Artificial Intelligence Review **21:** 293–316, 2004.

Interval Relations in Lexical Semantics of Verbs

MINHUA MA & PAUL MC KEVITT
School of Computing & Intelligent Systems, Faculty of Engineering, University of Ulster, Derry/Londonderry, Northern Ireland, BT48 7JL (E-mail: {m.ma, p.mckevitt}@ulster.ac.uk)

Abstract. Numerous temporal relations of verbal actions have been analysed in terms of various grammatical means of expressing verbal temporalisation such as tense, aspect, duration and iteration. Here the temporal relations within verb semantics, particularly ordered pairs of verb entailment, are studied using Allen's interval-based temporal formalism. Their application to the compositional visual definitions in our intelligent storytelling system, CONFUCIUS, is presented, including the representation of procedural events, achievement events and lexical causatives. In applying these methods we consider both language modalities and visual modalities since CONFUCIUS is a multimodal system.

Keywords: CONFUCIUS, knowledge representation, language visualisation, natural language understanding, temporal relations, verb semantics, visual semantics

1. Introduction

There are two main kinds of temporal reasoning formalism in artificial intelligence systems: point-based formalisms to encode relations between time points (moments), and interval based temporal calculus to encode qualitative relations between time intervals (Allen, 1983). Point-based linear formalisms are suitable for representing moments, durations, and other quantitative information, whilst interval-based temporal logic is useful for treating actual intervals and expresses qualitative information, i.e. relations between intervals. In the interval temporal logic, temporal intervals can always be subdivided into subintervals, with the exception of moments which are non-zero length intervals without internal structure. Allen argues that 'the formal notion of a time point, which would not be decomposable, is not useful' (Allen, 1983, p. 834) and the difference between interval-based and point-based temporal structures is motivated by different sources for intuitions: interval logic is meant to model time as used in natural language, whereas point-based formalisms are used in classical physics.

A common problem in the tasks of both visual recognition (image processing and computer vision) and language visualisation (text-to-graphics) is to represent visual semantics of actions and events, which happen in both space and *time continuum*. We use an interval-based formalism in the compositional predicate-argument representation discussed here to represent temporal relationships in visual semantics of eventive verbs. Our choice of temporal structure is motivated by our desire to analyze composition of actions/events and temporal relations between ordered pairs of verb entailment based on visual semantics. Since states and events are two general types of verbs and events often occur over some time interval and involve internal causal structure (i.e. change of state), it is convenient to integrate a notion of event with the interval logic's temporal structure, and event occurrences coinciding, overlapping, or preceding another, may easily be represented in interval temporal logic.

First, we begin with background to this work, the intelligent multimodal storytelling system, CONFUCIUS, and review previous work on temporal relations in story-based systems and natural language processing (Section 2). Then we investigate various temporal interrelations between ordered pairs of verb entailment using an interval-based formalism in Section 3. We turn next in Section 4 to discuss some attributes of interval relations and revise the conventions to indicate directions for causal relationship and backward presupposition.

Next in Section 5 we apply this method in our visual definitions of verbs in CONFUCIUS and discuss its applications in different circumstances such as procedural events, achievement events and lexical causatives. Following this, relations of our method to other work are considered (Section 6), and finally Section 7 concludes with a discussion of possible future work on adding quantitative elements to compositional visual representation.

2. Background and Previous Work

Our long-term objective is to create an intelligent multimedia storytelling platform called Seanchaí. Seanchaí consists of *Homer*, a storytelling generation module, and *CONFUCIUS*, a storytelling interpretation and multimodal presentation module (Figure 1).

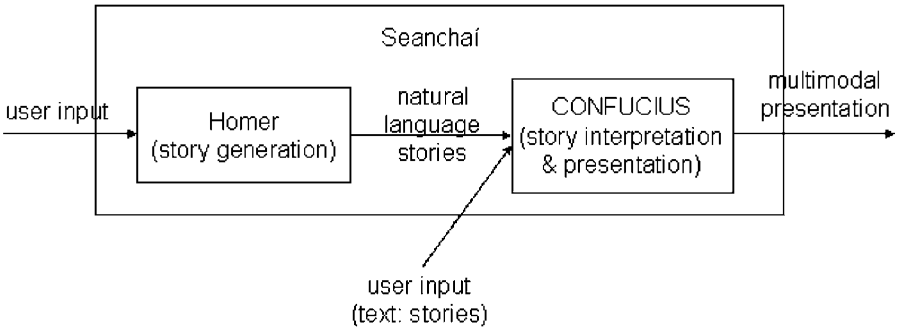

Figure 1. Seanchaí: an Intelligent MultiMedia storyteller.

2.1. *CONFUCIUS*

CONFUCIUS focuses on story interpretation and multimodal presentation and automatically generates multimedia presentations from natural language input. In the architecture shown in Figure 2, the boxed left part includes the graphic library such as characters, props, and animations for basic activities, which is used in the *Animation engine* (see Ma and Mc Kevitt, 2003b). The input sentences are processed by the *surface transformer* for transformations such as indirect to direct

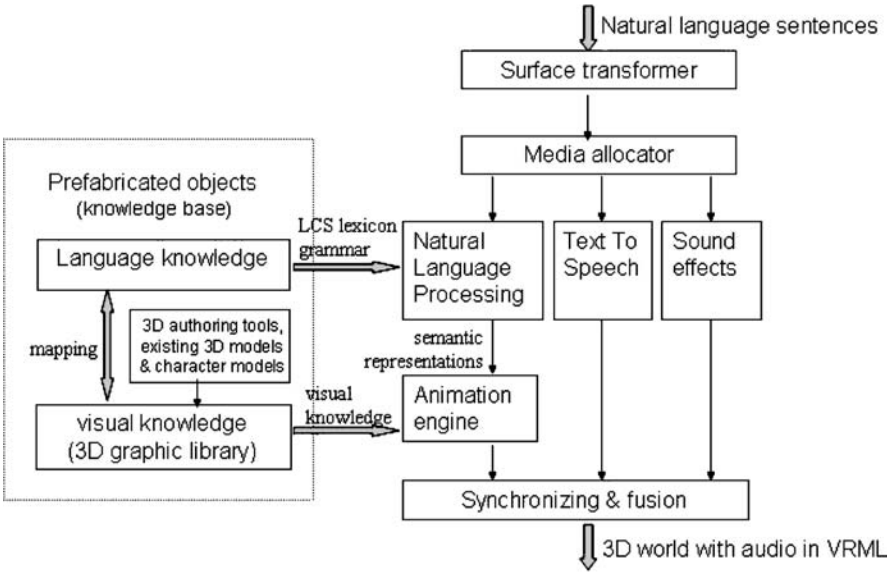

Figure 2. Architecture of CONFUCIUS.

quotation and passive to active voice. Then the *media allocator* allocates the contents to *natural language processing* (NLP), *Text to Speech* (TTS) and *sound effects* modules respectively. The three modules of NLP, TTS and sound effects operate in parallel. Their outputs converge at the *Synchronizing & fusion* module, which generates a holistic 3D world representation in VRML. It employs temporal media such as 3D animation and speech to present short stories. Establishing correspondence between language and animation is the focus of this research. This requires adequate representation and reasoning about the dynamic aspects of the story world, especially about events, i.e. temporal semantic representation of verbs.

Any multimodal presentation system like CONFUCIUS needs a multimodal semantic representation to allocate, plan, and generate presentations. Figure 3 illustrates the multimodal semantic representation of CONFUCIUS. Between the multimodal semantics and each specific modality there are two levels of representation: one is a high-level multimodal semantic representation which is *media-independent*, the other is a *media-dependent* representation which bridges the gap between general multimodal semantic representation and specific media realization and is capable of connecting meanings across modalities, especially between language and visual modalities. CONFUCIUS uses a compositional predicate argument representation (Ma and Mc Kevitt, 2003a) to connect language with visual modalities shown in Figure 3. The interval-based temporal logic we discuss here is applied to the compositional visual representation which is further discussed in Section 5. This method is suited for representing temporal relations and hence helping to create 3D dynamic virtual reality in language visualisation.

Figure 4 shows the knowledge base of CONFUCIUS that encompasses *language knowledge* for the natural language processor to extract

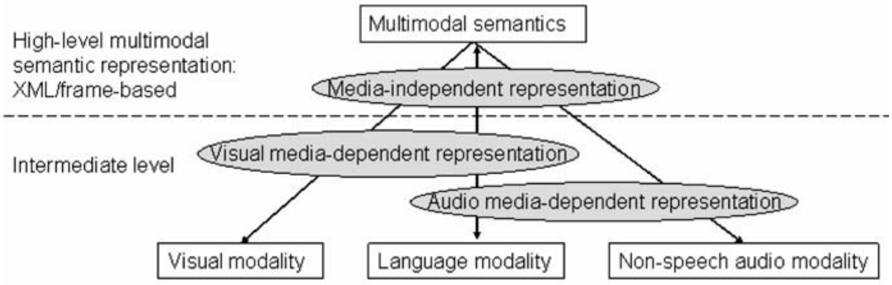

Figure 3. Multimodal semantic representation of CONFUCIUS.

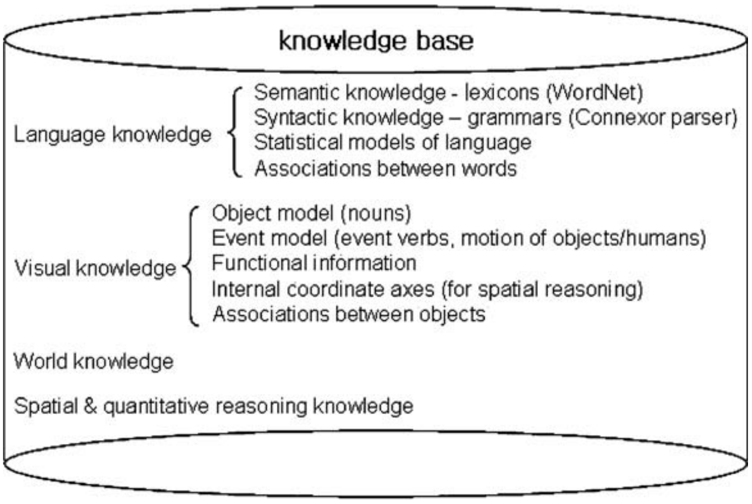

Figure 4. Knowledge base of CONFUCIUS.

semantic structures from text, *visual*, *world*, and *spatial reasoning knowledge*. We use the WordNet (Fellbaum, 1998) and LCS database (Dorr and Jones, 1999) resources in our language knowledge. Visual knowledge consists of the information required to generate 3D animation. It consists of *object model*, *functional information*, *event model*, *internal coordinate axes*, and *associations between objects*. The *Object model* includes visual representation of the ontological category (or conceptual 'parts of speech') – things (nouns), which consists of simple geometry files for props and places, and H-Anim files for human character models, which are defined in geometry & joint hierarchy files following the H-Anim specification (H-Anim, 2001).

The *event model* consists of visual representations of events (verbs) that contain explicit knowledge about the decomposition of high level acts into basic motions, and defines a set of basic animations such as *walk*, *jump*, *give*, *push* by determining key frames of corresponding rotations and movements of human joints and body parts involved. The current prototype of CONFUCIUS has 17 basic event models (verbs) which are able to visualise not only these basic actions but also their synonyms, hypernyms, troponyms, coordinate terms, and a group of verbs in corresponding verb classes (Levin, 1993). Additionally, the visual knowledge is capable of being expanded by appending more event models, object models and their functional and spatial information. The *internal coordinate axes* are indispensable in some primitive actions of *event models* such as rotating operations, which require spatial reasoning based on the object's internal axes. The *event model* in visual knowledge requires access to other parts of visual knowledge. For instance, in the event 'he cut the cake', the verb 'cut' concerns kinematical knowledge of the subject – a person, i.e. the movement of his hand, wrist, and forearm, and hence it needs access to the *object model* of a man who performs the action 'cut', and it also needs *function information* of 'knife', the *internal coordinate axes* information of 'knife' and 'cake' to decide the direction of the 'cut' movement. Here we focus on an efficient temporal representation for *event models* in this knowledge base, exploring how to apply interval relations in modeling the temporal interrelation between the subactivities in an event.

Action verbs are a major part of events involving humanoid performers (actor/experiencer) in animation. Action verbs can be classified into four types (see Figure 5): movement or partial movement, lexical causatives, verbs without distinct visualisation when out of context, and high level behaviours. In the movement verb group, there is an important class involving multimodal presentation – communication verbs. These verbs require both visual presentation such as lip movement (e.g., 'speak', 'sing'), facial expressions (e.g., 'laugh', 'weep') and audio presentation such as speech or other communicable sounds. Here we focus on the lexical causatives and high level behaviours (the two

2.2.1. Action verbs
  2.2.1.1. Movement or partial movement
    2.2.1.1.1. Biped kinematics, e.g. go, walk, jump, swim, climb
    2.2.1.1.2. Face expressions, e.g. laugh, angry } involve speech modality
    2.2.1.1.3. Lip movement, e.g. speak, say, sing, tell
  2.2.1.2. Lexical causatives
    2.2.1.2.1. Concerning single object, e.g. push, kick, bring, open
    2.2.1.2.2. Concerning multiple objects
      2.2.1.2.2.1. Bitransitive verbs, e.g. give, sell, show
      2.2.1.2.2.2. Transitive verbs with object & implicit instrument/goal/theme,
        e.g. cut, write, butter, pocket
  2.2.1.3. Verbs without distinct visualisation when out of context
    2.2.1.3.1. trying verbs: try, attempt, succeed, manage
    2.2.1.3.2. helping verbs: help, assist
    2.2.1.3.3. letting verbs: allow, let, permit
    2.2.1.3.4. create/destroy verbs: build, create, assemble, construct, break, destroy
    2.2.1.3.5. verbs whose visualisation depends on their objects,
      e.g. play (harmonica/football), make (the bed/trouble/a phone call), fix (a drink/a lock)
  2.2.1.4. High level behaviours (routine events)
    e.g. interview, eat out (go to restaurant), call (make a telephone call), go shopping

Figure 5. Categories of action verb.

boxed parts in Figure 5) since their internal structure can be specified in terms of interval temporal logic.

The usual way of representing a verb is in terms of the participants that it requires in the event. One standard method of characterising the participants in an event is by virtue of general event roles, referred to as theta roles, such as agent, patient, experiencer, source, goal, instrument, and theme. Jackendoff (1990) suggested Lexical-Conceptual Structure (LCS) in which theta roles are split into two groups: *thematic tier* and *action tier*. Thematic tier roles deal with physical disposition and change of (usually locational) state, e.g., source, theme, goal, location; action tier roles characterise the way the object is involved, e.g., agent, patient, beneficiary. He identified each role to a particular argument position in a conceptual relation of EVENT. Badler et al. (1997)'s Parameterized Action Representations (PARs) also provide a parameterized representation for actions in the technical order domain. The parameters involved are agent, objects, applicability conditions, culmination conditions, spatiotemporal, manner, and subactions.

Having reviewed Theta roles, Jackendoff's LCS EVENT parameters, and Badler et al.'s PARs, we specified the following parameters to represent an event in our knowledge base: *agent/experiencer*, *theme* (*object*), *spatiotemporal*, *manner*, *instrument*, *preconditions*, *subactivities*, and *result state* (see Figure 6). Preconditions are conditions that must exist before the action can be performed, e.g., test for reachability (the agent should be able to reach the light-fixture in the action of changing a light-bulb). Spatiotemporal information may use LCS-like PATH/PLACE predicates. We investigate 62 common English prepositions and define 7 PATH predicates and 11 PLACE predicates to interpret spatial relations and movement of objects and characters in 3D virtual worlds[1].

```
[EVENT
    agent:
    theme:
    space/time:
    manner:
    instrument:
    precondition:
    subactivities:
    result:
]
```

Figure 6. Parameters of EVENT.

The interval logic we propose is used in the *subactivities* parameter to represent the temporal relationship between subactivities.

The natural language processing component of CONFUCIUS (Figure 7) consists of syntactic parsing and semantic analysis. We use the Connexor Functional Dependency Grammar (FDG) parser (Järvinen and Tapanainen, 1997) for part-of-speech tagging, syntactic parsing and morphological parsing. For semantic analysis, we use WordNet (Fellbaum, 1998) and the LCS database (Dorr and Jones, 1999) to perform semantic inference, disambiguation, coreference resolution, and temporal reasoning. On temporal reasoning, we posit a distinction between lexical temporal relations and post-lexical temporal relations. Post-lexical level temporal analysis concerns phrase, sentence, and even discourse levels. Lexical temporal relations are within verb semantics, including temporal relations between pairs of verb entailment, complex actions and subactions, lexical causatives, and achievement and accomplishment events. Classical models of temporal reasoning study post-lexical temporal relations in terms of various grammar means of expressing verbal temporalisation such as tense, aspect, duration and iteration. However, lexical and post-lexical temporal relations influence each other. Post-lexical temporal relations may affect lexical processing by frequency, repetition, association, and orthographic legality, and lexical temporal relations can influence post-lexical processing by pre-

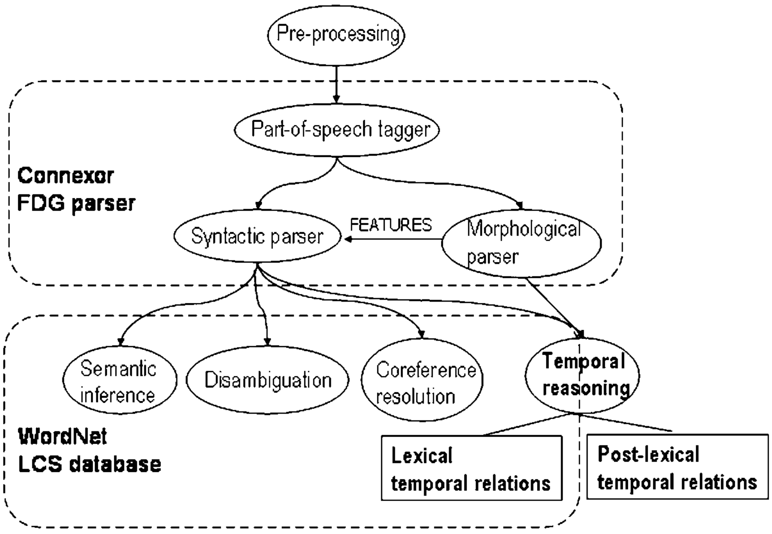

Figure 7. Natural language processing in CONFUCIUS.

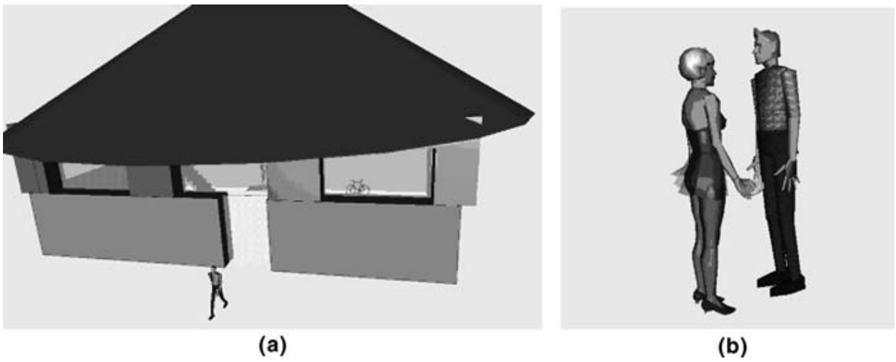

Figure 8. Snapshots of CONFUCIUS' 3D animation output. (a) Key frame of *John left the gym*. (b) Key frame of *Nancy gave John a loaf of bread*

dictability. The current prototype of CONFUCIUS visualises single sentences which contain action verbs with *visual valency* (Ma and Mc Kevitt, 2004) of up to three, e.g., (1) *John left the gym*, (2) *John gave Nancy a book*. Figure 8 shows example key frames of the 3D animation output of single sentences.

2.2. *Previous work on temporal relations*

Here we introduce Allen's (1983) thirteen basic interval relations (Table 1), which will be used in visual semantic representation of verbs in CONFUCIUS' language visualisation. Allen's interval relations have been employed in story-based interactive systems (Pinhanez et al., 1997) to express progression of time in virtual characters and handling linear/parallel events in story scripts and user interactions. In their stories, interval logic is used to describe the relationships between the time intervals which command actuators or gather information from sensors, which in turn decides the storyline. There are three types of interaction pattern in their interactive systems: linear, reactive, and tree-like. In reactive patterns, a story unfolds as a result of the firing of behaviors as a response to users' actions; in tree-like patterns, the user chooses between different paths in the story through some selective action. Linear, reactive, and tree-like interaction patterns can be modeled with interval logic.

On sentence level (or post-lexical level) temporal analysis within natural language understanding, there are extensive discussions on tense, aspect, duration and iteration, involving *event time*, *speech time*, and *reference time* (Reichenbach, 1947). To represent the relations

Table 1. Allen's thirteen interval relations ('e' denotes 'end point', 's' denotes 'start point')

Basic relations		Example	Endpoints
Precede	x p y	xxxx	$x_e < y_s$
Inverse precede	y p^{-1}x	yyyy	
Meet	x m y	xxxxx	$x_e = y_s$
Inverse meet	y m^{-1}x	yyyyy	
Overlap	x o y	xxxxx	$x_s < y_s < x_e \cap$
Inverse overlap	y o^{-1}x	yyyyy	$x_e < y_e$
During	x d y	xxxx	$x_s > y_s \cap$
Inverse during (include)	y d^{-1} x	yyyyyyyyy	$x_e < y_e$
Start	x s y	xxxxx	$x_s = y_s \cap$
Inverse start	y s^{-1} x	yyyyyyyyy	$x_e < y_e$
Finish	x f y	xxx	$x_e = y_e \cap$
Inverse finish	y f^{-1} x	yyyyyyyyy	$x_s > y_s$
Equal	x ≡ y	xxxxx	$x_s = y_s \cap$
	y ≡ x	yyyyy	$x_e = y_e$

among them, some use point-based metric formalisms (e.g., van Benthem, 1983), some use interval-based logic (e.g., Halpern and Shoham, 1991), others integrate interval-based and point-based temporal logic (Kautz and Ladkin, 1991) because of the complexity of temporal relations in various situations, for example, the distinction between punctual events and protracted events, achievements and accomplishments (Vendler, 1967; Smith, 1991), stative verbs and eventive verbs, states, events and activities (Allen and Ferguson, 1994). However, few of these are concerned with the temporal relations at the lexical level, e.g., between or within verbs. In lexical semantics, extensive studies have been conducted on the semantic relationships of verbs (Fellbaum, 1998), but few temporal relations have been considered. The closest work to that presented here was developed by Badler et al. (1997). They generalized five possible temporal relationships between two actions in the technical orders (instruction manuals) domain. The five temporal constraints are *sequential*, *parallel*, *jointly parallel* (the actions are performed in parallel and no other actions are performed until after both have finished), *independently parallel* (the actions are performed in parallel but once one of the actions is finished, the other one is stopped), and *while parallel* (the subordinate action is performed while the dominant action is performed; once the dominant action finishes, the subordinate action is stopped). In the following sections we investigate temporal relations at the lexical level since this work will facilitate our compositional visual definitions of verbs in language visualisation.

3. Temporal Relations in Verb Entailments

In this section various temporal relations between ordered pairs of verbs in which one entails the other are studied and their usage in visualisation is discussed. Verb entailment is a fixed truth relation between verbs where entailment is given by part of the lexical meaning, i.e. entailed meaning is in some sense contained in the entailing meaning. Verb entailment indicates an *implication* logic relationship: 'if x then y' ($x \Rightarrow y$). Take the two pairs *snore-sleep* and *buy-pay* as example, we can infer snore ⇒ sleep and buy ⇒ pay since when one is snoring (s)he must be sleeping, and if somebody wants to buy something (s)he must pay for it, whilst we cannot infer in the reverse direction because one may not snore when (s)he is sleeping, and one might pay for nothing (not buying, such as donation). In these two examples, the entailing activity could temporally *include* (i.e. d^{-1}) or *be included in* (i.e. d) the entailed activity. Fellbaum (1998) classifies verb entailment relations into four kinds, based on temporal inclusion, backward presupposition (e.g., the activity *hit/miss* supposes the activity *aim* occurring in a previous time interval) and causal structure (Figure 9).

Troponymy is one important semantic relation in verb entailment (Fellbaum, 1998) which typically holds between manner elaboration verbs and their corresponding base verbs, i.e. two verbs have the troponym relation if one verb elaborates the manner of another (base)

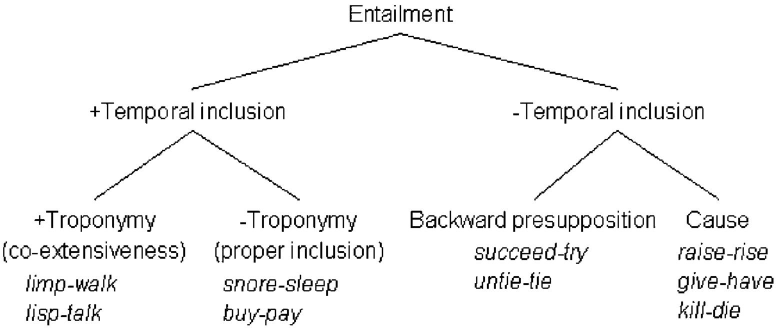

Figure 9. Fellbaum's classification of verb entailment.

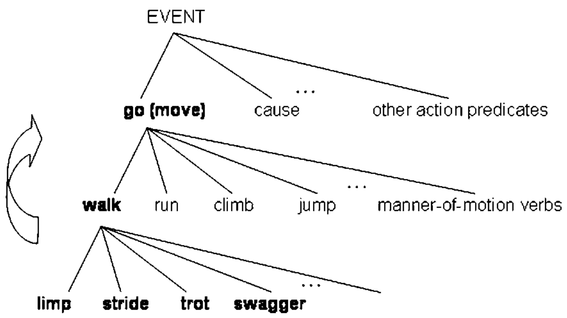

Figure 10. Troponymy tree.

verb. For instance, *mumble-talk* indistinctly, *trot-walk* fast, *stroll-walk* leisurely, *stumble-walk* unsteadily, *gulp-eat* quickly, the relation between *mumble* and *talk*, *trot/stroll/stumble* and *walk*, *gulp* and *eat* is troponymy. Figure 10 shows a tree of troponyms, where children nodes are troponyms of their parent node (e.g., the bolded route *limp/stride/trot-walk-go*). In CONFUCIUS, we use the method of `base verb` + `adverb` to present manner elaboration verbs, that is, to present the base verb first and then, to modify the manner (speed, the agent's state, duration, and iteration) of the activity. To visually present 'trot', we create a loop of walking movement, and then modify a cycle interval to a smaller value to present fast walking.

In Table 2 we analyze the possible temporal relations between these verb entailments and give some examples. We note that the interval relation between a troponym pair of verbs is $\{\equiv\}$, e.g., limp $\equiv$ walk. The relation set of $\{p,m,o,s,f^{-1},\equiv\}$ may hold in any pair with causal structure (i.e. lexical causatives), between the eventive verb and its result state (either stative verb or adjective), such as *give-have*, *eat-full*, *work-getPaid*, *heat-hot*. Thanks to the productive morphological rules in

Table 2. Temporal relations in verb entailments

Verb entailment relations	Temporal relations	Examples
Troponym	$\{\equiv\}$	limp $\equiv$ walk
Non-troponym (proper temporal inclusion)	$\{d,d^{-1}\}$	snore d sleep, buy d^{-1} pay
Backward presupposition	$\{p^{-1},m^{-1}\}$	untie p^{-1} tie $\cup$ untie m^{-1} tie
Cause	$\{p, m, o, s, f^{-1}, \equiv\}$	eat p fullUp $\cup$ eat o fullUp, give m have, build o exist

English deriving verbs from adjectives via affixes such as – en and –ify, these deadjectival verbs, e.g., 'whiten', 'shorten', 'strengthen', 'soften', often refer to a change of state or property and have the meaning (*make/become/cause* +corresponding adjective or its comparative form). The temporal relation between the pair of deadjectival verbs and the state of their corresponding adjectives is also $\{p,m,o,s,f^{-1},\equiv\}$. For instance, the possible interval relations set between *shorten-short/shorter* could be $\texttt{shorten}\ \{\texttt{p},\texttt{m},\texttt{o},\texttt{s},\texttt{f}^{-1},\equiv\}\texttt{short/shorter}$. Similarly, the relation set $\{p,m,o,s,f^{-1}\}$ is also applicable to cognate verbs and adjectives (or their comparative forms) such as *beautify-beautiful* and *clarify-clear/clearer*.

Eventive verbs, which have internal causal structure and are distinguished from stative verbs on this basis according to Gennari and Poeppel (2002), are our main concern in language visualisation.

4. Propagation within Interval Logic

Allen's interval algebra is convenient for describing the propagation algorithm used for logical inference through a collection of intervals, determining the most constrained disjunction of relations for each pair of intervals which satisfies the given relations.

Reversibility and transitivity are important attributes in temporal reasoning. They provide an algorithm which propagates the temporal relations through a collection of intervals, determining the most constrained disjunction of relations for each pair of intervals which satisfies the given relations and is consistent in time. By adding directions to interval relations we may denote the implication logic relationship between two events.

4.1. *The algebra of temporal relations: reversibility and transitivity*

The reversibility of an interval relation is:
$\forall \texttt{R}: \ \texttt{act1 R act2}, \qquad \texttt{R} \in \{\texttt{p},\texttt{p}^{-1},\texttt{m},\texttt{m}^{-1},\texttt{o},\texttt{o}^{-1},\texttt{d},\texttt{d}^{-1},\texttt{s},\texttt{s}^{-1},\texttt{f},\texttt{f}^{-1}\} \Leftrightarrow \texttt{act2 R}^{-1}\ \texttt{act1}$ For instance, $\texttt{untie p}^{-1}\ \texttt{tie} \Leftrightarrow \texttt{tie p untie}$. All interval relations are reversible, and the relation $\equiv$ is reflexive, symmetric, and transitive.

Transitivity of one interval relation is defined as:
if $\exists \texttt{R}: (\texttt{act1 R act2}) \cap (\texttt{act2 R act3}), \qquad \texttt{R} \in \{\texttt{p},\texttt{p}^{-1},\texttt{o},\texttt{o}^{-1},\texttt{d},\texttt{d}^{-1},\texttt{s},\texttt{s}^{-1},\texttt{f},\texttt{f}^{-1},\equiv\} \Rightarrow \texttt{act1 R act3}$

then this temporal relation R is transitive, to wit, the temporal relations between the pairs of intervals can be propagated through the collection of all intervals. For instance, $\texttt{born p age}, \texttt{age p die} \Rightarrow \texttt{born p die}$. Notice that m and m^{-1} are not in the set of possible transitive relations, because the nature of these two relations is not transitive, i.e. $(\texttt{act1 m act2}) \cap (\texttt{act2 m act3}) \Rightarrow\sim (\texttt{act1 m act3})$. All the other temporal relations $\{\texttt{p}, \texttt{p}^{-1}, \texttt{d}, \texttt{d}^{-1}, \texttt{s}, \texttt{s}^{-1}, \texttt{f}, \texttt{f}^{-1}, \equiv\}$ must be transitive except o and o^{-1} since $\texttt{act1 o act3}$ cannot be inferred from $(\texttt{act1 o act2}) \cap (\texttt{act2 o act3})$, though it might be true.

The temporal reasoning of the interval relations can be obtained by computing the possible relations between any two time intervals. For instance, $(\texttt{x d}^{-1}\texttt{y}) \cap (\texttt{y p z}) \Rightarrow \texttt{x R z}, \texttt{R} \in \{\texttt{p}, \texttt{o}, \texttt{d}^{-1}, \texttt{f}^{-1}, \texttt{m}\}$. In this case, x could be the activity ‘buy’, y could be the activity ‘pay’ and z could be ‘consume’.

4.2. *Revised interval relation conventions*

Here we revise Allen’s interval logic by adding directions of *implication* logic relationships to it, using R>, <R, or <R>, R$\in$ $\{p, p^{-1}, m, m^{-1}, o, o^{-1}, d, d^{-1}, s, s^{-1}, f, f^{-1}, \equiv\}$. Hence, $\texttt{limp} \equiv> \texttt{walk}$ indicates their troponomy relation, and $<\equiv>$ indicates synonym relations like speak $<\equiv>$ `say`, or same activity from different perspectives such as `teach` $<\equiv>$ `learn`, `buy` $<\equiv>$ `sell`.

By this facility we may also use $\texttt{build o} > \texttt{exist}$ to indicate causal relationship (in prediction), and use $\texttt{tie} < \texttt{p untie}$ to indicate backward presupposition (in planning).

5. Application of Interval Representation

The interval temporal logic discussed above can be combined with truth conditions of perceptual primitives such as support, contact, and attachment to represent simple spatial motion events for recognizing motion verbs in animation or vision input, such as Siskind’s (1995) event logic. The drawback of Siskind’s event logic is that it is limited to a reduced set of actions, such as *drop*, *place*, and *pick up*. In this section we apply the interval logic to a compositional predicate-argument model of visual definition (Ma and Mc Kevitt, 2003a), which is a Prolog-style semantic representation, to represent the temporal relationship between subactivities. Prolog uses comma as logic AND and semicolon as logic OR. Their denotation has been altered to express sequential and parallel temporal relations between subactivities in Ma and Mc Kevitt (2003a).

The relationship between the definiendum verb and the defining subactivities is temporal inclusion (whether proper inclusion or not), i.e. $\texttt{act1 R act2}$, $\texttt{R} \in \{\texttt{d}, \texttt{s}, \texttt{f}, \equiv\}$, act1 is part of, or a stage in, temporal realization of act2, and hence it could be one sub-activity in act2's visual definition[2]. $\equiv$ is a special case. If there is only one subactivity in a definition and the relation of this subactivity and its defined verb is $\equiv$ or $\equiv>$, the definition is rather an interpretation than a semantic decomposition, e.g., in the definition `slide() : -move()`, the temporal relation between the subactivity and definiendum is $\texttt{slide} \equiv \texttt{move}$. Because 'slide' is a troponym of 'move', i.e. $\texttt{slide} \equiv> \texttt{move}$, we use 'move' to define 'slide' but not 'slide' to define 'move'. The relationship between any subactivity and the verb sense it defines are {d,s,f,≡}:

```
act() : -

subact1(),

......

subacti(),

......
```

$\texttt{subacti R act},\ \texttt{i} \in \texttt{N},\ \texttt{R} \in \{\texttt{d}, \texttt{s}, \texttt{f}, \equiv\}$

In the proposal of Ma and Mc Kevitt (2003a) there are two symbols indicating temporal relations between subactivities. The comma separating two sub-activities in sequential order, e.g., '`act01`, `act02`', means that `act02` follows `act01`. This temporal relation could subsume several relations $\{\texttt{p}, \texttt{m}, \texttt{o}, \texttt{f}^{-1}, \texttt{d}^{-1}\}$ in interval logic, i.e. all temporal relations in $\texttt{x R y}$ while $\texttt{x}_\texttt{s} < \texttt{y}_\texttt{s}$, though `p` and `m` are the most frequently occurring

```
call(a):-
  pickup(a,tel.rcver,a.l_ear),
  dial(a, tel.keypad),
  speak(a, tel.rcver),
  putdown(a,tel.rcver, tel.set).
```

a. Original visual definition of "call"

```
call(a):-
  pickup(a,tel.rcver,a.l_ear){p,m,o,f^-1,d^-1}
  dial(a, tel.keypad) {p}
  speak(a, tel.rcver) {p,m}
  putdown(a, tel.rcver, tel.set).
```

b. Visual definition of "call" using interval logic

```
roll(obj,rollAngle,newPos):-
      moveTo(obj, newPos);
      rotate(obj,0,0,rollAngle).
```

c. Original visual definition of "roll"

```
roll(obj, rollAngle, newPos):-
      moveTo(obj, newPos) ≡
      rotate(obj,0,0,rollAngle).
```

d. Visual definition of "roll" using interval logic

Figure 11. Visual definitions using interval logic.

Table 3. Temporal relations between subactivities

Original proposal	Interval relations
`act01, act02`	`act01` R `act02`, $\mathtt{R} \in \{\mathtt{p},\mathtt{m},\mathtt{o},\mathtt{f}^{-1},\mathtt{d}^{-1}\}$
`act01; act02`	`act01` $\equiv$ `act02`

relations denoted by comma. Figure 11 shows the visual definition of 'call' in Ma and Mc Kevitt (2003a) (Figure 11a) and the improved definition using interval logic (Figure 11b). In addition, semicolon is used to indicate the *equal* temporal relation $\equiv$ between two activities which occur simultaneously. '`act01; act02`' means that `act01` and `act02` start and finish at the same time. This temporal relation is usable for defining verbs such as rolling of a wheel (Figure 11c and d).

Table 3 compares the original proposal of compositional visual definitions with the improved version we propose herein, which is an interval temporal extension of Prolog conventions. Though the logic denotations of Prolog's comma and semicolon have been altered to express temporal relations in Ma and Mc Kevitt (2003a), they do not have enough capability to indicate different temporal relations between events. Note that there is no means to distinguish between the five relations denoted solely by comma in the original proposal. For example, in the definition of 'turn' in 'turn a vehicle' (Figure 12), the activity of `slowdown` can also *overlap/include/be finished by* `changeGear` besides *preceding* or *meeting* `changeGear`, i.e. `slowDown` $\{\mathtt{p},\mathtt{m},\mathtt{o},\mathtt{f}^{-1},\mathtt{d}^{-1}\}$ `changeGear`. But there is no way to indicate this by our original representation using ','. It is necessary to distinguish the relation between `slowDown` and `changeGear` with the relation between `steer` and `straight`, because the latter relation is just a simple *precede* or *meet* relation[3] $\{\mathtt{p},\mathtt{m}\}$ (Figure 12b) whilst the former relation could be any of $\{p,m,o,f^{-1},d^{-1}\}$. The original representation (Figure 12a) obviously cannot distinguish between them.

```
turn():-
    … slowDown() ,
      changeGear() ,
    … steer() ,
      straight().
```

a. Original representation of "turn"

```
turn():-
    … slowDown() {p,m,o,f^-1,d^-1}
      changeGear() {p,m}
    … steer() {p,m}
      straight().
```

b. Improved representation of "turn"

Figure 12. Original and improved temporal representations of 'turn'.

```
eatOut():-
  bookASeat() ,
  goToRestaurant() ,
  orderDishes() ,
  eat() ,
  pay() ,
  leave().
```

a. Original visual definition

```
eatOut():-
 bookASeat() {p}
 goToRestaurant(){p,m}
 orderDishes() {p}
 eat() {p,m}
 pay() {p,m}
 leave().
```

b. "eatOut" in a restaurant

```
eatOut():-
  bookASeat() {p}
  goToRestaurant(){p,m}
  orderDishes() {p}
  eat() {p,p^-1,m}
  pay() {p,m}
  leave().
```

c. "eatOut" in a restaurant/fast food shop

```
eatOut():-
 [bookASeat() {p}]
 goToRestaurant(){p,m}
 orderDishes() {p}
 eat() {p,p^-1,m}
 pay() {p,m}
 leave().
```

d. Optional subactivities in definition

Figure 13. Visual definitions of 'eatOut'.

Another advantage of replacing comma and semi-colon with interval relation symbols is that this method can define multiple temporal relationships in one definition. For instance, one may argue the `eatOut` definition in Figure 13b that in fast food shops people pay first and then get the food they order. Figure 13c includes this circumstance by adding p^{-1} in the relation set between `eat()` and `pay()`, as opposed to defining another event describing `eatOut` in fast food shops. The distinction between (b) and (c) shows the nonlinear advantage on efficiency and flexibility, which is similar to partial-order planning[4] vs. total-order planning.

Either in Schank's scripts (Schank and Abelson, 1977) or in decomposition visual definitions (Ma and Mc Kevitt, 2003a), there may be some subactivities which are optional in the script/definition. In the eatOut example, `bookASeat()` is optional. We use square brackets to indicate optional subactivities (Figure 13d).

5.1. *Punctual events*

There is a group of verbs indicating punctual events which never hold over overlapping intervals, or two intervals one of which is a subinterval of the other, such as 'find', 'arrive', 'die'. Vendler (1967) classified them as *achievement* events (distinct from *accomplishment* events), which occur at a single moment and involve unique and definite time instants. Dowty (1979) draws attention to a major difference between achieve-

```
die():-          find():-             arrive():-
  fall().          search(),            go(),
                   eyesFixedOn().       stop().
```

Figure 14. Examples of punctual events' visual definitions.

ments and accomplishments: accomplishment verbs are telic, describing activities that normally lead to a result. Smith (1991) similarly proposes that achievements are *instantaneous events* that result in *a change of state*. It seems that point-based relations are more appropriate for these verbs. However, pragmatic, ontological, and practical cases for interval relations have been advocated. Some pragmatic approaches (Verkuyl, 1993) deny the semantic distinction between accomplishments and achievements. They hold that the length of the event is not a linguistic matter. Jackendoff (1991) and Pinon (1997) introduce the concept of *boundaries* into a temporal ontology for aspectual semantics to analogise achievement events. Boundaries are *ontologically dependent* objects: they require the existence of that to which they are bound.

These considerations are in respect of language modalities. When multimodal representation is concerned, we take visual representation into account, so punctual events could also be represented using interval-based relations. As stated in Pinon's boundaries analogy the existence of achievement events depends on the existence of their corresponding accomplishments, and in visual representation we cannot separate these events from their context, e.g., to separate 'find' from 'search', and 'arrive' from 'go'. In computer games and dynamic visual arts like movies, for example, the event 'die' is usually associated with a 'falling' movement. When we include context in their visual definitions (Figure 14), these events become intervals rather than moments. Therefore we can declare that all verbs are in time intervals, whether they indicate states, processes, or punctual events. Strictly speaking, the relationships between these punctual events and the subactivities in their visual definitions cover all five possible relations between a point and an interval: *starts*, *before*, *during*, *finishes*, and *after* since these are also the relations between punctual events and their contexts.

5.2. *Temporal relations in lexical causatives*

Visual definition should also include causative information which helps solve the 'frame problem', i.e. to determine the result state following a particular action (the effects of actions). Hence the visual definitions of

```
kill(killer, victim, weapon):-
      hit(killer, victim, weapon),
      die(victim).
```

Figure 15. Causative information in visual definition.

causative verbs like ‘kill’ must subsume their result states (stative verbs) like ‘die’ (Figure 15).

Moreover, interval relations can represent the distinction between *launching* and *entraining* causation. In the following sentences, (1–4) describe causation of the *inception* of motion (launching causative), whereas (5) describes *continuous* causation of motion (entraining causative). A disjunction set of interval relations between the cause and the effect is adequate to define the difference: {p,m,o,s} for launching causative verbs (1-4), and $\{\equiv, f^{-1}\}$ for entraining causatives (5).

Examples	Temporal relation between cause-effect
1. John threw the ball into the field.	{s}
2. John released the bird from the cage.	{p}
3. John gave the book to me.	{m}
4. John opened the door.	{o}
5. John pushed the car down the road.	$\{\equiv, \mathtt{f}^{-1}\}$

5.3. *Representing actions consisting of repeatable periods*

Herein we introduce a facility to represent repeatable periods of subactivities since many actions may be sustained for a while and consist of a group of repeatable subactivities. We use square brackets and a subscript R to indicate the repetition constructs in examples of Figure 16, which can also be nicely captured by Kleene iteration in finite state descriptions for temporal semantics[5] (Fernando, 2003). The activities bracketed by $[\,]_R$ are repeatable. Besides periodical repetition of subactivities, it can represent morphological prefix ‘re-’ as well, as the recalculate example in Figure 16, substituting the number of iterations (which is 2 in this case) for R. This facility of representing iteration may

```
walk():-      hammer(aPerson, aNail):-          Recalculate():-
  [step()]R.    [hit(aPerson,aNail,hammer)]R.     [calculate()]2.
```

Figure 16. Verbs defined by repeatable subactivities.

be used for post-lexical level repetition as well, e.g., events marked by 'again', 'continues to', or 'a second time'.

Due to the advantages in representing temporal relations between entailed verb pairs, punctual events and their contexts, lexical causatives, and iteration of sub-activities, we adopt interval logic rather than point-based logic. Though some may argue that in some cases like $\mathtt{kill}\ \mathtt{f}^{-1}$ die, the quantitative factor is critical. Fernando (2003) introduces a temporal granularity δ which is a non-empty observation interval $\delta \in \mathtt{R}$ greater than 0. An event with end-points $\mathtt{p}, \mathtt{q}$:

$$(\mathtt{p}, \mathtt{q}) = \{\mathtt{r} \in \mathtt{R} \,|\, \mathtt{p} < \mathtt{r} < \mathtt{q}\}$$

$\mathtt{p}, \mathtt{q}$ are real numbers indicating temporal instants of the event's end-points. R denotes the set of real numbers. $\mathtt{p}, \mathtt{q} \in \mathtt{R}$ and $\mathtt{q} - \mathtt{p} > \delta$. The requirement $\mathtt{q} - \mathtt{p} > \delta$ bounds the precision of a δ-observation and ensures that δ covers part of the event. Next, given $\mathtt{O}$ for a set of events, let Succ be the binary relation on O defined by

$$(\mathtt{p}, \mathtt{q})\ \mathtt{Succ}(\mathtt{p}', \mathtt{q}') \quad \mathtt{iff}\ 0 \leq \mathtt{p}' - \mathtt{q} \leq \delta$$

for all $(\mathtt{p}, \mathtt{q}), (\mathtt{p}', \mathtt{q}') \in \mathtt{O}$. The requirement $\mathtt{p}' - \mathtt{q} \geq 0$ ensures that each point in $(\mathtt{p}, \mathtt{q})$ is less than each point in $(\mathtt{p}', \mathtt{q}')$; while $\mathtt{p}' - \mathtt{q} \leq \delta$ precludes a gap between $(\mathtt{p}, \mathtt{q})$ and $(\mathtt{p}', \mathtt{q}')$ large enough to squeeze in an intervening δ-observation, and hence ensures δ covering both events (at least part of each). Therefore, `Succ` is exactly Allen's relation {p} with an observation seeing the end point of the former event and the start point of the latter event.

Let's analyse the intrinsic causal structure of the event *kill*. There is a nuance in meaning between *kill* and *cause to become not alive* (Arnold et al., 1994) in virtue of the quantitative factor of temporal relation, in particular, where a *killing* is a single event, a *causing to become not alive* involves two events: a *causing* and a *dying*. If the causal chain that links a particular event to dying is long enough, one may admit that the event caused the dying, but not want to say there has been a 'killing'. For the situation where the causing event is a *shooting*, i.e., `shoot Succ die` constructs the event *kill*. $\mathtt{p}' - \mathtt{q} \leq \delta$ makes sure that the time difference between $\mathtt{p}'$ (the start point of *die*) and $\mathtt{q}$ (the end point of *shoot*) is less than or equal to the observation interval δ, i.e., only when the both events are observed the event *kill* is fulfilled, which requires the visualisation of *killing* to subsume both the cause and the *dying* events (or part of each, at least). How long should the state of *wounded but not dead* ($p' - q$) last that we can say that it is one event *kill* rather than two events *cause* and *die*? We can see the importance of the quantitative factor for language generation, especially in the stage of sentence planning and surface realization. However, it might not be critical for language understanding (or visualization), e.g., the interpretation of kill includes shoot $\{p, m, o, f^{-1}\}$ die, and if the relation is `shoot p die` (or `shoot Succ die`), we can give a reasonable value to the $p' - q$, say, several minutes, before the victim finally dies. The value could even be 0 or negative in the case of `shoot` $\{m, o, f^{-1}\}$ `die`, i.e., the killer kept shooting until (or after) the victim died.

6. Relation to Other Work

Previous temporal representation, analysis and reasoning in syntax (e.g., tense and aspect) and pragmatics is at sentence level, while research on lexical semantics takes few temporal relations into consideration. All temporal relation research within natural language processing is limited within the language modality itself and does not take other modalities such as vision into account. The work we present here brings interval temporal logic, which is significantly more expressive and more natural for representing events and actions, to visual semantics of verbs at the lexical level and uses this methodology to enhance our compositional predicate-argument visual definition of action verbs for dynamic language visualisation.

Pinhanez et al. (1997) use interval logic in storytelling, but they use it to describe the relationships between the time intervals of events and interactions, i.e., the storyline. Table 4 shows a comparison with Badler's (Badler et al., 1997) temporal constraints for actions in the technical orders (instruction manuals) domain. We can use interval logic to represent all the five constraints they put forward. Their constraints are compositional (e.g., jointly parallel deals with three actions), and all the constraints are disjunctions of several interval relations. They also consider other non-temporal factors such as dominancy of action (e.g., while parallel). We claim that interval relations are more flexible and suitable for general purposes since they are 'minimal' relations of time intervals. For domain-specific applications such as technical instructions, their specific temporal representation may work well.

Table 4. Comparison with Badler's temporal constraints

Badler's temporal constraints (technical orders domain)	Interval relations
Sequential	$\{p,m\}$
Parallel	$\{s,s^{-1},\equiv\}$
Jointly parallel	(act1 $\{s,s^{-1}, \equiv\}$ act2) $\{p,m\}$ act3
Independently parallel	$\{f,f^{-1}, \equiv\}$
While parallel	Act_dominant $\{s^{-1},f^{-1}, \equiv\}$ act_indominant

7. Conclusion and Further Work

Temporal relation is a crucial issue in modelling action verbs, their procedures, contexts, presupposed and result states. In this paper we have discussed temporal relations within verb semantics such as various temporal interrelations between ordered pairs of verb entailment, accomplishment and achievement events, and lexical causatives. We propose an enhanced compositional visual definition of verbs based on Allen's interval logic and apply it to CONFUCIUS' animation generation. One of the limitations of this temporal representation is lack of quantitative information, which is due to our adoption of the interval-based relations: (1) the durations of activities cannot be specified, though repetition of activities could be indicated by defining one repeatable period and specifying its repeat attribute; (2) for overlapping events $x \; \{o, o^{-1}\} \; y$, our temporal representation only works when the exact start point of y is unimportant; (3) for events $x \; \{p, p^{-1}\} \; y$, it is difficult to relate the distance between the two intervals, i.e. the distance between the end point of x and the start point of y in the case of $x \; p \; y$. Future versions of the compositional visual representation will introduce quantitative elements to overcome these limitations. Future research should also address the issue of action composition for simultaneous subactivities which might influence each other, such as *waving as walking* changing the movement of an arm.

Acknowledgements

We are grateful for useful ideas and constructive comments on an earlier version of this paper from Dr. Tim Fernando of the Computer Science Department, Trinity College Dublin and Dr. David McSherry of the

School of Computing & Information Engineering, University of Ulster. We also would like to thank the anonymous reviewers for their valuable suggestions.

Notes

[1] The 7 PATH predicates are *to*, *from*, *toward*, *away_from*, *via*, *across*, *along*, and the 11 PLACE predicates are *at*, *behind*, *end_of*, *in*, *in_front_of*, *near*, *on*, *out*, *over*, *top_of*, *under*.

[2] act2 is the definiendum verb, and act1 is one of its defining subactivities.

[3] Because the activity ‘straight’ must happen after ‘steering’ finishes.

[4] Partial-order planning focuses on relaxing the temporal order of actions. Plans can be *totally ordered* if every action is ordered with respect to every other action, or *partially ordered* if actions can be unordered with respect to each other.

[5] The main idea behind the finite-state approach of Fernando (2003) is that events are not so much about temporal intervals but rather about samplings of such intervals, and very selective samplings at that, which may well leave holes in intervals. His approach casts doubt on the adequacy of Allen’s 13 interval relations in characterizing the temporal relations between events.

References

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM* **26**(11): 832–843.

Allen, J. F. & Ferguson, G. (1994). Actions and events in interval temporal logic. *Journal of Logic and Computation* **4**(5): 531–579.

Arnold, D. J., Balkan, L., Meijer, S., Humphreys, R. L. & Sadler, L. (1994). *Machine Translation: an Introductory Guide*. London: Blackwells-NCC.

Badler, N., Webber, B., Palmer, M., Noma, T., Stone, M., Rosenzweig, J., Chopra, S., Stanley, K., Dang, H., Bindiganavale, R., Chi, D., Bourne, J. & Di Eugenio, B. (1997). *Natural language text generation form Task networks*. Technical report, Computer and Information Science Department, University of Pennsylvania, USA.

Dorr, B. J. & Jones, D. (1999). Acquisition of Semantic Lexicons: using word sense disambiguation to improve precision. In Evelyne Viegas (ed.), *Breadth and Depth of Semantic Lexicons*, 79–98, Norwell, MA: Kluwer Academic Publishers.

Dowty, D. (1979). *Word Meaning and Montague Grammar*. Dordrecht, The Netherlands: Reidel Publishing Company.

Fellbaum, C. (1998). A semantic network of English verbs. In Fellbaum, C. (ed.), *WordNet: An Electronic Lexical Database*, 69–104. Cambridge, MA: MIT Press.

Fernando, T. (2003). Finite-state descriptions for temporal semantics. In Bunt, H. van der Sluis, I. & Morante, R. (eds.), *Proceedings of The Fifth International Workshop on Computational Semantics* (IWCS-5), 122–136, Tilburg, The Netherlands, January.

Gennari, S. & Poeppel, D. (2002). Events versus states: empirical correlates of lexical classes. In *Proceedings of CogSci2002*, 351–356. George Mason University, Virginia, USA, August.

H-Anim (2001). Humanoid animation working group. http://www.h-anim.org.

Halpern, J. Y. & Shoham, Y. (1991). A propositional modal logic of time intervals. *Journal of ACM*, **38**(4): 935–962.

Jackendoff, R. (1990). *Semantic Structures*. Current studies in linguistics series, Cambridge, MA: MIT Press.

Jackendoff, R. (1991). Parts and boundaries. *Cognition*, **41**(1): 9–45.

Järvinen, T. and Tapanainen, P. (1997). A Dependency Parser for English. Technical Report, No.TR-1, Department of General Linguistics, University of Helsinki, Finland.

Kautz, H. A. & Ladkin, P. (1991). Integrating metric and qualitative temporal reasoning. In *Proceedings of the Ninth National Conference on Artificial Intelligence* (AAAI-91), 241–246, Anaheim, California, USA, July.

Levin, B. (1993). *English verb classes and alternations: a preliminary investigation*. Chicago: The University of Chicago Press.

Ma, M. & Mc Kevitt, P. (2003a). Semantic representation of events in 3D animation. In Bunt, H., van der Sluis, I., & Morante, R. (eds.), *Proceedings of The Fifth International Workshop on Computational Semantics* (*IWCS-5*), 253–281, Tilburg University, The Netherlands, January.

Ma, M. & Mc Kevitt, P. (2003b). Building character animation for intelligent storytelling with the H-Anim standard. In McNeill, M. (ed.), *Eurographics Ireland Chapter Workshop Proceedings 2003*, 9–15, Coleraine, Northern Ireland, April.

Ma, M. & Mc Kevitt, P. (2004) Visual semantics and ontology of eventive verbs. In *Proceedings of the First International Joint Conference on Natural Language Processing* (IJCNLP-04), Su, K.-Y. & Tsujii, J.-I. (eds.), 278–285, Resort Golden Palm, Sanya, China, March.

Pinhanez, C., Mase, K. & Bobick, A. (1997). Interval scripts: a design paradigm for story-based interactive systems. *Proceedings of CHI'97*, 287-294, Atlanta, Georgia, USA, March.

Pinon, C. (1997). Achievements in an event semantics. *Proceedings of Semantics and Linguistic Theory 7*, Lawson, A. (ed.), 276–292, Ithaca, NY: CLC Publications, Cornell University.

Reichenbach, H. (1947). *The Elements of Symbolic Logic*. London: Macmillan.

Schank, R. C. & Abelson, R. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum.

Siskind, J. M. (1995). Grounding Language in Perception. In Mc Kevitt, P. (ed.), *Integration of Natural Language and Vision Processing (Volume I): Computational Models and Systems,* 207–227. London, UK: Kluwer Academic Publishers.

Smith, C. S. (1991). *The Parameter of Aspect*. Dordrecht, The Netherlands: Kluwer.

van Benthem, J. (1983). *The Logic of Time*. Dordrecht, The Netherlands: Reidel Publishing Company.

Vendler, Z. (1967). Verbs and times. In Vendler, Z. (ed.), *Linguistics in Philosophy*, 97–121, Ithaca, N.Y.: Cornell University Press.

Verkuyl, H. (1993). *A Theory of Aspectuality*. Cambridge: Cambridge University Press.