A Visualization Method of Time Expressions Using Starting/Ending Point Plane

Mitsunori Matsushita, Masakatsu Ohta, and Toshiyuki Iida
NTT Communication Science Laboratories,
1-1, Hikarinooka, Yokosuka-Shi, Kanagawa, Japan.
{mat, ohta, iida}@cslab.kecl.ntt.co.jp

Abstract

We propose a method that visualizes the time transitions in written time expressions in order to analyze how the author used them. The method represents time expressions as points on a plane and connects them with lines. We apply the method to an adventure story and a company's history story and discovered two characteristics: First, during the progression of the text, the durations implied by the time expressions change from long to short. Second, at the start of a new paragraph, expressions tend to refer to events or periods of longer duration again.

1 Introduction

We are engaged in basic research into a method of understanding time expressions in order to create a computer that can understand and represent subjects, including time concepts, the way a human does.

How do humans understand and represent the passage of time for occurrences and events? Usually, people don't describe time uniformly but describe it using various levels of detail depending on the situation. We can observe people's way of describing time in their writings. When an author writes a story, if he (or she) judges an occurrence or an event to be important, it will be described in detail; therefore, the stream of time will also be described in detail. In contrast, if he (or she) doesn't judge it to be important or judges it to be background, it will be described more coarsely and the stream of time will also be described more coarsely. In writing, these changes of the description granularity give a rhythm to the story[1]. Because of this rhythm, readers are captivated by the story and understand the author's intention more precisely.

That is, people change the viewpoint of occurrences to comply with their relative importance or interest value, and this change can be observed via the stream of time that is described in the story. Thus, it is obvious that time expressions in the story play an important role in creating the rhythm.

The analysis of written time expressions is a subject in linguistics that has received limited attention in artificial intelligence research. To date, most AI research has concentrated on the analysis of verb structure[2][3], temporal connectives[4] and temporal database query[5]. How granularity can be used in AI has been studied by Bettini[6] and Combi[7].

Yet from the simple observations presented above, it is obvious that such considerations are important.

If a computer can understand how a human chooses time expressions properly, communication between the computer and the human will be more fluent, and the computer will be able to better understand what the human intends.

Therefore, the way of using time expressions in writing needs to be analyzed and the effects of change of granularity also needed to be analyzed.

The visual representation of time expressions will help with these analyses. Thus, in this paper, we discuss a method of visualizing of time expressions.

2 Conventional Method and Its Problem

Usually, time expressions are located on a numerical line. For example, the time expression "January" is expressed as in Figure 1. If the scale of the numerical line is "month", "January" is located as shown in Figure 1(a) because, in this case, its starting point and ending point are "January". If the scale is "day", it is expressed as in Figure 1(b) because its starting point is "January 1st" and its ending point is "January 31st".

By this method, we can easily see the temporal lo-

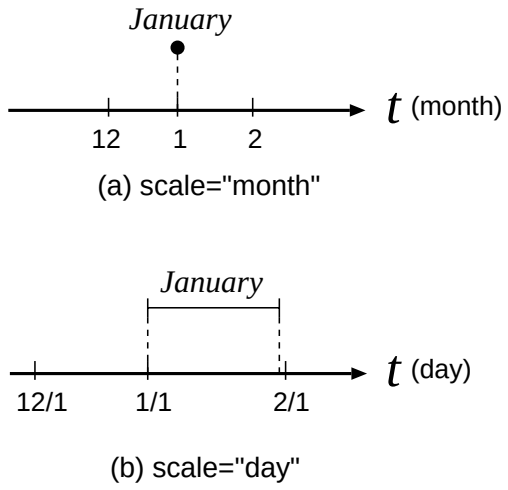

Figure 1: Visualizing "January" using conventional method.

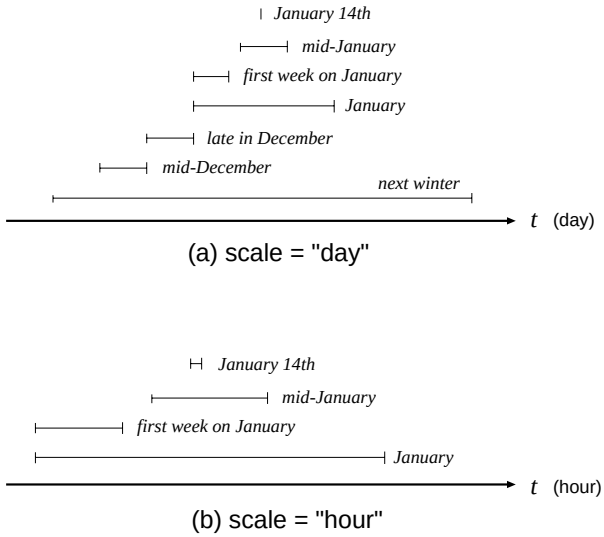

Figure 2: Visualizing several time expressions using conventional method.

cation and width of each expression. However, can the transition of the expressions also be easily seen? Here's an example conversation that includes time expressions.

A: " Let's go skiing *next winter.* "

B: " Ok, *mid-December* or *late in December* seems good to me."

A: " Oh, I may be not free. "

B: " How about *January* ?"

A: " As I have an appointment in *the first week on January*, *mid-January* is better."

B: " All right. Let's go on *January 14th.*"

This conversation includes six time expressions; "next winter", "mid-December", "late in December", "January", "mid-January", and "January 14th". How do these expressions change in this conversation? These expressions are shown in Figure 2(a) on the numerical line whose scale is "day".

For the sequence of time expressions in Figure 2(a), some might find it difficult to observe the characteristics of the transition of the expressions, each location and width can be understood easily though.

Moreover, these expressions are shown in different granularity by changing the scale of the numerical line. Figure 2(b) is the result of visualizing the last four expressions on the numerical line whose scale is "hour". Comparing Figure 2(a) and Figure 2(b), it is clear that it is difficult to determine the parts having identical transition when comparing scales of different granularities.

Thus, we propose a new visualization method in order to solve these problems.

3 Proposed Method and Its Characteristics

In this method, time expressions in a piece of text are located as follows on a plane whose axes represent the starting point and ending point. The scale of both axes is equal to the smallest scale in the time expressions. Regarding each of the time expressions as temporal interval $t_i = [a_i, b_i](a_i \leq b_i)$, they can be put on the plane. That is, the starting point and the ending point are regarded as the coordinates on the x-axis and y-axis respectively.

Let's think about the expression "January" as an example.

"January" is located using the proposed method as follows. If the scale of the axes of the plane is "month", "January" is located on the line $y = x$ as shown in Figure 3(a) because its starting point and ending point are the same. If the scale is "day", it is located as shown in Figure 3(b). Converting the scale of both axes corresponds to the change of granularity. Namely, expressions that are shown on the "month" scale correspond to a point of view at a coarser granularity, and expressions that are shown on "day" scale correspond

to a point of view at a finer granularity. Using the proposed method, every expression exists in the upper area, above the line $y = x$, in any scale because all expressions satisfy the relation $a_i \leq b_i$.

To see the transition of the time expressions, these are located on the plane and are connected in order of appearance in the text. The time expressions in the previous example conversation are expressed in Figure 4(a).

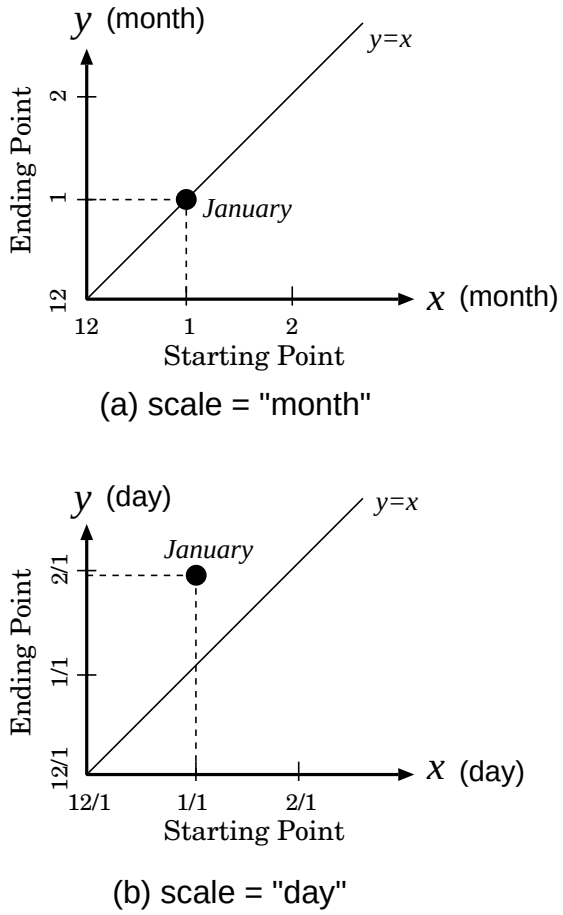

Figure 3: Visualizing "January" using proposed method.

These expressions are also shown at both granularities by changing the scale of the starting/ending point plane, as with the conventional method. The granularity is made finer by making the scale of both axes finer. All the time expressions recede from the line $y = x$ while keeping the same form. Figure 4(b) is the result of visualizing the last four expressions on the numerical line whose scale is "hour". Comparing Figure 4(a) and Figure 4(b), we find that the same time expressions maintain their shape regardless of the scale of the axes.

The graph has three characteristics as follows.

1. The absolute time increases from left to right.
2. The larger the time period is the further it lies from the straight line $y = x$.
3. Time expressions that have the same width line up on a straight line $y = x + k$ $(k \geq 0)$ regardless of the scale of axes. This variable k corresponds to

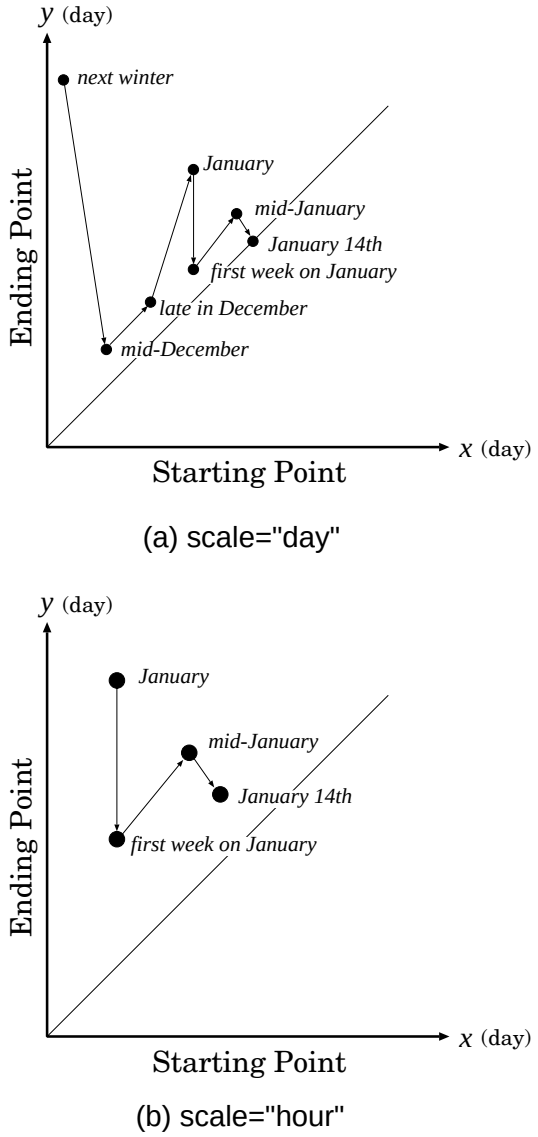

Figure 4: Visualizing several time expressions using proposed method.

the width of the time expressions that belong to the scale. For example, "Spring" and "Summer" line up on the same straight line because they have same width, but "Spring" and "January" don't line up on the same line regardless of the scale of the axes.

We classified time expressions that appear in written texts into four categories: fixed expression, quoted expression, expression of duration and expression of relation[8]. A fixed expression is a set of time expressions whose temporal location can be fixed themselves (ex. January, 4:00 p.m.). Quoted expression is a set of time expressions whose temporal locations are fixed by quoting fixed time expressions (ex. tomorrow, next year). Time expressions included these two categories are shown as points in the proposed method.

Expression of duration is a set of time expressions that indicates the distance between two temporal locations (ex. three weeks, one year). Time expressions included this category are shown as distances between two points on the proposed method.

Expression of relation is a set of time expressions that indicates the relationship among time expressions (ex.before, contains). Thirteen types of relations including "equal relation" defined by Allen[9] exist. Time expressions included this category are treated as constraints among time expressions. Figure 5 shows each relation of temporal location T. Relations from 1 through 6 are represented as regions, and relations from 7 through 12 are represented as line segments.

The temporal location of all expressions is not clearly definable. "One day" is one of such expressions even if it is used in a quoted expression. If a reference

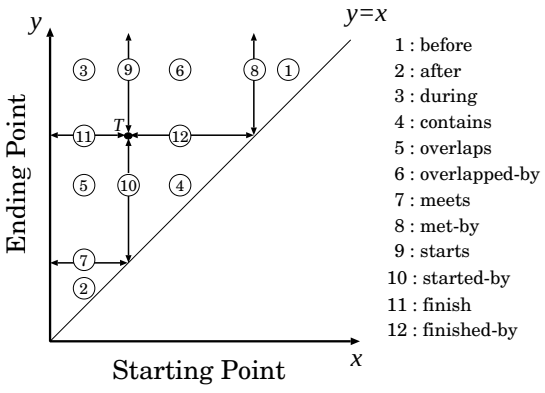

Figure 5: Temporal relationship on the starting/ending point plane.

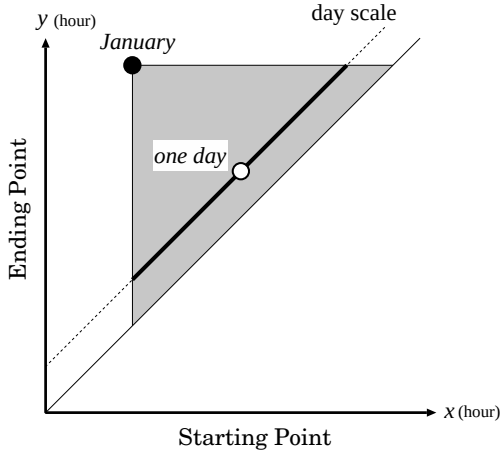

Figure 6: Example of determination of a region using constraints

point for an unknown time expression is fixed, the region of the unknown expression can be specified from its relationship to another expression. For example, "one day" was known to be contained in "January" from the context, so we can make two constraints for "one day":

(1) "one day" belongs to the "day" scale.

(2) "one day" is contained in "January".

From constraint (1), since "one day" is classified into a quoted expression, it is expressed as a point which is on the "day scale" line on the starting/ending point plane. From constraint (2), since the relationship between "January" and "one day" is "contain", the possible region of "one day" is restricted in a triangular region as shown in Figure 5. Figure 6 shows these two constraints. The dotted line shows constraint (1) and the gray region shows constraint (2). The location of "one day" needs to satisfy these two constraints, thus, the possible region is on the thick line in Figure 6. Of course, if it is possible to get more constraint regarding "one day" from a context, the possible region can be specified more accurately.

In dealing with a large number of time expressions, visualization using the proposed method is easier to understand than using a time-line because the transition can be seen as connected lines even if these expressions belong to different scales. Depending on these characteristics, we can analyze the transition of time expressions in writing.

4 Example

We applied the proposed method to an adventure story and a company's history story in order to test the expressive ability of the method.

4.1 Adventure Story Example

First, we targeted and analyzed a novel that was translated into Japanese suitable for the lower grades of elementary school. The novel that we focused on is an adventure story named "*Deux ans de vacances*" written by Jules Verne[10].

The reason we targeted an adventure novel is that it includes various expressions, and these expressions are used properly in response to the events in the story.

We focused on Chapters 4 and 6 and picked out time expressions and located them using the proposed method. Chapter 4 describes the details of one event, whereas Chapter 6 describes several events in turn. The maximum scales of the chapters are "day" and "year" respectively.

In this experiment, the starting point of the time expression that appears first in each chapter was placed at the origin, and we ignore time expressions included in conversation style because our aim is to see the transition of the story. Time expressions whose locations cannot be determined precisely like "one day" are located in the center of the possible region.

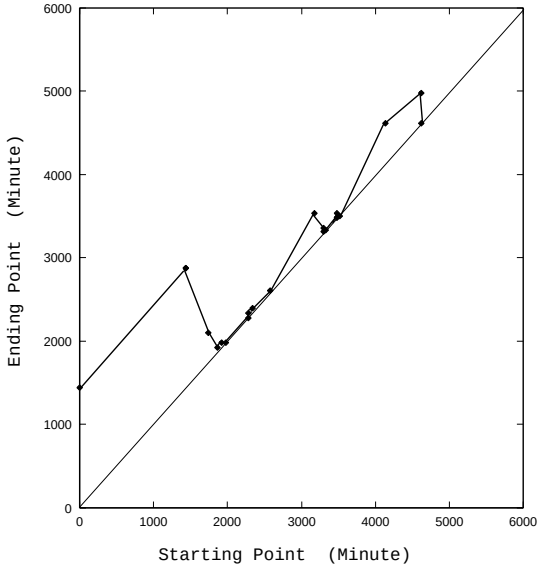

Figure 7: Transition of the Time Expressions in Chapter 4

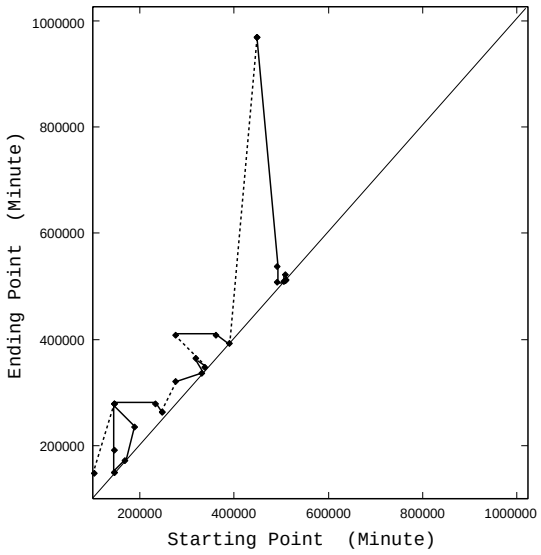

Figure 8: Transition of the Time Expressions in Chapter 6

Figure 7 shows the result of processing the time expressions included in Chapter 4 and Figure 8 shows the result of processing the time expressions included in Chapter 6. Dotted lines in these figures show the places where the topic changes in the text.

The change of scale in Chapter 4 is relatively small and the pattern of connected lines seems to be monotonous along the line $y = x$. In Chapter 6, on the other hand, there are time expressions on various scales and the pattern of the connected line is to repeatedly approach and recede from the line $y = x$.

The difference in the pattern between these two figures can be understood as a reflection of the characteristics of each chapter. That is, since every event is treated at almost the same granularity, few granularity change occurs in chapter 4 and the graph shows the transition of story monotonously. On the other hand, since Chapter 6 includes various events that are described in by various granularities, granularity change depends on the topic, and the graph shows the transition to a more complex story than in Chapter 4.

From these figures, we can see two characteristics. First, during the progression of the text, temporal width changes from long to short. Second, the width increases at the start of each new paragraph.

5 Company's History Story Example

We then analyzed a book of our company's history[11]. The content of Chapter 1 is the history of

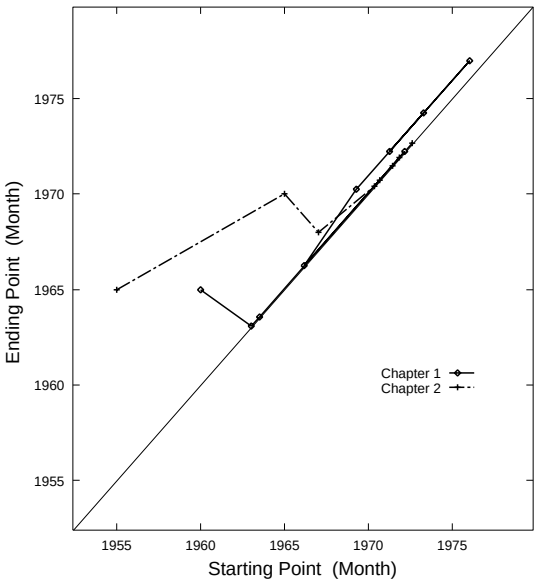

Figure 9: Transition of the Time Expressions in History story

the popularization of telephones, and Chapter 2 is the history of diversification of telephony services. Figure 9 shows the result of processing the time expressions and includes both chapters. The solid line indicates the result of chapter 1 and chained line indicates the result of chapter 2. The maximum scales of the chapters are "five year" and "decade year" respectively, and the minimum scale of both chapters is "month" scale.

The transition of both chapter seems to be similar pattern. The pattern starts at a far point from the line $y = x$ and then it approaches $y = x$ with the progress of the story. Furthermore, there are few severe time returns and scale change of time don't happen frequently.

From this figure, we can see two characteristics that are the same as found in the adventure story. First, during the progression of the text, temporal width changes from long to short. Second, the width increases at the start of each new paragraph. Thus, it is expected that the story that has such temporal structures is easy to understand for readers.

Comparing these two examples, the appearance pattern of time expressions seems to be different depending on the category of story. We think the rhythm of adventure story is switched every chapter in order to attract reader more active than company's history story.

6 Conclusion

We proposed a method that visualizes the transitions of time in written time expressions in order to analyze how the author used them. The method places the time expressions on a plane as points and connects them with lines. Using this method, the time expressions in a story can be visualized as a pattern. As an example, we applied the method to an adventure story and a company's history written in Japanese.

Our method falls into the category of research on the temporal structure in discourse[12][13]. We think it can be an effective tool in making a cognitive model for the use of time expressions and thus, its application to AI.

Now we are constructing a visualization tool based on the proposed method[14].

Presently, the proposed method can handle explicit time expressions like today and yesterday. Expressions that can be related to an explicit time can also be easily represented. For example, "When the accident happened" can be graphed if one knows the explicit time it happened. However, these kinds of expressions, as mentioned in [15] cannot be easily represented by themselves by the visualization method. In order to know the stream of time in text more exactly, a way of handling other time expressions must be developed.

We'll analyze which kind of expressions contribute to creating rhythm in writing and consider ways of visualizing such expressions. We are also interested in fusing the proposed method and pattern recognition techniques.

References

[1] Genette, G.: "Figures III," Seuil (1972).

[2] Moulin, B.: "The Representation of Linguistic Information in an Approach Used for Modelling Temporal Knowledge in Discourses," *Lect. Notes Comput. Sci.*, Vol.699 pp. 182–204 (1993).

[3] Moulin, B.: "The Temporal Structure of a Discourse and Verb Tense Determination," *Lect. Notes Comput. Sci.*, Vol.835 pp. 45–68 (1994).

[4] Lascarides, A. and Oberlander, J.: "Temporal Connectives in a Discourse Context," *Proc. of the 6th Conf. Eur. Chapter Assoc. Comput. Linguist.*, pp. 260–268 (1993).

[5] Chandra, R. Stonebraker, M. and Segev, A.: "Implementing Calendars and Temporal Rules in Next Generation Databases," *Proc. of the 10th Int. Conf. Data Eng.*, pp. 264–273 (1994).

[6] Bettini, C., Wang, X. C., and Jajodia, S.: "A General Framework and Reasoning Models for Time Granularity," *Proc. of TIME-96*, pp. 104–111 (1996).

[7] Combi, C., Pinciroli, F., and Pozzi, G.: "Managing Time Granularity of Narrative Clinical Information: The Temporal Data Model TIME-NESIS," *Proc. of TIME-96*, pp. 88–93 (1996).

[8] Matsushita, M., Makino, T. and Iida, T.: "A Method of Choosing Time Expressions with Granularity (in Japanese)" *Proc. of the 13th Fuzzy System Symposium*, pp. 425–428 (1997).

[9] Allen, J. F.: "Towards a General Theory of Action and Time," *Artif. Intell.*, Vol.23, No.2, pp.123–154 (1984).

[10] Verne. J.: "Deux ans de vacances (translated into Japanese by Segawa, M.)," Syueisha (1994).

[11] The editing committee of NTT's history: "The history of Nippon Telegraph and Telephone Corporation (in Japanese)," Information and Communication General Research Center (1986).

[12] Gagnon, M. and Lapalme, G. : "From Conceptual Time to Linguistic Time," *Comput. Linguist.*, Vol. 22, No. 1, pp. 91–127 (1996).

[13] Rodrigues, I. P. and Lopes, J. G. : "Temporal Structure of Discourse," *Proc. COLING-92*, pp. 331–337 (1992).

[14] Matsushita, M., Ohta, M. and Iida, T.: "A Visualization Tool for Granularity Using Time Expressions (in Japanese)," *Proc. of the 54th Annual Convention IPS Japan*, Vol.2, pp. 39–40 (1997).

[15] Gayral, F. and Grandemange, P. : "Une ontologie du temps pour le langage naturel (in French)," *Proc. COLING-92*, pp. 295–302 (1992).