A Model for Time Granularity in Natural Language

Gérard Becher, Françoise Clérin-Debart, Patrice Enjalbert
Greyc, CNRS UPRESA 6072
Université de Caen
14032 Caen Cedex, France
{becher, debart, patrice}@info.unicaen.fr

Abstract

We propose here a model for dealing with time granularity in natural language. In contrast with many other fields where granularity levels are essentially quantitative, in natural language we are confronted with a more subtle and qualitative kind of granularity where the different levels are not always precisely defined. Our model for representing such phenomena is based on time units and intervals. Time units are considered as indivisible chunks of time and correspond in many aspects to the usual notion of temporal points with the difference that the former are durative whereas the latter represent traditionally instantaneous events. Relations between time units and/or intervals are exhaustively studied and a relation algebra is defined which extends the traditional algebra on points and intervals. In particular, a containment relation is introduced in order to express changes in granularity. A logic with restricted quantifiers is proposed for formalizing temporal knowledge and some examples are discussed which show the relevance of the model for natural language.

1. Introduction

In this paper we propose a model—together with an appropriate logical formalism—for representing time-related knowledge resulting from an analysis of a text written in a given natural language.

There is a wide variety of temporal logics in the literature, ranging from propositional modal logic [8] to typed first-order logic [1] or first-order modal logic [2]. The choice of one of these formalisms, for a given application area, depends crucially on the nature of the underlying concepts of time associated with the selected logic.

Two kinds of time representations are usually coexisting in the literature: point-based and interval-based representations. They differ in the nature of their primitive entities: a temporal point is generally seen as something which is instantaneous (without any duration) whereas an interval[1] corresponds to something which is durative. Time points are often represented by integers, rational or real numbers whereas time intervals correspond to intervals in such a set of numbers.

Although there are specific domains where the use of such specialized representations may be sufficient, text analysis requires often both concepts of points and intervals. Moreover, many phenomena cannot be well represented according to an unique scale of time: typically, in natural language, one has the ability to expand any temporal entity — focusing on it — even if it was previously considered as punctual. Thus, a suitable representation of time should unify both approaches and should admit temporal entities with variable natures at different levels of granularity.

The problem of time granularity is of growing interest in the AI community and has been discussed for example in [5, 15, 14, 16, 17, 18, 19]. However, in most of the papers in the literature, granularity is considered from a more or less quantitative point of view. Intensive use of metrics (dealing with months, weeks or days) is rather common and an important concern is often how to switch from a granularity level to another.

We are interested here in a more qualitative point of view: instead of fixing absolute granularity levels we want to represent the relative position of things which are not necessarily on the same level. In other words, instead of choosing a level for reasoning, we will organize the knowledge in such a way that the differences of levels are genuine part of the representation itself. We claim that such a model is more appropriate for the representation of the different processes which occur in most sentences in natural language.

This paper is organized as follows: in section 2, we describe our intuition of time before defining formally in section 3 the notions of *time units* and of *temporal intervals*. The relations between such time units and intervals are ex-

[1] Under the designation 'intervals' one means generally convex intervals. Some authors consider also unions of convex intervals. See for example [12], [13], [3] and [2]

haustively studied in section 4 where we define the relation algebra A_T. The syntax and semantics of our proposed logical formalism is described in section 5. Applications of the formalism in concrete situations are discussed in section 6. Finally, in section 7 we define the notion of duration of time units or temporal intervals. Durations, jointly with the notion of scale factor which is defined in the same section, are tailored in order to give the ability to use metrics when necessary. Section 8 is the conclusion.

2. Two visions of time

The concept of time can be considered in different ways. At first glance, we can think about time as a succession of instants (let us call them *temporal points* or simply *points* for short). The set of these points may be discrete or dense. We will refer to such a vision of time under the designation of *physical time*. A common and convenient idea is to assimilate the flow of *physical time* points to a flow of numbers: if time is discrete, we can consider for instance the set of integers whereas rational or real numbers may be used in the dense case.

Although we can consider intervals in such structures in order to represent duration, the notion of a *physical time* point is too restrictive from our point of view. In fact, we believe that such a kind of points never occur in common discourse. Even if we consider for example that *"Mary left for Paris at 8 in the morning"*, we do *not* speak about a temporal *physical* point since the real instant of Mary's departure surely has a certain duration (probably very small). And it is clear that a lot of things may happen precisely during this moment (imagine for example that Mary's son was hurt and cried just when she leaved). Some authors overcome this problem by considering points as intervals of very short length.

Our proposition is different. We consider that the distinction — traditional in linguistics — between instantaneous (punctual in [7]) and durative events is perfectly relevant. But this distinction can only be relative and even instantaneous events have a (relatively small) duration. For example in *"John reached the top of the mountain and sat down for a rest"* the event *"John reached…"* is presented as instantaneous, but this alleged *instant* can be re-elaborated if we go on with: *"while reaching the summit he could admire a nice panorama"*. Indeed, it would be very difficult to decide which is the precise *instant* in question. In other words we think — according to the work of Laurent Gosselin in [7] also related to [10]— that text understanding consists in the elaboration of several processes and that each process has a temporal interval associated with it. However, we can never have an exact idea of such an interval and we deal only with *approximations* of it. At some level of approximation, a process P may be considered as punctual: this means that we consider the associated interval as an indivisible chunk of time. But this can be further refined in order to consider some new events or some durative activity occurring during P and mentioned later in the text, as in the above two examples.

Again, we feel that the notion of time point elaborated in physics (element of $\mathbb{Q}$ or $\mathbb{R}$) is not adequate to the cognitive notion of instant, which is rather a small (with respect to some scale) lap of time, considered (maybe temporarily) as indivisible. It is the purpose of our model to give a mathematical account for this kind of dynamic between instants and intervals. Another situation we can model that way is concerned with the locution *"A as soon as B"* where, according to L. Gosselin, *A* follows *B* within an infinitely small lap of time, *i.e.* an *instant* in our sense. This also will be discussed in section 7.

In order to avoid confusion with physical time points, we shall use the term *time unit* instead of *time point* when we are speaking about punctual approximations of a process. In contrast to physical time points, time units have a duration which can be potentially very huge. For example, in a certain context, we should need to consider a punctual approximation of the process corresponding to the disappearance of the dinosaurs! In section 7, we will give a way to specify the duration of time units. For that purpose, we shall project them onto the line representing physical time, and associate to each time unit two numbers, the difference of these numbers measuring the duration of the time unit.

3. Time units and intervals

In this section, we define formally the two notions of *time units* and *temporal intervals* and also precise the relations existing between them. By evidence, we cannot consider time simply as a linearly ordered set of time units. Since we will be able to assert that other events may occur *during* an interval which has been approximated by a (punctual) time unit, there must be another relation than the *precedence relation* between time units. For that reason, we will introduce the *containment relation* which is closely related with Allen's *during* relation but with the main difference that the containment relation concerns not only intervals but also time units. In our intuition, the precedence and containment relations, respectively denoted by $<$ and $\trianglelefteq$, are clearly disjoint and their union is a total ordering on the set of time units. The precedence relation is a strict partial ordering whereas the containment is a partial ordering whose strict part will be denoted $\triangleleft$. Thus, the set of times units can be seen as a two-dimensional structure: graphically, events may be ordered horizontally according to their precedence and vertically according to their level of granularity. For example, considering that q_1, q_2, $q_{1,1}$ and $q_{1,2}$ are time units, if q_1 precedes q_2 and $q_{1,1}$ precedes $q_{1,2}$ ($q_1 < q_2$

and $q_{1,1} < q_{1,2}$), then q_1 will be represented at the left of q_2 and $q_{1,1}$ at the left of $q_{1,2}$. If now $q_{1,1}$ and $q_{1,2}$ occur both during q_1 (written $q_1 \lhd q_{1,1}$ and $q_1 \lhd q_{1,2}$), then $q_{1,1}$ and $q_{1,2}$ will be drawn both above q_1 (figure 1). It is interesting to

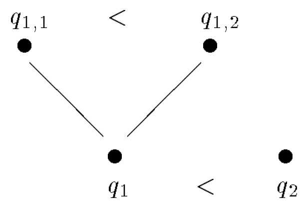

Figure 1. A two-dimensional structure

note that, if a time unit q is such that q precedes q_1, then it must also precede $q_{1,1}$ since $q_{1,1}$ occurs during q_1. On the other hand, if we know that $q_{1,2}$ precedes a time unit q', then we ignore if q_1 contains q' or if q_1 precedes q'. This means in particular that $< \circ \lhd$ is different from $\lhd \circ <$ and our model will reflect this difference.

A suitable way for capturing the intuition of these two relations is to represent time units by words on the set of rational numbers, the precedence relation corresponding then in a natural way to a subset of the lexicographical ordering and the containment relation being the prefix relation.

Definition 1 *The set of* time units *is the set* $\mathbb{Q}^\star$ *of all finite sequences of rational numbers.*

$$\mathbb{Q}^\star = \{ (q_1, q_2, \ldots, q_n) \mid n \in \mathbb{N} \ \ \forall i, 1 \leq i \leq n, \ q_i \in \mathbb{Q} \}$$

If a is an element of $\mathbb{Q}^\star$, the length of the sequence a is denoted by $|a|$. The empty sequence is the unique time unit of length 0. For every time unit a in $\mathbb{Q}^\star$, and for every integer i such that $1 \leq i \leq |a|$, we denote by a_i the element numbered by i in the sequence a.

The lexicographical ordering on such sequences of rational numbers is a linear ordering on $\mathbb{Q}^\star$. From this total ordering, we extract the two partial orderings corresponding to $\unlhd$ and $<$ in the following way:

Definition 2 *For all time units a and b in $\mathbb{Q}^\star$, $a \unlhd b$ if and only if a is an initial sequence of b.*

$$a \unlhd b \textit{ iff } |a| \leq |b| \textit{ and } \forall i, 1 \leq i \leq |a|, \ \ a_i = b_i$$

The relation $\unlhd$ is clearly reflexive, anti-symmetric and transitive and hence, it is a (partial) ordering relation on $\mathbb{Q}^\star$ whose strict part will be denoted by $\lhd$. We write also $a \unrhd b$ instead of $b \unlhd a$ and $a \not\trianglelefteq b$ whenever $a \unlhd b$ *not* holds.

Definition 3 *For all time units a and b in $\mathbb{Q}^\star$, $a < b$ if and only if there exists a strictly positive integer k such that:*

$$\begin{aligned} \forall\, i < k \quad & a_i = b_i \\ & a_k < b_k \end{aligned}$$

It is clear that the above definition can be reformulated, saying that $a < b$ if and only if $a <_{lex} b$ and $a \not\trianglelefteq b$, where $<_{lex}$ is the strict lexicographical ordering induced on $\mathbb{Q}^\star$ by the natural ordering on $\mathbb{Q}$. Obviously, the relation $<$ is a partial strict ordering on $\mathbb{Q}^\star$. Temporal intervals are defined by their bounds which consist in an ordered pair of time units:

Definition 4 *A* temporal interval *is a couple (a, b) of $\mathbb{Q}^\star$ satisfying $a < b$. The set of all temporal intervals is denoted by $\mathbb{L}$ and the union of $\mathbb{L}$ and $\mathbb{Q}^\star$ is denoted by $\mathbb{I}$.*

$$\mathbb{L} = \{ (a, b) \mid a \in \mathbb{Q}^\star, \ \ b \in \mathbb{Q}^\star, \quad a < b \}$$
$$\mathbb{I} = \mathbb{L} \cup \mathbb{Q}^\star$$

Time units and temporal intervals will be used to represent information at different levels of granularity. More precisely, if we approximate the interval corresponding to a certain process p by two different time units a and b, we will have either $a \unlhd b$ or $b \unlhd a$. On the other hand, if we approximate p by a time unit a and a temporal interval $i = (i', i'')$, we will have $a \unlhd i'$ and $a \unlhd i''$. This means that the level of the observation is reflected in a certain sense by the length of the sequence. However, the fact that two time units share the same length does not imply that they are observed at the same level of granularity. This is because we think that differences of level are only adequate if the corresponding facts are not completely independent (in which case they are represented by two sequences with a common initial segment). Thus, the notion of a common level is somewhat restricted as follows:

Definition 5 *For every sequences a and b in $\mathbb{Q}^\star$, we say that a and b are on the same level (denoted by $a \leftrightarrow b$) if and only if a and b are of the same length k and coincide until the level $k - 1$.*

$$a \leftrightarrow b \textit{ iff } |a| = |b| \textit{ and } \forall i \quad (1 \leq i < |a|) \rightarrow (a_i = b_i)$$

It is immediate to check that the relation $\leftrightarrow$ is an equivalence relation in $\mathbb{Q}^\star$. Moreover, each equivalence class is isomorphic to the set $\mathbb{Q}$ of rational numbers. Two time units which are in the same equivalence class are considered as being on the same level. We extend now this notion of levels to temporal intervals.

Definition 6 *Let c be a time unit and let $i = (i', i'')$ and $j = (j', j'')$ be two temporal intervals.*

We say that c is on the same level as i (alternatively i is on the same level as a) — and we write $c \leftrightarrow i$ or, respectively, $i \leftrightarrow c$ — if and only if i' and i'' are both on the same level as a.

We say also that i is on the same level as j if and only if i' and i'' are both on the same level as j' and j'', that is $i' \leftrightarrow i'' \leftrightarrow j' \leftrightarrow j''$.

4. The relation algebra A_T

In this section we define a relation algebra A_T and a connected weak representation $\mathcal{R}$ of A_T into the set $\mathcal{P}(\mathbb{I} \times \mathbb{I})$ of binary relations in the set $\mathbb{I}$. In section 5, we shall define a logical language with restricted quantifiers where these relations will be used for specifying the restrictions. For a formal definition of a relation algebra, see [11, 12]. Let us simply recall here that a relation algebra consists in a boolean algebra with two additional operations: the relative product and the converse. The relative product of a and b is denoted here by $a\,;b$ and the converse of a is denoted by $\breve{a}$. Of course, a number of additional properties involving these operations have also to be satisfied. A relation algebra is *proper* if its universe is a set of binary relations, and its operations coincide with the usual set-theoretic operations on these relations. A *representation* of a relation algebra $\mathcal{A}$ is an isomorphism Φ from $\mathcal{A}$ onto a proper relation algebra $\mathcal{B}$. In particular, Φ maps the relative product to the composition of the relations and we have: $\Phi(\alpha;\beta) = \Phi(\alpha) \circ \Phi(\beta)$. A *weak* representation is defined by weakening this last condition to $\Phi(\alpha;\beta) \supseteq \Phi(\alpha) \circ \Phi(\beta)$. A weak representation of A onto $\mathcal{P}(U \times U)$ is *connected* if it maps the unit of the boolean algebra A onto $U \times U$.

The relation algebra A_T is an atomic algebra defined by a set of 62 atoms partitioned in four classes and listed in figure 2. The tables of the relative product and converse operations are unfortunately too huge to be printed out here. However, the tables of their representation can be computed by a Caml-Light program, in which we coded only their restriction to the 5 atoms of type $R_1^1(i)$. This is due to the fact that relations on intervals can be expressed by using relations on their bounds, replacing for instance $(a', a'')\ R_2^1(i,j)\ b$ by $a'\ R_1^1(i)\ b$ and $a''\ R_1^1(j)\ b$.

The representation of an atom of A_T is a relation in $\mathbb{I}$, that is a subset of $\mathcal{P}(\mathbb{I} \times \mathbb{I})$. The intended meaning is that atoms of the form $R_1^1(i)$ (resp. $R_1^2(i)$, $R_2^1(i,j)$, $R_2^2(i,j)$) are mapped to subsets of $\mathbb{Q}^\star \times \mathbb{Q}^\star$ (resp. $\mathbb{Q}^\star \times \mathbb{L}$, $\mathbb{L} \times \mathbb{Q}^\star$, $\mathbb{L} \times \mathbb{L}$). These relations are pairwise disjoint subsets of $\mathbb{I} \times \mathbb{I}$ representing exhaustively all relative positions of two elements in $\mathbb{I}$. In other words, the set $\mathbb{I} \times \mathbb{I}$ can be partitioned into the set of these relations; thus the proposed representation is connected.

In order to give the intuition of the representation of an atom A_T in $\mathcal{R}$, we adapt a notation introduced by Gérard Ligozat in [13]. Let us consider for instance an arbitrary time unit y. The set $\mathbb{Q}^\star$ is partitioned by y in five zones (see figure 3) defined by :

$$
\begin{aligned}
Z_0(y) &= \{\, x \in \mathbb{Q}^\star \mid x < y \,\} \\
Z_1(y) &= \{\, x \in \mathbb{Q}^\star \mid y \lhd x \,\} \\
Z_2(y) &= \{y\} \\
Z_3(y) &= \{\, x \in \mathbb{Q}^\star \mid x \lhd y \,\} \\
Z_4(y) &= \{\, x \in \mathbb{Q}^\star \mid y < x \,\}
\end{aligned}
$$

Type	Relations
$R_1^1(i)$	$R_1^1(0)$, $R_1^1(2)$, $R_1^1(3)$, $R_1^1(1)$, $R_1^1(4)$
$R_1^2(i)$	$R_1^2(0)$, $R_1^2(2)$, $R_1^2(3)$, $R_1^2(4)$, $R_1^2(1)$, $R_1^2(5)$, $R_1^2(8)$, $R_1^2(6)$, $R_1^2(7)$, $R_1^2(9)$
$R_2^1(i,j)$	$R_2^1(0,0)$, $R_2^1(0,1)$, $R_2^1(0,2)$, $R_2^1(0,3)$, $R_2^1(0,4)$, $R_2^1(1,1)$, $R_2^1(1,4)$, $R_2^1(2,4)$, $R_2^1(3,4)$, $R_2^1(4,4)$
$R_2^2(i,j)$	$R_2^2(0,0)$, $R_2^2(0,1)$, $R_2^2(0,2)$, $R_2^2(0,3)$, $R_2^2(0,4)$, $R_2^2(0,5)$, $R_2^2(0,6)$, $R_2^2(0,7)$, $R_2^2(0,8)$, $R_2^2(0,9)$, $R_2^2(1,1)$, $R_2^2(1,5)$, $R_2^2(1,6)$, $R_2^2(1,7)$, $R_2^2(1,8)$, $R_2^2(1,9)$, $R_2^2(2,5)$, $R_2^2(2,6)$, $R_2^2(2,7)$, $R_2^2(2,8)$, $R_2^2(2,9)$, $R_2^2(3,5)$, $R_2^2(3,6)$, $R_2^2(3,7)$, $R_2^2(3,8)$, $R_2^2(3,9)$, $R_2^2(4,9)$, $R_2^2(5,5)$, $R_2^2(5,6)$, $R_2^2(5,7)$, $R_2^2(5,8)$, $R_2^2(5,9)$, $R_2^2(6,6)$, $R_2^2(6,9)$, $R_2^2(7,9)$, $R_2^2(8,9)$, $R_2^2(9,9)$

Figure 2. The 62 atoms of the algebra A_T

In the same spirit, if $y = (y', y'')$ is an interval of $\mathbb{L}$, we consider the following partition of $\mathbb{Q}^\star$ in ten zones pictured in figure 3:

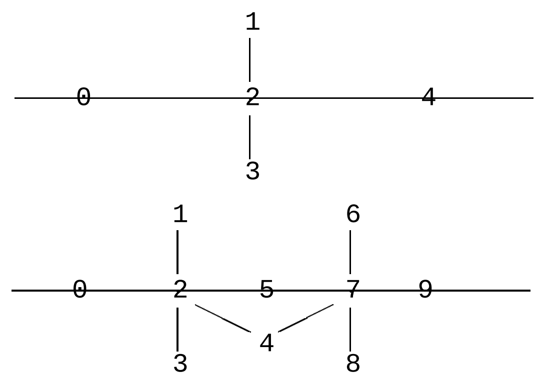

Figure 3. Two kind of partitions of $\mathbb{Q}^\star$

$$
\begin{aligned}
Z_0(y', y'') &= \{\, x \in \mathbb{Q}^\star \mid x < y' \,\} \\
Z_1(y', y'') &= \{\, x \in \mathbb{Q}^\star \mid y' \lhd x \,\} \\
Z_2(y', y'') &= \{y'\} \\
Z_3(y', y'') &= \{\, x \in \mathbb{Q}^\star \mid x \lhd y' \wedge x \ntrianglelefteq y'' \,\} \\
Z_4(y', y'') &= \{\, x \in \mathbb{Q}^\star \mid x \lhd y' \wedge x \lhd y'' \,\} \\
Z_5(y', y'') &= \{\, x \in \mathbb{Q}^\star \mid y' < x < y'' \,\} \\
Z_6(y', y'') &= \{\, x \in \mathbb{Q}^\star \mid y'' \lhd x \,\} \\
Z_7(y', y'') &= \{y''\} \\
Z_8(y', y'') &= \{\, x \in \mathbb{Q}^\star \mid x \lhd y'' \wedge x \ntrianglelefteq y' \,\} \\
Z_9(y', y'') &= \{\, x \in \mathbb{Q}^\star \mid y'' < x \,\}
\end{aligned}
$$

We define now formally the representation $\mathcal{R}(a)$ of an atom $a \in A_T$ as follows:

Definition 7 *If $R_1^1(i)$, $R_1^2(i)$, $R_2^1(i,j)$ and $R_2^2(i,j)$ are atoms in A_T, then their representation is defined by:*

$$\begin{aligned}
\mathcal{R}(R_1^1(i)) &= \{\, (x,y) \in \mathbb{Q}^\star \times \mathbb{Q}^\star \mid x \in Z_i(y) \,\} \\
\mathcal{R}(R_1^2(i)) &= \{\, (x,(y',y'')) \in \mathbb{Q}^\star \times \mathbb{L} \mid x \in Z_i(y',y'') \} \\
\mathcal{R}(R_2^1(i,j)) &= \{\, ((x',x''),y) \in \mathbb{L} \times \mathbb{Q}^\star \mid \\
&\qquad x' \in Z_i(y) \wedge x'' \in Z_j(y) \,\} \\
\mathcal{R}(R_2^2(i,j)) &= \{\, ((x',x''),(y',y'')) \in \mathbb{L} \times \mathbb{L} \mid \\
&\qquad x' \in Z_i(y',y'') \wedge x'' \in Z_j(y',y'') \,\}
\end{aligned}$$

It is easy to check that $<$ is the representation in $\mathcal{R}$ of the atom $R_1^1(0)$ and that $\trianglelefteq$ is the representation of the non-atomic element $R_1^1(2) + R_1^1(3)$. The representation contains also the 13 Allen's relations (for instance *meets* is the atom $R_2^2(0,2)$) although their composition does not respect Allen's composition table. The fact that $\mathcal{R}$ is a weak representation of A_T can be emphasized by noticing for example that $R_1^1(3);R_1^1(3) = R_1^1(3)$ but clearly we have $\lhd \circ \lhd \subsetneq \lhd$. Indeed, for any rational numbers $q_1, \ldots, q_n$, the two times units $a = (q_1, \ldots, q_{n-1})$ and $b = (q_1, \ldots, q_{n-1}, q_n)$ satisfy $a \lhd b$, but it is easy to check that (a,b) is not in the relation $\lhd \circ \lhd$.

5. The logical formalism

We describe here a formal language named the *time units language* (*TUL* for short) which might be suitable for representing time-related knowledge. The syntax of *TUL* is that of a first-order language with restricted quantifiers as described in [4]. Its signature consists in the union of two disjoint signatures: a temporal signature Δ and a standard first order signature Σ whose domain is the universe of the discourse. The temporal signature contains in particular the constant symbol now and all numbers for expressing fixed dates. In addition, a set of function symbols of arbitrary arity may be available for the sake of skolemization. Terms built over the signature Δ are called *temporal terms*. The predicate symbols in Δ are:

- two sort predicates *P* and *I* used to indicate the nature (time units or temporal intervals) of the temporal objects.
- the relation symbol $\leftrightarrow$ which will be used for specifying that two temporal objects are considered at the same level of granularity.
- the elements of the relation algebra A_T which will serve for indicating the relative positions of the temporal objects.

An *atomic restriction* is an atomic formula on the signature Δ. More complex restrictions can be expressed by combination of atomic restrictions in the following way: restrictions of type 1 are conjunctions of atomic restrictions using exclusively the predicates *P*, *I* and $\leftrightarrow$; restrictions of type 2 are formulas built with the relation symbols of the relation algebra A_T and the logical connectives $\wedge$, $\vee$ and $\neg$.

Formulas in *TUL* are standard first-order formulas built over the signature $\Delta \cup \Sigma$ with the exception that quantifiers may be indexed by a pair of a set of variables and two (possibly empty) restrictions. Such a restricted quantifier is of the form $Q\{x_1, \ldots, x_n\} : \{C_1\} :: \{C_2\}$ where Q is one of the quantifiers $\exists$ or $\forall$, $\{x_1, \ldots, x_n\}$ is a set of temporal variables (curly braces should be dropped when there is only one variable) and C_1 and C_2 are respectively restrictions of type 1 and 2.

The following is an example of a formula in *TUL*:

$$\forall\{t,i\} : (P(t_1) \wedge I(i)) :: (t\ R_1^2(3)\ i)\ \exists x\ \ A(t,i,x)$$

The interpretation of a formula of *TUL* causes no real problem since the semantics of the restricted quantifiers can be given by the usual relativization rules, transforming any formula of *TUL* into an equivalent formula of a language without restricted quantifiers by replacing $\forall\{x_1, \ldots, x_n\} : \{C_1\} :: \{C_2\}\ \ F$ by $\forall x_1, \ldots, \forall x_n((C_1 \wedge C_2) \rightarrow F)$ and by replacing $\exists\{x_1, \ldots, x_n\} : \{C_1\} :: \{C_2\}\ \ F$ by $\exists x_1, \ldots, \exists x_n(C_1 \wedge C_2 \wedge F)$.

All restrictions are interpreted in the fixed model consisting of the set $\mathbb{I}$ of time units and intervals. An interpretation for *TUL* is any extension of this fixed model obtained by adding a universe D disjoint from $\mathbb{I}$ and the classical interpretation function for the symbols in Σ. For the semantic notions of satisfiability and validity, see [4].

6. Applications

We give here some examples illustrating the adequacy of *TUL* when dealing with time-related knowledge in natural language. Following the authors in [6], we think that our vision of a process may change during the analysis of the text. The same process may be considered as punctual in a certain context, and seen a bit later as durative. In order to solve this problem, we associate an interval to each process of the text and elaborate several approximations of this interval. Durative approximations are reflected by temporal intervals and punctual approximations correspond to time units. Although the difference is sometimes subjective, let us briefly say that punctual approximations are used whenever a process is considered as a whole (even if it has a long duration) whereas structured processes are seen as durative. Here, structured means that other events are explicitly situated during the process or at its beginning or end.

For example, in the sentence *"Joe's father passed away three weeks ago after a long agony"*, we elaborate a durative approximation of the process corresponding to the agony since another process (the death) takes place at its end. On the other hand, the death is considered as a whole, in the absence of other precisions. Clearly, agony and death are in the relation $R_2^1(0,2)$.

In order to clarify the nature of the temporal concepts, we adopt in the remaining of the paper the following typographical conventions:

- a durative approximation of an interval i will be denoted by i indexed by an horizontal arrow ($i_\leftrightarrow$), whereas a punctual approximation is denoted by i indexed by a vertical arrow ($i_\downarrow$).
- if i is a temporal interval, we denote its left bound by i' and its right bound by i''. Hence $i = (i', i'')$.
- we shall write often $x < y$ instead of $x\ R_1^1(0)\ y$ and $x \lhd y$ instead of $x\ R_1^1(3)\ y$ or even $x\ R_1^2(4)\ y$ according to the context.

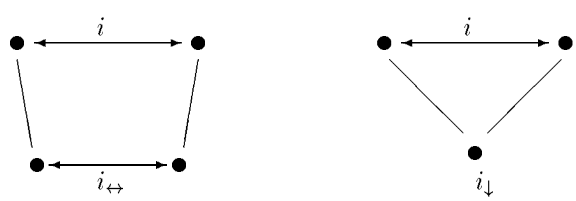

Figure 4. Durative and punctual processes

If i is an interval and $i_\leftrightarrow$ one of its durative approximations, then we consider that the two bounds of $i_\leftrightarrow$ are on the same level. Clearly we have $i_\leftrightarrow\ R_2^2(3,8)\ i$ and $i_\downarrow \lhd i$ for any durative (resp. punctual) approximation of i (see figure 4).

6.1 Mary's breakfast

Our first example is intended to show that knowledge may be situated at different levels, even if these levels are not made explicit in the text.

> *Yesterday Mary got up and had breakfast. As she finished, the phone rang suddenly. At the same time all lights went off.*

In the above text, we distinguish several temporal intervals. First of all, note that the enunciation interval is here the current interval also called *now* and which is considered as a temporal constant of the language. Then remark that *yesterday* (y for short) is an interval which precedes strictly the enunciation interval. At first glance, it is reasonable to consider *yesterday* and *now* as indivisible and to use punctual

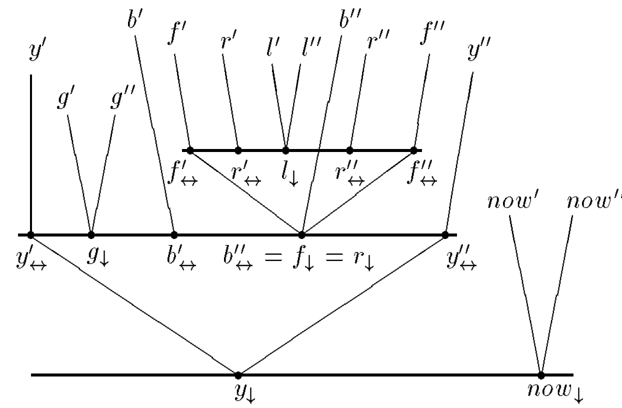

Figure 5. Mary's breakfast

approximations of *yesterday* and *now*. Thus, we choose to represent them by two time units $y_\downarrow$ and *now*$_\downarrow$ on the same granularity level. Of course, we will state that $y_\downarrow <$ *now*$_\downarrow$. However, there are other events occurring during the time unit $y_\downarrow$. Hence, we need also a durative approximation of *yesterday*, namely the temporal interval $y_\leftrightarrow = (y'_\leftrightarrow, y''_\leftrightarrow)$. Clearly, these two approximations verify $y_\downarrow \lhd y_\leftrightarrow \lhd y$.

Finally we use also five intervals g, b, f, r and l, corresponding respectively to the five processes *Mary got up*, *Mary had breakfast*, *Mary finished her breakfast*, *the phone rang* and *the lights went off*. In absence of further precisions, the process *Mary got up* is considered as punctual and approximated by the time unit $g_\downarrow$ at the same level as $y_\leftrightarrow$. On this same level, we shall place an approximation $b_\leftrightarrow = (b'_\leftrightarrow, b''_\leftrightarrow)$ of the breakfast which is seen as durative since other events will be situated at its end. The end itself becomes the punctual time unit $f_\downarrow$ when considered at this level. Of course, it verifies $f_\downarrow\ R_1^2(7)\ b_\leftrightarrow$. Due to the position of the verbs in the text, it is quite likely that Mary got up before having breakfast. Hence we can write that $g_\downarrow$ precedes $b_\leftrightarrow$: $g_\downarrow\ R_1^2(0)\ b_\leftrightarrow$.

At the level of $y_\leftrightarrow$, we cannot distinguish the phone ring from the end of the breakfast. Hence, this process is approximated by the time unit $r_\downarrow$ which is identical to $f_\downarrow$.

If we change now the granularity level of the observation — focusing on $f_\downarrow$ — we shall see the end of the breakfast and the phone's ring under a durative aspect and approximate them by two intervals $f_\leftrightarrow = (f'_\leftrightarrow, f''_\leftrightarrow)$ and $r_\leftrightarrow = (r'_\leftrightarrow, r''_\leftrightarrow)$ verifying $r_\leftrightarrow\ R_2^2(5,5)\ f_\leftrightarrow$. During the interval $r_\leftrightarrow$, we place also the punctual approximation $l_\downarrow$ of the process *the lights went off*.

The whole representation is pictured in figure 5 where horizontal lines represent levels whereas the sequences of rational numbers are pictured by vertical or oblique lines.

In *TUL*, the situation can be formalized as follows:

$$\exists\{y, y_\leftrightarrow, y_\downarrow\} : (I(y) \wedge I(y_\leftrightarrow) \wedge P(y_\downarrow) \wedge$$
$$y_\downarrow \leftrightarrow now_\downarrow) :: (y_\downarrow\ R_1^1(3)\ y \wedge y_\downarrow\ R_1^1(0)\ now_\downarrow)$$
$$\exists\{g, g_\downarrow\} : (I(g) \wedge P(g_\downarrow) \wedge g_\downarrow \leftrightarrow y_\leftrightarrow) ::$$
$$(g_\downarrow\ R_1^2(4)\ g \wedge g_\downarrow\ R_1^2(5)\ y_\leftrightarrow)$$
$$\exists\{b, b_\leftrightarrow\} : (I(b) \wedge I(b_\leftrightarrow) \wedge b_\leftrightarrow \leftrightarrow g_\downarrow) ::$$
$$(b_\leftrightarrow\ R_2^2(3,8)\ b \wedge b_\leftrightarrow\ R_2^2(5,5)\ y_\leftrightarrow \wedge g_\downarrow\ R_1^2(0)\ b_\leftrightarrow)$$
$$\exists\{f, f_\leftrightarrow, f_\downarrow\} : (I(f) \wedge I(f_\leftrightarrow) \wedge P(f_\downarrow)) ::$$
$$(f_\downarrow\ R_1^2(7)\ b_\leftrightarrow \wedge f_\downarrow\ R_1^2(4)\ f_\leftrightarrow \wedge f_\leftrightarrow\ R_2^2(3,8)\ f)$$
$$\exists\{r, r_\leftrightarrow, r_\downarrow\} : (I(r) \wedge I(r_\leftrightarrow) \wedge P(r_\downarrow) \wedge r_\leftrightarrow \leftrightarrow f_\leftrightarrow)$$
$$:: (r_\downarrow\ R_1^2(7)\ b_\leftrightarrow \wedge r_\downarrow\ R_1^2(4)\ r_\leftrightarrow \wedge r_\leftrightarrow\ R_2^2(3,8)\ r$$
$$\wedge\ r_\leftrightarrow\ R_2^2(5,5)\ f_\leftrightarrow)$$
$$\exists\{l, l_\downarrow\} : (I(l) \wedge P(l_\downarrow) \wedge l_\downarrow \leftrightarrow r_\leftrightarrow) :: (l_\downarrow\ R_1^2(5)\ r_\leftrightarrow)$$
$$got\_up(g, Mary) \wedge breakfast(b, Mary) \wedge$$
$$ring(r, phone) \wedge lights\_turned\_off(l)$$

6.2 The locution "as soon as"

Our second example of application deals with the use of adverbial locutions like "as soon as".

As soon as he rang the bell, the door opened

Two punctual processes are occurring in that example; let us call r and o the temporal intervals associated respectively with the processes *"ring the bell"* and *"open the door"*. In a classical interval-based representation, one would be tempted to employ two very small meeting intervals. But in that case, absolutely nothing can happen between the two processes! In our model, the intervals r and o are approximated by the same time unit $r_\downarrow = o_\downarrow$. As a consequence, at a certain level of granularity, the two processes are indiscernible. But if we look closer, we can see that there is a gap between them, and this (potentially huge) gap is clearly represented by the fact that r precedes strictly o. This situation is pictured in figure 6 and is described in *TUL* below. Of course, there are possible variants where ringing the bell or even opening the door are durative processes.

$$\exists\{r, r_\downarrow\} : (I(r) \wedge P(r_\downarrow)) :: (r_\downarrow\ R_1^2(4)\ r)$$
$$\exists\{o, o_\downarrow\} : (I(o) \wedge P(o_\downarrow)) :: (o_\downarrow = r_\downarrow \wedge o_\downarrow\ R_1^2(4)\ o$$
$$\wedge\ r\ R_2^2(0,0)\ o)\ \ Rings(r, bell) \wedge Opens(o, door)$$

7. Durations and scale factors

In this section, we propose an extension of our model intended to capture the intuition of durations of time units or

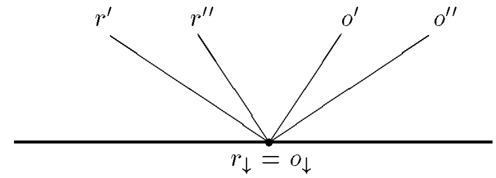

Figure 6. The door opened as soon as he rang the bell

intervals. For that sake, we assume that there exists an objective physical time, and we consider that each time unit can be projected onto the line representing this physical time. Such a projection defines two physical dates (or time stamps) corresponding respectively to the beginning and to the end of the time unit. Their difference corresponds to its duration. Evidently, such a projection function cannot be chosen arbitrarily: a number of conditions must be satisfied for guaranteeing the adequacy of the model to reality. However it is easy to check that there exist such functions.

Definition 8 *A projection function is a mapping p assigning to each time unit a a pair of rational numbers $p(a) = (\overleftarrow{a}, \overrightarrow{a})$ such that, for all time units a and b in $\mathbb{Q}^\star$ the following conditions holds:*

1. $\overleftarrow{a} < \overrightarrow{a}$ *and* $\overleftarrow{b} < \overrightarrow{b}$
2. *if* $a\ R_1^1(0)\ b$ *then* $\overrightarrow{a} < \overleftarrow{b}$
3. *if* $a\ R_1^1(1)\ b$ *then* $\overleftarrow{b} < \overleftarrow{a} < \overrightarrow{a} < \overrightarrow{b}$
4. *if* $a\ R_1^1(2)\ b$ *then* $\overleftarrow{a} = \overleftarrow{b}$ *et* $\overrightarrow{a} = \overrightarrow{b}$
5. *if* $a\ R_1^1(3)\ b$ *then* $\overleftarrow{a} < \overleftarrow{b} < \overrightarrow{b} < \overrightarrow{a}$
6. *if* $a\ R_1^1(4)\ b$ *then* $\overrightarrow{b} < \overleftarrow{a}$

If p is a projection function, the duration of a time unit a (w.r.t. p) is defined by: $d_p(a) = \overrightarrow{a} - \overleftarrow{a}$.

We extend now this definition to temporal intervals. Since intervals are ordered pairs of time units (which have their own duration), they cannot have a precise duration. However, minimal and maximal durations can be defined for intervals.

Definition 9 *The* duration *of an interval $i \in \mathbb{L}$ is the ordered pair of rational numbers $(\overleftarrow{b} - \overrightarrow{a}, \overrightarrow{b} - \overleftarrow{a})$ where $\overleftarrow{b} - \overrightarrow{a}$ is the* minimal *and $\overrightarrow{b} - \overleftarrow{a}$ the* maximal *duration of i.*

In text analyzing, durations of time units or intervals may be given either by temporal adverbs (as in *he looks briefly at the book*) or by absolute indication of time spaces (like in *he stays in Paris for two months*). But most often, assigning a duration to a time unit requires a large amount of implicit knowledge about the real world. In many cases, we don't know much about the absolute duration of a time unit (at least not in a quantified form), but we may have implicit or explicit information about the ratio between this time unit and the time units of lower levels. More formally, if a and b are in $\mathbb{Q}^\star$ and if $a \lhd b$, then we may know that $d(b) < k \times d(a)$. This ratio k, that we will denote by the term *scale factor*, may be quantified when time units correspond to days, hours or minutes, ... but most often we have only qualitative information about it. However, we postulate that it is worth integrating such a scale factor constraint in our model, even if it may only serve to perform occasionally some kind of fuzzy reasoning. Both notions of durations and scale factors may be reflected in the formalism by adding two restriction predicates to the temporal signature.

8. Conclusion and future work

As shown in the previous examples, there is not always a need to use metrics when we are dealing with time granularity. In particular, this is obvious in natural language sentences where concepts can be considered at different time levels without quantifying them. The model that we describe here is intended for representing such a kind of qualitative granularity: it allows essentially to express the relations between the temporal concepts. Situations which are considered as tricky by the linguists, such as the use of the locution *as soon as* — which is something intermediate between Allen's *meets* and *precedes* relations — can be well represented. Of course, since this is a work in progress, there remain unsolved problems. In particular, we have to precise how this representation can be exploited by an automated reasoner using a suitable constraint solver. This concern will be addressed in a further paper.

One may object that things are not only qualitative in natural language: there may also be quantitative information such as absolute dates or durations. Dates cause no real problem since they can be expressed by constant sequences of rational numbers and the language may be enriched allowing for example formulas of the form $P@t$ where t is a fixed date as in [9]. Concerning durations, we introduced the notions of scale factors and durations, but it remains to make precise how they can effectively be used in order to integrate both quantitative and qualitative aspects.

Acknowledgments

Many thanks are due to Heinrich Herre for his useful comments and suggestions on an earlier draft of this paper.

References

[1] F. Bacchus, J. Koomen, and J. Tenenberg. A non-reified Temporal Logic. *Artificial Intelligence*, 52:87–108, 1991.

[2] G. Becher. First Order Modal Temporal Logic with Generalized Intervals. In I. C. S. Press, editor, *Proceedings of the Third International Workshop on Temporal Representation and Reasoning*, pages 170–175, 1996.

[3] M. Bouzid and P. Ladkin. Simple Reasoning with Time-Dependent Propositions. In *International journal of the IGPL*, to appear.

[4] H. Bürckert. *A Resolution Principle for a Logic with Restricted Quantifiers*, volume 568 of *LNAI*. Springer-Verlag, 1991.

[5] J. Euzenat. An algebraic approach to granularity in qualitative space and time representation. In *IJCAI*, pages 894–900, 1995.

[6] F. Gayral and P. Grandemange. Événements : Ponctualité et Durativité. In *Actes 7ième RFIA*, pages 905–910, Paris, 1991.

[7] L. Gosselin. *Sémantique de la temporalité en français. Un modèle calculatoire et cognitif du temps et de l'aspect*. Champs linguistiques - recherches. Duculot, 1995.

[8] Y. Halpern, Joseph and Y. Shoham. A Propositional Modal Logic of Time Intervals. In *Proceedings of the 1st IEEE Symposium on Logic In Computer Science*, pages 279–292. Computer Society Press, 1986.

[9] H. Herre and G. Wagner. Stable Semantics of Temporal Deductive Databases. *Proceedings of the 4th Workshop on Deductive Databases*, 1996.

[10] W. Klein. *Time in Language*. Cambridge University Press, 1995.

[11] P. Ladkin and R. Maddux. On binary constraint networks. Technical report, Kestrel Institute Technical Report KES.U.88.8, 1988.

[12] G. Ligozat. Weak Representations of Interval Algebras. In *Proceedings AAAI*, pages 715–720, 1990.

[13] G. Ligozat. On generalized Interval Calculi. In *Proceedings AAAI*, pages 234–243, 1991.

[14] A. Montanari. A metric and layered temporal logic for time granularity, synchrony and asynchrony. In H.-J. Ohlbach, editor, *ICTL*, pages 49–58, 1994.

[15] A. Montanari, E. Maim, E. Ciapessoni, and E. Ratto. Dealing with Time and Granularity in the Event Calculus. pages 702–712, 1992.

[16] E. Mota, M. Haggith, A. Smaill, and D. Robertson. Time Granularity in Simulation Models of Ecological Systems. In *Proceedings of the Workshop on Executable Temporal Logics*. IJCAI, 1995.

[17] E. Mota and D. Robertson. Representing Interaction of Agents at Different Time Granularities. In I. C. S. Press, editor, *Proceedings of the Third International Workshop on Temporal Representation and Reasoning*, 1996.

[18] E. Mota, D. Robertson, and A. Smaill. NatureTime: Temporal Granularity in Simulation of Ecosystems. *Journal of Symbolic Computation - Special Issue on Executable Temporal Logics*, 22(5 and 6), 12 1996.

[19] G. Wiederhold, S. Jajodia, and W. Litwin. Dealing with Granularity of Time in Temporal Databases. *Lecture notes in computer science*, 498:124–140, 1991.