A System for Extraction of Temporal Expressions from French Texts Based on Syntactic and Semantic Constraints

Nikolai Vazov
Groupe sur l'asymetrie des langues naturelles - UQAM
Montréal (Québec) Canada H3C 3P8
c1144@er.uqam.ca

Abstract

We present a system for extraction of temporal expressions from French texts. The identification of the temporal expressions is based on a context-scanning strategy (CSS) which is carried out by two complementary techniques: search for regular expressions and left-to-right and right-to-left local chart-parsing. A number of semantic and distant-dependency constraints have been integrated to the chart-parsing procedure in order to improve the precision of the system.

1 Introduction

The identification and the interpretation of temporal and aspectual information plays an important role in text understanding. This information is encoded in the natural languages by a wide array of linguistic means ranging from grammatical (morpho-syntactic) to lexical (verbs and adverbials) or strictly syntactic phenomena (temporal anaphora (Webber, 1988) or argument structure of the verb (Verkuyl, 1993)).

In this paper we present a system for identification of lexical non-verbal means of expressing temporal information in French texts. The system detects adverbs (e.g. *yesterday*), adverbial noun phrases (e.g. *three days before the marriage of his youngest son*) and adverbial expressions (e.g. *at 10 o'clock*), regardless of their number and complexity. Unlike some other approaches in temporal annotation (Ferro et al., 2001), this system identifies exclusively temporal expressions answering *when*-questions. Hence, the system will flag expressions which specify the time of occurrence of the situations (e.g. *John received the letter on May 10th*) and will skip those modifying lexical units (e.g. *John received her letter of May 10th*).

The identification of the temporal expressions is based on a context-scanning strategy (CSS) (Desclés, 1997) which is carried out by two complementary techniques: search for regular expressions and left-to-right and right-to-left chart-parsing.

The first technique looks for a particular set of markers (regular expressions) encoding temporal information. These markers can be stand-alone or parsing-triggering. The stand-alone markers are inactive chart elements (not necessarily a single word) which represent autonomous temporal expressions. The adverb *then* in *Then, this was the only solution* is an example of a stand-alone marker.

The second technique of the CSS is launched, if the system identifies a parsing-triggering marker. The parsing-triggering markers are active chart elements which signal the presence of a larger temporal expression (for example *before* in *three weeks before Christmas*) and trigger a chart-parsing procedure for its identification. The chart-parsing scans the left and the right context of the marker in order to determine which lexical units from this context belong to the temporal expression. The identified entire temporal expression represents a temporal marker (inactive chart element) generated incrementally by association of chart elements situated to the left and to the right of the parsing-triggering marker (the initial active chart element) (Wonsever and Minel, 2001).

The present paper is organised in 5 sections. Section 2 is a short presentation of the CSS as a method for goal-oriented information extraction. Section 3 discusses the classification of the markers into stand-alone and parsing- triggering. Section 4 shows a procedure launched by stand-alone markers, namely, the identification of *exclusion zones*. In turn, section 5 presents the chart-parsing rules with the various semantic and syntactic constraints on their realisation. The architecture of the system is shown in section 6.

2 The underlying principles of the context-scanning strategy

The context-scanning strategy (CSS) (Desclés et al., 1997; Minel and Desclés, 2000) is based on the hypothesis that the semantic representations are not necessarily determined by world knowledge but can be considered as configurations integrating elements of local linguistic information. The interaction of these elements is controlled by a set of mechanisms which build the semantic representation by putting together all the pieces of local information. The elements of this representation can be derived from grammar or lexical semantics and from some syntactic phenomena which together account for the non-atomic (holistic) compositionality of semantic representations.

This hypothesis allows to formulate the two distinctive features of the CSS:

- The CSS is designed to carry out a very specific task: to identify and interpret all the surface linguistic elements which account for a particular semantic representation. It does not claim to provide a complete analysis of texts.
- The CSS extracts sets of interactive elements integrated in complex semantic representations. Unlike the approaches based on world knowledge, it *deduces* these representations using exclusively linguistic data found in the text.

The system presented in this paper satisfies both features. The specific task of the system (first feature) is to indentify all non-verbal linguistic units which can convey temporal information and to group them in sets. These sets are called temporal expressions and may contain one or more members.

All temporal expressions are considered as semantic functions whose argument is the semantic value of the grammatical tense in the analysed utterance. The result of the application of this function to its argument is the semantic (aspecto-temporal) value of the entire utterance. Thus, the identification of the temporal expressions is an important step in the deduction of a semantic representation which integrates information from different linguistic units (second feature). Utterances in 1 and 2 show the role of the temporal expressions in this deduction. Example 1 denotes an event which took place before the utterance time. In turn, the lack of temporal expression in 2 entails a double interpretation: the utterance can refer either to the fact of Jeanne's departure in a past moment, or to the fact that Jeanne is absent at the moment of utterance. This ambiguity is manifested in the English translation of 2.

(1) *Jeanne est partie trois minutes avant Pierre.*
Jeanne left three minutes before Pierre.

(2) *Jeanne est partie.*
Jeanne left/has left.

The discussion in this paper will focus only at the first feature of the implemented CSS, namely, the identification and regrouping of linguistic units into temporal expressions.

As it was mentioned in the introduction, the identification of the temporal expressions is carried out at two steps: identification of markers (stand-alone and parsing-triggering) and chart-parsing applied to the left/right context of the parsing triggering markers. These steps will be discussed in detail in the three following sections.

3 Selection of the markers

All temporal markers (regular expressions) detected by the system at the first step of the analysis fall into two sets: Σ and M ($\Sigma \cap M = \emptyset$).

3.1 Stand-alone markers

The set Σ contains the stand-alone markers which represent autonomous temporal expressions. These expressions can be represented by the following context-scanning rule:

$$TemporalExpression \rightarrow \sigma \mid \sigma \epsilon \Sigma \quad (1)$$

The stand-alone markers are grouped in two subsets ($\Sigma_1 \cap \Sigma_2 = \emptyset$), defined with regard to the structure of their elements.

- markers (Σ_1) representing constant strings (for example, *par la suite* 'after that', *le lendemain matin* 'the next day's morning')
- markers (Σ_2) specifying the initial element of a string, considered as a temporal expression (for example, *quand* 'when'), and a set of following lexical units bound to the right by a particular syntactic category

The set of lexical units following the markers of Σ_2 consitutes an *exclusion zone*. The identification of this zone is discussed in section 4.

3.2 Parsing-triggering markers

The markers included in the set M are both indicators and constituent elements of a larger temporal expression. They are the active elements of a chart-parsing procedure which they launch in order to determine the boundaries of this expression in the analysed sentence.

The temporal expressions signaled by the markers from this set can be represented by the following contextual rule:

$$TemporalExpression \rightarrow LC.m.RC \mid m \epsilon M \quad (2)$$

where

- LC is the left context of the marker *m*. It contains an array of strings (lexical units) whose syntactic categories have been authorised by the right-to-left chart-parser (see section 5)
- LC can be empty
- "." is the operator of concatenation
- RC is the right context of the marker *m*. It contains an array of strings (lexical units) whose syntactic categories have been authorised by the left-to-right chart-parser (see section 5)
- RC can be empty

All the markers in M have been organised in three subsets M_1, M_2 and M_3 ($M_1 \cap M_2 \cap M_3 = \emptyset$), according to the particularities of the chart-parsers they trigger:

- Markers triggering a left-to-right chart parser (M_1). These markers represent the leftmost lexical unit of the larger temporal expression and launch parsing rules applying exclusively to the right context of the marker. This set contains markers like *il y a* 'ago', *au cours* 'during', etc.

- Markers triggering a left-to-right *and* a right-to-left parser (M_2). The markers included in this set never occur at the leftmost or at the rightmost position in the larger temporal expression. Hence, the need to analyse both their left and right context in order to identify the boundaries of the temporal expression they belong to. This is the largest of the three sets and contains elements like the names of the months, temporal units (*minute, seconde,..*), names of the 4 seasons, etc.

- Markers requiring a left-to-right *and* a right-to-left parser (M_3). These are the markers which can occur at any position in the larger temporal expression. However, the left and right parsing rules may eventually rule out the left *and/or* the right context as not belonging to the temporal expression and return the detected marker as the only constructive element of the temporal expression. An example of such a marker is *après* 'after' in the following sentences:

  (3) *Jeanne est arrivée trois minutes après Pierre.*
  Jeanne came three minutes after Pierre.

  (4) *Jeanne est arrivée trois minutes après.*
  Jeanne came three minutes later.

  (5) *Jeanne est arrivée après Pierre.*
  Jeanne came after Pierre.

  (6) *Jeanne est arrivée après.*
  Jeanne came later.

Example 3 shows a larger temporal expression (*trois minutes après Pierre*) which extends to the left and to the right of the detected marker *après*. In example 4, the same triggering marker has only a left context (*trois minutes*), the right one being interpreted by the parser as empty (punctuation mark stands for the end of a temporal expression). Conversely, in example 5, the parsing rules will authorise the association of elements (*Pierre*) from the right context of the marker but will block the association of such elements (*arrivée*) from its left context. Finally, following the same line of reasoning as in examples 4 and 5, the parsing of example 6 will return the marker itself.

The above examples explain why it is important to distinguish between the three sets of triggering markers. The M_1 markers signal temporal expressions which only extend to the right of the marker's position. The M_2 markers signal temporal expressions which extend to the right *and* to the left of the markesr's position. Alone, M_2 markers do not constitute a temporal expression. The markers of M_3 can be used alone as temporal expressions *or* constitute such an expression together with other elements from their left *and/or* right context.

The total number of markers detected by the system is 121. They are subdivided as follows: Σ_1 (7), Σ_2 (36), M_1 (13), M_2 (51), M_3 (14).

4 Exclusion zones

The temporal expressions introduced by conjunctions like *lorsque* 'when', *en même temps que* 'in the same time as', etc (elements of the set Σ_2) are, with very few exceptions, subordinate temporal clauses. Thus, the extraction of these temporal expressions comes down to the identification of the boundaries of these subordinate clauses, in particular, of their right end, usually signalled by a punctuation mark. The lexical units between the conjunction and the punctuation mark form an *exclusion zone* (Wonsever and Minel, 2001), since they become integrated into the temporal expression under a negative condition: the elements of this set should *not* belong to the syntactic category of PONC [1].

[1] The system uses the shareware on-line tagger of LATL.ch S.A.(Société Anonyme) (www.latl.ch). The notation for the syntactic categories in this discussion follows the notation provided by the tagger: DET (determiner), PRE (preposition), ADJ (adjective), PRO (pronoun), NUM (numeral), NOC (common noun), ADV (adverb), INF (infinitive), PPA (past participle), PPR (present participle), CLI (clitic), COJ (conjunction), NOP (proper noun) and PONC (all punctuation marks).

(7) *J'aurais dû proposer les motions nos 49 et 50 en même temps que la motion no 44.*
I should have put Motions Nos. 49 and 50 at the same time as Motion No.44.

However, the subordinate clauses may contain inserted segments, separated by various punctuation marks which do not fix the end of the temporal expression. In order to handle such segments, the system will scan the exclusion zones in two steps.

The goal of the first step is to detect the verb (VERB category) in the right context. If detected, the VERB category starts the second step whose goal is to find a PONC (punctuation category) to the right of the verb. The latter signals the end of the exclusion zone (in curly braces in 8).

(8) *Lorsque {le premier ministre a annoncé*$_{[VERB]\,step1}$ *qu'il prenait sa retraite},*$_{[PONC]\,step2-endzone}$ *de nombreux hommages lui ont été rendus à la Chambre.*
When the Prime Minister announced his retirement there were many tributes paid in the House.

Processing of the exclusion zone in two steps allows to jump over potential inserted segments separated by punctuation marks, as in 9 (in curly braces).

(9) *J'ai été saisie d'une vive émotion ... lorsque{[, de ma place, dans les gradins du Centre municipal,]*$_{inserted\,segment}$ *j'ai*$_{[VERB]}$ *vu le député de Burnaby-Kingsway courir vers moi}*$_{endzone}$.
I had a moment of emotional enthusiasm ... when I was sitting in the stands in my section at the Civic Centre and I saw the hon. member for Burnaby-Kingsway racing toward me.

5 Left and right chart-parser rules

5.1 General principles

The left and right chart-parser rules triggered by the M_1, M_2 and M_3 markers are recursion-based algorithms. These rules scan the tagged right and left context of the marker and control the association of the adjacent lexical units to the body of the temporal expression signaled by this marker. The control is carried out by the following context-scanning parsing rules:

Rule for the left context:

```
while cat_{w^n} ∈ C_left
  m = w^n.m
  n = n - 1
```

where

- w is a lexical unit from the LC
- n is the position of w in LC
- *cat* is the syntactic category of w
- m is temporal expression containg only the parsing-triggering marker before the application of the rule and completed at each loop by a new element from the LC
- "." is the operator of concatenation
- C_{left} = {DET, PRE, ADJ, PRO, NUM, NOC} - the set of categories authorised in the positions preceding the marker m

Rule for the right context:

```
while cat_{w^n} ∈ C_right
  m = m.w^n
  n = n + 1
```

where

- w is a lexical unit from the RC
- n is the position of w in RC
- *cat* is the syntactic category of w
- m is the temporal expression containg only the parsing-triggering marker before the application of the rule and completed at each loop by a new element from the RC
- "." is the operator of concatenation
- C_{right} = {DET, ADJ, NUM, NOC, ADV, INF, PPA, PPR, CLI, COJ, NOP} - the set of categories authorised in the positions following the marker m

Each rule tries to recursively match a category from a built-in set (C_{left} and C_{right}) to the category of the next lexical unit in the context. If the two categories match, the lexical unit is associated to the temporal expression and the rule moves on to check the category of the next lexical unit. In turn, if the two categories do not match, the rule blocks the association of the unit to the temporal expression and signals the left, respectively, the right boundary of the temporal expression. The analysis of example 10 below shows the scanning procedure step by step (the marker *heures* has been detected at a previous stage):

(10) *Ils ont quitté Air France vers 10 heures pour une courte promenade sur les Champs-Elysées.*
They left Air France about 10 o'clock for a short walk on Champs-Elysées.

left pass 1 the NUM category of the lexical unit '10' matches the built-in category NUM in the scanning rule

left pass 2 the PRE category of the lexical unit 'vers' matches the built-in category PRE in the scanning rule and associates it to the already processed left context: *vers 10* left pass 3 the NOP category of the lexical unit 'France' is not in the rule's list of categories - the rule cannot move further left and the system returns: *vers 10*

right pass 1 the PRE category of the lexical unit 'pour' is not in the category list of the rule - the rule cannot move furhter right and the system returns an empty string

The string *vers 10 heures* constitutes the larger temporal expression signaled initially by the triggering marker *heures*. Formulated in terms of the rules above it will have the form: $w^6.w^7$.m.$\emptyset$ (7 & 6 are the positions of the lexical units in the LC).

5.2 Constraints in the chart-parsing rules

5.2.1 Adjacency constraints on the positions of syntactic categories

The matching of categories discussed in the previous section is achieved by context-free rules - one for the left and one for the right context of the parsing-triggering marker, respectively.

However, this context-free algorithm will fail to detect the boundaries of the temporal expression in a number of cases where, following a successful category matching, it will associate to this expression some elements which do not belong to its body. Without additional restrictions to the rules, the system will provide the following wrong right-context analysis for example 11 (the marker *année* was detected at a previous stage):

(11) *L'année dernière les actions de cette entreprise ont augmenté de 30%.*
Last year the shares of this company rose by 30%.

pass 1 category of *dernière* is ADJ : authorised

pass 2 category of *les* is DET (determiner) : authorised

pass 3 category of *actions* is NOC : authorised

pass 4 category of *cette* is DET : authorised

pass 5 category of *entreprise* is NOC : authorised

pass 6 category of *ont* is VAU (auxiliary): blocked

Since the syntactic categories DET and NOC are in the list of the authorised categories, they become accepted by the system regardless of their position. The result of this overgeneration will be the segment *l'année dernière les actions de cette entreprise* 'last year the shares of this company'.

To avoid this problem, the above context-free rules have been modified to check the syntactic categories which are adjacent to the syntactic category under analysis and to verify, if a number of adjacency constraints have been respected. An example of such a constraint is the following rule:

If ADJ is immediately preceded by a NOC or by an empty context, and if it is immediately followed by DET, the DET is not an element of the temporal expression.

In example 11 this constraint will block the association of new elements to the right of the ADJ *dernière* due to the empty context preceding the ADJ[2] and to the DET category (*les*) following the adjective.

5.2.2 Semantic constraints on the lexical units

Another constraint in the chart-parsing rules is the condition on the semantics of the lexical units in the context of the marker. This constraint affects exclusively the nouns in the left context of the markers from the set M_3 (see section 3). Actually, the left context of these markers is a modifier of the expression constitued by the marker and its right context. Thus, *three minutes before his departure* can be represented as (`three minutes`(`before his departure`)) where `before` $\epsilon\, M_3$ and `his departure` is its right context. The internal organisation of these configurations determines the semantics of the nouns occurring in the left context of the marker: they have to denote a time period, like minutes, seconds, seasons, etc. The set of these nouns is actually a subset of M_2 - the set of markers containing the names of months, temporal intervals, etc...(see section 3). In order to authorise the association of the noun to the temporal expression, the scanning rule checks if this noun is a member of this subset. If this is the case, the rule moves left to the next lexical unit. And conversely, if the noun is not a member of this set, the associaiton is blocked and the rule signals the boundary of the temporal expression.

5.2.3 Symmetric-dependency constraints

Symmetric-dependency constraints are indispensable for processing coordinate structures. First, they prevent the splitting of single temporal expressions like *au mois de mars et au mois d'avril* 'in March and in April' into *au mois de mars* and *au mois d'avril*, respectively:

(12) *Les syndicats ont organisé deux réunions au mois de mars et au mois d'avril.*
The trade unions organised two meetings in March and in April.

[2] The adjective is the leftmost element of the right context of the marker *année*.

Second, they allow to rule out the possessive use of *de* or *du*, as in 13:

(13) *Toutes les conditions sont réunies pour vous permettre de gagner un billet avec vos Miles de cet hiver.*
We do our best to let you win a free ticket for the Miles from (gathered) this winter.

and to keep coordinate structures, as in 14:

(14) *De 1980 à 1985, des milliers de Canadiens ont pu être contaminés par le VIH.*
From 1980 to 1985 thousands of Canadians could have become contaminated by HIV.

The identification of the expression in 12 is carried out by a rule looking for a symmetric dependency between the left and the right context of the parsing-triggering marker (*mars*). It checks if its right context element, immediately following the coordinate conjucntion (*et*), matches the beginning of its left context. If the beginning of the two strings matches (*au ... au*), the parser will move to the next element in the context without splitting the temporal expression. Otherwise, the system returns an empty right context.

Example 14 will be processed by a similar rule, except that the contexts will be checked in a reversed order. If an occurrence of *de/du* has been detected in the left context of the marker, this rule will check for a "symmetrical" preposition ({*à, jusqu'à, jusqu'au*}) in the marker's right context. If such a preposition is present, the system will continue the chart-parsing to the right. Otherwise, it will skip the marker and will resume the search in its left and right context.

5.2.4 Relative clause constraints

Identification of the boundaries of *qui* 'who' relative clauses in the right context of the markers plays an important role in the extraction of temporal expressions.

If the relative clause modifies a grammatical subject, the latter is not included in the temporal expression (example 15)[3]:

(15) *La réunion a commencé et {3 minutes après} son porte-parole qui avait déjà annoncé la bonne nouvelle est parti pour la capitale.*
The meeting began, and 3 minutes later his spokesman who had already announced the good news left for the capital.

In turn, if the relative clause modifies another argument of the predicate, this argument is included in the temporal expression:

(16) *Le ministre est venu {3 minutes après son porte-parole} qui avait déjà annoncé la bonne nouvelle.*
The minister came 3 minutes after his spokesman who had already announced the good news.

To identify the role of the modified element, the system counts the number of tensed verbs (VERB categories) to the right of the relative pronoun.

If $n_{verbs} = 1$, then the detected verb belongs to the relative clause. Hence, the modified element is not a subject (example 16) and is included into the temporal expression.

If $n_{verbs} > 1$, the system performs one more test - it checks if the *last detected verb* is preceded by *et qui* or *et*. This last test reflects the empirical observation that the coordinated predicates in a relative clause can occur exclusively in the following schemata (*et* may or may not be followed by *qui*):

- qui .. Verb_1 .. et (qui) .. Verb_2 ..
- qui .. Verb_1 .. , (qui) Verb_2 ..et (qui) .. Verb_3 ..

If the last test detects *et qui* or *et* to the left of the last verb, the latter is a member of a pair of coordinated predicats of the relative clause, modifying a non-subject argument. Such a construction would be processed similarly to example 16, the only difference being the number of verbs in the relative clause.

Else, if the last verb is not preceded by one of the above elements, it belongs to the main clause and stands for the predicate of the grammatical subject - the element modified by the relative clause (example 15).

6 Architecture of the system

The overall architecture of the system is shown in Figure 1. The input text is tokenized into sentences and a search for stand-alone markers is launched.

If the system detects such a marker, it extracts it from the sentence and resumes the search within its left and right context. If both contexts are empty, the index of sentences is incremented and the same procedure repeats for sentence $_{n+1}$. In turn, if the system fails to detect a stand-alone marker (during the first or any of the following loops), it scans the sentence for parsing-triggering markers.

If such a marker has been identified, the system launches the chart-parsing rules with various constraints in order to identifiy the larger temporal expression signaled by the parsing-triggering marker. Once identified, the larger temporal expression is extracted from the sentence and its

[3] The punctuation, like commas, is an important source of information about the temporal expressions' boundaries. Naturally, the system's scanning rules take it into account. The constraints discussed in section 5.2.4 are needed to process constructions which do not contain punctuation marks.

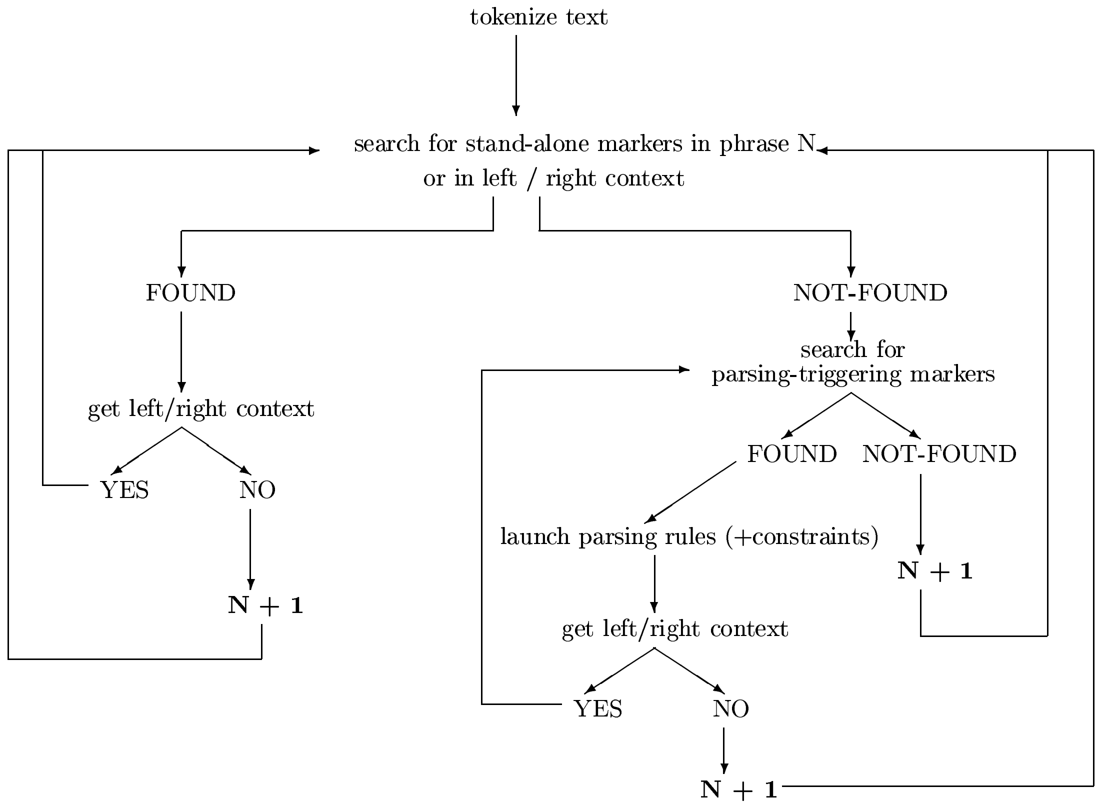

Figure 1: The flowchart of the system

left and right contexts are scanned again for other markers of the same type. This loop will repeat until there is no remaining left and right context. Then the system increments the index of the sentences and resumes the search for stand-alone markers in the next sentence.

7 Evaluation and conclusion

The system presented in this paper is developed in Java (JDK 1.3) and extracts non-verbal temporal expressions from non-annotated French texts using exclusively linguistic markers and local chart-parsing techniques.

So far, approximately 25% of the markers have been tested on a large corpus (Hansard Canadien 1986-1993, available in www-rali.iro.umontreal.ca/TransSearch/ and containing about 50 mln words).

The performance of the system has been evaluated using the following procedure:

1. A set of sentences containing occurrences of markers in Σ_1, Σ_2, M_1, M_2 and M_3 have been selected manually from the corpus. This set contains approximately 100 sentences for each marker.

2. The relevant temporal expressions in these sentences have been identified manually. Then, the same set of sentences has been processed by the system.

The tests yield a recall of approximately 95% and a precision of about 85%. The recall results are determined mainly by the list of markers (keywords) which unfailingly point to a relevant temporal expression. Hence, if the system fails to identify a temporal expression detected by a human, this would mean that this expression contains a keyword which is missing in the built-in list. Adding this keyword to the list will allow to update the system without any major modifications. Some minor problems in the search for regular expressions reduce the recall rate to 95%.

The lower rate of the precision results (85%) is due to two problems:

- the extracted expression contains a temporal marker but does not represent a relevant temporal expression (e.g. *un billet primes avec vos Miles* ***de cet hiver*** 'a winning ticket with your miles of this winter')

- the boundaries of the extracted expression are not correctly identified (this problem occurs in processing exclusion zones as well as during the chart-parsing)

The processing time for a sentence containing an average number of temporal expressions (2) ranges from 0.2 to 7.9 seconds. This considerable variation is due to three factors:

- type of identified marker which may or may not require a chart-parsing
- number of elements in the sentence which have to be tested as potential temporal expressions due to their formal characteristics (like seasons' names, digits, etc.)
- connection speed (for the on-line POS tagger)

The following real example, containing 2 standalone markers (`samedi` and `vendredi`) and 1 parsing-triggering one (`lors d'`) is processed in 6 sec. approximately.

*****TEXTE*****

```
Trois mille séminaristes, jeunes prêtres
et étudiants ont souhaité vendredi un bon
anniversaire au pape qui fêtera samedi ses
21 ans de pontificat, en lui chantant en
polonais " stolat " ( " cent ans " ) , lors
d'une messe dans la basilique Saint-Pierre
```

*****FIN*****

samedi
vendredi
lors d'une messe
PHRASE 1 ANALYSEE ** NOMBRE D'EXPRESSIONS REPEREES: 3 ** TEMPS DE TRAITEMENT 5.89 sec

The results obtained by the system can be further used in four different fields:

- Information extraction: the temporal expressions retrieved by the system provide substantial information about the temporal localisation of events described in the analysed texts (Ben Hazez and Minel, 2000).
- Annotating temporal information and time tagger development: TEMPEX v 1.03 Time Tagger (Mani et al., 2001)
- Computational semantics: the semantics of the temporal expressions (for example, the set of expressions referring to past events) along with the semantics of grammatical tenses is a constituent element of the aspecto-temporal value of the sentence (Battistelli, 2000; Vazov, 1999)
- Parsing: the temporal expressions are syntactic constituents whose identification by parsing techniques is a difficult task. The identification of this constituent will fix its boundaries and hence simplify the task of the parser

8 Acknowledgements

This research has been supported by the Social Sciences and Humanities Research Council of Canada (grant # 412-97-0016), for the "Assymmetry and Natural Language Processing Project", awarded to professor Anne-Marie Di Sciullo, in the Département de Linguistique at Université du Québec à Montréal (UQAM).

I would also like to thank Professor Guy Lapalme from Université de Montréal for his valuable comments and remarks on the draft of this paper.

References

D. Battistelli. 2000. *Passer du texte à une séquence d'images, analyse spatio-temporelle de textes, modelisation et réalisation informatique (système SPAT)*. Ph.D. thesis, Université Paris - Sorbonne.

S. Ben Hazez and J.-L. Minel. 2000. Designing tasks of identification of complex patterns used for text-filtering. In *RIAO*, pages 1558–1567.

J.-P Desclés, E. Cartier, A. Jackiewicz, and J.-L. Minel. 1997. Textual processing and contextual exploration method. In *CONTEXT'97*, pages 189–197, Brasil. Universidade Federal do Rio de Janeiro.

J.-P. Desclés, 1997. *Systèmes d'exploration contextuelle. Co-texte et calcul du sens*, pages 215–232. Presses Universitaires de Caen.

L. Ferro, I. Mani, B. Sundheim, and G. Wilson. 2001. Tides temporal annotation guidelines. Technical Report MTR 00W0000094, The MITRE Corporation.

I. Mani, L. Ferro, B. Sundheim, and G. Wilson. 2001. Guidelines for annotating temporal information. In *Notebook Proceedings of Human Language Technology Conference 2001*, pages 299–302, San Diego, California, March 18-21.

J-L. Minel and J-P. Desclés. 2000. *Résumé Automatique et Filtrage des textes.* Ingénierie des langues. Editions Hermès, Paris.

N. Vazov. 1999. Context-scanning strategy in temporal reasoning. In *Modeling and Using Context (Second International and Interdisciplinary Conference, CONTEXT'99)*, pages 389–402. Springer-Verlag.

H. Verkuyl. 1993. *A theory of aspectuality. The interaction between temporal and atemporal structure.* Cambridge Studies in Lingustics. Cambridge University Press.

B. Webber. 1988. Tense as discourse anaphor. *Computational Linguistics*, 14(2):61–73.

D. Wonsever and J.-L. Minel. 2001. Contextual rules for text analysis. In *2nd International conference on Intelligent Text Processing and Computational Linguistics CICLing-2001*, Mexico City.