical heading of ["object 1"]	not change	not change
the vertical position of ["object 1"]	increase	increase
the vertical speed of ["object 1"]	not disappear	not disappear

["time 1"] ["time 2"] ["time 3"]

The vertical heading of ["object 1"] at ["time 1"] is up.

The speed of ["object 1"] at ["time 3"] exceeds 0.800 m/s.



Q/A Explanation Summarization

["human 1"] touches ["object 1"] from ["time 1"] to ["time 4"].

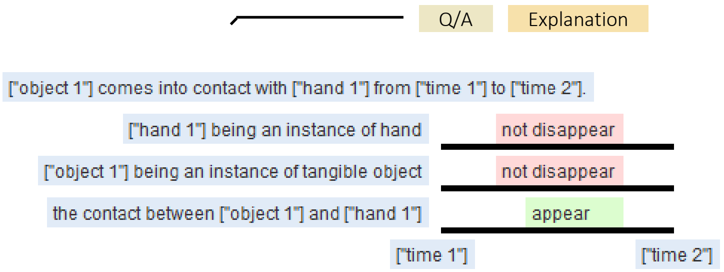
["human 1"] being an instance of human	not disappear	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	appear	not disappear	disappear

["time 1"] ["time 2"] ["time 3"] ["time 4"]

The elapsed time from ["time 2"] to ["time 3"] does not exceed 00:00:01.500.

The speed of ["object 1"] at ["time 2"] does not exceed 0.800 m/s.

The speed of ["object 1"] at ["time 3"] does not exceed 0.800 m/s.

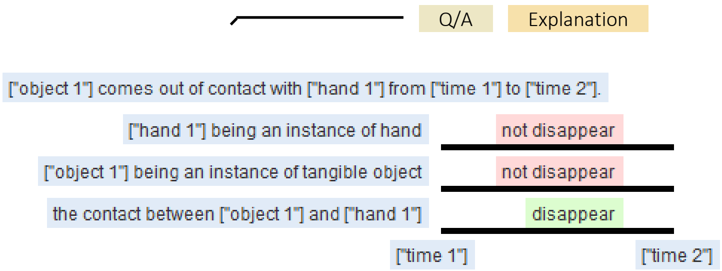


Q/A Explanation

["object 1"] comes into contact with ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["object 1"] being an instance of tangible object	not disappear
the contact between ["object 1"] and ["hand 1"]	appear

["time 1"] ["time 2"]



Q/A Explanation

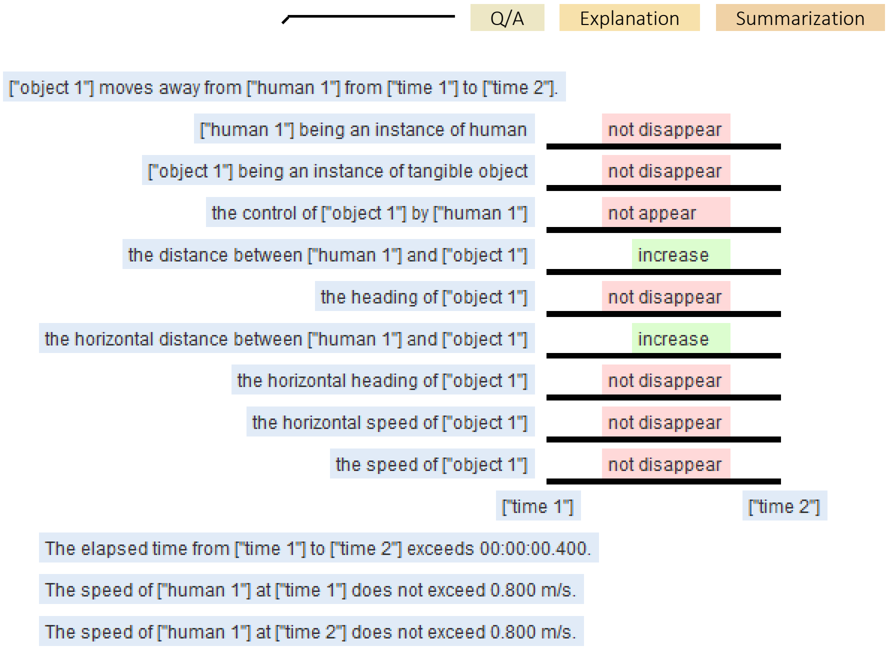
["object 1"] comes out of contact with ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand not disappear

["object 1"] being an instance of tangible object not disappear

the contact between ["object 1"] and ["hand 1"] disappear

["time 1"] ["time 2"]



Q/A Explanation Summarization

["object 1"] moves away from ["human 1"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human not disappear

["object 1"] being an instance of tangible object not disappear

the control of ["object 1"] by ["human 1"] not appear

the distance between ["human 1"] and ["object 1"] increase

the heading of ["object 1"] not disappear

the horizontal distance between ["human 1"] and ["object 1"] increase

the horizontal heading of ["object 1"] not disappear

the horizontal speed of ["object 1"] not disappear

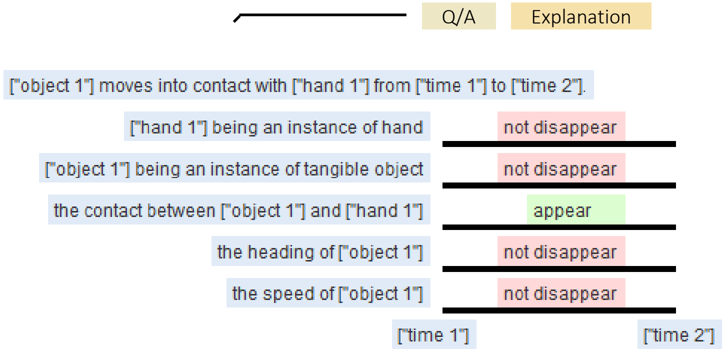
the speed of ["object 1"] not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The speed of ["human 1"] at ["time 1"] does not exceed 0.800 m/s.

The speed of ["human 1"] at ["time 2"] does not exceed 0.800 m/s.



Q/A Explanation

["object 1"] moves into contact with ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand not disappear

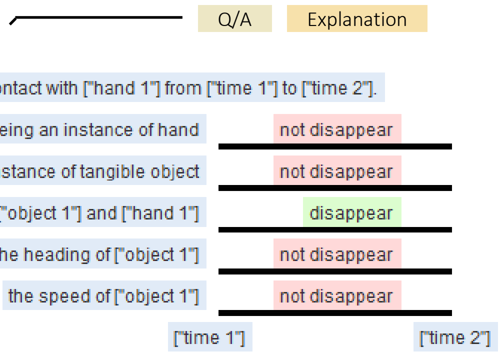
["object 1"] being an instance of tangible object not disappear

the contact between ["object 1"] and ["hand 1"] appear

the heading of ["object 1"] not disappear

the speed of ["object 1"] not disappear

["time 1"] ["time 2"]

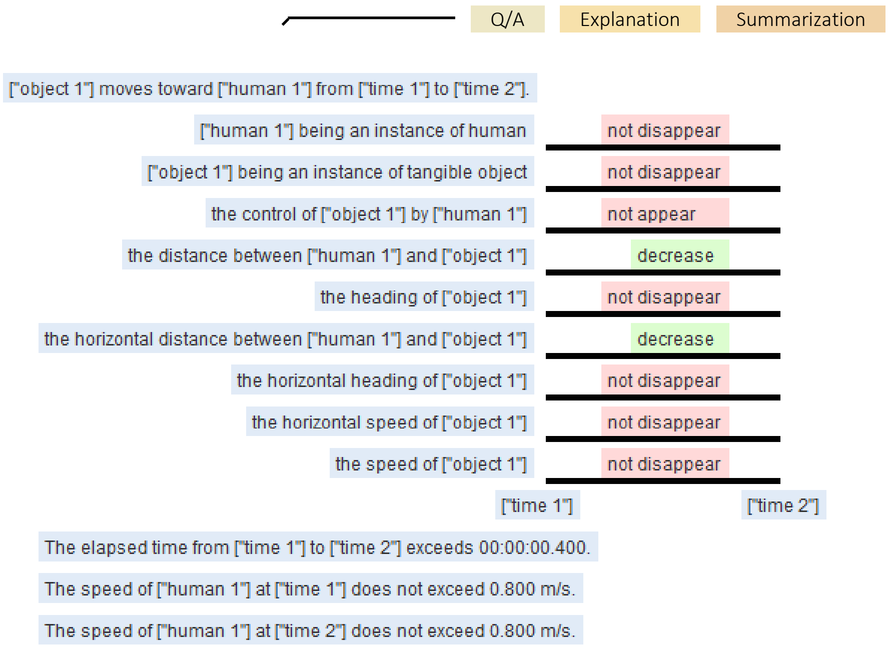


Q/A Explanation

["object 1"] moves out of contact with ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["object 1"] being an instance of tangible object	not disappear
the contact between ["object 1"] and ["hand 1"]	disappear
the heading of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear

["time 1"] ["time 2"]



Q/A Explanation Summarization

["object 1"] moves toward ["human 1"] from ["time 1"] to ["time 2"].

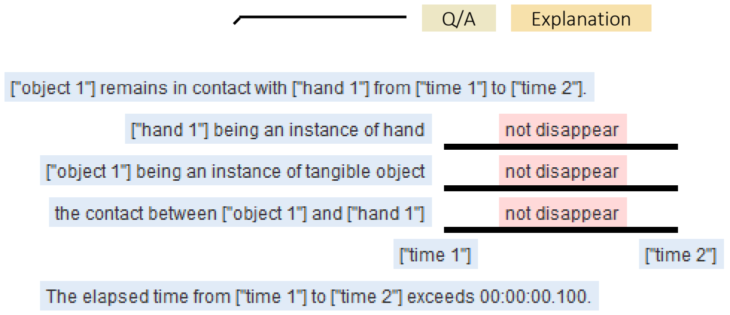
["human 1"] being an instance of human	not disappear
["object 1"] being an instance of tangible object	not disappear
the control of ["object 1"] by ["human 1"]	not appear
the distance between ["human 1"] and ["object 1"]	decrease
the heading of ["object 1"]	not disappear
the horizontal distance between ["human 1"] and ["object 1"]	decrease
the horizontal heading of ["object 1"]	not disappear
the horizontal speed of ["object 1"]	not disappear
the speed of ["object 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The speed of ["human 1"] at ["time 1"] does not exceed 0.800 m/s.

The speed of ["human 1"] at ["time 2"] does not exceed 0.800 m/s.



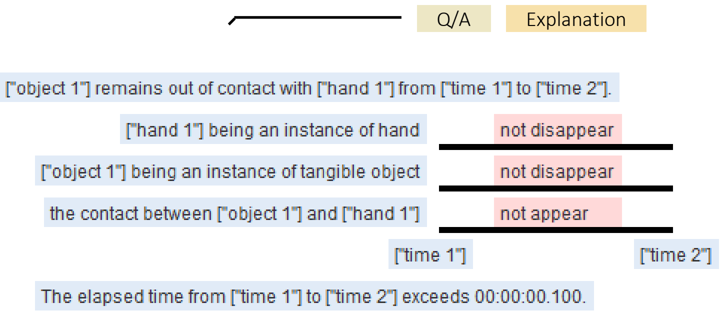
Q/A Explanation

["object 1"] remains in contact with ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["object 1"] being an instance of tangible object	not disappear
the contact between ["object 1"] and ["hand 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



["object 1"] remains out of contact with ["hand 1"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand — not disappear

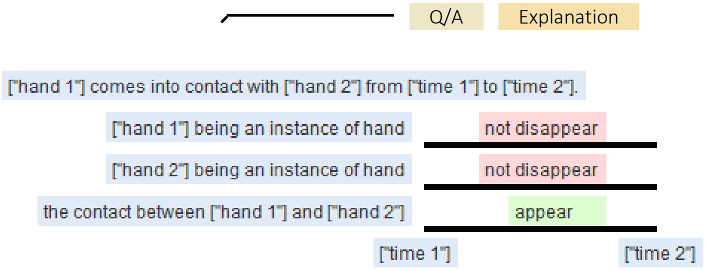
["object 1"] being an instance of tangible object — not disappear

the contact between ["object 1"] and ["hand 1"] — not appear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.

2 Humans, 0 Objects



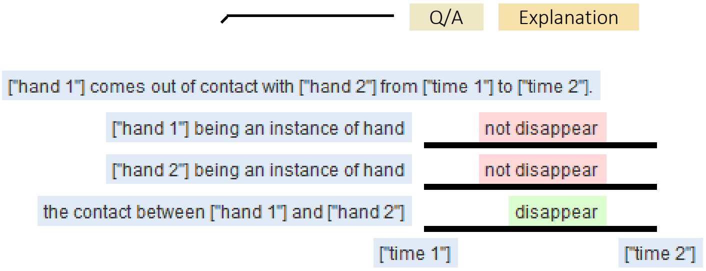
["hand 1"] comes into contact with ["hand 2"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand — not disappear

["hand 2"] being an instance of hand — not disappear

the contact between ["hand 1"] and ["hand 2"] — appear

["time 1"] ["time 2"]



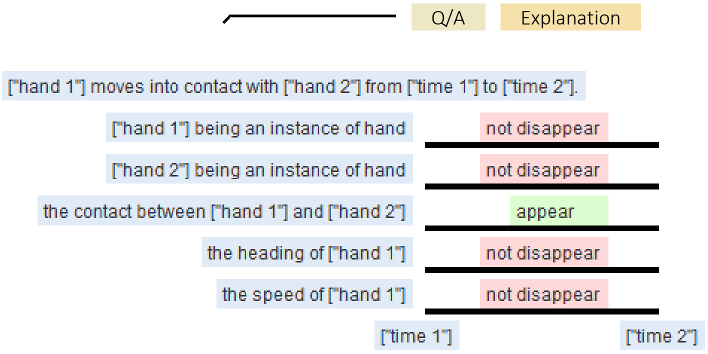
["hand 1"] comes out of contact with ["hand 2"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand — not disappear

["hand 2"] being an instance of hand — not disappear

the contact between ["hand 1"] and ["hand 2"] — disappear

["time 1"] ["time 2"]



["hand 1"] moves into contact with ["hand 2"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand — not disappear

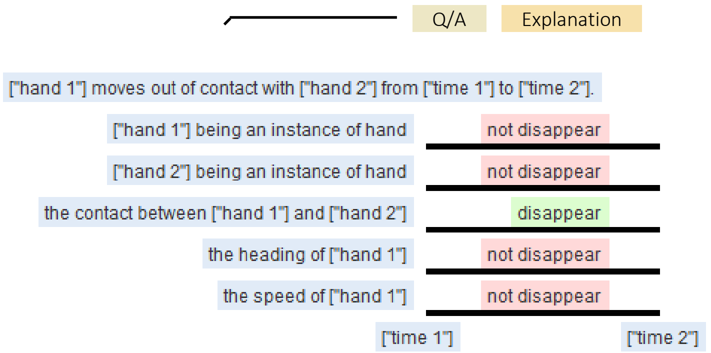
["hand 2"] being an instance of hand — not disappear

the contact between ["hand 1"] and ["hand 2"] — appear

the heading of ["hand 1"] — not disappear

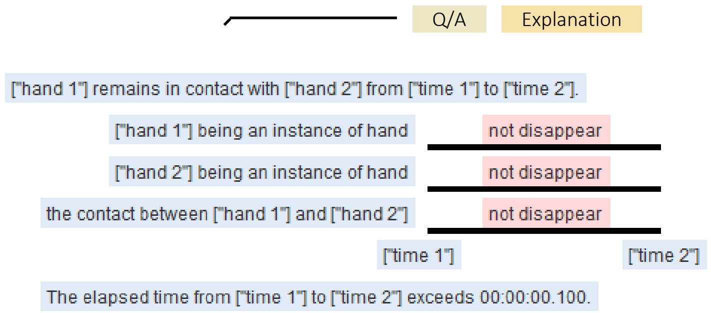
the speed of ["hand 1"] — not disappear

["time 1"] ["time 2"]



["hand 1"] moves out of contact with ["hand 2"] from ["time 1"] to ["time 2"].

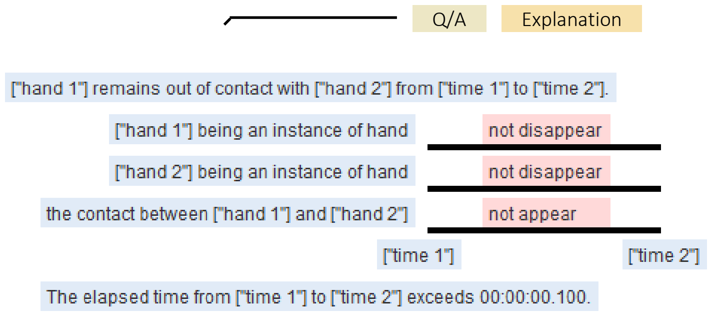
["hand 1"] being an instance of hand	not disappear
["hand 2"] being an instance of hand	not disappear
the contact between ["hand 1"] and ["hand 2"]	disappear
the heading of ["hand 1"]	not disappear
the speed of ["hand 1"]	not disappear
	["time 1"] ["time 2"]



["hand 1"] remains in contact with ["hand 2"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["hand 2"] being an instance of hand	not disappear
the contact between ["hand 1"] and ["hand 2"]	not disappear
	["time 1"] ["time 2"]

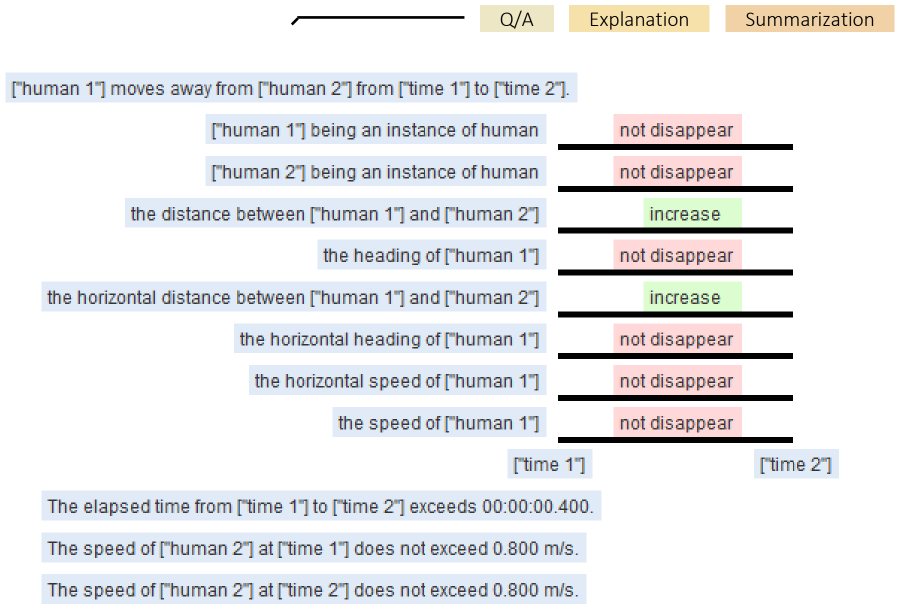
The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



["hand 1"] remains out of contact with ["hand 2"] from ["time 1"] to ["time 2"].

["hand 1"] being an instance of hand	not disappear
["hand 2"] being an instance of hand	not disappear
the contact between ["hand 1"] and ["hand 2"]	not appear
	["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



Q/A Explanation Summarization

["human 1"] moves away from ["human 2"] from ["time 1"] to ["time 2"].

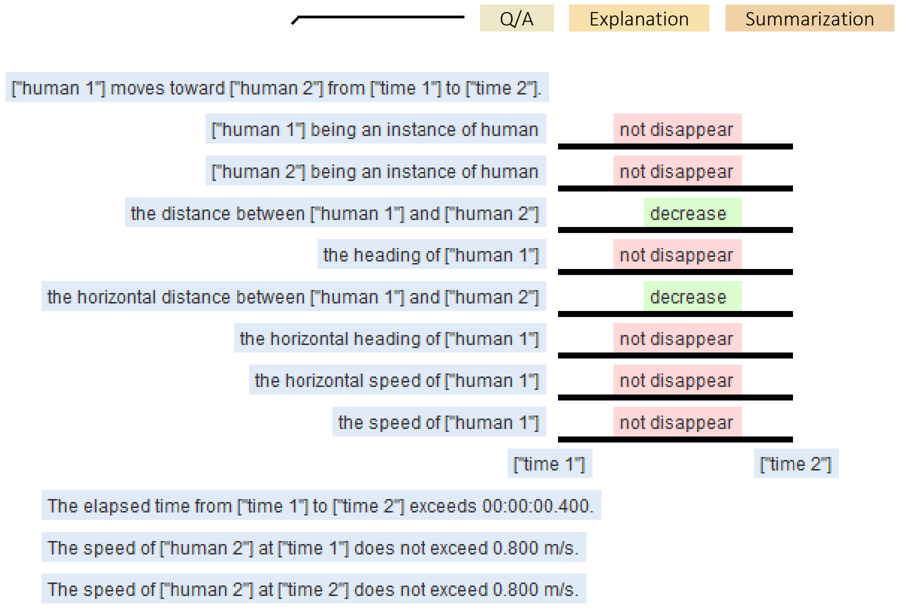
["human 1"] being an instance of human	not disappear
["human 2"] being an instance of human	not disappear
the distance between ["human 1"] and ["human 2"]	increase
the heading of ["human 1"]	not disappear
the horizontal distance between ["human 1"] and ["human 2"]	increase
the horizontal heading of ["human 1"]	not disappear
the horizontal speed of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear

["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The speed of ["human 2"] at ["time 1"] does not exceed 0.800 m/s.

The speed of ["human 2"] at ["time 2"] does not exceed 0.800 m/s.



Q/A Explanation Summarization

["human 1"] moves toward ["human 2"] from ["time 1"] to ["time 2"].

["human 1"] being an instance of human	not disappear
["human 2"] being an instance of human	not disappear
the distance between ["human 1"] and ["human 2"]	decrease
the heading of ["human 1"]	not disappear
the horizontal distance between ["human 1"] and ["human 2"]	decrease
the horizontal heading of ["human 1"]	not disappear
the horizontal speed of ["human 1"]	not disappear
the speed of ["human 1"]	not disappear

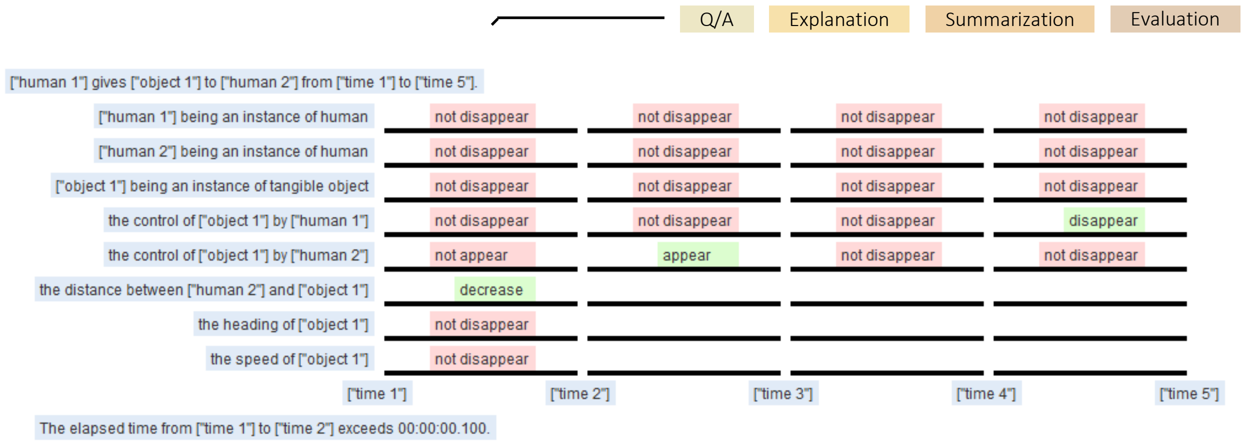
["time 1"] ["time 2"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.400.

The speed of ["human 2"] at ["time 1"] does not exceed 0.800 m/s.

The speed of ["human 2"] at ["time 2"] does not exceed 0.800 m/s.

2 Humans, 1 Object

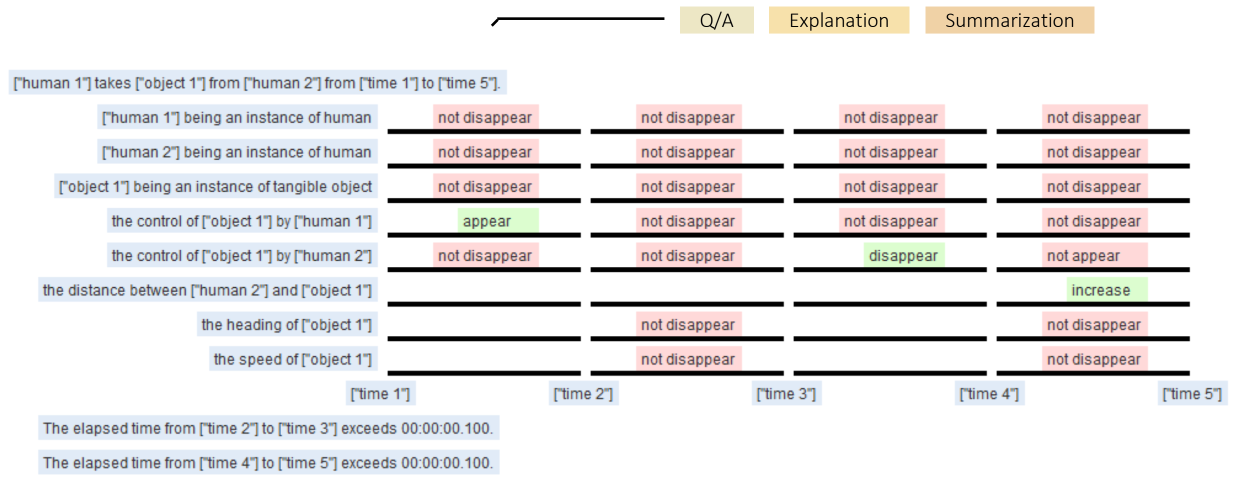


["human 1"] gives ["object 1"] to ["human 2"] from ["time 1"] to ["time 5"].

["human 1"] being an instance of human	not disappear	not disappear	not disappear	not disappear
["human 2"] being an instance of human	not disappear	not disappear	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	not disappear	not disappear	not disappear	disappear
the control of ["object 1"] by ["human 2"]	not appear	appear	not disappear	not disappear
the distance between ["human 2"] and ["object 1"]	decrease			
the heading of ["object 1"]	not disappear			
the speed of ["object 1"]	not disappear			

["time 1"] ["time 2"] ["time 3"] ["time 4"] ["time 5"]

The elapsed time from ["time 1"] to ["time 2"] exceeds 00:00:00.100.



["human 1"] takes ["object 1"] from ["human 2"] from ["time 1"] to ["time 5"].

["human 1"] being an instance of human	not disappear	not disappear	not disappear	not disappear
["human 2"] being an instance of human	not disappear	not disappear	not disappear	not disappear
["object 1"] being an instance of tangible object	not disappear	not disappear	not disappear	not disappear
the control of ["object 1"] by ["human 1"]	appear	not disappear	not disappear	not disappear
the control of ["object 1"] by ["human 2"]	not disappear	not disappear	disappear	not appear
the distance between ["human 2"] and ["object 1"]				increase
the heading of ["object 1"]		not disappear		not disappear
the speed of ["object 1"]		not disappear		not disappear

["time 1"] ["time 2"] ["time 3"] ["time 4"] ["time 5"]

The elapsed time from ["time 2"] to ["time 3"] exceeds 00:00:00.100.

The elapsed time from ["time 4"] to ["time 5"] exceeds 00:00:00.100.

References

Bendix, E. H. (1966). *Componential Analysis of General Vocabulary: The Semantic Structure of a Set of Verbs in English, Hindi and Japanese*, Mouton.

Borchardt, G. C. (1992). “Understanding Causal Descriptions of Physical Systems.” In *Proceedings of the AAAI Tenth National Conference on Artificial Intelligence*, 2–8.

Borchardt, G. C. (1994). *Thinking between the Lines: Computers and the Comprehension of Causal Descriptions*, MIT Press.

Borchardt, G. C. (2014). *Möbius Language Reference, Version 1.2*, Technical Report MIT-CSAIL-TR-2014-005, MIT Computer Science and Artificial Intelligence Laboratory.

Borchardt, G., Katz, B., Nguyen, H.-L., Felshin, S., Senne, K. and Wang, A. (2014). *An Analyst's Assistant for the Interpretation of Vehicle Track Data*, Technical Report MIT-CSAIL-TR-2014-022, MIT Computer Science and Artificial Intelligence Laboratory.

Katz, B. (1990). "Using English for Indexing and Retrieving." In *Artificial Intelligence at MIT: Expanding Frontiers, Vol. 1*, Cambridge, Massachusetts, 134–165.

Katz, B. (1997). "Annotating the World Wide Web Using Natural Language." In *Proceedings of the 5th RIAO Conference on Computer Assisted Information Searching on the Internet (RIAO '97)*, Montreal, Canada, 136–155.

Miller, G. A. and Johnson-Laird, P. N. (1976). *Language and Perception*, Harvard University Press.

Riggall, A. C. and Postle, B. R. (2012). "The Relationship between Working Memory Storage and Elevated Activity as Measured with Functional Magnetic Resonance Imaging," *Journal of Neuroscience*, 32(38), 12990-12998.

Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A. and Blake, A. (2012). "Efficient Human Pose Estimation from Single Depth Images," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 35, No. 12, 2821-2840.

Song, S. and Xiao, J. (2014). "Sliding Shapes for 3D Object Detection in Depth Images". In Proceedings of the European Conference on Computer Vision (ECCV 2014), 634-651.