UCL Université catholique de Louvain

Faculté de philosophie, arts et lettres (FIAL)

FIAL
Un éventail de formations, une multitude d'horizons

La temporalité dans le domaine médical : extraction et normalisation

Mémoire réalisé par
Pauline Neuzy

Promoteur(s)
Cédrick Fairon

Année académique 2016-2017
Master en linguistique, finalité en traitement automatique du langage

Université Catholique de Louvain

Faculté de Philosophie, Arts et Lettres

Master en linguistique, finalité en Traitement Automatique du Langage

La temporalité dans le domaine médical

extraction et normalisation

Mémoire réalisé par :

Pauline Neuzy

Promoteur :

Professeur Cédrick Fairon

Année académique 2016-2017

Résumé

À travers les âges, nous sommes passés d'un temps observé à un temps fabriqué, puis, aujourd'hui, à un temps mesuré. Et même, informatisé. Nous vivons dans une société où tout va toujours plus vite, où tout s'automatise chaque jour davantage, et où la technologie joue un rôle essentiel.

Le domaine médical offre un terrain particulièrement propice à l'utilisation des nouvelles technologies : nous voulons d'un seul clic avoir accès à l'historique médical du patient, aux antécédents familiaux,... La reconnaissance des expressions temporelles, leur normalisation et l'identification des événements médicaux liés à ces informations temporelles sont devenus des sujets de recherches très répandus depuis quelques années.

Dans ce dessein, nous proposons Time extractor, un système à base de règles qui permet la reconnaissance et la normalisation des expressions temporelles en langue française dans un corpus médical. Nous avons aussi abordé la contextualisation de ces expressions avec les événements médicaux auxquels elles sont liées.

Mots-clés : Reconnaissance d'informations ; extraction d'informations ; extraction temporelle ; annotation temporelle ; normalisation temporelle ; domaine médical ; dossier patients informatisés ; système à base de règles ; français

Abstract

Throughout the ages, we went from an observed time to a built time, and today, to a measured time. Even computerised. We are living in a society where everything moves always faster, where everything becomes each time more automatic, and where technology plays a crucial part.

The medical field is especially suitable for the new technologies : we want to have direct access to the patient's medical history. Temporal expression recognition, their normalisation and the identification of medical events linked to these timexes have therefore been studied a lot these past few years.

To this end, we present Time extractor, a rules-based system which recognises et normalises french temporal expression in patients health records. This system also approaches the contextualisation of the timexes with medical events.

Keywords : Information recognition ; information extraction ; temporal extraction ; temporal annotation ; temporal normalisation ; medical domain ; patients health records ; rules-based system ; French

Remerciements

Tout d'abord, j'adresse mes remerciements au professeur Cédrick Fairon, pour avoir accepté de m'accompagner, de me guider tout au long de ce mémoire. Je le remercie pour ses conseils avisés, sa patience et pour le temps qu'il a bien voulu me consacrer.

J'exprime toute ma reconnaissance à monsieur Brice de Behault et monsieur Thibault Bughin pour m'avoir permis de réaliser chez EarlyTracks le stage qui a débouché sur la création de Time extractor, pour leur aide précieuse, pour la confiance qu'ils m'ont accordée lors de l'annotation de documents confidentiels, ainsi que pour leur souplesse et leur compréhension face aux impératifs de cours.

Je remercie également celles qui ont partagé, dans la joie et le stress, ce master avec moi. Elise, Alix, Danica, Jen, les choses auraient été bien différentes sans vous. Merci aussi à Adeline et Marie, nos prédécesseurs, pour tous leurs conseils.

Enfin, je souhaite aussi témoigner toute ma gratitude à mon entourage pour son soutien et ses encouragements et pour le travail de relecture et de critique qu'il a accepté de mener.

Table des matières

I Introduction **17**

II Expression du temps et extraction d'informations **23**

1 Introduction **25**

2 Le temps **27**

1 La temporalité 27

2 La temporalité en français 28

2.1 Les expressions temporelles absolues et relatives 28

2.2 Le marquage du temps en langue française 29

2.3 L'expression du temps en chinois 41

2.4 Conclusion 43

3 L'extraction d'informations **45**

1 Quelques définitions 45

2 L'EI et le TAL 46

3 L'extraction de la temporalité 48

4 Temporalité dans le domaine médical : État de l'art **49**

1 TimeML - un langage de balisage 50

1.1 Systèmes à base de règles 53

1.2 Méthodes statistiques 58

1.3 Systèmes hybrides 62

1.4 Conclusion de l'état de l'art 69

1.5 SNOMED-CT, un outil de nomenclature des termes cliniques 71

III Application : Time extractor **73**

5 Introduction **75**

6 Annotation du corpus **79**

1 L'annotation 81

1.1 Les erreurs et problèmes rencontrés 85

7 Unitex **87**

1 Extraction des dates 89

1.1 Les graphes 89

2 Contextualisation des dates 92

2.1 Les dictionnaires 92

8 Normalisation des dates et envoi à l'interface **97**

9 Time extractor **99**

IV Évaluation et perspectives **103**

10 Evaluation de Time extractor **105**

1 Evaluation de la contextualisation et de la normalisation 108

11 Perspectives **113**

1 Graphies 113

2 Normalisation 114

3 Améliorations avec TimeML ? 115

4 Réflexion sur les différentes approches pour l'extraction de la temporalité dans le domaine médical 116

5 Comparaison avec les travaux de Hamon et Grabar (2014) 117

V Conclusion générale **119**

VI Bibliographie **123**

VII Annexes **137**

Table des figures

2.1 Déictiques de localisation temporelle (Kerbrat-Orecchioni, 1980). 33

4.1 Les balises TimeML utilisées en français. 50
4.2 Exemple d'un modèle de Markov caché. 59
4.3 Temporal event processor system architecture (Li & Patrick). 61
4.4 Deux tables de résultats du modèle statistique. 61
4.5 Architecture du HeidelTime GATE wrapper. 63
4.6 Architecture du système TN-TIES. 65
4.7 Architecture de la reconnaissance des événements (Event). 67
4.8 Architecture de la reconnaissance temporelle (Timex3). 68
4.9 Architecture TLink (liens temporels entre événements). 68

5.1 Schéma récapitulatif des différentes phases de conception de Time extractor. . . 76
5.2 Schéma récapitulatif des différentes phases d'analyse, extraction, contextualisation, normalisation et affichage de Time extractor. 77

6.1 Annotation du corpus médical. 81

7.1 Exemple d'un graphe Unitex (à gauche) et du sous-graphe *jour_semaine* à droite. 87
7.2 Fonctionnement de l'extraction et de la contextualisation des expressions temporelles (zoom de la figure 5.2). 88
7.3 Exemple d'une partie d'un graphe d'extraction et sa sortie, suivi de ses zooms. . 90
7.4 Graphe d'extraction d'âge (en années). 92
7.5 Exemples d'un dictionnaire utilisé pour le repérage des expressions médicales. . 93
7.6 Exemple d'une partie d'un graphe d'extraction et sa sortie. 94
7.7 Zooms du graphe de contextualisation de la figure 6. 95

7.8 Graphe *procedure_time* - Repérage des expressions temporelles. 95

9.1 Interface de Time extractor. 99
9.2 Interface de Time extractor, avec un texte prêt à être analysé. 100
9.3 Interface de Time extractor, avec un texte analysé (zoom sur le texte). 100

10.1 Exemple de contextualisation de tous les types d'événements. 109
10.2 Exemple de contextualisation et de non-reconnaissance des paires procedure - timex. 109
10.3 Exemple de contextualisation et de non-reconnaissance des paires substance - timex. 110
10.4 Exemple de contextualisation et de non-reconnaissance des paires function - timex.110
10.5 Exemple de contextualisation et de non-reconnaissance des paires disorder - timex.110
10.6 Exemple de contextualisation et de non-reconnaissance des paires avec une timex typée « UNKNOWN ». 111
10.7 Exemple de contextualisation et non-reconnaissance des paires disorder - time. . 111
10.8 Exemple de contextualisation d'un triple événement time - procedure - disorder. 112
10.9 Exemple de contextualisation d'événements non soulignés de manière homogène. 112

11.1 Exemple de contextualisation d'événements non soulignés de manière homogène. 115

Que de temps perdu à gagner du temps !

Paul Morand

Première partie

Introduction

Le Temps a toujours suscité l'intérêt et même exercé une certaine fascination. Saint Augustin disait : *« Quand on ne me le demande pas, je sais ce qu'est le temps ; quand on me le demande, je ne le sais plus ».* Le temps a sans cesse été évalué, de manière différente selon la communauté, la région du monde, l'époque...

Ainsi, durant l'Antiquité, les Grecs mesurent le temps en observant les astres et en conceptualisent la notion avec les mathématiques [1]. Chaque cité grecque politiquement indépendante possède son propre calendrier. Suite aux conquêtes d'Alexandre le Grand et à la rencontre des savants avec l'astronomie égyptienne, la journée est divisée en douze heures nocturnes et douze heures diurnes.

Quant au calendrier du monde romain, il découpe l'année en douze mois (de vingt-neuf ou trente-et-un jours) divisés en trois périodes liées au phases de la lune : les calendes (la nouvelle lune), les nones (le premier quart de la lune) et les ides (la pleine lune). Une période intercalaire est ajoutée tous les deux ans afin de combler l'écart entre les calendriers tropique et lunaire. Finalement, Jules César lance une réforme du calendrier, qui débouche sur le calendrier julien lequel sera adopté dans tout l'Empire, des Gaules à L'Egypte. Il s'impose finalement (presque) partout *« pour une question de commodité du déroulement de la vie civile »* [2].

1. Luc Brisson, cité par Formoso, (2013 : 30).
2. John Scheid, cité par Formoso (*Ibid.,* : 33).

Au Moyen-Âge et jusqu'au XIVe siècle, le repère dans le temps est le son de la cloche (puisque chaque village possède son église). Le temps est divisé pour organiser le travail et la prière. La journée comporte ainsi huit heures canoniales [3] : les Matines (milieu de la nuit), les Laudes (l'aurore, les Laudes sont une louange au soleil qui se lève), le Prime (première heure du jour), la Tierce (troisième heure du jour), la Sexte (aussi appelée Sixte, sixième heure du jour, ou midi), la None (neuvième heure du jour), les Vêpres (le soir, avant que le soleil ne se couche) et les Complies (qui marquent l'heure du coucher).

L'horloge mécanique est inventée fin du XIIIe siècle. Le temps est nouveau : il est unifié et précis, différent des heures canoniales, sonnées de façon approximative. La journée est dorénavant divisée en heures égales, dans l'artisanat, les apprentis sont payés à l'heure de travail. Peu à peu, le temps se libère de toute connotation religieuse : nous entrons dans un temps moderne.

Dans le monde oriental, la mesure du temps est signe de pouvoir : la mise en place du calendrier officiel de l'année nouvelle revient à l'État. Par ailleurs, à chaque nouvelle dynastie, une des préoccupations premières est d'établir un calendrier. En effet, si le pouvoir permet au temps de se désordonner, les conséquences seraient catastrophiques. Les Chinois étudient le ciel pour établir le calendrier et observent la nature pour déterminer les saisons. Le calendrier solaire (agricole) divise l'année en douze mois, eux-mêmes divisés en deux périodes [4]. À chaque début de mois correspond la lune nouvelle, à chaque moitié de mois, la pleine lune. Un seul événement reste certain : le solstice d'hiver durant le onzième mois. Une année compte 353 à 355 jours. Tous les deux trois ans, un treizième mois est intercalé, formant ainsi une année de 383 à 385 jours. Le calendrier chinois est définitivement formalisé en 1645. Il se calcule avec le mouvement du soleil, affinant ainsi l'arrêt du treizième mois [5].

À travers les âges, nous passons d'un temps observé à un temps fabriqué, puis, aujourd'hui, à un temps mesuré. Et même, informatisé. Nous vivons dans une société où tout va toujours plus vite, où tout s'automatise chaque jour davantage, et où la technologie joue un rôle essentiel. Nous escomptons naturellement rencontrer cette rapidité et ce fonctionnement automatique dans toutes les spécialités.

Le domaine médical offre un terrain particulièrement propice à l'utilisation des nouvelles tech-

3. Les heures canoniales sont les moments quotidiens dédiés à la prière
4. Notons qu'en chinois, le caractère 月 (yuè) désigne le *mois* et la *lune*.
5. Ce treizième mois explique que le Nouvel An chinois a lieu entre fin janvier et fin février.

nologies. En effet, si les dossiers patients[6] sont aujourd'hui enregistrés électroniquement, ils ont la forme de textes bruts parfois forts longs, et où les informations pertinentes sont disséminées dans tout le document. Un apport considérable et non-négligeable sera donc de mettre en exergue dans les rapports patients les événements médicaux et l'information temporelle correspondante. Cela permettra une lecture rapide mais néanmoins complète de l'historique médical du patient.

Plusieurs outils existent déjà pour la reconnaissance ou l'annotation des faits médicaux. SNOMED-CT, présenté aux pages 71-72, en fait partie. C'est une nomenclature internationale de quelques 350 000 concepts médicaux, partagés entre 19 axes qui permettent d'exposer les symptômes, les maladies, les interventions, etc.

D'autres outils ont également été créés pour la reconnaissance et la normalisation des expressions temporelles dans le domaine médical. Nous en présenterons quelques-uns dans le chapitre 4. Cependant, la plupart de ces outils ont été conçus pour la langue anglaise, or nous nous intéressons au français.

Ainsi, Kevers (2011 : 99) confie trois tâches au *traitement automatique de l'information temporelle* : *repérer les éléments qui véhiculent ce type d'information [l'expression du temps], [...] les interpréter afin de leur donner une valeur univoque dans l'espace du temps, et, finalement, [...] en fournir une représentation normalisée.*

Un événement est toujours lié à une information temporelle, peu importe la forme de cette information. La reconnaissance, l'extraction et la normalisation de la temporalité sont le sujet de nombreuses recherches. Notre propos concerne les expressions de temps dans le domaine médical. Cependant, comme Styler *et al.* l'ont indiqué,

> *Robust temporal understanding of such [clinical] narratives , however, has been difficult to achieve, due to the complexity of determining temporal relations among events, the diversity of temporal expressions, and the interaction with broader computational linguistic issues.* (Styler *et al.*, 2014 : 143).

6. Ces documents patients incluent les rapports de sorties, les résultats d'examens, les consignes per- et post-opératoire, etc.

Nous avons développé un système - Time extractor - dont le but premier est de reconnaître et extraire les expressions de temps dans un corpus médical. Ces dates sont ensuite normalisées [7]. Dans un second temps, nous avons également développé un outil permettant de corréler ces dates avec un événement médical [8].

Le propos de ce mémoire et les résultats de ces recherches que nous présentons en ces pages débute par une introduction, suivie d'une évocation de l'expression du temps et de l'extraction de l'information. Nous parlons ensuite de l'état de l'art de l'extraction de la temporalité dans le domaine médical. La troisième partie présente notre système, Time extractor. La partie suivante évalue notre outil et les perspectives avant de conclure.

7. Dans ce cas, c'est-à-dire qu'elles sont remplacées par leurs dates relatives (donc sous le format JJ-MM-AAAA).

8. Nous verrons que cet outil est encore au stade de développement et doit encore être amélioré.

Deuxième partie

Expression du temps et extraction d'informations

Chapitre 1

Introduction

La vision et l'évaluation du temps ont bien changé depuis l'aube de l'humanité. L'expression du temps, elle, varie selon la langue parlée :

> *L'expérience humaine du temps se manifeste à travers la langue. Or la langue n'est pas un calque de la réalité ; chaque langue offre sa propre construction du réel et élabore son système temporel, qui n'est ni le temps du monde (temps physique) ni celui du calendrier (temps chronique).* (Tomassone, 2001 : 172).

Donc, suivant la langue et la vision du monde qu'elle fonde, la façon d'exprimer le temps change. Nous verrons dans ce chapitre que l'expression de la temporalité est différente en français et en chinois. Puisque notre outil Time extractor a été conçu pour le français, nous allons tenter de décrire plus précisément la formulation du temps dans la langue de Voltaire. Une dernière section proposera différentes façon de signifier la temporalité en chinois.

Cette deuxième partie consacrée au temps met donc en exergue les difficultés à reconnaître la temporalité et à extraire toutes les expressions du temps dans un texte. Le chapitre 3 définit ensuite le domaine de l'extraction de l'information. Nous finissons cette partie par un état de l'art de l'extraction de l'information temporelle, et plus précisément la reconnaissance et la normalisation de la temporalité dans le domaine médical.

Chapitre 2

Le temps

1 La temporalité

L'indication temporelle se rapporte à une autre information, ce qui lui donne une dimension particulière. Un événement est toujours lié à une information temporelle, quelle qu'en soit la forme ; la notion de temps ne dépend pas que de la linguistique.

La compréhension du temps nécessite une certaine **connaissance du monde** qui s'apparente à la culture générale. Par exemple, « les jours suivant la chute du mur de Berlin » concernent les jours suivant le 9 novembre 1989.

Il arrive qu'on donne **une date pour désigner un événement** : « le 11 septembre » se réfère aux attentats du World Trade Center et du Pentagone du 11 septembre 2001. Le « 6 juin 44 » évoque le débarquement des Alliés en Normandie. Il est aussi des **temporalités implicites** : si nous parlons d'un « discours du Président Obama », nous renvoyons à un discours prononcé lors de ses années de mandat.

Parfois, les **événements** eux-mêmes **suggèrent une temporalité**. De fait, des termes comme « tomber », « explosion » impliquent une action quasi instantanée, qui se calculent en secondes, voire en millisecondes. Tandis que d'autres indiquent une action qui prend un certain temps à s'accomplir : « ébullition » en minutes, « grossesse » en mois, « mandat » en années [1].

1. Kevers (2011 : 105).

Les mots sont porteurs de temporalité, le **temps linguistique** relève des termes et des traits grammaticaux qui situent un événement sur l'axe du temps. Selon Benveniste, le temps linguistique[2] est lié à l'instance du discours. Il affirme que tous les systèmes temporels ont en commun le fait que la parole est centrée autour du **présent de l'énonciation**. Ce dernier se déplace en même temps que le discours et délimite les formes verbales. Ainsi, quand nous parlons de « hier », « aujourd'hui », « demain », ces termes sont à comprendre en fonction du présent. « Aujourd'hui » présente un événement simultané à l'énonciation, « demain » le situe à un temps proche postérieur à l'énonciation, tandis que « hier » à un temps proche antérieur à l'énonciation. Si l'énonciation est reportée, « aujourd'hui » n'aura plus le même sens et ne devient intelligible qu'avec une date. Dans un récit, on parle du **temps de l'énoncé**, plutôt que de présent de l'énonciation.

2 La temporalité en français

2.1 Les expressions temporelles absolues et relatives

Deux types d'expressions temporelles sont communément utilisées : les **expressions temporelles absolues** et **expressions temporelles relatives**. Avec ces deux types de timex[3], nous remarquons que le temps peut être linguistiquement marqué :

« lundi » ;

« 21 novembre » ;

« cette semaine » ;

etc.

Les *expressions temporelles absolues* se réfèrent directement à la date du calendrier :

« 21 novembre » ;

« jeudi 11 mai 2017 » ;

« 18 avril 1999 » ;

« en 2016 ».

2. Dans son ouvrage *Problèmes de la linguistique générale*, Paris, Gallimard, 1974.
3. Dans ce mémoire, nous utilisons le terme *timex* pour désigner les expressions temporelles.

Le second type, les *expressions temporelles relatives*, reprend les expressions temporelles vagues. Elles ne permettent pas de récupérer directement la date absolue des faits. La date du présent de l'énonciation doit être connue pour pouvoir normaliser [4] la timex relative :

« lundi » ;

« cette semaine » ;

« dans trois jours » ;

« à 8h du matin ».

Dans un exemple concret, datant du vendredi 18 avril 2014 :

« Elle est venue à l'hôpital lundi dernier suite à une chute de cheval survenue la veille. »

Pour situer sur le calendrier la chute de cheval, il faut se rendre au lundi précédant la date du document (ce qui donne lundi 14 avril 2014 [5]) et en prendre la veille, donc le dimanche 13 avril 2014.

Cet exemple montre que nous pourrions proposer plusieurs catégories de dates relatives, suivant leur degré d'imprécision. En effet, les expressions temporelles de l'exemple ci-dessus permettent de retrouver la date absolue des l'arrivée à l'hôpital et la date de la chute. Cependant,

« Elle s'est cassée la clavicule droite il y a des années. »

ne permet pas de récupérer la date absolue - ou de donner une idée de l'année - de cette fracture. Certaines expressions de temps sont trop vagues et imprécises, et elles ne peuvent être normalisées efficacement.

2.2 Le marquage du temps en langue française

Le temps peut être exprimé en français de diverses façons : les adverbiaux temporels, les connecteurs temporels, les formes verbales, la morphologie, l'agencement des phrases et des propositions, les compléments de verbes, les circonstanciels, les constructions syntaxiques,... [6].

4. La normalisation d'une timex consiste à retrouver la date absolue complète de l'expression temporelle.

5. En prenant comme hypothèse que *lundi dernier* se rapporte au lundi le plus proche dans le passé. Nous pourrions cependant décider qu'il s'agit du lundi précédant le dernier lundi en date. Ceci est particulièrement vrai et d'application si la date d'énonciation est un mardi, un mercredi et peut-être même un jeudi. Nous gardons donc une certaine imprécision et une marge d'erreur est possible.

6. Pour un compte rendu précis et complet, lire le chapitre *L'expression du temps dans le langage naturel*, aux pages 111 à 140 de la thèse de doctorat de Laurent Kevers (2011), disponible à l'adresse

Les adverbiaux temporels

> *Un moyen très courant d'exprimer le temps dans les textes est d'utiliser des expressions qui ont, de manière assez explicite et directe, une valeur temporelle. Ces expressions sont regroupées sous le nom d'adverbiaux temporels. Le terme adverbial est à mettre en relation avec la notion de complément adverbial (Gross [1986] [...])* [7]*, c'est-à-dire d'un ensemble d'éléments qui peuvent prendre la fonction d'adverbe* [8] (Kevers, 2011 : 111-112).

Rappelons que pour un grammairien, un adverbe n'est pas une fonction, mais une classe grammaticale précise. Certains exemples donnés par Kevers (« lundi », « décembre », « au vingt-et-unième siècle »,...) sont des noms communs ou des groupes nominaux et non des adverbes stricto sensu. Kevers, conscient de ces subtilités, choisit dans un souci de simplification la désignation *adverbes* et *adverbiaux temporels* pour des termes qui expriment une valeur temporelle. Nous avons opté pour cette terminologie.

Kevers (2011 : 183-187) organise les adverbiaux temporels selon quatre critères qui clarifient et complètent la classification de Borillo [9].

Critère 1 Le premier critère affirme que chaque adverbial temporel est **ponctuel** ou **duratif**. L'expression ponctuelle signale un espace temporel indivisible et désigne un point précis sur la ligne du temps. Une expression durative représente un espace délimité par une fourchette : elle a donc besoin d'un temps de début et d'un temps de fin. L'expression durative représente un intervalle sur la ligne du temps. Exemples :

Ponctuel : « vendredi 21 juillet 2017 » ;

Ponctuel : « au vingt-et-unième siècle » ;

Duratif : « de septembre à décembre » ;

Duratif : « depuis 2015 ».

`https://halshs.archives-ouvertes.fr/tel-00568089/`. Le travail de Kevers étant incontournable pour l'extraction de la temporalité en français, nous y ferons régulièrement référence, et particulièrement dans ce chapitre.

7. *Chez Gross, le nom d'adverbe généralisé, d'adverbe ou de complément adverbial est donné aux compléments qui, dans la terminologie traditionnelle, sont repris sous les catégories d'adverbes, de compléments circonstanciels et de propositions subordonnées circonstancielles.* (Kevers, 2011 : 111-112)

8. *Pour plus de commodité, nous utiliserons dès maintenant de manière indifférente, sauf indication contraire, les termes adverbes et adverbiaux.* (*Ibid.*)

9. Pour la classification de Borillo, voir le chapitre 4 de Kevers (2011 : 111-118).

Critère 2 Selon le deuxième critère, chaque adverbial temporel est **absolu** ou **relatif**. Ce critère classe les adverbes selon que leur localisation sur la ligne du temps est indépendante d'un autre point (localisation absolue), ou si elle a plutôt besoin d'un repère temporel (localisation relative). Ce critère rejoint la définition des dates absolues et relatives évoquées précédemment (pages 28-29). Exemples :

Absolu : « le 21 juillet 2017 » ;

Absolu : « 2014 » ;

Relatif : « vendredi » ;

Relatif : « la semaine précédente ».

Critère 3 Le troisième critère dit que les adverbes de temps sont **précis** ou **flous**. L'expression sera dite précise si elle comprend toute la zone désignée, à l'exclusion du reste :

Précis : « le lundi 15 mai 2017 » ;

Précis : « à midi » ;

Précis : « l'année 2017 ».

Une expression sera floue si elle comprend davantage que la zone déterminée, ou si elle n'en comprend qu'une petite partie. Il y a plusieurs manières d'arriver à ce caractère flou :

Flou : dans une plage temporelle de granularité élevée[10] : « durant le mois de novembre » ;

Flou : dans les alentours d'un repère précis : « lundi 24 juillet vers midi » ;

Flou : spécification relative imprécise : « depuis environ deux mois ».

Le caractère précis ou flou d'une expression varie selon sa granularité. *Ainsi, toute expression dont la granularité est inférieure ou égale au jour est interprétée comme étant précise, à moins qu'elle ne soit accompagnée d'une marque d'imprécision. A l'inverse, toute expression de granularité supérieure au jour est considérée comme floue (« au 20e siècle », « en décembre »), à moins qu'une marque de précision soit explicitement présente (« le 20e siècle », « tout le mois de décembre »)*[11].

10. La granularité est le fait que le temps peut être envisagé selon des unités plus ou moins grandes (la seconde, le jour, l'année, le millénaire, etc.). On parle de *granularité naturelle des événements* quand certains événements sont naturellement exprimés avec une unité particulière. Par exemple, on s'exprimera en secondes pour une course de 400 mètres, mais en années pour un mandat présidentiel. Kevers (2011 : 143) parle de la granularité comme *un ordre de grandeur à partir duquel la dimension temporelle des événements peut être exprimée*. Pour plus d'informations, voir Kevers (2011 : 143 - 144) et Euzenat (1994).

11. Kevers (2011 : 184 - 185).

Kevers attire également notre attention sur la différence entre les expressions floues et les expressions précises sous-spécifiées. Par exemple, « jeudi » est une expression précise sous-spécifiée, puisque nous ne pouvons pas savoir, sans autres spécifications, de quel jeudi il s'agit.

Critère 4 Enfin, selon le quatrième critère, les adverbes de temps sont **uniques** ou **répétitifs** :

Unique : « vendredi soir » ;

Répétitif : « chaque vendredi » ;

Répétitif : « les trois derniers week-ends ».

Ces quatre critères peuvent être combinés entre eux et forment huit catégories d'expressions de temps possibles. À ces catégories, quatre autres plus particulières sont ajoutées : les **durées**, les **âges**, les **durées imprécises** et les **âges imprécis**.

Les connecteurs temporels

Le terme *connecteur* est une appellation générique ; il désigne les termes qui, en établissant des relations entre les phrases et les propositions, permettent l'organisation des textes. Ils sont particuliers parce qu'ils ne reprennent aucun élément du contexte.

Les connecteurs temporels permettent d'indiquer et d'organiser l'enchaînement chronologique des faits :

Type de succession	connecteurs
Succession pure et simple	adverbes : *alors, après, ensuite, puis,...* conjonction de coordination : *et*
Succession par paliers : distinction de stades dans le déroulement	*d'abord* *ensuite, puis,...* *enfin, finalement,...* *premièrement, deuxièmement,...*
Succession ordonnée **par rapport au moment** **de l'énonciation**	*hier, aujourd'hui, demain,* *la semaine prochaine, le mois dernier,* *dans un mois, dans un an,...*

Table 2.1 – Tableaux de connecteurs temporels (Tomassone, 2002 : 76).

Les formes verbales

Le *temps* du verbe, ce terme, *temps*, est ambigu en français : il désigne à la fois la notion de temps (la temporalité) ou le temps des verbes. En anglais, on utilise les termes *time* et *tense* pour respectivement renvoyer à la notion et aux verbes. Damourette et Pichon [12] ont proposé l'expression de *tiroirs verbaux* pour désigner les formes verbales. Dans ce travail, nous utiliserons également ce terme afin d'éviter toute ambiguïté.

Gosselin (1996 : 23 - 24) distingue **les temps absolus** - passé, présent, futur - à l'aide de circonstanciels de temps déictiques [13] :

« Je viendrai *demain*. » ;

« Je viendrai *dans trois semaines*. » ;

« *Je viendrai *hier*. » ;

« *Je viendrai *en ce moment*. ».

Le tableau ci-dessous reprend les déictiques de localisation temporelle, il est proposé par Kerbrat-Orecchioni en 1980 [14].

	Déictiques / Référence T_0	**Relatifs au cotexte** **Référence : Y exprimé dans le cotexte**
Simultanéité	*en ce moment, maintenant*	*à ce moment-là, alors*
Antériorité	*hier, l'autre jour, récemment*	*la veille, la semaine précédente, peu avant*
Postériorité	*demain, l'année prochaine, dans deux jours*	*le lendemain, l'année suivante, deux jours plus tard*
Neutres	*aujourd'hui, ce matin, cet été, tout à l'heure*	*un autre jour*

FIGURE 2.1 – Déictiques de localisation temporelle (Kerbrat-Orecchioni, 1980).

12. Cités par Riegel *et al.* (2006 : 289).
13. Pour rappel, les **déictiques**, parfois aussi appelé *embrayeurs* sont *« des unités linguistiques dont le sens implique obligatoirement un renvoi à la situation d'énonciation pour trouver le référent visé »* (G. Kleiber, 1986, cité par Jakobson `http://lexgram.over-blog.fr/article-25284697.html`, page consultée le 21 juillet 2017). Ainsi, *je* est un déictique désignant le locuteur.
14. Du site Internet `http://lexgram.over-blog.fr/article-25284697.html`, consulté le 21 juillet 2017.

Le temps relatif - antérieur, simultané, ultérieur - est reconnu par des circonstanciels temporels anaphoriques :

« Il disait qu'il était venu *la veille*. » ;

« *Il disait qu'il était venu *le lendemain* .» ;

« *Il disait qu'il était venu *à ce moment*[15]. ».

L'aspect

La forme verbale est indubitablement liée à **l'aspect**[16], que Tomassone (2001 : 174) définit de la manière suivante :

> *L'aspect est tout ce qui permet au locuteur de donner, indépendamment de toute considération chronologique, une représentation du déroulement d'un procès exprimé par le verbe, de sa durée, de son achèvement.*

a) Aspect accompli et non-accompli Tomassone parle d'ailleurs d'un « double jeu de tiroirs verbaux » pour le système verbal du français. Ce double jeu se réfère aux formes simples et composées à l'intérieur de chaque mode[17]. Ces deux types de formes introduisent les notions d'accompli / de non-accompli.

En effet, une forme simple signifie que l'événement est toujours en cours, il est donc non-accompli. Par contre, une forme composée indique qu'il est accompli au moment donné :

« Il faut que vous partiez demain matin. » ;

« Il faut que vous soyez parti demain matin. ».

Dans la première phrase, le départ doit voir lieu le lendemain matin, tandis que la deuxième phrase signifie que le départ doit avoir eu lieu avant le lendemain matin. Ces notions d'accomplissement et de non-accomplissement ne sont pas dépendantes de l'époque passé/présent/futur. Le *présent de l'indicatif* annonce un événement qui est en cours d'accomplissement (et dont l'aspect est donc non-accompli). Le passé composé indique un aspect accompli.

15. Ces circonstanciels sont corrects si l'anaphore temporelle renvoie à un instant antérieur au présent de l'énonciation.

16. Cette section utilise principalement les travaux de Tomassone (2001 : 174-175) et (279-286).

17. Pour rappel, *« à chaque forme simple correspond une forme composée dont l'auxiliaire est au même temps que la forme simple »* (*Ibid.*, 2001 : 174).

b) Aspect borné et non-borné (ou global et sécant) L'imparfait, temps du passé, s'emploie avec le passé simple dans le plan de l'Histoire et avec le passé composé dans le temps du Discours.

(1) « Comme le soir tombait, l'homme sombre arriva... » ;

(2) « (Quand) le soir tomba, l'homme sombre arriva. » ;

(3) « Le soir tomba : l'homme sombre arrivait. » [18].

Ces trois phrases présentent les mêmes événements, passés, simultanés et non-accomplis. Cependant, ils ne sont pas perçus de la même façon : tandis que le passé simple saisit la structure temporelle dans son intégralité, l'imparfait la prend à un moment de son déroulement, une partie est déjà terminée tandis que l'autre est encore à réaliser. On appelle l'opposition entre passé simple et imparfait une opposition d'aspect **borné** (ou *global*) et **non-borné** (ou *sécant*). Un événement exprimé au passé simple représente un tout indivisible, peu importe sa durée. Il est délimité et peut donc être accompagné d'un complément temporel indiquant un laps de temps délimité (1). L'imparfait, par contre, dissimule le début et la fin du fait, il ne peut donc pas se combiner à un complément temporel signalant un début et une fin (2), mais bien avec un complément temporel qui ne signale pas la fin d'un événement (3) [19].

(1) « Il dormit pendant 6 heures. » ;

(2) « *Il dormait pendant 6 heures. » ;

(3) « Il dormait depuis 6 heures. ».

L'expression lexicale de l'aspect L'aspect peut être marqué par un affixe ou par une périphrase verbale. Par exemple, les auxiliaires *aller* et *venir* suivis d'un infinitif transmettent un aspect prospectif (*aller*) et rétrospectif (*venir*). L'aspect itératif peut être signalé par des adverbes ou compléments circonstanciels tels que *souvent, fréquemment, rarement, tous les jours/mois/ans, une fois par semaine, parfois,....* Le tableau ci-dessous (proposé par Tomassone, 2011 : 175) résume l'expression lexicale de l'aspect :

18. Exemples et analyse viennent de Tomassone (2001 : 174-175).
19. Tomassone (2001 : 175).

	Inchoatif	**Itératif**	**Terminatif**	**Progressif**
Périphrase verbale	*commencer à* *se mettre à*		*finir de* *cesser de* *achever*	*aller* + participe passé *être en train de*
Préfixe/suffixe Formation du mot	verbes en *-ir* : jaunir *-iser* : automiser, stabiliser *en-* : endormir	*reprendre*, *voleter* *toussoter* *sautiller*		forme pronominale : *se nourrir* *s'endormir*

TABLE 2.2 – Expression lexicale de l'aspect (Tomassone, 2011 : 175).

Il existe également une opposition entre **perfectif** et **imperfectif**. Le premier prend en compte la fin de l'action (*mourir, sortir, entrer, partir, naître, exploser, casser, tomber,...*). Le second, par contre, ne comprend pas de limite intrinsèque (*aimer, durer, marcher, attendre,...*).

Présent, passé, futur

Les temps de l'indicatif possèdent une valeur aspectuelle exprimée par l'opposition temps simples et temps composés - que nous venons d'évoquer - et une information temporelle : le passé, le présent et le futur. Ces derniers sont définis par deux facteurs : le moment de l'énonciation (le présent du locuteur) et le moment de l'événement narré. Dans un discours historique, les tiroirs verbaux de l'indicatif sont localisés par rapport au temps des événements, la narration définissant le temps de l'énonciation [20].

Présent Le présent, ou moment de l'énonciation, est utilisé pour signaler que ce que le locuteur raconte est vrai, existe au moment énoncé. C'est le *temps du commentaire*, avec lequel on rapporte ce qui est en train de se passer :

« Il pleut. » ;

« Vous jouez. » ;

« Ils sont en train de danser. ».

20. Cette section est principalement basée sur les travaux de Tomassone (2001 - 176-179) et (2002 : 286-293).

Il est particulièrement intéressant puisque les verbes conjugués à la première personne de l'indicatif présent accomplissent le fait marqué :

« Je mange. » ;

« Nous marchons. » ;

« Je vous remercie. ».

Conjugués à une autre personne, un certain décalage se remarque entre l'action et l'énonciation : « Tu les remercies. ». En général, le temps référé par le présent tourne autour du moment de l'énonciation : « Il épouse Sarah aujourd'hui. ».

Le présent *d'habitude* exprime un fait vrai au moment de l'énonciation, tout comme il était vrai avant et il le sera encore après :

« Jérôme se lève tous les jours à 6 heures. » ;

« Après l'école, je vais à la piscine. ».

Le présent d'habitude est également à comprendre via le contexte ou avec des informations connues des interlocuteurs. En effet, « Demain, je vais à la piscine. » ou « Je vais à la piscine. » pourraient ne pas être des présents d'habitude, mais plutôt une circonstance. De plus, le sens intervient aussi puisque certaines expressions sont plus enclines à former des habitudes que d'autres : « Je vais à la piscine. » serait ainsi plus facilement une habitude que « Je tombe. ». Ce présent d'habitude n'est pas à confondre avec un présent *atemporel*, aussi appelé le présent *gnomique*. Il s'agit du présent des proverbes, dans lequel l'acte d'énonciation est sans importance puisque le fait est vrai dans tous les temps : « La Terre tourne. ».

Le présent *historique* est utilisé pour signaler des événements qui ne se situent pas au moment de l'énonciation, mais au temps de l'énoncé. Le présent est alors substitut du passé simple. La localisation des faits est opérée avec des repères temporels explicites (des dates). Le présent *de narration* est différent du présent historique, puisqu'il crée un effet dramatisant.

Futur Le futur présente le moment de l'événement postérieur au moment de l'énonciation. Le *futur simple* signale que cette action est non accomplie, tandis que le *futur antérieur* (la forme composée du futur) l'indique comme accomplie à un moment du futur :

« Demain matin, je serai à la mer. » ;

« Demain à huit heures, j'aurai mangé et je marcherai sur la plage. ».

Le futur simple ne saisit pas l'événement dans sa globalité, mais le présente à un moment de son déroulement, sans donner d'indications sur le début ou la fin de l'action. Ainsi, dans le premier exemple ci-dessus, le « je » sera à la mer le lendemain matin, mais nous ne savons pas combien de temps le locuteur y restera, ni ce qu'il fera avant et/ou après. En ce sens, le futur s'oppose au passé simple et s'approche de l'imparfait (aspect non-borné).

Cependant, futur et passé simple se ressemblent en ce qu'ils sont coupés du présent. Le futur simple non-accompli est ainsi opposé au *futur périphrastique* (aussi appelé *futur immédiat* ou *futur proche*) exprimant un événement imminent :

« Je vais partir. » vs. « Je partirai. »

Le futur proche signale une certitude, un acte très proche du présent et donc vrai. En revanche, le futur simple annonce plutôt une éventualité. Notons cependant que *aller + verbe* est en réalité une forme aspectuelle, et non un forme temporelle. Elle présente le moment événement à un stade antérieur à sa réalisation, que ce soit dans le passé (« J'allais partir. »), ou dans le présent (« Je vais partir. »). Notons également que cette forme du verbe *aller* ne se retrouve que dans l'indicatif présent et imparfait. Le futur est utilisé pour former une hypothèse plus ou moins sûre, que l'avenir vérifiera : « Elle sera en retard. » produit l'hypothèse de son retard, qui sera ou non confirmé dans le temps postérieur au moment de l'énonciation.

Le conditionnel présent (forme simple) et le conditionnel passé (forme composée) sont reconnus en fonction du moment de l'événement, *elles réfèrent à un moment postérieur à celui des événements principaux (le temps de l'énoncé) et antérieur à celui de l'énonciation : « Il pensait (pensa) qu'il n'aurait aucun mal à le convaincre. ».* (Tomassone, 2001 : 176).

Passé Le tableau ci-dessous, proposé par Tomassone (2001 : 177), résume les temps passés de l'indicatif :

		Accompli	**Non-accompli**
Borné	**Histoire**	passé antérieur	passé simple
	Discours	passé surcomposé	passé composé
Non-borné		plus-que-parfait	imparfait

TABLE 2.3 – Relations entre les tiroirs verbaux du passé de l'indicatif (Tomassone).

> *Au-delà des formes aspectuelles accompli (formes composées) / non-accompli (formes simples), les tiroirs verbaux du passé se répartissent entre les deux modes d'énonciation historique (passé simple et passé antérieur) et discursive (passé composé et passé surcomposé), le couple imparfait/plus-que-parfait se rencontrant dans les deux* (Tomassone : 2001 : 176-177).

Le passé simple est situé en fonction du temps de l'énoncé. Ainsi, dans une narration, les verbes se succèdent en fonction de l'avancée du récit : « Je m'assis à table, bus une longue gorgée d'eau et commençai à manger. ». Un lien de cause à effet se déduit parfois : « Elle participa activement aux travaux de la commission et en fut élue présidente. ».

Le passé composé, par contre, ne renvoie pas deux faits comme étant temporellement liés : « Elle a participé activement aux travaux de la commission et en a été élue présidente. » n'indique pas nécessairement une relation causale. Les événements coordonnés au passé composé sont localisés en fonction du présent de l'énonciation.

L'imparfait, par son aspect non-borné, propose un événement dont les limites de début et de fin dépassent sur celles d'un autre événement, lui au passé simple ou composé, et dont il constitue le cadre : « Je mangeais quand Lucie m'a appelé. ».

D'autres indicateurs temporels

Il existe d'autres manières de marquer le temps en français. Nous avons déjà évoqué **l'agencement des phrases et des propositions** : Kevers (2011 : 103-104) propose les exemples suivants, pour indiquer que les propositions juxtaposées, coordonnées ou subordonnées portent une information temporelle sur les événements :

> « Il traversa la rue prudemment car les voitures roulaient rapidement. Il prit un taxi pour se rendre à l'aéroport. » ;
> « Il a raté son bus car il s'est levé trop tard. » ;
> « Il a pris le bus après avoir atteint Bruxelles en train. » ;
> « Il a rattrapé le bus qu'il avait raté. ».

Les **indicateurs chronologiques** et les **repérages par rapport à l'énonciation ou à l'énoncé** sont également très utilisés. Les premiers sont définis par rapport aux repères objectifs (comme les calendriers) qui permettent de localiser les événements. Le *temps est divisé en unités régulières et hiérarchisées (seconde, minute, heure, jour, semaine, mois) qui réfèrent à certains phénomènes cosmiques (le cycle du jour et de la nuit) ou qui sont établies arbitrairement (un mois a 30 ou 31 jours)* [21]. Ces repérages permettent de donner une valeur temporelle objective aux événements.

Le repérage par rapport à l'énonciation rapporte au présent de l'indicatif, qui indique la concomitance du fait avec son énonciation [22]. Ainsi, « Je mange. » peut être dit « Maintenant je mange. ». « Maintenant » est donc un *embrayeur* qui renvoie au moment de l'énonciation. En français, des mots sont utilisés pour indiquer la simultanéité, l'antériorité, ou la postériorité par rapport au moment de l'énonciation, en fonction de l'unité de temps qui est utilisée. Le tableau ci-dessous reprend quelques termes indiquant ces repérages et l'essentiel des différences entre les deux systèmes (énonciation vs. énoncé). Deux types de simultanéité y sont évoqués : la *simultanéité stricte*, qui reprend « maintenant », « en ce moment » ; et une *simultanéité dilatée*, qui englobe le moment de l'énonciation : « aujourd'hui », « ce matin », « cette année ». D'autres termes sont repris sous la section *neutres*. Ils renvoient au passé, au futur et au présent. Seul le contexte permet de savoir de quelle époque il s'agit.

21. Tomassone (2001 : 180).
22. *Ibid.*

Antériorité	hier avant-hier lundi dernier (passé) la semaine dernière (passée) il y a (voilà, voici) un an récemment	la veille l'avant-veille **le** lundi précédent la semaine précédente un an auparavant (avant) peu avant
Simultanéité (stricte ou dilatée)	maintenant en ce moment aujourd'hui cette année ce lundi	alors à ce moment-là ce jour-là cette année-là ce lundi-là
Postériorité	demain après-demain lundi prochain la semaine prochaine dorénavant bientôt	le lendemain le surlendemain **le** lundi suivant la semaine suivante dès lors peu après
Neutres	lundi tout à l'heure cet hiver, cette nuit	le lundi précédent/suivant un moment avant/après l'hiver précédent/suivant la nuit précédent/suivante

TABLE 2.4 – Repérage par rapport à l'énonciation et à l'énoncé (Tomassone, 2001 : 180).

Au terme de cet exposé, un constat s'impose : l'expression du temps en langue française se révèle multiple et complexe. Par ailleurs, elle envisage des aspects qui n'ont que peu (ou pas) d'importance dans d'autres langues.

2.3 L'expression du temps en chinois

Les langues romanes et germaniques (leur forme de conjugaison, leurs adverbes temporels, leur aspect,...) nous sont relativement familières. Mais dans d'autres langues l'expression du temps est très différente. Cette section évoque quelques façons d'exprimer le passé, le présent et le futur en chinois.

Le présent en chinois

着 (zhe) est un suffixe verbal qui signale une action en cours, seule ou simultanée à une autre : « 我正忙着呢。» (« Je suis occupée. », action seule en cours) ; « 我笑着吃。» (« Je mange en riant. », deux actions simultanées).

Cette particule a plusieurs prononciations, différentes selon le sens qu'elle porte. Ce sens peut également varier selon les autres caractères associés. Ainsi, lorsque le caractère 着 est prononcé « zhuó » et est accompagné de 穿, il signifie « porter un vêtement » [23].

正在 (zhèngzài), en revanche, est un adverbe (« maintenant, être en train de ») qui se place devant le verbe pour décrire le déroulement d'une action : « 我正在做吃的。» (« Je [suis en train de] prépare[r] à manger. »).

Le futur en chinois

Le futur est généralement exprimé de façon implicite par le contexte. Cependant, deux caractères peuvent apporter une indication supplémentaire : 要 yào et 会 huì. À noter que ces deux sinogrammes peuvent avoir beaucoup de sens différents.

Ainsi, 要 yào est à la fois un adjectif qui signifie « important », un verbe « vouloir », « demander », un auxiliaire « devoir », « falloir », « avoir besoin de »,... Quand il exprime le futur, 要 indique une action qui va avoir lieu : « 我明天要去上课。» (« Demain j'irai en cours. »)

Cependant, cette phrase peut également être traduite avec le semi-auxiliaire « devoir » qui marque l'obligation, « Demain je dois aller en cours. ».

En revanche, 会 huì indique plutôt une probabilité qu'un futur :
« 明天会下雨。» (« Demain j'irai en cours. »). Il est parfois utilisé comme futur éloigné. À l'écrit, on utilise généralement le caractère 将 (jiāng) tandis que 会 est utilisé à l'oral.

Le passé en chinois

过 (guò) indique un passé révolu, c'est-à-dire quelque chose qui a commencé, s'est déroulé et s'est terminé dans un temps passé au moment de l'énonciation. Un exemple typique est d'indiquer quelque chose qu'on a déjà fait : « 我去过比利时。» (« Je suis (déjà) allé en Belgique. »), ou pas encore fait : « 我不去过中国。» (« Je ne suis jamais allé en Chine. »).

23. Pour plus d'exemples, se rendre à la page Internet d'un dictionnaire chinois-anglais `http://ce.linedict.com/#/cnen/entry/24b90203b7a743f7a8ac576c0ec77007` (cette page renvoie à la définition de着 zhuó).

了 est une particule qui indique un changement d'état : « 我知道了。» (« J'ai compris. »). Elle se place alors en fin de phrase, mais quand elle indique une action passée, elle se place immédiatement après le verbe : « 我买了一本书。» (« J'ai acheté un livre. »).

是... 的 (shì... de), cette structure accentue un aspect en particulier qui s'est passé. Elle est typiquement utilisée pour mettre en exergue le lieu ou le temps : 他是昨天去上海的。, « Hier, il est allé à Shanghai. » insiste sur le *hier*, 昨天.

2.4 Conclusion

Ce chapitre a balayé quelques-uns des aspects de l'expression de la temporalité en français et a permis de nous rendre compte de la difficulté que représentent la reconnaissance et la compréhension précises de toutes les informations temporelles dans un texte.

Nous avons privilégié une approche, celle de Kevers (2011), et présenté la classification des adverbiaux temporels qu'il avait établie en modifiant celle de Borillo [24]. Notons que les travaux de Gosselin [25] sont également une référence dans l'expression du temps en français. En outre, il y a d'autres grammairiens et chercheurs qui ont étudié cet aspect de la langue de Voltaire. Parmi eux, nous citerons Benveniste [26], Confais [27], Pinchon [28],...

Dans notre outil, nous identifierons principalement les adverbes, les connecteurs, les indicateurs chronologiques et les termes de repérages par rapport à l'énonciation ou à l'énoncé.

24. Borillo, A. (1983). Les adverbes de référence temporelle dans la phrase et dans le texte. *DRLAV : Revue de linguistique*, 29 : 109-131.
Borillo, A. (1986). Lexique et syntaxe : Les emplois adverbiaux des noms de temps. In *Actes du séminaire "Lexique et Traitement Automatique des Langues"*, Toulouse, France. Université Paul Sabatier.
Borillo, A. (1998). L'expression de la durée : construction des noms et des verbes de mesure temporelle. *Linguisticae Investigationes*, XII(2) : 363–396.

25. Gosselin, L. (1996). *Sémantique de la temporalité en français*. Champs linguistiques. De Boeck, Bruxelles.
Gosselin, L. (2005). *Temporalité et modalité*. Champs linguistiques. De Boeck, Bruxelles.

26. Benveniste, E. (1974). *Problèmes de linguistique générale*. Paris, Gallimard.

27. Confais, J.P. (1995). *Temps, mode, aspect*. Toulouse, Presses universitaires du Mirail-Toulouse.

28. Pinchon, J. (1969). Problèmes de classification. Les adverbes de temps. *Langue Française*.

Chapitre 3

L'extraction d'informations

1 Quelques définitions

> *Information extraction is the automatic identification of selected types of entities, relations, or events in free text.* (Grishman *in* ed. Mitkov, 2005 : 545).

L'extractions d'informations (EI) reconnaît des éléments spécifiques pré-établis dans des textes libres (non-structurés) rédigés en langage naturel. L'EI a été développée dans le cadre des conférences MUC (Message Understanding Conference) et ACE (Automatic Content Extraction). Selon Cunningham (2006 : 665),

> *Information extraction (IE) is a technology based on analyzing natural language in order to extract snippets of information. The process takes texts (and sometimes speech) as input and produces fixed-format, unambiguous data as output.*

La mise en place de systèmes utilisant l'EI peut être longue et fastidieuse : une grande connaissance de la langue est nécessaire pour que le texte soit efficacement analysé et qu'il ne manque pas de l'information après extraction. De plus, les systèmes EI sont parfois spécifiques à certains domaines. Notre système a d'ailleurs été construit pour le domaine médical. Et, s'il peut être utilisé avec d'autres textes (après tout, l'expression du temps en français est semblable partout), peut-être n'obtiendrions-nous pas d'aussi bons résultats qu'avec un rapport de sortie clinique. De fait, le domaine médical comprend plusieurs expressions temporelles qui lui sont propres, et l'usage des abréviations est très répandu.

2 L'EI et le TAL

L'extraction d'informations est très utilisée dans le traitement automatique du langage (TAL). Jurafsky & Martin (2009 : 759-761) comptent six étapes dans les systèmes d'EI. (1) La première est la **reconnaissances des entités nommées** (named entity recognition, ou NER), c'est-à-dire la reconnaissance des noms propres (les personnes, les lieux, les organisations). (2) La deuxième étape concerne la **résolution des références**. Elle permet d'inclure une série de termes reconnus dans le texte comme référant à une seule et même entité. Par exemple, si les termes « hôpital de la Citadelle » et « Citadelle » apparaissent à différents endroits du textes, la résolution des références va permettre de dire que ces deux syntagmes renvoient à la même entité. (3) La troisième étape permet de **détecter et de classer les relations sémantiques** entre des entités. (4) La quatrième se rapporte à la **détection et classification des événements**. Ainsi, dans le texte nous avons déjà souligné les entités et les relations qu'elles entretiennent entre elles. Cette étape permet de trouver et de classer les événements auxquels ces entités participent. Les dernières étapes relèvent de la temporalité, avec (5) la **reconnaissance des expressions temporelles** et (6) **l'analyse temporelle des événements**.

Appelt & Israel (1999 : 7) distinguent deux approches spécifiques de développement de systèmes d'EI : la *Knowledge Engineering Approach*, se basant sur les connaissances des experts, et l'*Automatic Training Approach*, se basant sur l'entraînement automatique.

> *The **Knowledge Engineering Approach** is characterized by the development of the grammars used by a component of the IE system by a "knowledge engineer," i.e. a person who is familiar with the IE system, and the formalism for expressing rules for that system, who then, either on his own, or in consultation with an expert in the domain of application, writes rules for the IE system component that mark or extract the sought-after information. Typically the knowledge engineer will have access to a moderate-size corpus of domain-relevant texts (a moderate-size corpus is all that a person could reasonably be expected to personally examine), and his or her own intuitions. The latter part is very important. It is obviously the case that the skill of the knowledge engineer plays a large factor in the level of performance that will be achieved by the overall system.*

In addition to requiring skill and detailed knowledge of a particular IE system, the knowledge engineering approach usually requires a lot of labor as well. Building a high performance system is usually an iterative process whereby a set of rules is written, the system is run over a training corpus of texts, and the output is examined to see where the rules under- and overgenerate. The knowledge engineer then makes appropriate modifications to the rules, and iterates the process.
(Applet & Israel, 1999 : 7).

The ***Automatic Training Approach*** *is quite different. Following this approach, it is not necessary to have someone on hand with detailed knowledge of how the IE system works, or how to write rules for it. It is necessary only to have someone who knows enough about the domain and the task to take a corpus of texts, and annotate the texts appropriately for the information being extracted. Typically, the annotations would focus on one particular aspect of the system's processing. For example, a name recognizer would be trained by annotating a corpus of texts with the domain-relevant proper names. A coreference component would be trained with a corpus indicating the coreference equivalence classes for each text.*

Once a suitable training corpus as been annotated, a training algorithm is run, resulting in information that a system can employ in analyzing novel texts. Another approach to obtaining training data is to interact with the user during the processing of a text. The user is allowed to indicate whether the system's hypotheses about the text are correct, and if not, the system modifies its own rules to accomodate the new information.
(Applet & Israel, 1999 : 7).

3 L'extraction de la temporalité

Le précédent chapitre nous a permis de clarifier ce qu'inclut la temporalité. Créer un système qui extrait efficacement et sans erreurs toutes les marques temporelles est long, complexe et difficile. L'outil que nous avons développé se concentre avant tout sur les expressions du temps explicites (les adverbes, connecteurs, etc.).

Pour rappel, Kevers (2011 : 149) détermine trois facettes principales à l'extraction de l'information temporelle :

> *[1] La première [...] consiste à* ***repérer et délimiter*** *à l'aide d'un marquage* ***les expressions temporelles****. [2] La deuxième consiste à* ***interpréter ces expressions*** *dans le but d'y associer de manière explicite et univoque la valeur temporelle qu'elles véhiculent, et cela selon un format normalisé. Enfin, [3] le troisième aspect concerne la* ***détection d'événements****, l'****identification des liens*** *qui relie ceux-ci aux expressions temporelles, et finalement leur* ***organisation chronologique****.*

Concernant l'interprétation des expressions temporelles, Mani (2004 : 2) évoque le problème du *default inference*, qui indique que pour comprendre une phrase, un certain sens logique est nécessaire. Cependant, une connaissance du contexte [1] peut amener à une interprétation différente.

Kevers reprend également (2011 : 150) Ahn *et al.* (2005) en listant trois aspects de l'information temporelle :

(1) Le premier aspect indique que la temporalité est transmise par des groupes de mots reconnaissables et qui peuvent être annotés ;

(2) Ces groupes de mots sont interprétables (on leur donne une certaine valeur temporelle) ;

(3) Ces valeurs temporelles sont liées aux faits du discours et les organisent.

1. Le terme *contexte* inclut autant le contexte d'énonciation, que le milieu, les antécédents médicaux familiaux et de la personne concernée, etc.

Chapitre 4

Temporalité dans le domaine médical : État de l'art

Précédemment, nous avons noté qu'il existe différentes façon de transmettre une information temporelle et relevé que le domaine médical recourt à des expressions temporelles spécifiques. Ainsi,

« J+1 », « J3 », « H2 » ;

« 8W4D » ;

« per-opératoire », « postopératoire » ;

« 2x/j 5j/7 » ; ...

sont toutes des expressions que nous ne retrouverons pas (ou très rarement) dans un autre domaine.

Les possibilités qu'offrent l'extraction automatique des événements et l'identification de leur information temporelle dans le domaine médical font de ce dernier, depuis quelques années, le sujet de nombreuses recherches.

Quelques-unes seront présentées dans ce chapitre. Ainsi, le TimeML est un langage de spécification qui a été développé pour l'annotation et la normalisation de l'information temporelle d'un texte. Les systèmes évoqués ensuite sont ceux à base de règles, les approches statistiques, et enfin les systèmes dits hybrides (qui reprennent des règles et des méthodes statistiques).

1 TimeML - un langage de balisage

TimeML est un langage de balisage conçu pour la reconnaissance et la normalisation des expressions temporelles et des événements présents dans un texte. Ce langage propose un guide d'annotation initialement prévu pour un texte libre en anglais (Pustejovsky et al. 2003). Cependant, TimeML a par la suite été développé dans d'autres langues, dont le français [1].

TimeML est spécialisé dans les expressions temporelles, les événements et les relations qu'ils entretiennent. Il a été développé dans le but d'annoter et de standardiser ces éléments dans les textes écrits en langue naturelle.

L'annotation est prévue comme suit [2] :

<EVENT>	permet de marquer les événements, ponctuels ou ayant une certaine durée, qui se produisent ou qui se sont produits
<TIMEX3>	permet de marquer les heures, les dates, les intervalles et les durées. Ils peuvent être complètement spécifiés (ex : « le 3 juin 1950 ») ou non (ex : « le 3 juin »)
<SIGNAL>	permet d'annoter les mots indiquant comment les informations temporelles sont reliées (ex : avant, après, durant, . . .)
<TLINK>	Temporal link : permet de décrire les relations temporelles entre les balises <EVENT>, entre les balises <TIMEX3>, ou entre les balises <EVENT> et <TIMEX3>
<ALINK>	Aspectual link : permet de décrire les relations entre les balises <EVENT> correspondant à un événement aspectuel (commencer, finir, . . .) et les balises <EVENT> correspondant à l'événement argument
<SLINK>	Subordination link : permet de décrire le contexte introduisant les relations entre deux balises <EVENT>

FIGURE 4.1 – Les balises TimeML utilisées en français.

TimeML permet à l'information temporelle d'être exprimée de manière structurée. Ce langage a été développé dans le but de répondre à quatre attentes : (1) identifier un événement et déterminer son espace temps, (2) ordonner les événements les uns par rapport aux autres, (3) permettre de contextualiser les expressions temporelles non spécifiées (du type « la semaine passée »), et enfin, (4) permettre de calculer la durée d'un événement et les répercussions qu'a cet événement [3].

1. Bittar (2010) a d'ailleurs rédigé un guide d'annotation pour le français. Il est à noter que ce guide est en réalité pour l'ISO-TimeML, extension de TimeML. Cette extension et ses différences par rapport à TimeML sont également présentées dans ce chapitre.

2. CAIRN.INFO, *La composante temps dans l'information géographique textuelle*. `http://www.cairn.info/revue-document-numerique-2007-2-page-129.html`, page consultée le 16 juillet 2016.

3. `http://www.timeml.org/`, page consultée le 16 juillet 2016.

Par exemple :

L'expression « avant septembre 2017 » serait analysée comme suit :

> <TIMEX3><SIGNAL> avant </SIGNAL> septembre 2017 </TIMEX3>

La phrase « Luc a commencé à préparer une choucroute. » [4].

> Luc a <EVENT eid="e1" eiid="ei1" class="ASPECTUAL" pos="VERB" tense="PAST"> commencé </EVENT> à <EVENT eid="e2" eiid="ei2" class="OCCURRENCE" pos="VERB" vform="INFINITIVE"> préparer </EVENT> une choucroute.
> <ALINK lid="l1" eventInstanceID="ei1" relatedToEventInstance="ei2" relType="INITIATES"/>

Ce second exemple montre l'annotation d'une relation aspectuelle entre un verbe matrice aspectuel [5] (« a commencé ») et son complément événementiel (sous la forme de l'infinitif « préparer »).

Le verbe « commencer » est annoté par la balise **<EVENT>**, la classe **<ASPECTUAL>** et les traits grammaticaux **<VERB>** (**POS** signifie *part-of-speech* et **<PAST>** (***tense***). Le verbe « préparer » est également un **<EVENT>**, mais de classe **<OCCURRENCE>**. La balise `vform` signale la forme verbale (un infinitif dans ce cas). Les attributs `eid` et `eiid` permettent d'identifier les événements et leurs instances. Finalement **<ALINK>** (identifiée par l'attribut `lid`) désigne une relation aspectuelle. Pour rappel, la figure 4.1 définit ce type de relation comme *permet[tant] de décrire les relations entre les balises* **<EVENT>** *correspondant à un événement aspectuel [...] et les balises* **<EVENT>** *correspondant à l'événement argument*. Cette balise comprend plusieurs attributs : `eventInstanceID` et `relatedEventInstance` servent à sélectionner les instances des événements concernés : `ei1` pour « commencer » et `ei2` pour « préparer ». Enfin, l'attribut `relType` désigne le type de relation aspectuelle concernée. Dans cet exemple, il s'agit d'une relation **INITIATES**, qui indique qu'`ei1` se localise avant `ei2` [6].

L'avantage de TimeML est donc qu'il peut représenter les complexités des relations entre les expressions temporelles et les événements. Llorens *et al.* (2012 : 3045) donne deux types de

4. L'exemple et son analyse viennent de Bittar & Danlos (2009 : 4).

5. Verbes qui signalent une phase initiale, médiane ou terminale, tels que « commencer », « entamer », « continuer », « poursuivre », « terminer », « arrêter », etc. (*Ibid.*).

6. Pour plus d'explications sur les balises et attributs et pour plus d'exemples, se rendre sur la page Internet `https://www.researchgate.net/publication/265349449_ISO-TimeML_Annotation_Guidelines_for_French_Version_10`, reprenant le guide d'annotation d'André Bittar (2010).

notations de normalisation d'une timex[7], *(i) a Gregorian calendar time or date formatted in the ISO 8601 standard as in [October 2012 :* `2012-10`*], or (ii) a period formatted as* `P`*, a* `number` *and an abbreviated* `temporal unit`*, as in [Two weeks :* `P2W`*].*

Une extension a été apportée au langage TimeML, l'ISO-TimeML. C'est un langage de spécification formel utilisé pour marquer l'information temporelle dans le langage naturel. La norme ISO-TimeML ne diffère pas tellement du langage TimeML, mais elle a été développée dans un esprit multilingue et ne contient pas certaines redondances de TimeML (par exemple, la balise `MAKEINSTANCE`, utilisée pour les différentes instances d'un événement, est supprimée) (Bittar, 2008 : 6).

Un autre changement important entre les deux langages est le passage d'une annotation par ligne (TimeML) à une annotation "en confrontation" (ISO-TimeML). La raison première d'étendre le TimeML à l'ISO est la production de ressources annotées durables du langage pour des applications pratiques. Trois éléments doivent être précisés : la position de l'intervalle par rapport aux autres intervalles (*order*), la taille de l'intervalle (*measure*) et le nombre d'intervalles (*quantity*).

La représentation temporelle dans ce type de corpus est cependant difficile, à cause de la diversité des expressions de temps, la complexité à déterminer les relations temporelles entre les événements, et la difficulté à gérer les *granularités* temporelles d'un événement. De manière plus générale, elle est également ardue dans le traitement de la langue naturelle.

Ainsi, Styler et al.[8] proposent une implémentation et extension d'ISO-TimeML, le Thyme-TimeML :

> *considerable effort has gone into both describing and proscribing the annotation of temporal orderings that are inferable only through domain-specific temporal knowledge*[9].

Un des buts de l'extension est l'identification automatique de l'historique médical d'un patient.

7. Ils définissent d'ailleurs une *timex* comme étant *a linguistic expression referring to a time, period, or recurring pattern in time* (Llorens *et al.*, 2012 : 3044). Pour rappel, nous utilisons le terme *timex* dans ce travail pour renvoyer à n'importe quelle expression temporelle.
8. Styler et al. (2014).
9. *Ibid.*, : 144.

De plus,

> the *[t]emporal expression (Timex3) in the clinical domain function the same as in the general linguistic community, with two notable exceptions. ISO-TimeML Sets (statements of frequency) occur quite frequently in the medical domain [...], the second major change for the medical domain is a new type of Timex3 which we call prepostexp. This covers temporally complex terms.* [10].

Enfin,

> *the process of normalizing those Timex3s is significantly more complex relative to the general domain, because many temporal expressions are anchored not to dates or times, but to other Events* [11].

1.1 Systèmes à base de règles

Les systèmes à base de règles sont mis en place à partir de ressources linguistiques (dictionnaires, grammaires, thesaurus, etc.). Le système s'appuie sur ces ressources développées à la main par un expert. Ce type de système est défini par Gupta (2015 : 4) de la manière suivante :

> *Rule based or knowledge-driven methods exploits human knowledge about the contents of the text to be processed in addition to linguistic and lexicographic knowledge . This knowledge is encoded in the form of patterns that express rules which are used to extract desired information from the text. Information is mined from corpora by using predefined or discovered linguistic patterns, which can be either lexico-syntactic patterns or lexico-semantic patterns. While lexico-syntactic exploits lexical representation with syntactical information, lexico-semantic uses semantic or meaning of the information along with lexical representations.*

10. *Ibid.,* : 147.
11. *Ibid.*, : 148.

1 - HeidelTime

HeidelTime [12] a été créé par Jannik Strötgen et Michael Gertz dans le cadre du défi TempEval-2 [13]. Il a été conçu pour l'extraction et la normalisation des expressions temporelles en anglais dans le but d'utiliser l'information temporelle dans d'autres applications. C'est un système à base de règles, défini comme un

> *multilingual, domain-sensitive temporal tagger developed at the Database Systems Research Group at Heidelberg University. It extracts temporal expressions from documents and normalizes them according to the TIMEX3 annotation standard* [14].

HeidelTime a été élaboré en UIMA (Unstructured Information Management Architecture) [15] afin de proposer un système utilisable dans différents contextes et domaines sans (ou avec peu de) modifications. Le système est aujourd'hui développé et abouti dans quelque treize langues et des recherches sont menées afin de l'étendre automatiquement à toutes les langues du monde [16]. Moriceau & Tannier [17] ont élargi le système au français.

Un des aspects importants et intéressants de cet outil est la **normalisation** ;

> *the operation of turning a temporal expression into a formated, fully specified representation (this includes finding the absolute value of relative dates)* [18].

HeidelTime est reconnu comme un des systèmes les plus performants. Il est par ailleurs régulièrement utilisé à côté d'autres outils ou de méthodes (notamment GATE-Time (pages 62-63), MedTime (page 65), le système de Tang *et al.* (page 69), systèmes présentés dans cet état de l'art).

Concernant son fonctionnement, HeidelTime

> *is a rule-based system with a separation between language-dependent resources and generic Java code. Resources consist in extraction rules [...] and lexicons for*

12. (Strötgen & Gertz, 2010 ; Jindal & Roth, 2013 ; Zweigenbaum & Tannier, 2013 ; Névéol & Tannier, 2014 ; Tapi Nzali *et al.*, 2015 ; Tapi Nzali & Tannier, 2015).
13. Voir annexe 1 page 139 pour la présentation des défis TempEval.
14. `https://github.com/HeidelTime/heideltime`, consulté le 4 avril 2016.
15. *[S]oftware systems that analyze large volumes of unstructured information in order to discover knowledge that is relevant to an end user [...] UIMA enables applications to be decomposed into components.* `https://uima.apache.org/`, page consultée le 29 juillet 2017.
16. Strötgen & Gertz (2015).
17. Moriceau & Tannier (2014).
18. *Ibid.*, : 1.

> *normalization [...]. Absolute temporal expressions [...] are extracted and normalized by the extraction rules. Relative temporal expressions [...] are extracted by the rules but are left underspecified [...]. Then, a normalization (Java code) is applied [...] As a reference time, normalization can use the document creation time [...] or the previously mentioned date. HeidelTime required sentence, token and part-of-speech information* [19].

HeidelTime utilise le temps des verbes et des phrases afin de déterminer si l'expression temporelle se réfère au passé ou au futur. Le TreeTagger français permet de donner cette information lors du pré-traitement des documents. Les résultats obtenus sont satisfaisants, mais quelques erreurs persistent. C'est souvent dû à la mauvaise identification du temps de la phrase : HeidelTime prend en compte le verbe le plus proche de l'expression temporelle, mais il ne s'agit pas toujours du verbe dont dépend le complément de temps :

> « François Hollande **assure** que le prochain président de la République **devra** "être l'inverse de Nicolas Sarkozy", dans un entretien à Libération **mercredi**. » [20]

Dans cet exemple, l'expression temporelle (« mercredi ») n'est pas normalisée correctement. En effet, le verbe conjugué au présent « assure » devrait être utilisé pour la normalisation, tandis que le verbe au futur « devra » est un verbe secondaire. Cependant, dû à la proximité de la timex avec ce verbe secondaire, c'est lui qui est utilisé pour la normalisation.

Les ressources françaises sont accessibles sur le site d'HeidelTime [21]. Quelques modifications ont été apportées et d'autres recherches ont été réalisées afin d'élargir le domaine des textes traités. Tapi Nzali & Tannier [22] ont soumis un corpus médical à HeidelTime, insérant à l'outil quelque quatorze règles d'extraction et de normalisation pour le français. Par exemple, « J+1 » désigne le lendemain d'une opération, ou « J11 », onze jours après le début d'un traitement.

Les ressources de l'outil comprennent trois types de fichiers : *pattern*, *normalizations* et *rules* [23]. Les fichiers *pattern* contiennent les mots et expressions utilisées dans les expressions temporelles. Les fichiers du second type, *normalizations*, contiennent l'information normalisée sur les

19. *Ibid.*
20. Exemple et analyse proviennent de *Ibid.,* : 5.
21. L'outil est disponible gratuitement en ligne, à l'adresse `https://github.com/HeidelTime/heideltime`.
22. Tapi Nzali & Tannier (2015).
23. Sur le site GitHub/HeidelTime, les dossiers des fichiers en français gardent le nom anglais (en fait, "pattern" est en réalité appelé "repattern", `https://github.com/HeidelTime/heideltime/tree/master/resources/french`, page consultée le 23 mai 2016.).

patterns (par exemple, le mois de février équivaut au chiffre 02). Enfin, les fichier *rules* comprennent les règles de date (une centaine de règles), de temps (vingt règles), de durée (vingt-cinq règles) et d'expressions figées (douze règles). Chacune des règles contient deux parties : une partie extraction et une partie normalisation. La première définit les expressions qui doivent être retrouvées dans le document, tandis que la seconde normalise les expressions extraites. Par exemple, l'expression « November 2001 » sera analysée de la sorte :

> « `RULENAME="date_r7a",EXTRACTION="(%reMonthLong|%reMonthShort)( of | )%reYear4Digit", NORM_VALUE="group(5)%normMonth(group(1))"` »[24].

Dans cet exemple, RULENAME est un nom unique donné à une règle. *Extraction* contient l'expression régulière pour extraire l'expression désirée. Les attributs *reMonthLong, reMonthShort, reYear4Digit* sont utilisés comme des ressources de pattern. NORM_VALUE est l'expression normalisée, qui utilise *normMonth* comme ressource. Des ressources de normalisation ressemblent à :

```
normMonth("November") = "11"
normSeason("summer") = "SU"
```

Jindal & Roth[25] affirment qu'HeidelTime est un des deux meilleurs systèmes d'extraction temporelle, avec SUTime (présenté au point suivant). Ils ont pointé comme défaut le fait qu'ils ne soient disponibles que pour des textes généraux, mais grâce aux efforts de Tapi Nzali & Tannier[26], les règles nécessaires ont été ajoutées au logiciel HeidelTime afin de pouvoir extraire la temporalité dans le domaine médical.

2 - SUTime

SUTime[27] est présenté comme

> *a temporal tagger for recognizing and normalizing temporal expressions in English text*[28].

24. L'exemple et son explication viennent de Gupta (2015 : 5).
25. Jindal & Roth (2013).
26. Tapi Nzali & Tannier (2015).
27. Angel & Manning, (2012). SUTime est accessible sous la forme d'une librairie Java implémentée comme annotateur dans le Stanford CoreNLP pipeline `http://nlp.stanford.edu/software/sutime.shtml`.
28. Angel & Manning, (2012 : 1).

C'est un système à base de règles qui permet la reconnaissance d'expressions temporelles grâce à des modèles d'expressions régulières.

> *Temporal expressions are bounded in their complexity, so many of them can be captured using finite automata [...] [using] a cascade of finite automata can be very effective at extracting information from text* [29]

La stratégie suivie par Angel & Manning dans la conception de SUTime a (1) d'abord été de réaliser des modèles pour les expressions numériques, (2) ensuite des modèles afin de trouver les expressions temporelles simples, et (3) enfin de former des modèles composés sur les expressions déjà trouvées pour reconnaître les expressions plus compliquées. L'utilisation d'expressions régulières facilite l'ajout de règles.

SUTime permet principalement trois actions :

(1) extraire des expressions temporelles de textes, l'annotation se faisant avec TIMEX3 (qui, pour rappel, fait partie de TimeML) ;

(2) représenter les objets temporels en classes Java (donner en une suite logique les expressions temporelles extraites) ;

(3) déterminer les dates réelles auxquelles les expressions font référence (en utilisant les dates des documents).

Le système a plusieurs limitations, dues notamment à l'absence de règles pour détecter les expressions temporelles ambigües. De plus, la date du document ne permet pas toujours de déterminer correctement la date réelle d'un jour de la semaine : si le document a été écrit un lundi, l'expression temporelle « vendredi » ne permet pas de savoir s'il s'agit du vendredi précédent ou suivant. De plus, les vacances ne sont pas reconnues et les périodes de temps (« from 3 to 4 pm », « 12-13 March 2011 ») ne sont pas bien reconnues. Enfin, nous noterons que SUTime n'analyse pas de texte en français (d'autres règles devraient alors être implémentées) et le corpus n'est pas non plus axé sur le domaine médical.

29. *Ibid.*

1.2 Méthodes statistiques

Les méthodes statistiques utilisent des techniques empiriques pour apprendre des textes (ou d'autres données) et ainsi construire des modèles de langues. Ces derniers seront ensuite utilisés dans des tâches de pré-traitement. Ce type de recherches date des années nonante. Dans les années 2000, leur utilisation s'est généralisée. Gupta (2015 : 7) en donne la définition suivante :

> *Statistical methods are class of algorithm that learn a model by looking at annotated training examples. Among the supervised learning algorithms for NER, considerable work has been done using Hidden Markov Model (HMM), Decision Trees, Maximum Entropy Models (ME), Support Vector Machines (SVM) and Conditional Random Fields (CRF). Typically, supervised methods either learn disambiguation rules based on discriminative features or try to learn the parameter of assumed distribution that maximizes the likelihood of training data.*

Dans cette section, nous allons très brièvement présenter deux approches statistiques susnommées par Gupta (les modèles de Markov cachés et les conditional randorm fields models) [30].

Modèle de Markov Caché (HMM, Hidden Markov Model)

> *Un modèle de Markov caché est basé sur un modèle de Markov [modèle statistique composé d'états et de transitions. Une transition [unidirectionnelle] matérialise la possibilité de passer d'un état à un autre], sauf qu'on ne peut pas observer directement la séquence d'états : les états sont cachés. Chaque état émet des "observations" qui, elles, sont observables. On ne travaille donc pas sur la séquence d'états, mais sur la séquence d'observations générées par les états.* [31]

Le principe d'un HMM est qu'un processus contient un ensemble d'états possibles $\{S_1, S_2, ...S_N\}$. Ce processus passe d'un état à un autre autre et génère ainsi une séquence du type : $S_{i1}, S_{i2}, ...,$ $Sik, ...$ Le principe même des HMM réside dans la probabilité d'occurrence d'un état, qui dépend entièrement de l'état précédent.

30. Voir ses travaux (2015 : 7-14) pour de plus amples informations.
31. Rodon, `http://igm.univ-mlv.fr/~dr/XPOSE2012/HiddenMarkovModel/description.html`, page consultée le 31 juillet 2017.

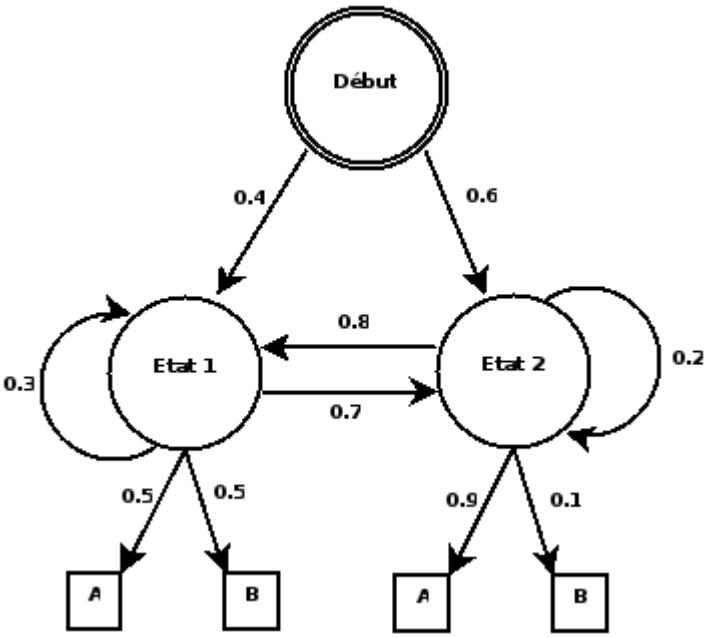

FIGURE 4.2 – Exemple d'un modèle de Markov caché.

L'image ci-dessus[32] représente un modèle de Markov caché. L'état « Début » permet de représenter la probabilité de commencer dans chacun des états du modèle. Il est à noter que la somme des probabilités de chaque état équivaut à 1. Chaque état a par ailleurs deux observations, A et B. Il s'agit d'observations que les états peuvent émettre, chacune avec une certaine probabilité.

Conditional Random Fields (CRF) Based Models Les modèles CRF sont des systèmes probabilistes pour l'étiquetage et la segmentation de données structurées.

> *The underlying idea is that of defining a conditional probability distribution over label sequences given a particular observation sequence, rather than a joint distribution over both label and observation sequences*[33].

Les CRF sont donc des modèles qui indiquent une mesure de certitude pour l'étiquetage des données, ce qui en fait un des avantages premiers de ce type de modèle. Ils sont généralement plus fiables que les modèles de Markov cachés :

32. Image et explications viennent de `http://igm.univ-mlv.fr/~dr/XPOSE2012/HiddenMarkovModel/description.html`.

33. Prateek Joshi, `https://prateekvjoshi.com/2013/02/23/what-are-conditional-random-fields/`, page consultée le 31 juillet 2017.

The primary advantage of CRFs over hidden Markov models is their conditional nature, resulting in the relaxation of the independence assumptions required by HMMs in order to ensure tractable inference. Additionally, CRFs avoid the label bias problem, a weakness exhibited by maximum entropy Markov models (MEMMs) and other conditional Markov models based on directed graphical models. CRFs outperform both MEMMs and HMMs on a number of real-world tasks in many fields, including bioinformatics, computational linguistics and speech recognition[34]

Système de Li & Patrick (2012)

Li & Patrick ont développé un système qui utilise une méthode statistique pour trouver les informations temporelles liées à des événements médicaux dans des rapports électroniques de patients en anglais. L'annotation utilisée est une extension de celle de Zhou et al.[35]. Elle comprend cinq classes :

(1) les dates (spécifiques et relatives) ;

(2) la fréquence ;

(3) la durée ;

(4) les moments temporellement clés ;

(5) les expressions temporelles vagues (du style « since then », « at a later time »).

Les événements médicaux pris en compte sont les expressions indiquant une condition de santé, les processus médicaux (« use propofol », « given filling »), les changements dans la condition du patient, le comportement du patient (« sitting out of bed », « slept well »), le plan médical (« aiming for neutral balance ») et quelques autres événements comme « social visit », « wash up », etc.

La figure 4.3 montre l'architecture du système de Li & Patrick. Le début du diagramme correspond à la première étape du traitement du corpus. Leur méthode d'extraction utilise un

"highly restricted training sample expansion" and "the structure distance between the temporal expressions & related event expression" as effective learning features[36].

34. *Ibid.* Pour plus d'informations, consulter la page web de Wallach, H.M., *conditional random fields*. 2005. `http://www.inference.org.uk/hmw26/crf/`.

35. Hripcsak et al (2005), Zhou et al. (2005), Zhou et al. (2006).

36. Li & Patrick. (2012 : 550).

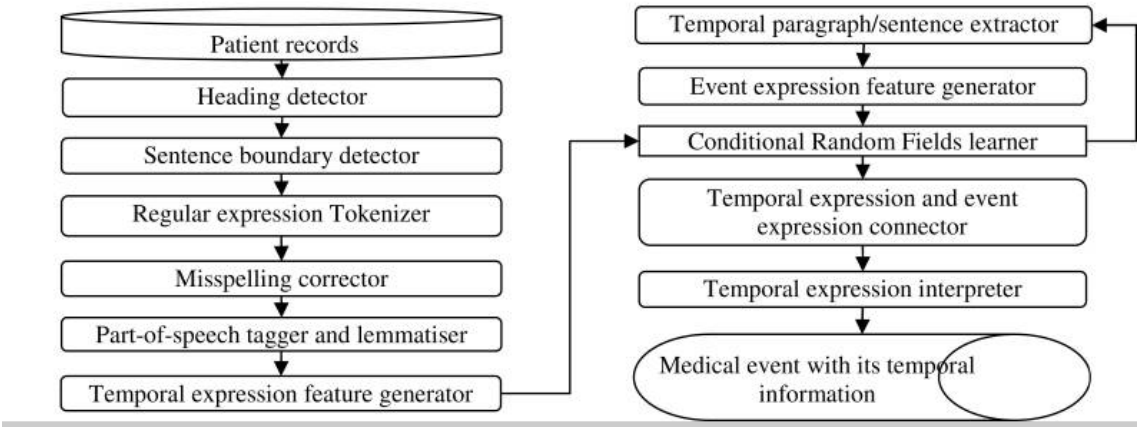

FIGURE 4.3 – Temporal event processor system architecture (Li & Patrick).

Leurs résultats sont encourageants. L'image ci-dessous reprend des tables présentant les résultats obtenus par leur système [37]. Ces deux tables permettent de montrer que le même (ou presque) F-score peut être obtenu tant par leur meilleur modèle [38] que par la méthode à base de règles [39]. La méthode est alliée à leur meilleur modèle. Cela permet une amélioration du rappel et de la précision.

Temporal Class	Frequency	Baseline (model #1) [R/P/F]	Best model (model #22) [R/P/F]
Date and time	362	0.377/0.870/0.526	0.674/0.915/0.776
Event dependent temporal	72	0.000/0.000/0.000	0.042/0.500/0.077
Fuzzy time	243	0.432/0.846/0.572	0.697/0.884/0.780
Duration	116	0.273/0.833/0.412	0.543/0.909/0.680
Frequency	103	0.035/0.750/0.066	0.200/0.955/0.331
Micro Average	896	0.331/0.855/0.477	0.586/0.904/0.711

Table 3. The temporal expressions performance for each class (with frequency) by applying the baseline model and the best model[4].

Temporal Class	**Rule method** (highly restricted training sample expansion mapping) [R/P/F]	Combined model [R/P/F]
Date and time	0.687/0.777/0.729	0.749/0.935/0.831
Key event	0.214/0.882/0.345	0.236/0.944/0.378
Fuzzy time	0.729/0.854/0.787	0.798/0.898/0.845
Duration	0.438/0.875/0.583	0.655/0.883/0.753
Frequency	0.732/0.740/0.736	0.699/0.960/0.809
Micro Average	0.634/0.805/0.710	0.703/0.920/0.797

Table 4. The scores of the highly restricted training sample expansion mapping and the combined learning model for temporal events[4].

Temporal Class	Baseline model (model #1) [R/P/F]	Best model (model #33) [R/P/F]
Temporal related medical event	0.381/ 0.666/0.485	0.712/0.778/0.743

Table 5. The scores of baseline model and the best model in the temporal related medical event extraction[4]

FIGURE 4.4 – Deux tables de résultats du modèle statistique.

37. Pour les détails concernant les modèles, se rendre aux pages 545 à 546 de l'article de Li & Patrick.
38. Les différents modèles mis en place par Li & Patrick ont été testés et leur modèle 22 obtient les meilleurs résultats. Ce modèle 22 correspond à *Bag-of-lemma with correction + part-of-speech + chunk + SCT concept + SCT category + orthography + prefix + suffix* tandis que le baseline model consiste simplement en un *bag of words*. Li & Patrick (2012 : 548).
39. R/P/F dans les deux tables correspondent à Recal/Precision/F-score.

À cause de la faiblesse grammaticale des EPRs (Electronic Patient Records), une méthode d'extraction robuste a dû être utilisée afin d'obtenir des performances élevées. Li & Patrick [40], ont ainsi proposé

> *a sophisticated clinical parser with grammar correction working in concert with domain specific knowledge.*

1.3 Systèmes hybrides

Un système hybride est un système reprenant des méthodes symboliques (à base de règles) et statistiques. Aujourd'hui, ce sont ces systèmes qui sont les plus prometteurs.

1 - GATE-Time

Il existe des logiciels apparentés à Gate, notamment spécialisés dans l'extraction et la reconnaissance de l'information :

> *The GATE family is intended to minimise time and effort in developing and maintaining rich information extraction, retrieval and management systems, while staying at or near to the state of the technological art, partly by favouring interoperation and reuse over reinvention* [41].

Nous nous intéresserons ici à Gate-Time [42] un

> *integrated, open-source toolkit for the annotation of times and events in ISO-TimeML* [43].

Cet outil utilise HeidelTime pour l'étiquetage temporel. Gate-Time comporte également un reconnaisseur et un classifieur d'événements rapides qui peuvent être combinés avec différentes annotations de pré-traitement.

Les dates et expressions temporelles peuvent être explicites, implicites, sous-spécifiées ou relatives. La difficulté de leur normalisation dépend du type d'occurrence des expressions et du

40. Li & Patrick (2012 : 550).
41. Cunningham *et al,* 2013.
42. Derczynski *et al.*, 2016.
43. *Ibid*, : 3702.

domaine des textes dans lequel elles apparaissent. Les expressions explicites peuvent être normalisées avec une connaissance standard du temps (« 2016/01 » équivaut à « Janvier 2016 », par exemple), tandis que les expressions implicites doivent l'être avec des connaissances non standard (« 2016-03-17 » pour le jour de la Saint Patrick en 2016). Par contre, pour les expressions relatives (« il y a deux semaines ») et sous-spécifiées (« novembre »), une référence temporelle est nécessaire à la normalisation. En effet, sans indication chronologique explicite, il est impossible de déterminer de quel mois de novembre nous parlons, ou à quoi se réfère les « deux semaines auparavant ».

Le jumelage entre HeidelTime et GATE reste délicat. HeidelTime a été développé avec une architecture IUMA [44] ; GATE offre seulement quelques aspects similaires à cette architecture. Dans la pratique, combiner des éléments venant de GATE ou de l'IUMA n'est pas toujours recommandé : HeidelTime pré-traite le texte automatiquement. Or, dans l'idée de GATE, il faudrait un pré-traitement flexible. L'image suivante représente l'architecture du *HeidelTime GATE wrapper* imaginé pour lever cette difficulté :

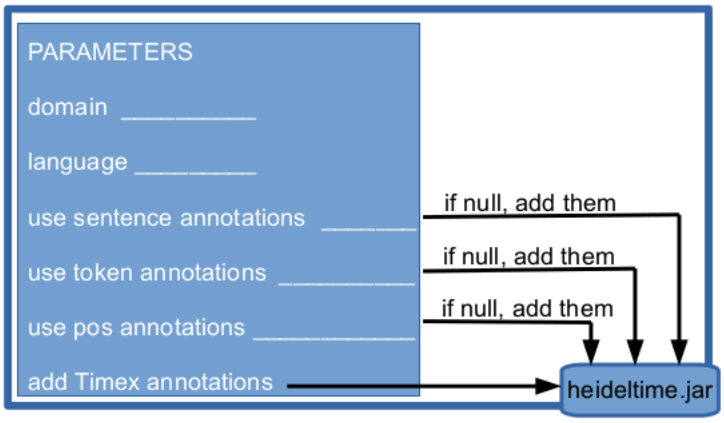

FIGURE 4.5 – Architecture du HeidelTime GATE wrapper.

Ce *wrapper* peut être utilisé avec ou sans une étape de pré-traitement si l'annotation de la phrase, du token et la catégorie grammaticale sont disponibles. L'utilisateur se doit alors de spécifier les noms d'annotation, et HeidelTime procure l'annotation des TIMEX3. En revanche, si l'utilisateur ne fixe pas ces paramètres, le pré-traitement s'exécute avec HeidelTime. (Derczynski *et al.*, 2016 : 3704).

44. Définition à la page 54, note en bas de page 15.

2 - TN-TIES

Le TN-Ties (Triage Note Temporal Information Extraction System) [45] est un système développé pour l'anglais qui

> *systematically processes triage note text and outputs a human and machine readable interpretation of the timing of the events leading up to the ED [Emergency Department] visit* [46].

Ce système s'est largement inspiré des travaux réalisés par Zhou *et al.* [47], étant donné la similarité entre les corpus (domaine médical) et les types de relations temporelles :

Class	**Example of temporal expression**
Relative Date and Time (RDT)	involved in MVC 1 week ago
Duration (DR)	fever of 103 for 3 days
Key Event (KE)	Pain level now : 8/10
Date and Time (DT)	Discharged from hospital 11–17 for same
Fuzzy Time (FT)	history of constipation
Other Event (OE)	took two Vicodin after accident
Recurring Time (RT)	dizzy one day every month

TABLE 4.1 – Types de relations temporelles et des exemples concrets en anglais (Zhou et al., 2005, cités par Irvine et al., 2008 : 3).

TN-TIES comporte trois étapes principales de traitement : un *chunker*, un *classifier* et un *interpreter*. Le système réalise une analyse grammaticale partielle du texte afin d'identifier des syntagmes cohérents. Le *chunker* utilise alors des mots-clés et la ponctuation pour extraire des segments de phrases. Le *classifier* permet ensuite de prédire à quelle classe temporelle appartiennent ces segments. Pour ce faire, il emploie des caractéristiques lexicales et des expressions régulières. Enfin, une série d'étapes d'interprétations (*interpreter*) permet d'identifier la zone temporelle du segment et d'estimer un point temporel plus ou moins précis.

45. Irvine *et al.*, 2008.
46. *Ibid.,* : 328.
47. Zhou *et al.* (2005 & 2006).

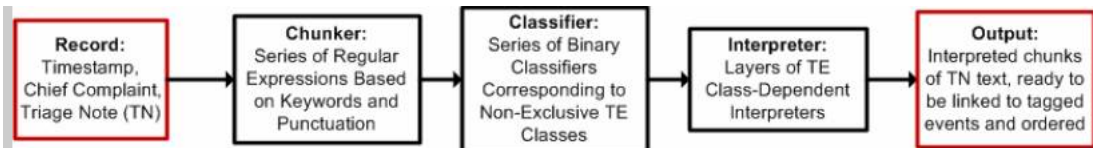

FIGURE 4.6 – Architecture du système TN-TIES.

Un avantage important de ce système est qu'il permet d'extraire et d'interpréter des informations temporelles qui proviennent de notes plutôt que de textes en anglais standard.

3 - MedTime

MedTime [48] est un système hybride qui utilise des modèles à base de règles et une approche par machine learning comme procédures de reconnaissance. Présenté pour le défi i2b2 2012 [49], cet outil comporte six étapes majeures :

(1) le pré-traitement, lui-même constitué de différentes étapes dont le repérage des dates d'admission et de décharge des patients ;

(2) le tag des expressions temporelles du texte par HeidelTime ;

(3) FREQUENCY TE tagging, qui annote les abréviations propres au monde médical qu'HeidelTime ne parvient pas à traiter correctement ;

(4) le *sequence labelling*, réalisé à l'aide de la reconnaissance temporelle basée sur le modèle CRF [50] ;

(5) la normalisation temporelle clinique (les auteurs utilisent différents algorithmes dont *Contextual Alias Registry* et *JChronic*) ;

(6) une étape de post-traitement.

Une des limitations de MedTime est son manque d'identification des relations temporelles (les relation du type TLINK, selon l'annotation de TimeML). Pour rappel, ces relations TLINK *permet[tent] de décrire les relations temporelles entre les balises* `<EVENT>`*, entre les balises* `<TIMEX3>`*, ou entre les balises* `<EVENT>` *et* `<TIMEX3>` [51].

48. Lin *et al.*, 2013.

49. Voir annexe 1 pages 139-140 pour la présentation du défi i2b2.

50. Conditional Random Field, ou champs aléatoires conditionnels *are a probabilistic framework for labeling and segmenting structured data* (`http://www.inference.phy.cam.ac.uk/hmw26/crf/`, page consultée le 28 juillet 2016. Pour plus d'informations, retourner aux pages 59-60 où le fonctionnement des modèles CRF est un peu plus détaillé).

51. CAIRN.INFO, *La composante temps dans l'information géographique textuelle*. `http://www.cairn.info/revue-document-numerique-2007-2-page-129.html`, page consultée le 16 juillet 2016.

4 - A Semantic Web Onthology

A semantic Web Onthology [52] est - en français - une ontologie [53] sémantique du Web. Il existe deux ontologies temporelles, dont une - la SWRL [54] - qui utilisée dans la recherche clinique. Cependant, cette dernière se concentre uniquement sur un événement ponctuel dans le temps et ne prend pas en compte les assertions temporelles présentes dans les rapports écrits des médecins. Des modèles tels que TimeML, qui visent à représenter l'information temporelle qui se trouve dans le langage naturel, ne sont pas compatibles avec l'OWL (Web Ontology Language) [55]. Une solution a été trouvée avec le CNTemporalRelation (Clinical Narrative Temporal Relation) Ontology qui peut analyser l'information temporelle trouvée dans des bases de données structurées et dans le langage naturel. Cette ontologie peut être utilisée pour analyser les instances et les relations temporelles dans les comptes-rendus cliniques. Elles peuvent être enregistrées en RDF Triples (qui comportent un sujet, un prédicat et un objet). Néanmoins, jusqu'à présent, elles n'existent pas en français.

5 - Autres outils et systèmes

Quelques autres systèmes ont également été mis en place lors de différentes recherches. Nous en verrons deux : celui de Roberts *et al.* (2013) et celui de Tang *et al.* (2013). Tous deux ont été réalisés dans le cadre du défi i2b2 2012 [56].

a - Le système de Roberts *et al.* (2013) Ce sytème porte principalement sur des *supervised classifiers* [57], bien que des modèles à base de règles et des *unsupervised learning* aient également été utilisés pour sa création. Ce système se distingue des autres par l'élaboration d'une méthode flexible facilement adaptable aux changements de données (les noms des hôpitaux, des médecins, etc.).

52. Tao *et al.*, (2010).
53. Une ontologie est une *structuration des concepts d'un domaine. Ces concepts sont rassemblés pour fournir les briques élémentaires et exprimer les connaissances dont on dispose dans ce domaine* (Technoloangues.net, `http://www.technolangue.net/article.php3?id_article=280`, page consultée le 29 juillet 2017).
54. Semantic Web Rule Language, it is composed of rules with the format `antecedent -> consequent`.
55. Utilisé dans les ontologies sémantiques du web pour fournir *un langage de représentation de la logique [...] restreinte pour [...] exprimer les connaissances ontologiques* (*Ibid.*).
56. Voir annexe 1 pages 139-140.
57. Utilisé pour le *supervised learning*, ou *apprentissage supervisé*. Il s'agit d'une technique d'apprentissage automatique pour la production de règles à partir de données dites d'apprentissage.

Trois éléments importants composent les *supervised machine learning methods* : (1) les particularités des inputs pour représenter les données, (2) un modèle pour agir en tant que fonction de classification et (3) un processus d'apprentissage afin d'estimer les paramètres du modèle.

Les images ci-dessous représentent l'architecture de la reconnaissance d'événements - ou Event - (figure 4.7), de la reconnaissance temporelle - Timex3 - (figure 4.8) et des liens entre événements - Tlink (figure 4.9).

Pour délimiter les événements, le système créé par Roberts et al. utilise un classifier CRF (Conditional Random Field) et des clusters Brown. Le *type*, la *modalité* et la *polarité* sont alors déterminés en utilisant des classifiers *Support Vector Machine* (SVM) [58] (Roberts et al., 2013 : 867-870).

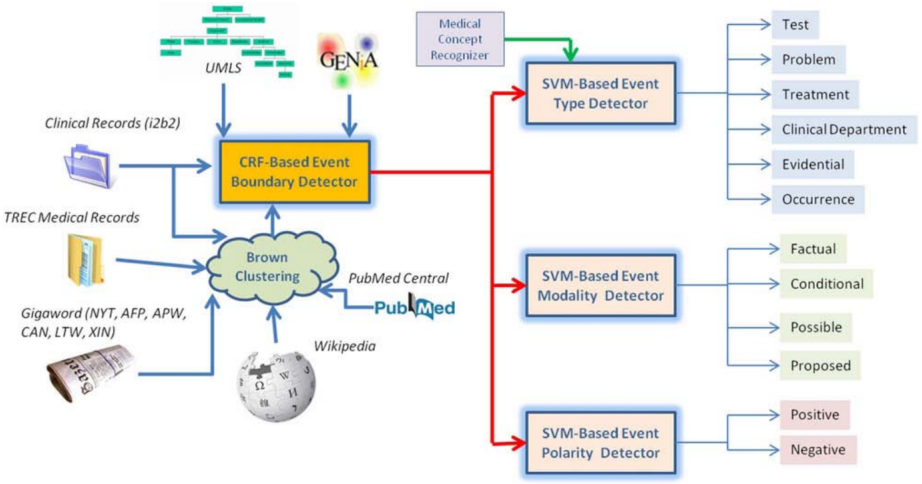

FIGURE 4.7 – Architecture de la reconnaissance des événements (Event).

La reconnaissance temporelle est similaire à celle des événements. Un classifier CRF reconnaît les délimitations Timex3 tandis que deux SVM en déterminent le type et la modalité. Une méthode à base de règles se charge alors de normaliser le Timex3. La normalisation Timex3 s'allie à des règles de TIMEN [59]. En plus de cela, deux types de règles sont utilisés : des règles "pre-TIMEN" qui normalisent complètement les expressions temporelles et les "post-TIMEN" qui corrigent quelques fautes et annotations. (Roberts *et al.*, 2013 : 871-872)).

58. *A support vector machine (SVM) is machine learning algorithm that analyzes data for classification and regression analysis. SVM is a supervised learning method that looks at data and sorts it into one of two categories. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible. SVMs are used in text categorization, image classification, handwriting recognition and in the sciences. A support vector machine is also known as a support vector network (SVN).* (techopedia, `https://www.techopedia.com/definition/30364/support-vector-machine-svm`. Page consultée le 18 juillet 2017). Il s'agit donc de techniques d'apprentissage supervisées.

59. Timen est un normaliseur temporel open source (Roberts *et al.* (2013 : 871).

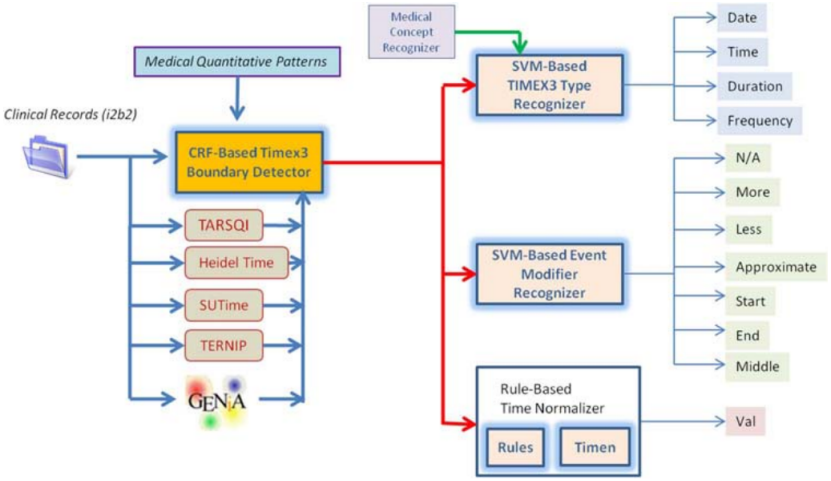

FIGURE 4.8 – Architecture de la reconnaissance temporelle (Timex3).

Pour la reconnaissance des liens temporels entre les événements, il y a une séparation entre la détection d'Event et de Timex3 dans le document des sections temporelles TLink. Un *ranker SVM* repère les liens entre Event et Timex3 dans les textes tandis que, dans les sections temporelles, une règle permet de relier les événements aux dates d'admission ou de sortie. Dans les deux cas, un classifier multi-classes SVM détermine ensuite les types des TLinks. (Roberts et al. (2013 : 872)).

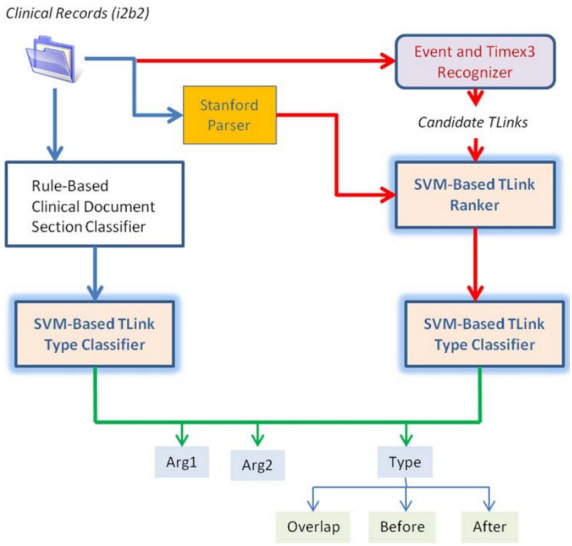

FIGURE 4.9 – Architecture TLink (liens temporels entre événements).

Si ce système parvient à reconnaître les événements et les expressions temporelles, davantage de données d'apprentissage dans le domaine médical seraient nécessaires pour améliorer l'outil. De plus, il faudrait rédiger des règles de normalisation des timex typiquement cliniques. Enfin, dernière remarque, ce système est développé pour l'anglais.

b - Le système de Tang *et al.* (2013) Ce système comporte trois modules : (1) l'extraction d'événements, (2) l'extraction d'expressions temporelles et (3) l'extraction de relations temporelles (ou TLink).

L'extraction d'événements se base sur le système conçu pour le défi i2b2 de 2010, à deux points près : le nombre d'événements - reconnus via un classifier à double étape basé sur des CRFs - est plus important et, une *polarité* et une *modalité* sont assignées aux événements (grâce à un système basé SVM).

Pour l'extraction d'expressions temporelles, un système à base de règles en Python a été développé. Il extrait les expressions temporelles et normalise leurs valeurs simultanément. Tang et al., ont utilisé beaucoup de règles du système HeidelTime en Java.

L'extraction TLink a été divisée en trois tâches : (1) premièrement, une extraction entre événements et sections de temps (qui sont la date d'admission et la date de décharge) - pour chaque événement, le type de TLink (before, after, overlap) doit être connu - ; (2) deuxièmement, le TLink entre événements et temps dans une même phrase ; et (3) troisièmement, le TLink entre des événements et temps à travers plusieurs phrases. Pour ce faire, une procédure de *TLink candidate generation* doit avoir lieu afin de limiter le nombre de paires événements/temps à traiter.

1.4 Conclusion de l'état de l'art

Dans cette section, nous réalisons un tableau récapitulatif des différents outils présentés.

Outil	**Compétences**	**Langue d'application**	**Type de système**	**Domaine médical ?**
HeidelTime	extraction timex normalisation timex	anglais (étendu au français)	règles	oui
SUTime	reconnaissance & normalisation timex	anglais	règles	non
Li & Patrick (2012)	reconnaissance timex liée à des événements médicaux	anglais	statistiques (CRF)	oui
Gate-Time	annotation timex et événements avec ISO-TimeML (et HeidelTime) ; reconnaisseur et classifieur d'événements ; normalisation timex			
TN-TIES	analyse, classification et interprétation des événements et timex liées aux événements précédant l'arrivée à l'hôpital	anglais	hybride	oui
MedTime	reconnaissance et annotation des timex (HeidelTime et FREQUENCY TE) ; normalisation timex	anglais	hybride (règles et CRF)	oui
Semantic Web Ontology	analyse des événements médicaux et des timex	anglais	hybride	oui
Roberts *et al.*, 2013	reconnaissance des événements ; reconnaissance timex ; liens TLINK	anglais	hybride : supervised machine learning	oui
Tang *et al.*, 2013	extraction des événements ; extraction des timex ; extraction des TLINK	anglais	hybride (CRF, SVM) HeidelTime	oui

TABLE 4.2 – Tableau récapitulatif des outils présentés dans l'état de l'art.

1.5 SNOMED-CT, un outil de nomenclature des termes cliniques

SNOMED-CT, Systematized NOmenclature of MEDicine - Clinical Terms, est défini par le site de la santé publique belge comme suit :

> *SNOMED CT® est un système international de terminologie clinique. Il comprend une liste de termes cliniques contrôlés, détaillés et validés médicalement avec leurs synonymes. Cette terminologie est utilisée dans les soins directs aux patients pour documenter les plaintes physiques et psychologiques, les symptômes, les circonstances, les maladies, les interventions chirurgicales, les diagnostics, les résultats et les décisions thérapeutiques. SNOMED CT® permet dès lors aux professionnels de la santé et aux chercheurs d'adopter un langage commun, tant sur le plan national qu'international. A l'heure actuelle, le système est utilisé dans plus de 50 pays, ce qui permet une description cohérente et exploitable de l'information médicale dans les dossiers patient électroniques*[60].

SNOMED-CT, très utilisé pour l'annotation des concepts médicaux[61], contient quelque 350 000 items cliniques, partagés entre 19 axes qui permettent d'exposer les symptômes, maladies, interventions, etc. Ce dispositif fournit les codes, les synonymes ainsi que les définitions des termes cliniques. Ce système est multilingue. Il est possible de combiner les codes pour un concept plus compliqué :

> Ex. : *"Venous insufficiency great saphenous vein both legs"*[62].
>
> **SNOMED-CT :**
>
> 234050002|Venous insufficiency of leg
>
> | :363698007|Finding site|=181402008|Great saphenous vein|,
>
> 272741003|Laterality|=51440002|Bilateral|
>
> **ICD-10-CM**[63] **:**
>
> I87.2 Venous insufficiency (chronic) (peripheral)

60. Service public fédéral. Santé publique, sécurité de la chaîne alimentaire et environnement. https ://www.health.belgium.be/fr/terminologie-et-systemes-de-codes-snomed-ct, consultée le 23 mai 2017.
61. Les concepts médicaux sont considérés comme des événements dans le domaine médical (Tannier, 2014 : 35).
62. Exemple de De Meyere démontrant l'efficacité de SNOMED-CT pour l'annotation de concepts médicaux compliqués, versus le système ICD-10-CM. Séminaires du Cental (28 avril 2017).
63. L'ICD (International Classification of Diseases) est traduit dans 43 langues et utilisé dans quelque 117 pays. Ce plurilinguisme lui permet d'être très exploité dans les recherches actuelles, bien que sa couverture ne soit pas très étendue.

A l'heure actuelle, SNOMED-CT est la méthode la plus complète qui soit. Mais jusqu'à présent, peu de recherches ont été conduites sur le codage automatique des événements médicaux. De plus, si SNOMED-CT est disponible en français, ce n'est que partiellement : il est conçu pour une classification des termes et non pour une extraction systématiquement des expressions et tournures telles qu'utilisées dans les documents médicaux.

Troisième partie

Application : Time extractor

Chapitre 5

Introduction

Dans cette partie, nous présentons les étapes du développement de notre outil - Time extractor - pour l'extraction et la normalisation des expressions temporelles dans un corpus médical.

Cet outil a été mis au point lors d'un stage effectué chez EarlyTracks, spin-off de l'UCL et spécialisée dans le text mining et dans l'extraction d'informations. Depuis 2015, elle est particulièrement active *dans le développement de solutions techniques et commerciales destinées à répondre aux besoins propres de gestion de la connaissance dans le secteur hospitalier*[1].

Le stage avait comme objectif de créer un outil pour extraire automatiquement les expressions temporelles dans un corpus médical et d'en normaliser quelques-unes ; de contextualiser des dates avec des maladies ou des traitements qui se côtoient dans le texte ; et de créer une interface qui permet d'entrer un texte (et sa date de création) et de le récupérer avec les dates extraites, normalisées et contextualisées.

Ainsi, dans la phrase

> « Le patient souffre d'une pneumonie depuis une semaine. »,
> rédigée le 18 mars 2017,

« depuis une semaine » serait la date extraite, normalisée en 11 mars 2017 et reliée à la maladie « pneumonie ».

1. WBC. Nos sociétés - EarlyTracks. `http://www.wbc-incubator.be/fr/incubes/earlytracks`. Page consultée le 22 mai 2017.

Pour atteindre cet objectif, il a d'abord fallu annoter un corpus, composé de 2405 rapports patients. Ces documents ont été confiés à EarlyTracks (qui les a anonymisés) par différents hôpitaux belges francophones. Nous soulignons que les documents patients étant confidentiels, les quelques exemples de ce rapport sont théoriques et n'ont aucun lien avec les documents patients étudiés.

Après l'étape d'annotation, des graphes Unitex ont été réalisés pour, dans un premier temps, extraire les dates et autres expressions temporelles et, dans un second temps, contextualiser les dates avec les maladies, traitements, etc. qui les côtoient dans le texte.

Une dernière phase, la normalisation, permet de calculer la date réelle d'une expression temporelle. Ainsi, dans l'exemple ci-dessus, *depuis une semaine* serait remplacé par *depuis le 11 mars 2017*.

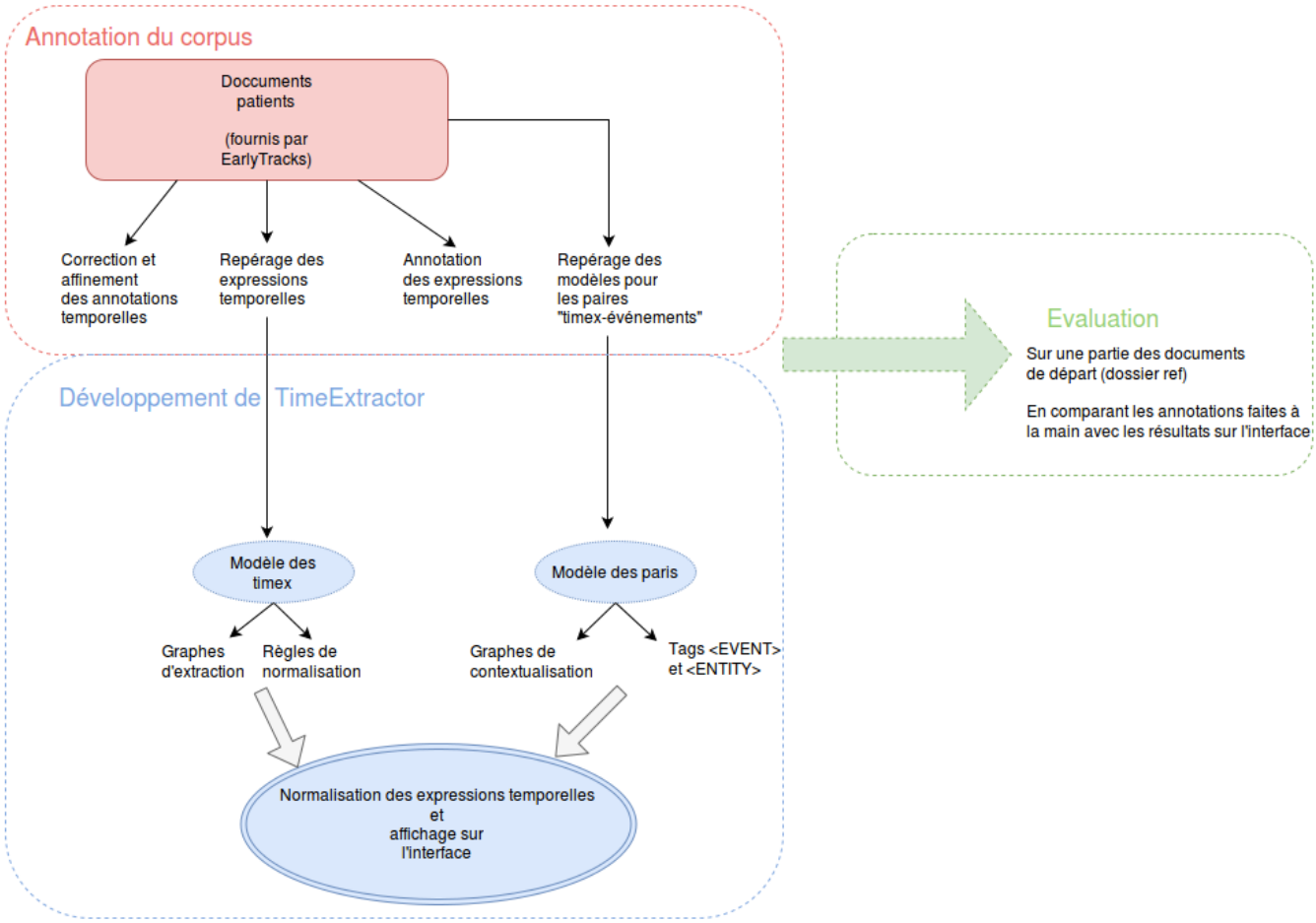

FIGURE 5.1 – Schéma récapitulatif des différentes phases de conception de Time extractor.

Cette figure donne un aperçu des différentes phases de conception de notre outil : nous avons commencé par une étape d'annotation du corpus, durant laquelle nous avons établi nos modèles d'expressions temporelles et de paires timex-événements. La seconde phase a été le développement de l'outil en soi, en commençant par les modèles de timex, puis par les modèles de paires, et enfin la normalisation des expressions temporelles. L'étape finale a été l'évaluation.

La figure ci-dessous décrit les étapes qu'une phrase simple franchit pour se retrouver dans l'interface :

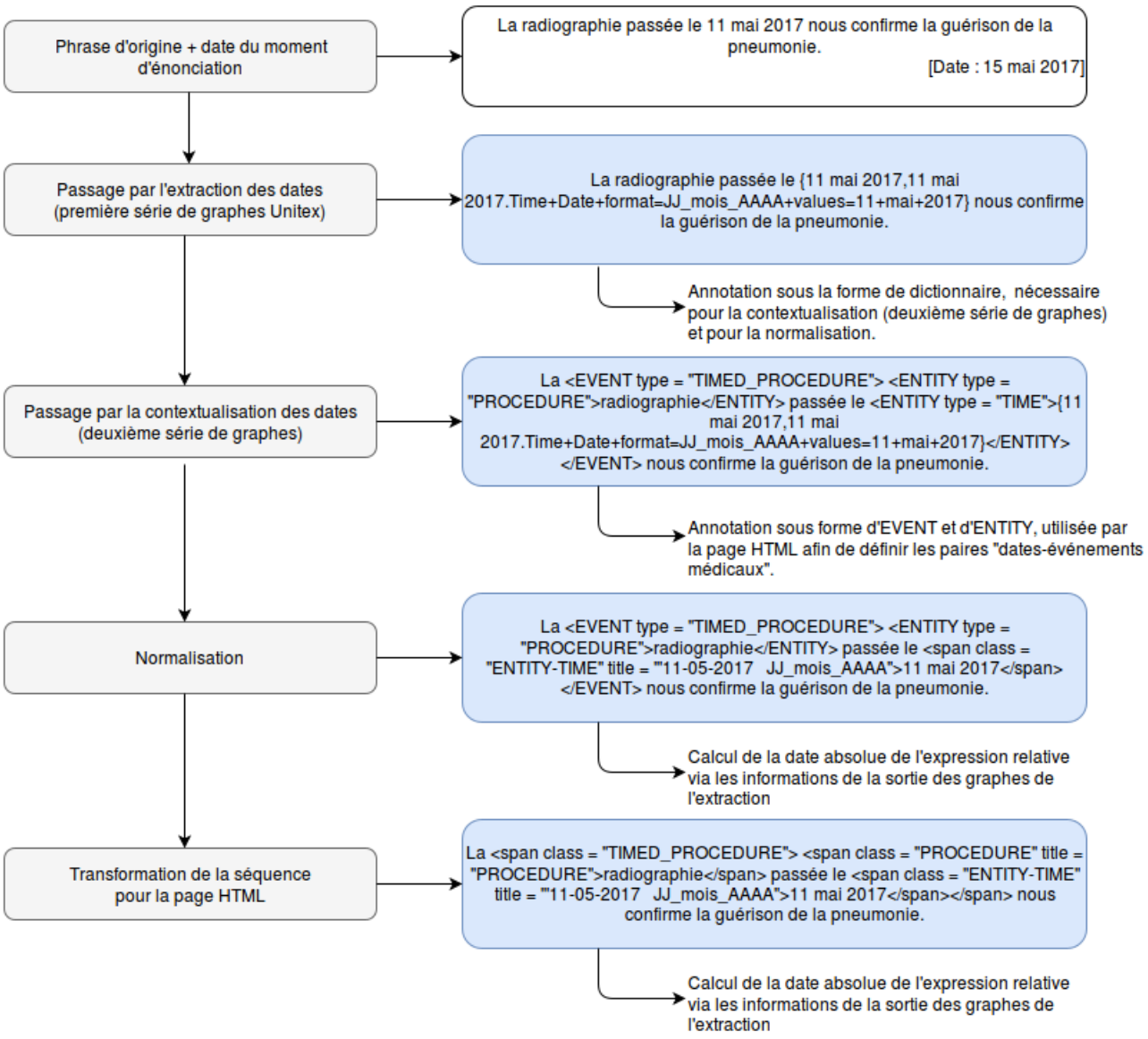

FIGURE 5.2 – Schéma récapitulatif des différentes phases d'analyse, extraction, contextualisation, normalisation et affichage de Time extractor.

Chapitre 6

Annotation du corpus

Le corpus annoté est essentiellement composé de rapports de sortie. Ces documents sont écrits en texte libre et brut, et contiennent souvent des abréviations. Ces fichiers répètent plusieurs fois la même information (certains médecins font un compte-rendu des antécédents du patient et reprennent ces informations dans le résumé final du rapport de sortie). D'Hondt *et al.* (2015) affirment que la redondance dans les rapports patients est relativement peu élevée. Nous n'avons analysé qu'un nombre relativement limité de documents et nous ne partageons pas leur avis.

Les autres documents reprennent des lettres informant les patients d'un rendez-vous (généralement peropératoire) ou d'un rendez-vous manqué. Enfin, nous avons également analysé des rapports d'imagerie (radiologies, scanners, etc.). Ces deux derniers types de documents ne comptent que peu d'expressions temporelles (avec pour certains rapports d'imagerie, une à aucune date !). De plus, ces expressions temporelles sont généralement sous le format JJ-MM-AAAA [1]. En revanche, les timex figurant dans les rapports patients se rencontrent sous de nombreuses formes et souvent abrégées. Leur reconnaissance et leur identification peuvent poser problème.

Enfin, nous tenons à souligner que ces documents provenaient de diverses institutions médicales belges, sises à Liège, Namur, Bruxelles,... Nous avons remarqué que chaque hôpital propose une mise en page qui lui est propre.

1. Ou JJMMAAAA, AAAA-MM-JJ,... La principale difficulté dans ce cas est de n'oublier aucun format de date.

Le corpus utilisé comprend 4788 fichiers, dont 2405 documents textes et 2383 fichiers d'annotations [2]. Le corpus est divisé en deux dossiers, *ref* et *dev*, lui-même divisé en deux sous-dossiers, *train* et *eval* :

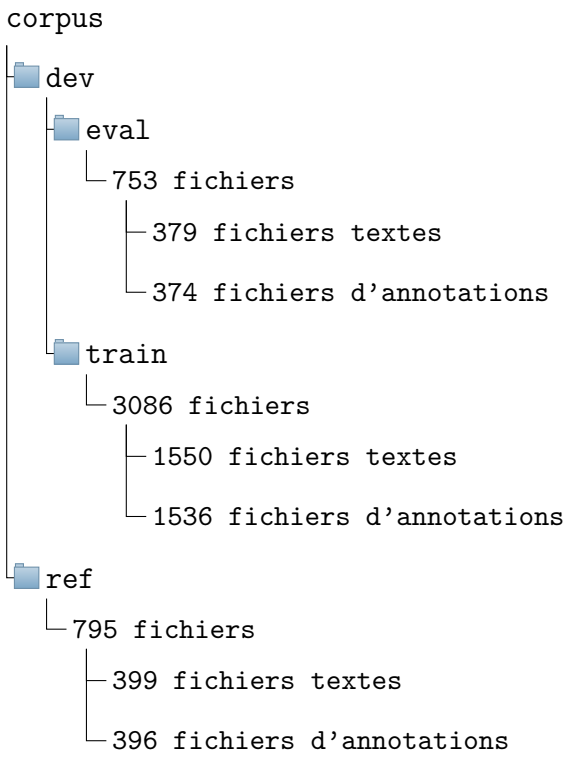

TABLE 6.1 – Division en dossiers et sous-dossiers du corpus, avec le nombre de fichiers, fichiers textes et fichiers d'annotations par sous-dossier.

Tous les fichiers présents dans ces dossiers sont donc annotés. En nous basant sur les fichiers de *train*, nous développons et entraînons nos grammaires Unitex. *Eval* est ensuite utilisé pour une première évaluation des grammaires. Enfin, *ref* est utilisé pour l'évaluation finale de ce projet. Afin de ne pas biaiser l'évaluation finale [3], tous les fichiers sont mélangés pour l'annotation, puis ils sont ré-organisés dans leurs dossiers et sous-dossiers respectifs.

Enfin, chaque texte a un fichier xml ".annotations" comprenant les informations d'annotation du fichier texte.

2. Certains fichiers textes sont vides de toute annotation. Ces fichiers textes n'ont donc pas de fichier d'annotations.

3. Dans la section suivante, nous explicitons notre méthodologie d'annotation, mais en résumé, nous avons mélangé tous les documents à annoter dans un seul dossier. Nous avons ensuite annoté tous ces fichiers, en prenant note des expressions temporelles typiquement cliniques. Sachant déjà que notre évaluation finale se ferait sur les fichiers de *ref*, le mélange de ces textes avec les textes d'entraînement nous a permis de ne pas porter une attention trop particulière aux expressions temporelles de *ref*.

1 L'annotation

Elle consiste en une amélioration d'une annotation déjà effectuée. En effet, comme affiché dans la figure ci-dessous, un outil d'annotation d'EarlyTracks permet de mettre en exergue les expressions temporelles, mais également les expressions médicales (du type maladies, parties du corps, médicaments,...) [4].

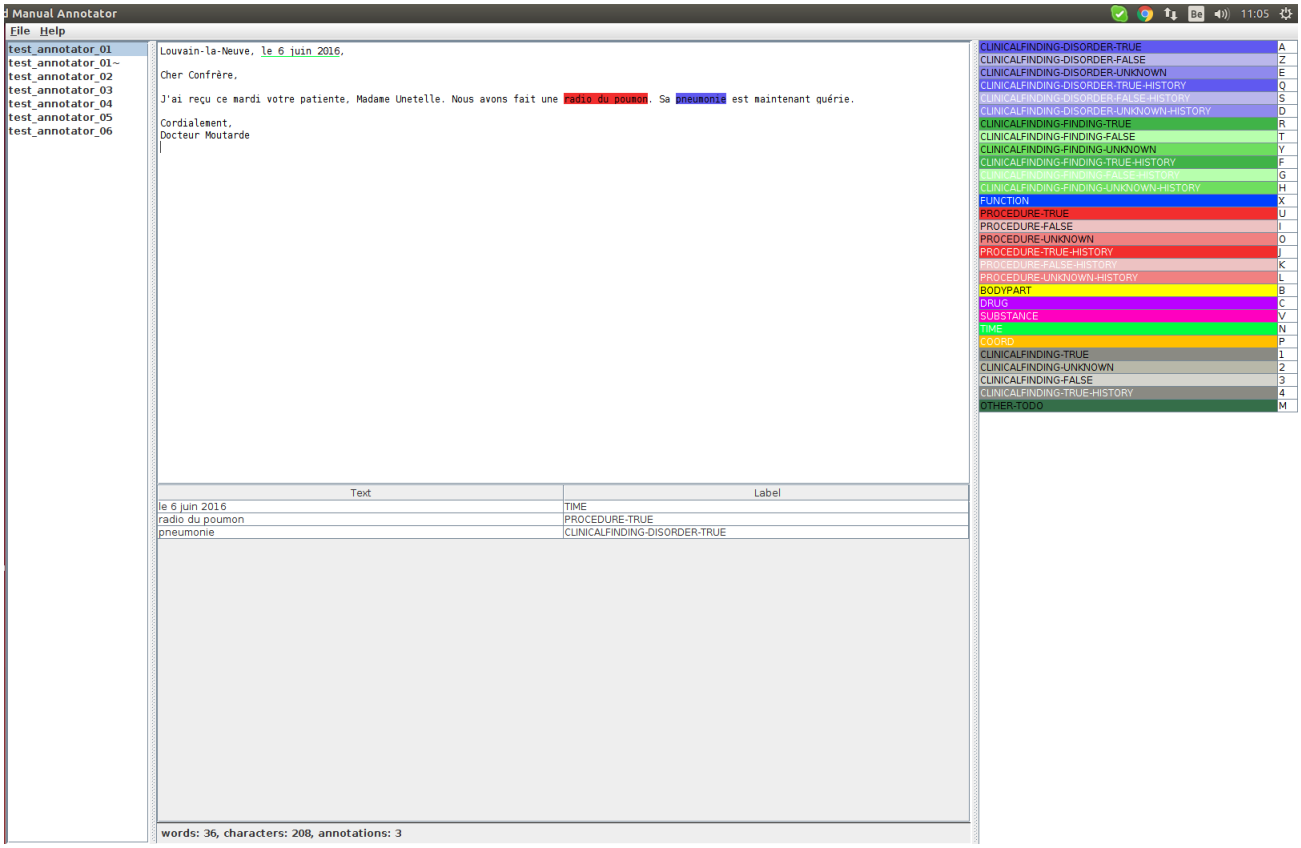

FIGURE 6.1 – Annotation du corpus médical.

4. Pour avoir accès à l'image grandeur nature et en couleur, consulter l'annexe électronique.

Le tableau ci-dessous reprend les différentes étiquettes disponibles sur cet outil :

CLINICALFINDING-DISORDER-TRUE	PROCEDURE-TRUE
CLINICALFINDING-DISORDER-FALSE	PROCEDURE-FALSE
CLINICALFINDING-DISORDER-UNKNOWN	PROCEDURE-UNKNOWN
CLINICALFINDING-DISORDER-TRUE-HISTORY	PROCEDURE-TRUE-HISTORY
CLINICALFINDING-DISORDER-FALSE-HISTORY	PROCEDURE-FALSE-HISTORY
CLINICALFINDING-DISORDER-UNKNOWN-HISTORY	PROCEDURE-UNKNOWN-HISTORY
CLINICALFINDING-FINDING-TRUE	FUNCTION
CLINICALFINDING-FINDING-FALSE	BODYPART
CLINICALFINDING-FINDING-UNKNOWN	DRUG
CLINICALFINDING-FINDING-TRUE-HISTORY	SUBSTANCE
CLINICALFINDING-FINDING-FALSE-HISTORY	TIME
CLINICALFINDING-FINDING-UNKNOWN-HISTORY	COORD
	OTHER-TODO
CLINICALFINDING-TRUE	
CLINICALFINDING-FALSE	
CLINICALFINDING-UNKNOWN	
CLINICALFINDING-TRUE-HISTORY	

TABLE 6.2 – Liste des labels dans l'outil d'annotation des documents médicaux.

Nous effectuons deux étapes d'annotation : la première se compose de la correction de l'annotation temporelle déjà effectuée [5], et la deuxième l'affine.

Dans la première étape, nous passons donc tous les documents textes en revue dans l'outil présenté ci-dessus et nous corrigeons les annotations TIME. Afin d'éviter d'en oublier pendant le développement des grammaires Unitex, nous prenons également note des expressions temporelles spécifiques au domaine médical. Le tableau ci-dessous en reprend quelques exemples :

5. Des étudiants de médecine avaient préalablement complètement étiqueté (avec les étiquettes de la table 6.2) les documents patients.

EXPRESSIONS MÉDICALES	SIGNIFICATIONS
J5 ou *J+5*	cinquième jour après une opération/intervention
J18+	dix-huitième jour et les suivants après une opération
H3 ou *H+3*	troisième heure après une opération/intervention
2x/j ou *2/j*	un médicament à prendre deux fois par jour
18 5/7 sa	enceinte de 18 semaines et 5 jours
1J/2 ou *1jour/2*	une fois tous les deux jours
post-op, *post-partum*, *post-chimio*,...	

TABLE 6.3 – Quelques exemples d'expressions temporelles spécifiques au domaine médical.

La deuxième étape consiste en l'amélioration des étiquettes d'annotation temporelle. Nous avons initialement décidé d'en appliquer 18 :

TIME-PRES	TIME-PASSE	TIME-FUT
TIME-PRES-REL	TIME-PASSE-REL	TIME-FUT-REL
TIME-PRES-DEP1	TIME-PASSE-DEP1	TIME-FUT-DEP1
TIME-PRES-DEP1-REL	TIME-PASSE-DEP1-REL	TIME-FUT-DEP1-REL
TIME-PRES-DEP2	TIME-PASSE-DEP2	TIME-FUT-DEP2
TIME-PRES-DEP2-REL	TIME-PASSE-DEP2-REL	TIME-FUT-DEP2-REL

TABLE 6.4 – Liste des labels pour l'annotation temporelle.

Le tag ***dep1*** signifie que l'expression temporelle porte sur une autre, il s'agit d'une expression relative. Par exemple :

« ce jour » ;

« lundi dernier » ;

« le 11 mai ».

sont toutes liées à la date du document. En effet, sans elle, comment savoir de quel jour, quel lundi ou du 11 mai de quelle année nous parlons.

Dep2 induit que l'expression temporelle soulignée porte sur un événement médical. Pour faire le parallèle avec TimeML, il s'agit d'une relation TLINK entre une timex et un événement. Par exemple, un examen particulier, une maladie, un médicament... :

« Elle a passé ce matin une radiographie des poumons. » ;

« Elle va prendre tel médicament pendant telle période à raison de x fois par jour. ».

Enfin, ***rel*** relève d'une expression temporelle liée à un événement médical. Il s'agirait ainsi d'une relation TLINK sans événement dans le texte. Par exemple :

> « Âgée de 24 ans. » (alors qu'il n'y a aucune mention d'une date de naissance dans le document)

Cependant, ***rel*** et **dep2** nous semblent relativement proches ; la différence entre les deux peut susciter la confusion. Nous n'avons finalement pas utilisé ***rel*** dans cette seconde étape d'annotation.

Pour plus de facilité dans la vision des concepts médicaux et des expressions temporelles, nous décidons de limiter les CLINICALFINDING-FINDING, les CLINICALFINDING-DISORDER et les PROCEDURE de la table 6.2 à un seul et unique tag. Dès lors, nous modifions les fichiers d'annotations de la première étape. Ainsi, on retire à ces trois types de tags les sous-étiquettes -TRUE, -FALSE, -UNKNOWN, -TRUE-HISTORY, -FALSE-HISTORY et -UNKNOWN-HISTORY. Par conséquent, les nouveaux labels pour l'affinement des tags temporelles sont les suivants [6] :

CLINICALFINDING-DISORDER	TIME-PRES	TIME-PRES-DEP1-REL
CLINICALFINDING-FINDING	TIME-PASSE	TIME-PASSE-DEP1-REL
PROCEDURE	TIME-FUT	TIME-FUT-DEP1-REL
BODYPART	TIME-PRES-REL	TIME-PRES-DEP2
DRUG	TIME-PASSE-REL	TIME-PASSE-DEP2
SUBSTANCE	TIME-FUT-REL	TIME-FUT-DEP2
FUNCTION	TIME-PRES-DEP1	TIME-PRES-DEP2-REL
COORD	TIME-PASSE-DEP1	TIME-PASSE-DEP2-REL
TIME	TIME-FUT-DEP1	TIME-FUT-DEP2-REL
OTHER-TODO		

TABLE 6.5 – Liste des labels de l'outil pour la deuxième couche d'annotations.

6. Cette table représente toutes les étiquettes, y compris les *rel* que nous avons par la suite décidé de ne pas utiliser.

1.1 Les erreurs et problèmes rencontrés

Les documents patients comportent parfois dans les dates des erreurs et des fautes de frappe :

- *2020* à la place de *2002* comme date du document (ce qui ne nous posera pas de problème puisque, dans notre interface, l'utilisateur devra lui-même rentrer la date du document) ;
- des dates de documents antérieures aux examens (pour la même raison que ci-dessus, cela ne sera pas un souci dans notre interface finale) ;
- des dates sous la forme *0206/2015* à la place de *02/06/2015*, ou *20020* pour *2002* : ces dates ont été soulignées puisqu'il s'agissait de constructions relevant de la temporalité, mais nous pouvons déjà affirmer qu'elles ne seront pas reprises par nos graphes Unitex ;
- l'âge du patient a été trouvé sous la forme *0 ans 0 mois 0 jours*, sans date de naissance présente dans le texte.

Lors de la deuxième annotation, nous rencontrons quelques problèmes par rapport à la position à adopter. En effet, nous avions décidé que la date du document serait en TIME-PRES et que l'annotation passé-présent-futur serait faite à partir de cette date de document. Ainsi, si la date de l'examen est antérieure à celle du rapport, elle est indiquée comme étant un TIME-PASSE. Cependant, dans ces cas-là, qu'en est-il des expressions du type *ce jour, pour le moment, actuellement, etc.* ? Nous remarquons que sont parfois inclus dans les rapports des résultats d'examens passés. Pour cette raison, toutes les expressions relevant d'une époque présente sont annotées comme référant à la date du document.

Chapitre 7

Unitex

Un outil permettant d'efficacement extraire des données linguistiques est le logiciel open-source Unitex [1]. Il s'appuie sur un principe d'automates à états finis [2] et fonctionne avec un système de graphes :

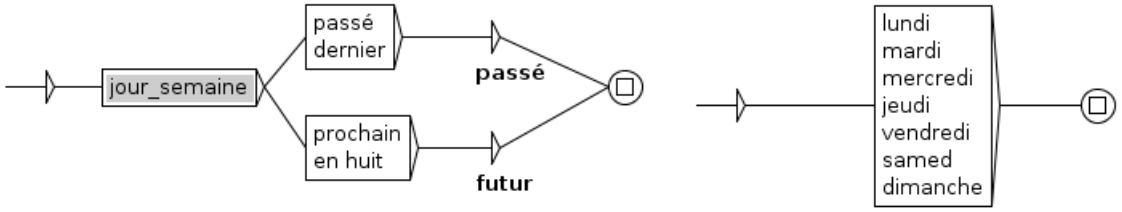

FIGURE 7.1 – Exemple d'un graphe Unitex (à gauche) et du sous-graphe *jour_semaine* à droite.

Ce graphe permet d'extraire les expressions telles que « lundi dernier » ou « vendredi prochain ». Ce graphe propose l'ajout d'une information supplémentaire (à la sortie du graphe) indiquant s'il s'agit d'une expression passée (« lundi dernier ») ou future (« vendredi prochain »).

Cet outil est régulièrement utilisé pour l'extraction des expressions temporelles : Kevers (2011), Gross (2013), Drat (2014), De Meyere et al. (2015) et l'encodage (manuel) de règles linguistiques.

1. Disponible à l'adresse `http://unitexgramlab.org/`.

2. En anglais, finite state automata. Il s'agit d'une construction abstraite, composée d'un ensemble fini d'états possibles, d'un ensemble fini de symboles en entrée et d'une fonction de transition entre les états. Le passage de l'état courant à l'état suivant se fait par une transition et est dirigé par une condition (Jurafsky & Martin, 2009 : 60).

Les grammaires d'extraction et de contextualisation de dates reposent sur les expressions temporelles observées dans le corpus médical. Une première série de graphes extrait les dates et autres expressions temporelles. Nous tentons ensuite d'extraire le contexte médical. Le passage par la première série de graphes permet d'insérer en sortie un format d'annotation en dictionnaire, utile pour le passage par la seconde série de graphes et pour la normalisation. Ce passage par Unitex est géré par le script *Unitex.py*[3]. Concrètement :

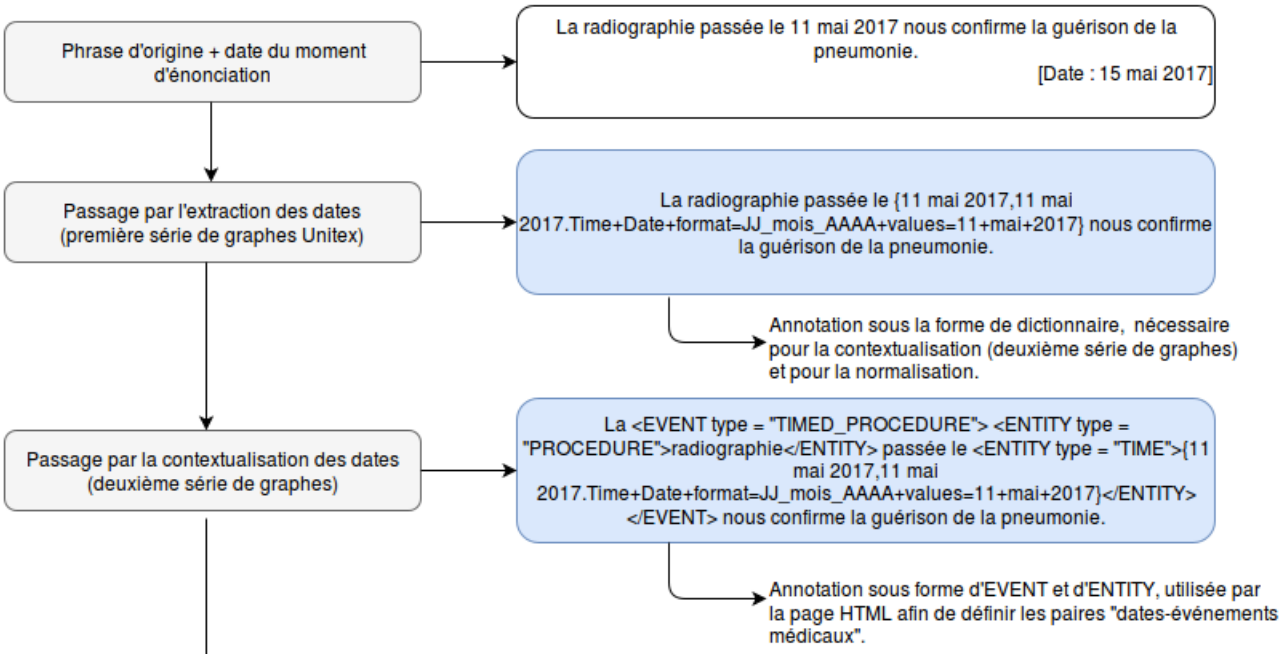

FIGURE 7.2 – Fonctionnement de l'extraction et de la contextualisation des expressions temporelles (zoom de la figure 5.2).

3. Disponible dans les annexes électroniques dans le dossier *server/services/*.

1 Extraction des dates

1.1 Les graphes

Le développement de la grammaire d'extraction des expressions temporelles est progressif et exploratoire. L'extraction se base sur les expressions annotées antérieurement dans le corpus médical (comme expliqué dans le chapitre 6, aux pages 79-85). Cela signifie qu'il est possible que d'autres tournures liées à la temporalité ne soient pas reconnues par notre grammaire.

Nous identifions la plupart des dates chiffrées [4], les expressions temporelles typiques du domaine médical relevées lors de l'annotation du corpus, et d'autres expressions plus textuelles (telles que « ce jour », « dans 3 mois », « début mars »). Nous repérons également des abréviations, comme *vdd* pour *vendredi*, des erreurs (*dcembre, decembre* pour *décembre*). Nous pouvons aussi détecter les mois en anglais (dans le corpus, certaines dates de document et de naissance sont écrites sous la forme *May 11, 2017*).

Pour être efficaces, les graphes Unitex doivent allier précision et souplesse : seules les expressions temporelles doivent être repérées, mais il doit également être possible d'aisément ajouter d'autres constructions temporelles et d'autres graphes pour étendre le système. Les graphes sont agencés judicieusement, de façon à y arriver facilement.

Pour répondre à ces exigences, nous utilisons le format des dictionnaires en sortie de graphe. Ce format consiste en la répétition, entre {}, de l'expression extraite, suivie de l'étiquette lexicale TIME. Pour être exactes, les formes dans le dictionnaire sont sous le format *{forme fléchie, lemme.*ÉTIQUETTE_LEXICALE*}*. Dans notre cas, les informations qui précèdent cette étiquette nous sont utiles pour la contextualisation des dates et ce qui la suit pour la normalisation.

4. Le point suivant, *Erreurs et décisions*, présente quelques types de dates chiffrées que nous avons choisis de ne pas reconnaître. Se rendre à la 90-91 pour de plus amples détails.

La figure ci-dessous reprend un exemple d'un morceau de graphe utilisé dans notre outil [5] :

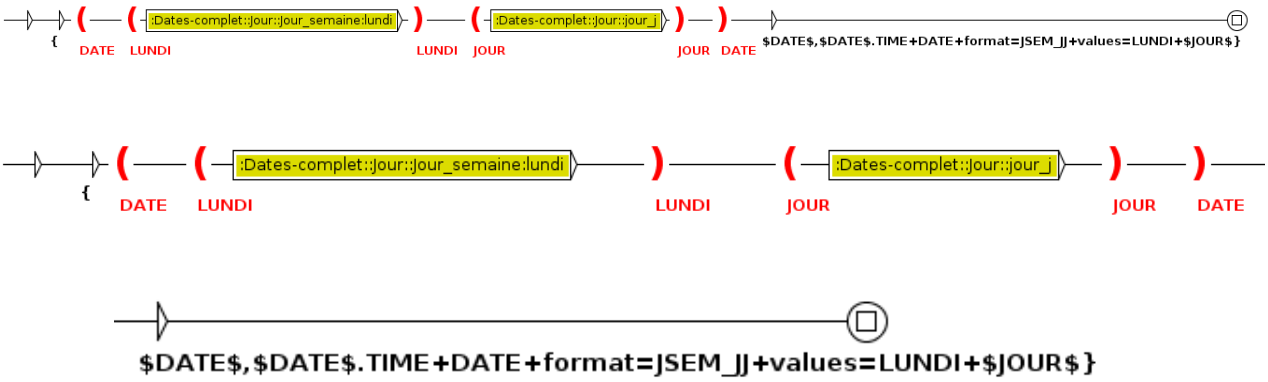

FIGURE 7.3 – Exemple d'une partie d'un graphe d'extraction et sa sortie, suivi de ses zooms.

Les parenthèses rouges (appelées *variables d'entrée* dans Unitex) permettent de garder en mémoire les termes pris entre les parenthèses et de les restituer dans la sortie du graphe. Ainsi, dans la phrase

« [...] lundi 15, [...] »

« lundi » est retenu entre les premières parenthèses et « 15 » entre les deuxièmes. Toute l'expression est retenue dans la variable d'entrée « DATE ». Le passage par ce graphe donne en sortie :

« [...] {lundi 15,lundi 15.TIME+DATE+format=JSEM_JJ+values=LUNDI+15} [...] »

Cette sortie reprend en première position la forme fléchie (il s'agit du premier « lundi 15 »), en deuxième position, le lemme [6], et enfin, en dernière position l'étiquette lexicale [7], TIME, suivie des informations nécessaires pour former les règles de normalisation. Le fichier envoyé à la grammaire de contextualisation de dates est préalablement passé par la grammaire d'extraction. Les graphes de cette seconde série reconnaissent alors les dates via l'appel <TIME> de l'étiquette lexicale.

5. Il s'agit du graphe *Format Jsem-jour* (disponible dans les annexes électroniques *graphes/extraction/Graphs/Dates-complet/Dates-Format/*. L'image complète du graphe est également disponible dans le dossier *graphes/images/*).

6. Un lemme correspond à la forme canonique d'un mot (la forme du mot utilisée pour l'entrée du dictionnaire), donc pour un verbe, son infinitif ; pour un nom, sa forme au singulier, etc.

7. Nous attirons l'attention sur le fait que cette sortie est mise en place pour les dictionnaires. Ainsi, si nous avions reconnu le terme « lundis », la forme fléchie serait « lundis » et le lemme « lundi ». Cependant, nous utilisons ici les dictionnaires pour indiquer qu'une expression est d'une catégorie « Time », utilisée lors du passage par la seconde grammaire de graphes.

Les graphes sont organisés en quatre points :

- l'âge
- les dates
- les dates médicales
- les expressions indéfinies

Dans la première section, nous récupérons les tournures concernant l'âge d'une personne, dans la seconde, toutes les constructions de dates qui ne sont pas médicales (le graphe de la figure 7.3 fait partie de cette catégorie). Les dates médicales sont gérées par les graphes du troisième point. Finalement, la dernière partie repère les constructions difficilement normalisables telles que les formulations concernant les délais, quelques durées (quand elles ne sont pas prises dans les graphes de dates médicales), les moments de la vie d'une personne (*l'enfance, l'adolescence, la jeunesse*, etc.) et les saisons [8].

Erreurs et décisions

De nombreuses dates apparaissent sous la forme :

« 05/10 » [5 octobre] ;
« 18.09 » [18 septembre] ;
« 12/2010 » [mois de décembre 2010] ;
« 09 :39 » [à 9h39].

En reconnaissant ces constructions temporelles, nous extrayons également des informations non temporelles, telles que des morceaux de numéros de téléphone, des informations médicales telles que les taux de pression artérielle, les numéros des dossiers,... Nous décidons donc de ne pas reconnaître ces dates car elles provoquent plus de bruits que de résultats concluants. Ces dates n'ayant pas de contextes réguliers, nous ne pouvions en forcer un.

Nous repérons tout de même les dates « 12/2010 ; 2010 » dans la deuxième grammaire, lorsque celles-ci sont suivies ou précédées d'un CLINICALFINDING ou d'une PROCEDURE.

8. Pour plus de détails et une présentation plus approfondie des graphes, consulter l'annexe 2 pages 141-144.

Un autre problème récurrent est d'éviter de reconnaître « 4430 ANS », le code postal et la ville d'Ans. En effet, « 30 ANS » est souvent repéré comme étant un âge. Nous réglons ce problème avec les contextes négatifs de droite et gauche (marqués par « ![ ] », comme présenté dans la figure ci-dessous). Si cette contextualisation n'affiche pas d'erreurs sur nos documents, nous ne garantissons cependant pas de manière absolue que cela ne puisse se produire sur d'autres.

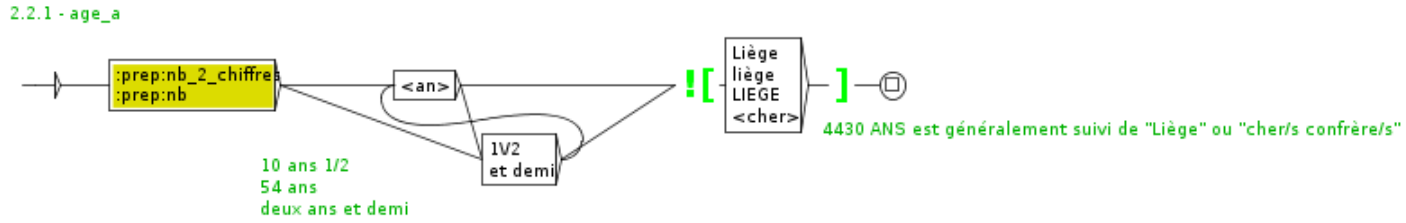

FIGURE 7.4 – Graphe d'extraction d'âge (en années).

Enfin, dans le corpus annoté, plusieurs constructions temporelles comprennent une date suivie d'une heure. Nous reconnaissons toutes ces dates et la plupart de ces heures (excepté celles sous la forme *hh :mm*) indépendamment l'une de l'autre. Nous les repérons en tant qu'une seule expression dans un seul cas :

« Jeudi 11 mai 2017 à 09h20 »

est repéré comme suit :

« {Jeudi 11 mai 2017 09h20,Jeudi 11 mai 2017 09h20.TIME+DATE-HEURE+format =JSEM_JJ_MM_AAAA_à_HHhMMm+values=JEUDI+11+MM05+2017+09h20} »

Hormis dans ce cas-ci, les heures et les dates sont reconnues de manière distincte.

2 Contextualisation des dates

2.1 Les dictionnaires

Après passage par les grammaires d'extraction d'expressions temporelles, le corpus est traité par la grammaire de contextualisation des événements médicaux.

Ces graphes sont organisés par les tags d'annotation non temporels vus à la table 6. 5, reprenant

les labels pour l'affinement de l'annotation [9]. Les constructions temporelles sont retrouvées par la sortie des graphes sous le format dictionnaire. Pour rappel, une boîte avec <TIME> (TIME étant l'étiquette lexicale) reconnaît une séquence {date,date.TIME} comme étant un TIME.

Des dictionnaires sont utilisés dans cette grammaire afin de reconnaître les formulations médicales. Ces dictionnaires permettent de récupérer, de la même manière que les dates, les tournures médicales et ont comme forme :

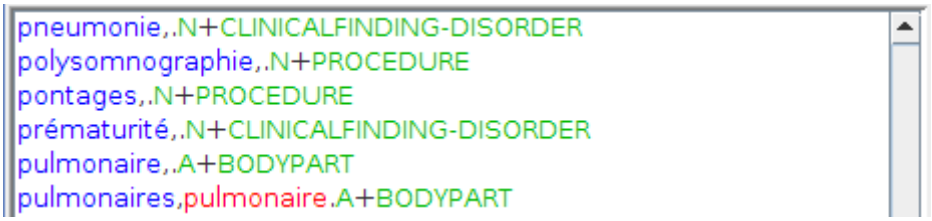

FIGURE 7.5 – Exemples d'un dictionnaire utilisé pour le repérage des expressions médicales.

Les graphes

Les graphes sont organisés selon les labels d'annotation non temporels vus à la table 5. Les dictionnaires sont utilisés pour retrouver les termes médicaux et les constructions temporelles. La phrase :

> « La radiographie passée le {11 mai 2017,11 mai 2017.Time+Date+format=JJ_mois_AAAA+values=11+mai+2017} nous confirme la guérison de la pneumonie. »

à la sortie de cette deuxième grammaire de graphes prend la forme :

> « La <EVENT type = "TIMED_PROCEDURE"><ENTITY type = "PROCEDURE">
> radiographie</ENTITY> passée le <ENTITY type = "TIME">{11 mai 2017,11 mai 2017.Time+Date+format=JJ_mois_AAAA+values=11+mai+2017}</ENTITY> </EVENT> nous confirme la guérison de la pneumonie. »

Ainsi, la procédure *radiographie* et la date *11 mai 2017* sont incluses dans un événement médical TIMED_PROCEDURE. La procédure fait partie d'une entité PROCEDURE et la date d'une entité TIME.

9. La table 7.1 à la page suivante donne tous les types d'événements reconnus.

Nous avons plusieurs événements et entités :

<EVENT type = TIMED__PROCEDURE></EVENT>
<EVENT type = TIMED__SUBSTANCE></EVENT>
<EVENT type = TIMED__FUNCTION></EVENT>
<EVENT type = TIMED__CLINICALFINDING-DISORDER></EVENT>
<EVENT type = TIMED__CLINICALFINDING-FINDING></EVENT>
<EVENT type = TIMED__DRUG></EVENT>
<EVENT type = TIMED__FINDING__PROCEDURE></EVENT>
<EVENT type = TIMED__DISORDER__PROCEDURE></EVENT>
<EVENT type = TIMED__BODYPART></EVENT>
<EVENT type = TIME></EVENT>
<ENTITY type = PROCEDURE></EVENT>
<ENTITY type = SUBSTANCE></EVENT>
<ENTITY type = FUNCTION></EVENT>
<ENTITY type = CLINICAL DISORDER></EVENT>
<ENTITY type = CLINICAL FINDING></EVENT>
<ENTITY type = DRUG__NAME></EVENT>
<ENTITY type = BODYPART></EVENT>
<ENTITY type = TIME></EVENT>

TABLE 7.1 – Liste des évenements et entités lors de la sortie de la grammaire de contextualisation.

Un graphe typique de contextualisation se présente comme suit :

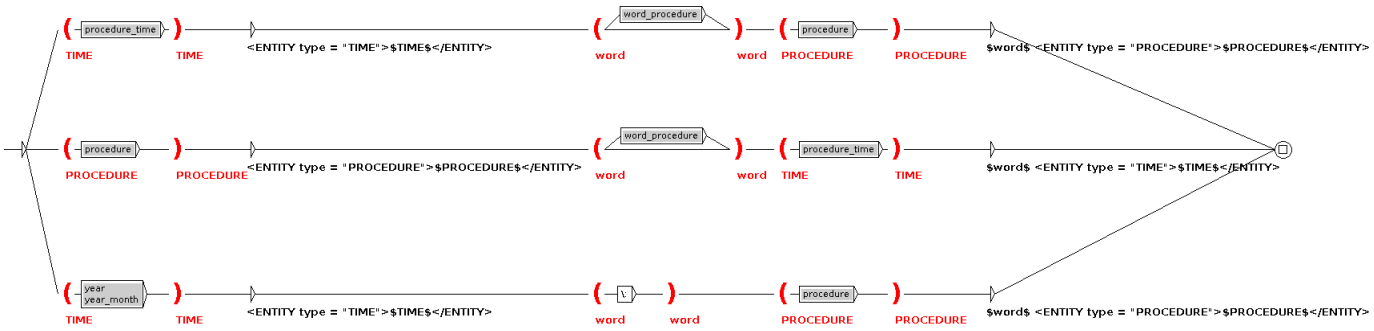

FIGURE 7.6 – Exemple d'une partie d'un graphe d'extraction et sa sortie.

où les sous-graphes de récupération des éléments des dictionnaires ressemblent à ceci :

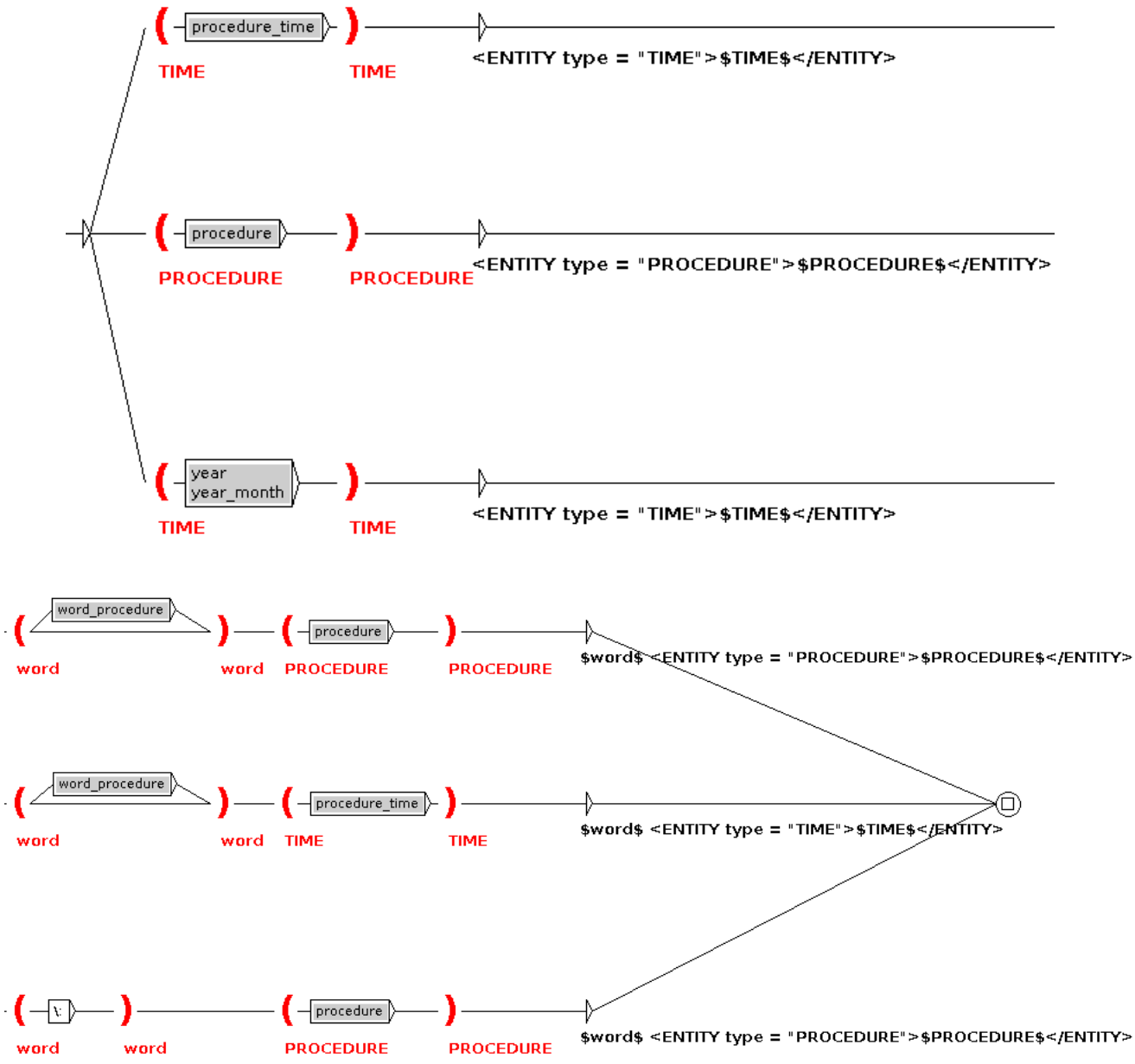

FIGURE 7.7 – Zooms du graphe de contextualisation de la figure 6.

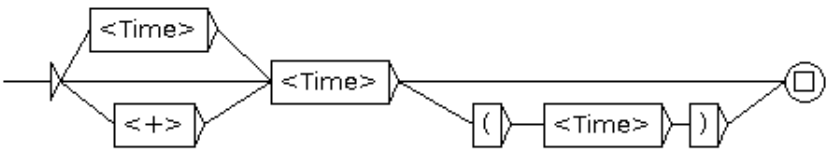

FIGURE 7.8 – Graphe *procedure_time* - Repérage des expressions temporelles.

Les sous-graphes nommés *word-* permettent de récupérer les événements médicaux-temporels, même s'ils sont séparés par plusieurs mots.

Erreurs et décisions

Notre contextualisation trouve parfois un événement médical-temporel qui n'est en réalité pas lié. De plus, de nombreux événements ne sont pas reconnus. Nous pouvons supposer que certaines formulations médicales ne se trouvent pas dans les dictionnaires Unitex utilisés, ou que les sous-graphes *word* ne repèrent pas tous les mots séparant les deux entités de l'événement.

Chapitre 8

Normalisation des dates et envoi à l'interface

Notre perception du temps repose sur des repères, tels que les calendriers. Ces derniers ont notamment pris des événements religieux comme point de départ : ainsi le calendrier chrétien débute à la naissance de Jésus-Christ, tandis que l'Hégire en 622 marque le début de l'arabe, etc. Aujourd'hui notre temps est divisé en unités régulières et hiérarchisées : les secondes, les minutes, heures, journées, semaines, mois, années, décennies, siècles,... Elles se réfèrent à des phénomènes cosmiques (comme le cycle du jour et de la nuit), ou elles ont été constituées arbitrairement (pourquoi un mois comporte-t-il 30, 31, 28 (et parfois 29) jours ?). Notre système de datation repose entièrement sur ces repères "objectifs" [1].

C'est avec ce système que nous normalisons nos dates extraites par Unitex. La normalisation indique que nous récupérons la date absolue (donc sous la forme JJ MM AAAA) d'une expression temporelle relative. Ainsi, dans la phrase

> « Je l'ai vu lundi dernier. », rédigée le mardi 16 mai 2017
> et dont la sortie d'Unitex est :
> « Je l'ai vue {lundi dernier,lundi dernier.Time+**format**=JSEM_passé+**values**=LUNDI+passé}. »,

la normalisation permettra de dire que « lundi dernier » renvoie au lundi 8 mai 2017.

1. Tomassone (200 : 180).

Nous attirons l'attention sur le fait qu'après l'étiquette *Time*, nous donnons le **format** de l'expression temporelle et les **valeurs**. Au début de la normalisation, nous récupérons toutes les constructions temporelles extraites par la grammaire d'extraction d'Unitex, ainsi que les informations de format et les valeurs. Enfin, nous récupérons l'expression du fichier source. Dans l'exemple ci-dessus, cette expression textuelle du fichier est *lundi dernier*. Grâce au format [2] nous connaissons la construction de la date et pouvons ainsi correctement récupérer les informations (dans values) dont nous avons besoin pour calculer la date réelle de l'expression.

Ainsi, dans notre exemple, avec le format (« JSEM_passé »), nous savons que nous avons affaire à une construction *jour de la semaine dans un contexte passé*. Nous allons donc dans la fonction gérant ces cas puis nous extrayons des "values" (« LUNDI+passé ») le jour de la semaine *lundi*. Nous allons ensuite dans une fonction qui calcule la date de *lundi passé* par rapport à la date du document. Le jour de la semaine de la date du document est d'ailleurs calculée au début du programme.

Le script *Normalisation.py* [3] récupère les formats et les valeurs en utilisant des expressions régulières. Il serait probablement possible et judicieux de le simplifier et l'automatiser, évitant ainsi d'avoir pour chaque type de format sa fonction de normalisation propre.

Une fois que la date réelle de l'expression est retrouvée, le script renvoie au texte l'expression textuelle et les informations de format et de date réelle. Elles sont modifiées pour apparaître en couleur et soulignées sous le format html.

2. Liste disponible à l'annexe 5 pages 151-173 et sur l'annexe électronique.
3. Disponible sur l'annexe électronique dans le dossier *server/services/*.

Chapitre 9

Time extractor

Des fichiers sont créés dans un dossier temporaire appelé *Time extractor*. Chaque passage par une grammaire crée un fichier : le fichier *temp_file_output* reprend le texte source modifié après passage par la grammaire d'extraction, *temp_file_final* après contextualisation et temp_file_norma après normalisation. Dans ce dossier, figure également le fichier *temp_file* qui reprend le fichier source rédigé sur l'interface. L'interface utilisateur ressemble à ceci [1] :

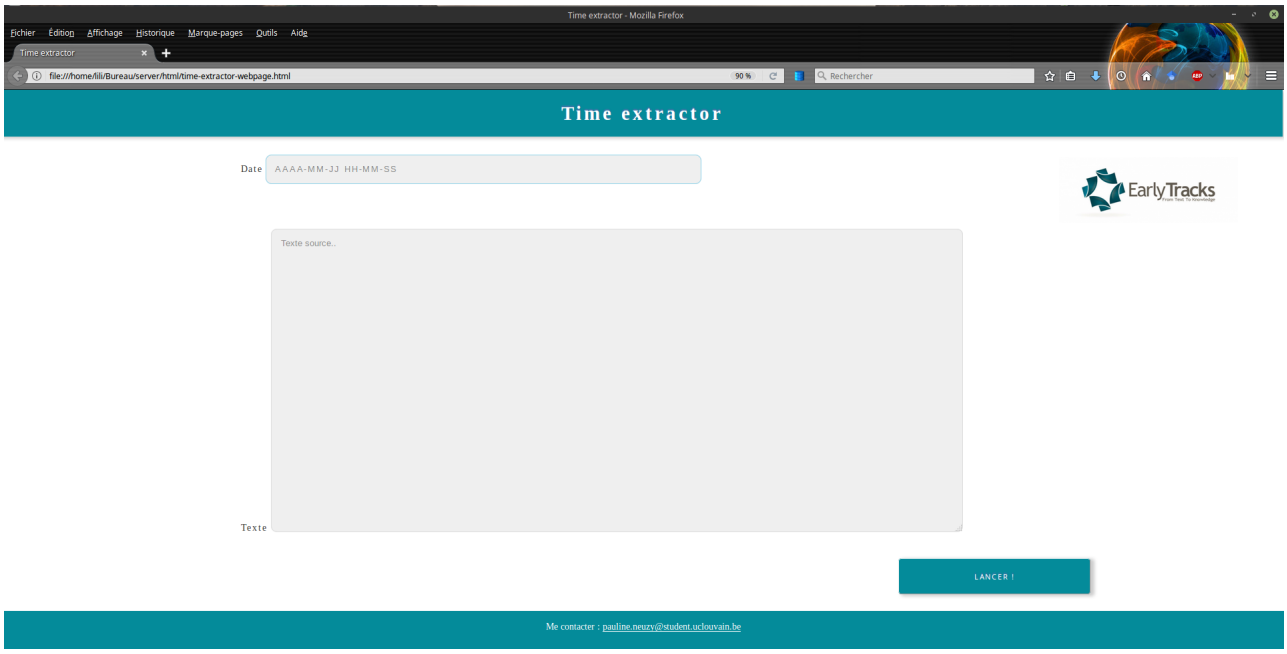

FIGURE 9.1 – Interface de Time extractor.

1. Retrouver ces captures en plus grand dans l'annexe 6, présentation de time-extractor.

Il suffit d'entrer une date (sous la forme AAAA-MM-JJ HH-MM-SS, année-mois-jour heures-minutes-secondes) dans le cadre prévu à cet effet et un texte à analyser dans l'autre.

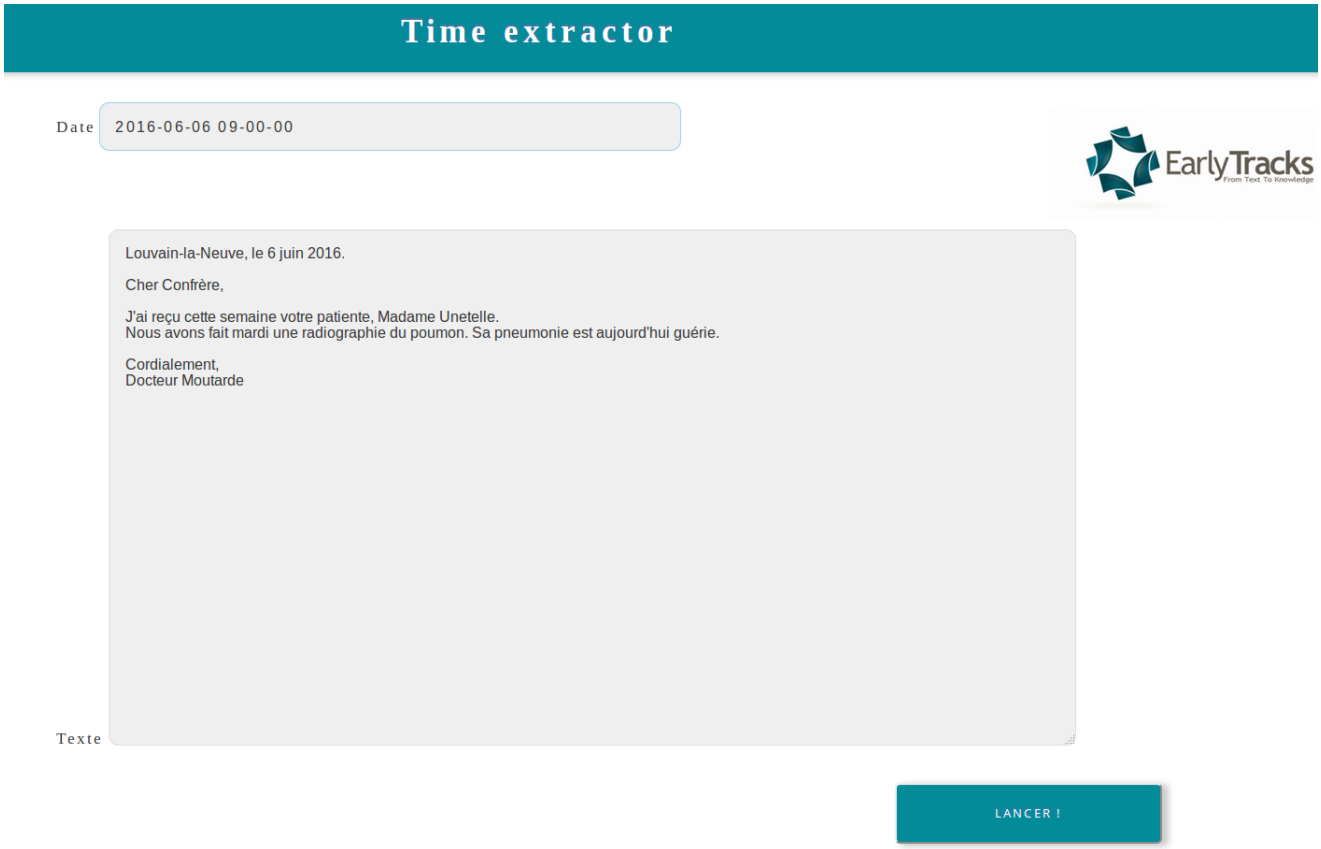

FIGURE 9.2 – Interface de Time extractor, avec un texte prêt à être analysé.

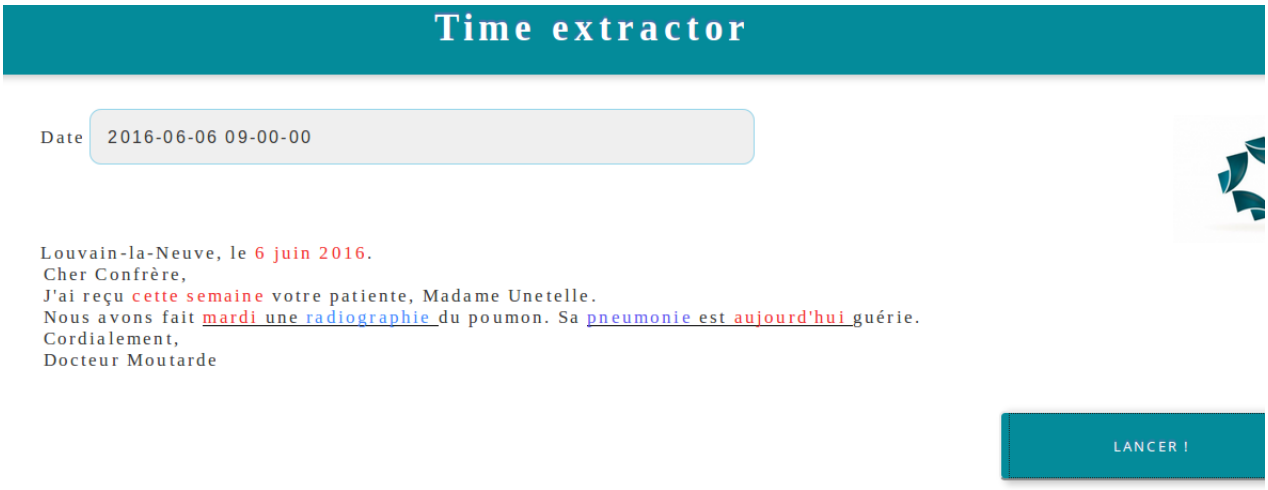

FIGURE 9.3 – Interface de Time extractor, avec un texte analysé (zoom sur le texte).

Nous recevons en résultat un texte, avec les dates inscrites en rouge, les événements médicaux-temporels soulignés, et les entités médicales en couleurs (dans ce cas-ci, nous avons une PROCEDURE en bleu).

Le texte inséré est identique après passage par notre programme. En effet, nous avons décidé de garder l'expression temporelle telle qu'inscrite dans le fichier de départ, permettant ainsi de ne pas afficher d'erreurs dans le texte si des fautes se sont glissées lors de la normalisation ou de la contextualisation. Cependant, si l'utilisateur passe sa souris sur la date ou l'événement médical coloré, il pourra voir la date réelle de l'événement, ainsi que le format qu'avait la construction temporelle de départ.

Quatrième partie

Évaluation et perspectives

Chapitre 10

Evaluation de Time extractor

Notre évaluation porte sur l'exactitude des extractions de constructions temporelles dans les documents du dossier *ref*. La contextualisation et la normalisation ne seront pas évaluées. En effet, les rapports médicaux n'ayant pas été annotés selon la forme *événement médical - date*, il nous serait difficile d'estimer avec précision la justesse de la grammaire de contextualisation. De plus, nous avons déjà souligné qu'elle manque de finesse puisqu'elle prend en considération des événements médicaux non liés à l'expression temporelle. Ainsi, 70 [1] fichiers textes ont été examinés.

Durant l'évaluation, nous avons compté

1. les expressions temporelles extraites par Unitex ;
2. les expressions temporelles annotées manuellement ;
3. les expressions non temporelles extraites par Unitex (les erreurs d'Unitex) ;
4. les expressions temporelles non reconnues par Unitex ;
5. les expressions temporelles non reconnues lors de l'annotation manuelle ;
6. les expressions temporelles reconnues différemment dans Unitex et par l'annotation manuelle.

1. Sur 399 fichiers textes. Nous avions cependant calculé le rappel et la précision après 35 documents, et les résultats n'ont que très peu changé entre 35 et 70 fichiers.

Le dernier point évoque les constructions différemment reconnues. Nous avons abordé ce sujet dans le chapitre sur l'annotation : en effet, le corpus annoté repère les tournures *dates à heure* comme une seule et même expression, tandis qu'Unitex les reconnaît distinctement, *la date* et *l'heure*[2].

Sur les 70 fichiers textes analysés, nous avons un total de :

822	expressions temporelles extraites par Unitex
802	expressions temporelles annotées manuellement
18	erreurs d'Unitex
53	expressions temporelles non reconnues par Unitex
47	expressions temporelles non reconnues lors de l'annotation manuelle
40	expressions temporelles reconnues différemment

TABLE 10.1 – Evaluation de Time extractor - le total d'expressions extraites par Unitex, les erreurs et différences par rapport au corpus annoté.

Nous constatons tout d'abord qu'Unitex a reconnu plus d'expressions temporelles que l'annotation manuelle. Cela est en partie dû aux erreurs et aux expressions différemment reconnues. En effet, 47 constructions temporelles n'ont pas été reconnues lors de l'annotation manuelle. Celles-ci sont le plus souvent « actuellement, actuel, ce jour ». Unitex étant plus automatique qu'une annotation manuelle, ces expressions sont systématiquement repérées par le programme. D'autres tournures temporelles ne sont pas reconnues.

Ensuite, Unitex a reconnu 18 expressions non temporelles. Elles incluent :

- « Hôpital de jour »
- « repas » (hors d'une liste de médicaments)
- « 1934-06-03 », prise comme une date sans le « 19 »[3]
- « Dr LU ANH », le « LU » étant repéré comme étant un « lundi »
- « *2 fois 1.8 gray* » ; « *pouls 16x4* »

Ces erreurs (et particulièrement celles du dernier point) nous indiquent que nous devons encore améliorer les contextes négatifs de gauche et de droite dans nos graphes Unitex. Parmi les 53

2. Hormis dans le cas du *Jeudi 11 mai 2017 à 09h20*, comme explicité page 75.
3. Cette erreur nous a d'ailleurs permis de voir une erreur récurrente dans l'annotation manuelle : les annotateurs n'ont jamais souligné ces dates.

expressions non reconnues par Unitex, 32 sont des expressions textuelles du type « au petit déjeuner », « dans un premier temps », « très prochainement », facilement incorporables à la grammaire d'extraction d'Unitex.

Enfin, les expressions reconnues différemment comprennent les constructions composées d'une date suivie d'une heure, d'une prise de médicament par jour à telle heure : « 2x/j à 9h »), mais également quelques autres expressions. En voici une liste non exhaustive :

Unitex	**Annotation manuelle**
novembre	dans le courant du mois de novembre
08.10.2015	du 26.09 au 08.10.2015
soirée	dans la même soirée
le matin	le matin de l'arrivée aux Urgences
01/10/2015 à 15h20[9]	01/10/2015 à 15h20
3x/jour	3x/jour jusqu'au 09.10 inclus
le matin et le soir	le matin et le soir
lors de son hospitalisation en 2011	lors de son hospitalisation en 2011

TABLE 10.2 – Liste non exhaustive d'expressions différemment reconnues par Unitex et par l'annotation manuelle.

Nous proposons ci-dessous deux évaluations [4] : l'une, *stricte*, qui considère que les erreurs d'Unitex comprennent les 18 erreurs marquées comme telles dans le tableau ci-dessus, mais également les 40 expressions différemment reconnues ; l'autre, *moins stricte*, qui ne prend en compte que les 18 erreurs.

	évaluation stricte (58 erreurs)	**évaluation moins stricte** (18 erreurs)
rappel	0.9526	1.002
précision	0.9294	0.9781

TABLE 10.3 – Rappel et précision de l'évaluation.

9. Les éléments soulignés dans ce tableau marque les expressions temporelles reconnues par Unitex vs. l'annotation manuelle dans les cas où une phrase contient deux expressions.

4. *[...] recall is the ratio of the number of correctly labeled responses to the total that should have been labeled ; precision [P] is the ratio of the number of correctly labeled responses to the total labeled. [...]* (Jurafsky et Martin, 2009 : 766-767).

Notons que le rappel moins strict a un résultat de 1.002, résultat qui semble improbable, sauf si on souligne que certaines expressions n'étaient pas reconnues dans le corpus annoté, mais bien dans Unitex. En effet, Unitex étant automatique, toutes les instances des constructions temporelles étaient reconnues, alors que certaines ont été oubliées par l'annotateur dans le corpus annoté.

Enfin, soulignons que l'outil fonctionne de manière plutôt satisfaisante, par rapport au service qui lui est demandé. Si toutes les dates ne sont pas extraites entièrement, au vu de l'application et de l'interface actuelle, cela ne pose pas de problème. Ainsi, les quelques erreurs signalées à la table 10.2 sont sans conséquences. Les 18 erreurs d'expressions non temporelles doivent, elles, être corrigées : elles peuvent porter préjudice à une lecture complète et rapide du rapport.

1 Evaluation de la contextualisation et de la normalisation

Cette évaluation [5] porte sur la reconnaissance des paires événement médical - expression temporelle. Les différents exemples d'application de notre outil montre que l'identification des paires doit encore être améliorée - rappelons seulement que le repérage des expressions temporelles (et leur normalisation) était l'objectif premier de l'outil -. À cet effet, nous présentons quelques exemples de la normalisation et envisageons les évolutions nécessaires.

Pour rappel, nos graphes de contextualisation ont été agencés et créés selon quelques modèles repérés lors de l'annotation des documents patients. Chaque type d'événement [6] a son propre graphe de contextualisation.

N'importe quelle paire d'événement médical - timex est reconnue si les éléments sont placés l'un à côté de l'autre dans le texte, comme le montre la figure ci-dessous qui reprend des exemples de toutes les paires d'événements possibles :

5. Cette évaluation est faite sans données chiffrées et se base sur des exemples d'application.
6. Voir table 7.1 page 93 pour les différents types d'événements de contextualisation.

FIGURE 10.1 – Exemple de contextualisation de tous les types d'événements.

Notre outil, dans les cas simples, fonctionne donc. Cependant, les figures suivantes reprennent des exemples qui paraissent tout aussi simples, mais qui ne sont pas encore reconnus par Time extractor.

Ainsi, les paires procédures - timex sont repérées lorsqu'un mot sépare les éléments. Cependant, s'ils sont espacés de plus d'un mot, la procédure n'est pas soulignée. Par ailleurs, dans la deuxième phrase, Time extractor aurait dû reconnaître une paire bodypart - timex [7] :

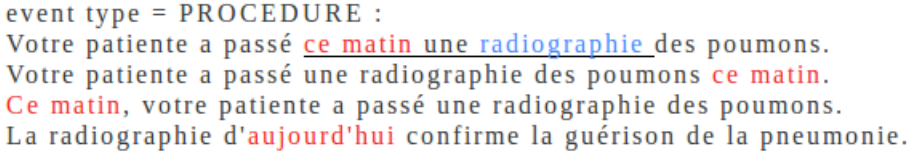

FIGURE 10.2 – Exemple de contextualisation et de non-reconnaissance des paires procedure - timex.

7. Il est probable que *poumons* ne soit pas noté (à sa forme pluriel et/ou singulier) dans nos dictionnaires.

De même, dans les événements de substance, les paires ne sont pas reconnues si les éléments sont séparés par plus d'un mot. L'exemple ci-dessous montre également que la reconnaissance des timex « tous les x temps » fait défaut. Nous pourrions aussi ajouter l'identification des chiffres préalable aux substances, ou tout du moins, au lemme « cigarette » :

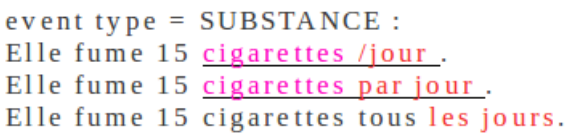

FIGURE 10.3 – Exemple de contextualisation et de non-reconnaissance des paires substance - timex.

De plus, les éléments de ponctuation coupent la reconnaissance des paires. Nous devons aussi permettre l'identification des paires séparées par un sujet et un verbe (du type « elle souffre ») : nous voyons ci-dessous que « depuis 3 ans » et « problèmes respiratoires » ne sont pas reliés. Aussi, les éléments médicaux doivent comprendre les termes tels que « difficultés », « problèmes »,... :

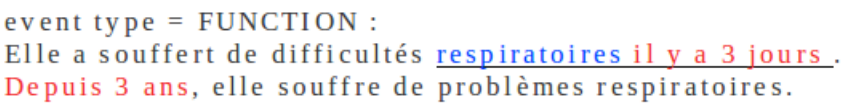

FIGURE 10.4 – Exemple de contextualisation et de non-reconnaissance des paires function - timex.

La figure 10.5 sur la reconnaissance des paires clinicalfinding-disorder - expression temporelle soulève deux problèmes de normalisation. Le premier duo d'exemples (guérison de la pneumonie) montre que les timex typées « UNKNOWN » (voir figure 10.6) ne sont pas reliée à un événement médical. La figure 10.7 (exemple de thrombose veineuse) expose un souci de normalisation des expressions typées « TIME » et le manque de reconnaissance d'événements "triples" :

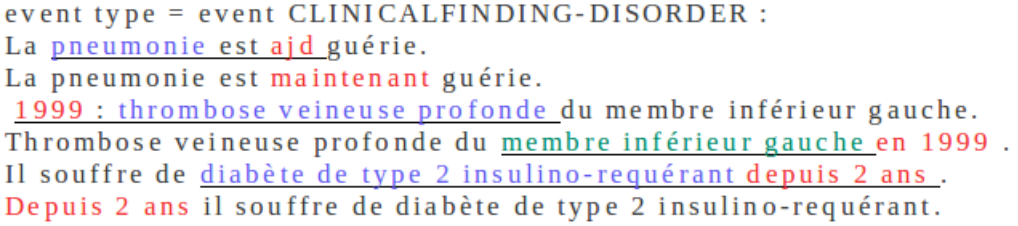

FIGURE 10.5 – Exemple de contextualisation et de non-reconnaissance des paires disorder - timex.

L'étiquette « UNKNOWN » est donnée à toutes les expressions temporelles extraites qui n'ont pas de règles de normalisation. Nous assurons ainsi que ces timexes soient tout de même notées comme telles et apparaissent correctement dans le texte :

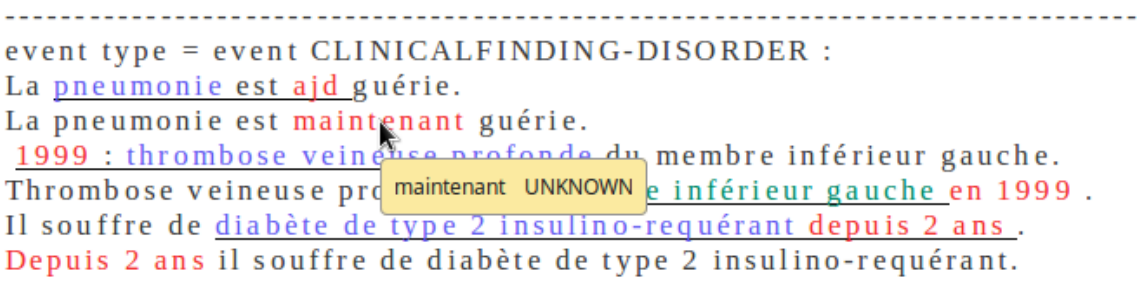

FIGURE 10.6 – Exemple de contextualisation et de non-reconnaissance des paires avec une timex typée « UNKNOWN ».

Quand des années ou des expressions chiffrées de type « MM AAAA » précèdent un événement médical, ces expressions sont reconnues en tant que « TIME ». Cependant, nous ne normalisons pas encore ces timex et elles gardent leur étiquette « TIME ».

Ensuite, « du membre inférieur gauche » est un événement de type « BODYPART ». Cependant, lorsqu'un « CLINICALFINDING-DISORDER est suivi ou précédé d'un « BODYPART », tout l'événement médical n'est pas reconnu. Une évolution de l'outil serait donc de permettre l'identification de plusieurs événements médicaux qui se suivent. Notons que les événements "triples" « TIME - PROCEDURE - CLINICALFINDING-FINDING » et « TIME - PROCEDURE - CLINICALFINDING-DISORDER » sont déjà reconnus (voir figure 10.8).

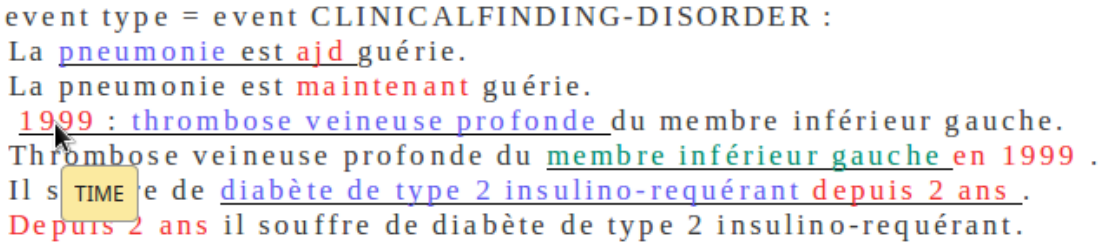

FIGURE 10.7 – Exemple de contextualisation et non-reconnaissance des paires disorder - time.

Nous avons donc ci-dessous un exemple de reconnaissance d'un événement "triple", une expression médicale, une procédure et un clinicalfinding-disorder :

FIGURE 10.8 – Exemple de contextualisation d'un triple événement time - procedure - disorder.

Enfin, les paires et les événements triples ne sont pas toujours correctement soulignés dans le texte. Une petite retouche s'avère nécessaire pour une visualisation complète et homogène des événements médicaux temporalisés :

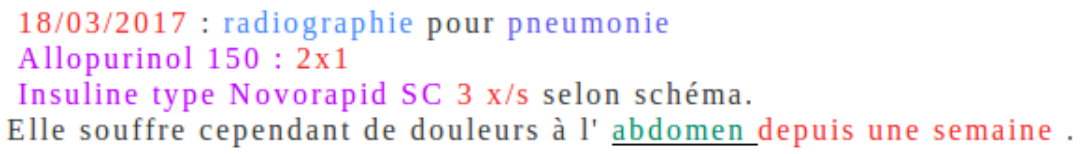

FIGURE 10.9 – Exemple de contextualisation d'événements non soulignés de manière homogène.

Chapitre 11

Perspectives

Dans la section précédente, nous avons évalué notre programme sous différents aspects. Plusieurs erreurs et problèmes ont été soulignés. Nous allons reprendre ces insuffisances et envisager comment les améliorer.

Pour commencer, l'analyse d'un autre corpus (varié et provenant d'autres hôpitaux, dans des domaines médicaux plus variés) pourrait permettre d'inclure de nouvelles graphies et de nouvelles expressions. En outre, nous avons repéré les expressions médicales dans le corpus, mais il serait intéressant de disposer d'une liste pré-établie lors d'autres recherches.

1 Graphies

Le corpus médical compte parfois plusieurs graphies pour un même mot ou une expression temporelle, notamment des abréviations. Nous avons, autant que possible, inclus ces abréviations dans les graphes d'extraction. Cependant, il n'était pas possible de les noter dans tous les cas. En effet, si nous prenons la graphie « mais » correspondant au mois de « mai », l'outil risque d'inclure la conjonction dans la liste des éléments reconnus. Autre exemple : « mardi » est parfois abrégé « ma », il y a confusion possible avec le possessif. Le « Lu » dans « Dr Lu Anh » est reconnu comme étant « lundi ».

Il faudrait pouvoir inclure ces graphies en analysant les contextes dans lesquels elles se trouvent et inclure des contextes négatifs de gauche et de droite. De plus, un graphe spécial devrait être utilisé pour les expressions temporelles seules (dans ce cas-ci, les jours de la semaine et les mois).

Nous avons également rencontré des mots ou de dates (partiellement) collés. Ces cas ne sont pas gérés par nos graphes. Ce problème ne nous a pas semblé très répandu, mais si ces occurrences devaient apparaître en nombre significatif, nous pourrions créer un module spécial ou une normalisation propre à ces cas.

Des graphes trop flexibles peuvent entraîner des erreurs, aussi doit-on être prudent dans l'ajout des graphies. Il faut aussi conditionner les expressions avec des contextes positifs et négatifs de gauche et de droite (via Unitex). Ainsi, « Hôpital de jour, pouls 16x4 », ... ne seraient plus retenus comme date.

2 Normalisation

Dans l'évaluation, nous avons vu qu'il y a parfois des erreurs de normalisation. En effet, notre programme utilise une annotation de sortie de graphes relativement minime [1]. L'utilisation d'une norme internationalement reconnue et pratiquée - comme TimeML - pourrait régler quelques soucis de normalisation. De plus, il serait intéressant de reconnaître le temps des verbes (ce qu'Unitex permet très facilement, en utilisant les dictionnaires) pour optimiser le système.

Dans notre programme, nous avons introduit la reconnaissance des événements médicaux liés aux expressions temporelles. De nombreux autres travaux se sont déjà penchés sur le sujet [2]. Néanmoins, ils sont pour la plupart basés sur l'anglais. En français, nous pouvons principalement évoquer HeidelTime, avec l'extension au français de Tapi Nzali et al. (2015) et le système SYNODOS [3] et le projet iMediate [4].

1. Ces sorties de graphes se trouvent dans l'annexe électronique.
2. Arnulphy *et al.*, 2014 ; Tang *et al.*, 2013, , Li *et al.*, 2012, UzZaman *et al.*, 2012 ; Tao *et al.*, 2012 ; Roberts *et al.*, 2013,...
3. SYstème de Normalisation et d'Organisation de Données médicales textuelles pour l'Observation en Santé. Bittar, 2008 ; Metzger *et al.* 2014
4. Qui reconnaît les événements médicaux dans un texte en se basant sur la nomencatre SYNODOS. Cependant,

Notre programme ne reconnaît que les événements médicaux textuellement proches d'une expression temporelle. Cependant, nous avons vu que cela provoquait des erreurs dans la reconnaissance d'un événement et de sa date si un autre événement médical est textuellement plus proche de l'expression temporelle que l'événement réellement lié. L'exemple ci-dessous montre ainsi des paires erronées, qui auraient dû être reconnues comme : « Date examen : 06/10/2015 » et « 06/10/2015 Type examen : CONS » :

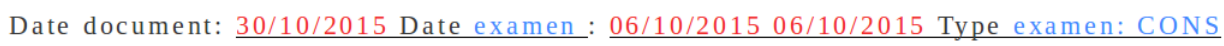

FIGURE 11.1 – Exemple de contextualisation d'événements non soulignés de manière homogène.

Le projet iMediate identifie tous les éléments médicaux d'un texte, mais peu d'expressions temporelles. La juxtaposition de notre système à ce type de programme pourrait amener une meilleure reconnaissance événementielle : tous les événements médicaux et toutes les expressions temporelles seraient soulignés. Un module pourrait alors analyser le texte et lier les événements et leurs dates.

3 Améliorations avec TimeML ?

Nous venons d'évoquer nos sorties de graphes minimes et simplistes. L'utilisation d'un langage de balisage tel que TimeML pourrait résoudre cette imperfection [5]. De plus, un autre avantage de TimeML est qu'il repère également les événements en plus des expressions temporelles, ainsi que les relations entre les deux (relations TLink).

Cependant, comme souligné par Bittar (2008 : 6) *TimeML [...] s'accroche aux relations superficiques et calculables de façon opérationnelle*. Ainsi, de nombreuses expressions temporelles telles que « il y a quelques années », des expressions typiquement médicales (« J5 », « 2x/j », etc.), ou encore les expressions contenant des abréviations ne seraient plus reconnues. Cela diminuerait l'efficacité de notre programme dont l'une des forces est justement la reconnaissance des expressions telles qu'utilisées réellement et fréquemment dans un corpus médical.

la reconnaissance des expressions temporelles est encore assez basique (De Meyere *et al.*, 2015 ; De Meyere, 2017).

5. Voir un exemple d'annotation en utilisant ISO-TimeML à la page 51.

Il y aurait plusieurs problèmes liés à l'utilisation d'une norme d'annotation. D'abord, elle doit pouvoir prendre en compte le domaine médical en français. Sans cela, trop d'informations seraient perdues. De plus, les abréviations doivent impérativement être reconnues : les fréquences de prises de médicaments, les durées, parfois même les dates (avec les jours de la semaine et les mois) sont toutes des informations susceptibles d'être abrégées. Une des difficultés serait alors de ne pas intégrer des expressions telles que « 16x4 » dans « pouls 16x4 » que notre système reconnaît erronément comme une fréquence. Notons également que si un système repère le terme « lundi », les termes précédant l'expression temporelle peuvent être nécessaires à la normalisation de la timex : ainsi, « avant lundi » induit une limite. Enfin, les expressions vagues et celles à plus larges fourchettes (« dans quelque temps », « l'année prochaine », « durant l'enfance », etc.) doivent également être reconnues et annotées pour en permettre la normalisation. Il nous paraît essentiel qu'un système prenne en considération ces divers points.

4 Réflexion sur les différentes approches pour l'extraction de la temporalité dans le domaine médical

La méthode descriptive que nous avons élaboré demande énormément de temps de conception. Dès qu'une nouvelle expression apparaît, il faut l'ajouter aux graphes Unitex et inclure sa normalisation dans le script *Normalization.py*. Cependant, le bruit est minime et l'ajout des éléments est facile et relativement rapide. Une meilleure normalisation et annotation en sortie des graphes permettraient d'augmenter cette facilité et cette rapidité, mais le processus pourrait s'avérer long et répétitif au cas où plusieurs expressions sont à insérer.

Une méthode par machine learning basée sur des ressources lexicales peut trouver beaucoup plus d'expressions temporelles que notre système. Néanmoins, une très grande quantité de données d'entraînement est nécessaire, à cause d'une part de la grande variabilité des expressions à reconnaître, d'autre part afin de rencontrer tous les cas possibles. Autre exigence : ces données lexicales doivent être spécifiques au domaine médical pour récupérer les expressions typiques. Arnulphy *et al.* soulignent *the necessity of providing semantic information large enough to cover the great number of ways to express events, especially for the nominal events*[6].

6. Arnulphy *et al.* (2013 : 23).

Un modèle statistique serait également avantageux pour les contextes de droite et de gauche de l'expression analysée : avec des ressources lexicales médicales, les dates « 110593 », « 11/05 », « 11.05 », « 05.93 », etc. pourraient être récupérées sans provoquer de bruit.

Évidemment, pour régler ce souci, on pourrait également imaginer utiliser un dictionnaire médical et l'indiquer comme contexte négatif. Mais pour ce faire, il faudrait analyser le contexte de toutes les expressions médicales reprenant une séquence chiffrée qui ne doit pas être reconnue par nos graphes.

5 Comparaison avec les travaux de Hamon et Grabar (2014)

Hamon et Grabar (2014) ont étudié l'identification des expressions temporelles dans le domaine médical. Ces recherches ont été menées avec l'outil HeidelTime[7], qu'ils ont choisi grâce à sa *best combination of performance and adaptability*[8] et qu'ils ont étendu au domaine médical. Tout comme notre outil, leur système est à base de règles.

Ils utilisent les normes d'annotations de Timex3, c'est-à-dire qu'ils divisent les expressions temporelles en quatre catégories : (1) *date*, (2) *duration*, (3) *frequency* et (4) *time*. Leurs recherches se découpent en trois étapes : l'extension d'HeidelTime au domaine clinique en langues française et anglaise, l'évaluation des résultats et la visualisation des événements médicaux avec leur information temporelle. Nous nous intéressons particulièrement à l'évaluation de leur reconnaissance des expressions temporelles en français.

Leur évaluation en français porte sur 26 documents patients provenant d'un hôpital français. Le tableau ci-dessous reprend la précision et le rappel de notre outil et de celui d'Hamon et Grabar. La précision est le pourcentage des expressions temporelles pertinentes reconnues divisé par le nombre total des expressions temporelles identifiées ; le rappel est le pourcentage des expressions temporelles pertinentes reconnues divisé par le nombre de relations temporelles à identifier.

7. Présenté aux pages 54 à 56.
8. Hamon et Grabar (2014 : 102).

	Rappel	**Précision**
Time extractor (évaluation stricte)	0.9526	0.9294
Hamon et Grabar	0.9504	0.9341

TABLE 11.1 – Rappel et précision de l'évaluation de Time extractor et d'Hamon et Grabar.

Comme l'indique la table ci-dessus, les résultats entre les deux outils sont très proches. Nous n'avons cependant pas accès au nombre d'expressions temporelles identifiées, le nombre d'expressions correctement reconnues, ni le nombre de timex à repérer dans leur corpus. Nous ne pouvons donc pas faire une analyse et une comparaison approfondies des résultats de leur système. Néanmoins, Hamon et Grabar nous informe des principales erreurs de leur outil :

> *the main limitation of the system is due to the incomplete coverage of the linguistic expressions (e.g. « au cours de, mensuel » [...]). Among the current false positive, we can find ratios (« 2/10 » is considered a date, while it means lab results), polysemous expressions (« Juillet » in « rue du 14 Juillet » [...]), and segmentation errors (« few days » detected instead of « the next few days »)*[9].

Ces erreurs rejoignent certaines des nôtres : l'expression « rue du 14 Juillet », par exemple, est semblable à notre souci de « 4430 Ans ». D'autre part, nous avons décidé de ne pas reconnaître les dates de type « 2/10 » à cause du bruit que cela provoquait, de même que nos « 16x4 » de « pouls 16x4 ». Nous rencontrons également plusieurs erreurs de segmentation.

Les travaux d'Hamon et Grabar ont donc, avec un système à base de règles, des résultats et des erreurs proches des nôtres. Les erreurs rencontrées sont liées à la tournure des expressions propres à chacun, aux variétés de langage, au panel d'options qu'offre la langue française pour exprimer le temps et aux spécificités du domaine médical.

9. *Ibid.,* : 104.

Cinquième partie

Conclusion générale

Un des aspects particuliers de l'information temporelle est qu'elle est toujours en lien avec une seconde indication, que cette dernière soit de type événementiel ou autre. L'expression de la temporalité est complexe et variée, et sa compréhension l'est tout autant : une certaine connaissance du monde, un événement porteur de temporalité, des expressions temporelles, des marqueurs linguistiques,... tous ces aspects doivent être pesés et réfléchis dans le développement d'un système d'extraction d'expressions temporelles. Par ailleurs, un domaine spécifique peut avoir sa propre façon d'exprimer le temps, comme c'est le cas en médecine. Ces éléments multiples doivent être pris en considération.

La dimension temporelle dans une langue précise est relativement stable et fait preuve d'une certaine régularité. Un système à base de règles peut donc être développé pour efficacement repérer et extraire ce type de données dans un texte, dans un document. C'est ce que nous avons démontré avec notre outil Time extractor, bien qu'il soit encore à un stade de développement.

Dès lors, nous avons parcouru et étudié l'état de l'art de l'extraction de la temporalité dans les domaines général et clinique, ainsi que l'état de l'art de la normalisation des données temporelles. Cette entreprise visait à répondre aux questions suivantes : *Comment extraire les expressions temporelles ? Quelles différences entre la sphère médicale et la sphère générale ?* Une étude détaillée de l'expression du temps en français a apporté des nuances et des approfondissements aux réponses déjà collectées, et révélé la complexité du français en cette matière.

L'analyse de documents patients a ensuite permis de créer des grammaires d'extraction des items temporelles avec le logiciel *Unitex* et de mettre au point un modèle de normalisation des dates. Cette dernière a été élaborée avec les sorties de graphes d'extraction. Nous avons ainsi pu établir notre propre système d'annotation pour la normalisation, ce qui a débouché sur l'établissement des règles d'interprétation.

Dans le domaine médical, le lien entre l'expression temporelle et l'événement clinique est essentiel pour l'accès rapide et l'affichage des informations et antécédents des patients. Nous avons ainsi tenté une approche (également par règles) de reconnaissance des paires fait médical - timex. Cet aspect de l'outil est un prototype qui doit encore être amélioré et développé. Il est pour l'instant une preuve du concept, plutôt qu'un produit (presque) fini.

Notre système actuel fonctionne comme suit : l'outil traite le texte de départ, met en exergue les informations temporelles reconnues ainsi que les paires événement médical - timex identifiées. Nous pourrions imaginer dans un développement futur, l'ajout d'une ligne du temps annexe qui reprendrait de manière suivie tous les événements médicaux, ainsi que leur date, qu'a connus un patient depuis sa naissance.

Les caractéristiques de Time extractor consistent en la prise en compte des imprécisions et des abréviations temporelles, ainsi qu'en la reconnaissance des timex typiquement cliniques. Rendre accessible, en un clic ou deux, le parcours d'un patient ; coupler une date et un événement médical afin de visualiser en un instant la chronologie des faits qui émaillent les données d'un dossier clinique sont l'objectif d'outils tels que Time extractor.

Bien que notre système ne soit pas encore parfait et que des retouches et améliorations soient nécessaires, il répond aux exigences de départ et il effectue déjà efficacement la tâche demandée.

Sixième partie

Bibliographie

[1] Ahn, D., Fissaha Adafre, S., & De Rijke, M. (2005a). Towards task-based temporal extraction and recognition. In *Dagstuhl Seminar Proceedings. Schloss Dagstuhl-Leibniz-Zentrum für Informatik*.

[2] Ahn, D., Adafre, S.F., de Rijke, M. (2005b). Extracting temporal information from open domain text : A comparative exploration. *Journal of Digital Information Mangement*, 3(1) :14-20.

[3] Angeli, G., Manning, C. D., & Jurafsky, D. (2012, June). Parsing time : Learning to interpret time expressions. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies* (pp. 446-455). Association for Computational Linguistics.

[4] Appelt, D. E., & Israel, D. J. (1999). Introduction to Information Extraction Technology. A Tutorial Prepared for the Int. Joint Conf. on Artificial Intelligence (IJCAI-99).

[5] Arnulphy, B., Claveau, V., Tannier, X., & Vilnat, A. (2014, March). Techniques d'apprentissage supervisé pour l'extraction d'événements TimeML en anglais et français. In *Conférence en recherche d'information et applications, CORIA 2014* (pp. 16-p).

[6] Arnulphy, B., Claveau, V., Tannier, X., & Vilnat, A. (2015, June). Supervised machine learning techniques to detect timeml events in french and english. In *International Conference on Applications of Natural Language to Information Systems* (pp. 19-32). Springer, Cham.

[7] Bittar, A. (2008, June). Annotation des informations temporelles dans des textes en français. In *RECITAL 2008*.

[8] Bittar, A., & Danlos, L. (2009a). Intégration des constructions à verbe support dans TimeML. In *TALN 2009*.

[9] Bittar, A. (2009b). Annotation of events and temporal expressions in French texts. In *Proceedings of the Third Linguistic Annotation Workshop* (pp. 48-51). Association for Computational Linguistics.

[10] Bittar, A. (2010). ISO-TimeML Annotation Guidelines for French Version 1.0. *ALPAGE & Université Paris*, 7. Disponible à l'adresse Internet `https://www.researchgate.net/publication/265349449_ISO-TimeML_Annotation_Guidelines_for_French_Version_10`.

[11] Bittar, A., Amsili, P., & Denis, P. (2011, June). French TimeBank : un corpus de référence sur la temporalité en français. In *TALN 2011-Traitement Automatique des Langues Naturelles* (Vol. 1, pp. 259-270). Laboratoire d'Informatique de Robotique et de Microélectronique.

[12] Bittar, A. & Hagège, C. (2012). Un annotateur automatique d'expressions temporelles du français et son évaluation sur le TimeBank du français. In *Actes de la conférence TALN 2012*.

[13] Bittar, A., Hagege, C., Moriceau, V., Tannier, X., & Teissedre, C. (2012, May). Temporal Annotation : A Proposal for Guidelines and an Experiment with Inter-annotator Agreement. In *LREC* (pp. 3741-3745).

[14] Bittner, T. (2002). Approximate qualitative temporal reasoning. *Ann. Math. Artif. Intell., 36*(1-2), 39-80.

[15] Boguraev, B., & Ando, R. K. (2006, May). Analysis of TimeBank as a Resource for TimeML parsing. In *Proceedings of LREC* (pp. 71-76).

[16] cairn.info, *La composante temps dans l'information géographique textuelle*. `http://www.cairn.info/revue-document-numerique-2007-2-page-129.html`, page consultée le 16 juillet 2016.

[17] Chang, A. X., & Manning, C. D. (2012, May). Sutime : A library for recognizing and normalizing time expressions. In *LREC* (Vol. 2012, pp. 3735-3740).

[18] Cheng, Y., Anick, P., Hong, P., & Xue, N. (2013). Temporal relation discovery between events and temporal expressions identified in clinical narrative. *Journal of biomedical informatics*, 46, S48-S53.

[19] Chikka, V. R. (2016). CDE-IIITH at SemEval-2016 Task 12 : Extraction of Temporal Information from Clinical documents using Machine Learning techniques. In *SemEval@ NAACL-HLT* (pp. 1237-1240).

[20] Cowie, J., & Lehnert, W. (1996). Information extraction. *Communications of the ACM, 39*(1), 80-91.

[21] Cunningham H (2006), Information Extraction, Automatic. In : Keith Brown, (Editor-in-Chief) *Encyclopedia of Language & Linguistics, Second Edition*, volume 5, pp.665-677. Oxford : Elsevier.

[22] Cunningham, H., Tablan, V., Roberts, A., & Bontcheva, K. (2013). Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS computational biology, 9*(2), e1002854.

[23] Derczynski, L., Strötgen, J., Maynard, D., Greenwood, M. A., & Jung, M. (2016). GATE-Time : Extraction of Temporal Expressions and Event. In *10th Language Resources and Evaluation Conferenc*e (pp. 3702-3708). European Language Resources Association (ELRA).

[24] Deléger, L., Grouin, C., & Zweigenbaum, P. (2010, September). Extracting medication information from French clinical texts. In *MedInfo* (pp. 949-953).

[25] De Meyere, D., Klein, T., François, T., Debongnie, J.C., Radulescu, C., Mbengo, N., Koura, M.O., Coppieters't Wallant, Y., Fairon, C. (2015). Automatic annotation of medical reports using SNOMED-CT : a flexible approach based on medical knowledge databases. 7th Language & Technology Conference : Human Language Technologies as a Challenge for Computer Science and Linguistics (Poznań, Poland, du 27/11/2015 au 29/11/2015). In *Zygmunt Vetulani & Joseph Mariani, Proceedings of the 7th Language & Technology Conference, Fundacja Uniwersytetu im. A. Mickiewicza : Poznań, Poland2015*, p. 519-523.

[26] De Meyere, D., *L'annotateur iMediate, un outil pour l'encodage de dossiers médicaux en SNOMED-CT* : Séminaires du Cental, 28 avril 2017, Louvain-la-Neuve.

[27] D'hondt, E., von Neuman, R. J., Tannier, X., & Névéol, A. (2015). Redundancy in French electronic health records : A preliminary study. *Proc of the 6th Health Text Mining and Information Analysis Work* (LOUHI), 21-30.

[28] Lucie Drat. Projet TourInFlux. Annotation des expressions temporelles. Linguistique. 2014.

[29] Euzenat, J. *Granularité dans les représentations spatio-temporelles.* RR-2242, INRIA. 1994. <inria-00077218>.

[30] Fiorin, J. L. (1998). Temps : entre la langue et le discours. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (10), 121-148.

[31] Formoso, E. (2013, Janvier). Des cadrans grecs au calendrier romain. *Les cahiers Sciences & vie, numéro 134.* pp. 29-33.

[32] Frèches J., (2013). *Dictionnaire amoureux de la Chine.* Plon, Paris.

[33] Fries, J. A. (2016). Brundlefly at SemEval-2016 Task 12 : Recurrent neural networks vs. joint inference for clinical temporal information extraction. *arXiv preprint arXiv :1606.01433.*

[34] Gicquel, Q., Tvardik, N., Bouvry, C., Kergourlay, I., Bittar, A., Segond, F., Darmoni, S., & Metzger, M. H. (2014). Annotation methods to develop and evaluate an expert system based on natural language processing in electronic medical records. *Studies in health technology and informatics, 216*, 1067-1067.

[35] Gosselin, Laurent, (1996), *Sémantique de la temporalité en français. Un modèle calculatoire et cognitif du temps et de l'aspect*, Louvain-la-neuve, Duculot, col. Recherches 292p.

[36] Gross, Elia (2003). Le temps grammatical et le temps qui s'écoule. In *Il Verbo italiano. Studi diacronici, sincronici, contrastivi, didattici. Atti del XXXV Congresso internationale di studi* (Parigi, 20-22 settembre 2001), pp. 379–396.

[37] Gross, M. (1986). *Grammaire transformationnelle du français. 3 - Syntaxe de l'adverbe.* Asstril, Paris.

[38] Guda, V., & Sanampudi, D. S. K. (2016). Extraction of Temporal Events in Natural Language Text. *Extraction, 6*(1).

[39] Gupta, N. (2015). Temporal Information Extraction, Extracting Events and Temporal Expressions, A Literature Survey.

[40] Gupta, N., Joshi, A., & Bhattacharyya, P. (2015). A temporal expression recognition system for medical documents by taking help of news domain corpora.

[41] Grouin, C., Grabar, N., Hamon, T., Rosset, S., Tannier, X., & Zweigenbaum, P. (2013). Eventual situations for timeline extraction from clinical reports. *Journal of the American Medical Informatics Association, 20*(5), 820-827.

[42] Hagège, C., Bittar, A.,(2012). Un annotateur automatique d'expressions temporelles du français et son évaluation sur le TimeBank du français.

[43] Hamon, T., & Grabar, N. (2014). Tuning HeidelTime for identifying time expressions in clinical texts in English and French. *EACL 2014*, 101-105.

[44] Hao, T., Rusanov, A., & Weng, C. (2013, August). Extracting and normalizing temporal expressions in clinical data requests from researchers. In *International Conference on Smart Health* (pp. 41-51). Springer, Berlin, Heidelberg.

[45] Hitzeman, J. (1995, November). Text Type and the Position of a Temporal Adverbial within the Sentence. In *University of Edinburgh*.

[46] Hobbs, J., & Pustejovsky, J. (2003, March). Annotating and reasoning about time and events. In *Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning* (Vol. 3).

[47] Joshi, P. PERPETUAL ENIGMA - PERENNIAL FASCINATION WITH ALL THINGS TECH. *What Are Conditional Random Fields ?*. 2013. `https://prateekvjoshi.com/2013/02/23/what-are-conditional-random-fields/`, page consultée le 31 juillet 2017.

[48] Irvine, A. K., Haas, S. W., & Sullivan, T. (2008). Tn-ties : A system for extracting temporal information from emergency department triage notes. In *AMIA Annual Symposium proceedings* (Vol. 2008, p. 328). American Medical Informatics Association.

[49] Jindal, P., & Roth, D. (2013). Extraction of events and temporal expressions from clinical narratives. *Journal of biomedical informatics, 46*, S13-S19.

[50] Jung, S., Kim, S., Yoo, S., & Choi, J. (2009). Toward the automatic generation of the entry level CDA documents. *Journal of Korean Society of Medical Informatics, 15*(1), 141-151.

[51] Jurafsky, D. & Martin, J.H. (2009). *Speech and Language Processing : An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, New Jersey.

[52] Kate, R. J. (2013). Towards converting clinical phrases into SNOMED CT expressions. *Biomedical informatics insights*, 6(Suppl 1), 29.

[53] Kessler, R., Tannier, X., Hagège, C., Moriceau, V., & Bittar, A. (2012). Extraction de dates saillantes pour la construction de chronologies thématiques. *TAL, 53(2)*, 57-86.

[54] Laurent Kevers. Accès sémantique aux bases de données documentaires. Techniques symboliques de traitement automatique du langage pour l'indexation thématique et l'extraction d'informations temporelles. Linguistique. Université Catholique de Louvain, 2011. Français. <tel-00568089>

[55] Khalifa, A., Velupillai, S., & Meystre, S. (2016). UtahBMI at SemEval-2016 Task 12 : Extracting temporal information from clinical text. *Proceedings of SemEval*, 1256-1262.

[56] Khare, R., An, Y., Li, J., Song, I. Y., & Hu, X. (2012, January). Exploiting semantic structure for mapping user-specified form terms to SNOMED CT concepts. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium* (pp. 285-294). ACM.

[57] Kovačević, A., Dehghan, A., Filannino, M., Keane, J. A., & Nenadic, G. (2013). Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives. *Journal of the American Medical Informatics Association, 20*(5), 859-866.

[58] Le blog de Lexgram. *Les déictiques 1.* http ://lexgram.over-blog.fr/article-25284697.html. Page consultée le vendredi 21 juillet 2017.

[59] Lee, D. H., Lau, F. Y., & Quan, H. (2010). A method for encoding clinical datasets with SNOMED CT. *BMC medical informatics and decision making, 10*(1), 53.

[60] Lelong, R., Merabti, T., Grosjean, J., Joulakian, M. B., Griffon, N., Dahamna, B., Cuggia, M., Pereira, S., Grabar, N., Thiessard, F. & Massari, P. (2014). Moteur de recherche sémantique au sein du dossier du patient informatisé : langage de requêtes spécifique. In *JFIM* (pp. 139-151).

[61] Li, M., & Patrick, J. (2012). Extracting temporal information from electronic patient records. In *AMIA Annual Symposium Proceedings* (Vol. 2012, p. 542). American Medical Informatics Association.

[62] Lin, Y. K., Chen, H., & Brown, R. A. (2013). MedTime : A temporal information extraction system for clinical narratives. Journal of biomedical informatics, 46, S20-S28.

[63] Llorens, H., Derczynski, L., Gaizauskas, R. J., & Saquete, E. (2012, May). TIMEN : An Open Temporal Expression Normalisation Resource. In *LREC* (pp. 3044-3051).

[64] Madkour, M., Benhaddou, D., & Tao, C. (2016). Temporal data representation, normalization, extraction, and reasoning : A review from clinical domain. *Computer methods and programs in biomedicine, 128*, 52-68.

[65] Mani, I. (2004, November). Recent developments in temporal information extraction. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP'03)* (pp. 45-60).

[66] Mani, I. (2007). Chronoscopes : a theory of underspecified temporal representations. *Annotating, Extracting and Reasoning about Time and Events*, 127-139.

[67] Metzger, M. H., Gicquel, Q., Rabarijaona, D., Segond, F., Bittar, A., Dini, L., Soualmia, L.F. & Darmoni, S. J. (2014). Usage secondaire du dossier médical informatisé à des fins épidémiologiques et d'évaluation de la qualité des soins : le projet SYNODOS. In *JFIM* (pp. 152-161).

[68] Mitkov, R. (Ed.). (2005). *The Oxford handbook of computational linguistics*. Oxford University Press.

[69] Mitsa, T. (2010). *Temporal data mining*. CRC Press.

[70] Moriceau, V., & Tannier, X. (2014). French resources for extraction and normalization of temporal expressions with heideltime. *extraction, 26*.

[71] Muller, P., & Tannier, X. (2004). Une méthode pour l'annotation de relations temporelles dans des textes et son évaluation. *Actes TALN'04*.

[72] Névéol, A., Grouin, C., Tannier, X., Hamon, T., Kelly, L., Goeuriot, L., & Zweigenbaum, P. (2015, September). CLEF eHealth Evaluation Lab 2015 Task 1b : Clinical Named Entity Recognition. In *CLEF (Working Notes)*.

[73] Nguyen, K. H., Tannier, X., & Moriceau, V. (2014). Ranking Multidocument Event Descriptions for Building Thematic Timelines. In *COLING* (pp. 1208-1217).

[74] Nzali, M. D. T., Névéol, A., & Tannier, X. (2015a). Analyse d'expressions temporelles dans les dossiers électroniques patients. In *TALN : Traitement Automatique des Langues Naturelles*.

[75] Nzali, M. D. T., Tannier, X., Névéol, A., & von Neuman, R. J. (2015b). Automatic Extraction of Time Expressions Across Domains in French Narratives.

[76] Pan, J. Z., Stamou, G., Tzouvaras, V., & Horrocks, I. (2005, September). f-SWRL : a fuzzy extension of SWRL. In *International Conference on Artificial Neural Networks* (pp. 829-834). Springer, Berlin, Heidelberg.

[77] Parent, G., Gagnon, M., & Muller, P. (2008). Annotation d'expressions temporelles et d'événements en français. *Actes de TALN*.

[78] Patrick, J., Wang, Y., & Budd, P. (2007, January). An automated system for conversion of clinical notes into SNOMED clinical terminology. In *Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68* (pp. 219-226). Australian Computer Society, Inc..

[79] Pereira, S., Névéol, A., Massari, P., Joubert, M., & Darmoni, S. (2006). Construction of a semi-automated ICD-10 coding help system to optimize medical and economic coding. In *MIE* (pp. 845-850).

[80] Pereira, S., Massari, P., Buemi, A., Dahamna, B., Serrot, E., Joubert, M., & Darmoni, S. J. (2009). F-MTI : outil d'indexation muititerminologique : application à l'indexation automatique de la SNOMED International. *Risques, Technologies de l'Information pour les Pratiques Médicales*, 57-68.

[81] Pratt-Hartmann, I. (2005). From timeml to TPL. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

[82] Pustejovsky, J., Ingria, B., Sauri, R., Castano, J., Littman, J., Gaizauskas, R., Setzer, A., Katz, G. & Mani, I. (2005). The specification language TimeML. *The language of time : A reader*, 545-557.

[83] Pustejovsky, J., Lee, K., Bunt, H., & Romary, L. (2010, May). ISO-TimeML : An International Standard for Semantic Annotation. In *LREC* (Vol. 10, pp. 394-397).

[84] Rabut I., Yongyi W., Hong L. (2009), *Méthode de chinois, premier niveau.* L'Asiathèque, maison des langues du monde, Paris.

[85] Riegel, M., Pellat, J.-C., Rioul, R. (2006, 3e éd.). *Grammaire méthodique du français*. Quadrige - Puf, Paris.

[86] Roberts, A., Gaizauskas, R., Hepple, M., Demetriou, G., Guo, Y., Roberts, I., & Setzer, A. (2009). Building a semantically annotated corpus of clinical texts. *Journal of biomedical informatics*, 42(5), 950-966.

[87] Roberts, K., Rink, B., & Harabagiu, S. M. (2013). A flexible framework for recognizing events, temporal expressions, and temporal relations in clinical text. Journal of the American Medical Informatics Association, 20(5), 867-875.

[88] Rodon, J. *Hidden Markov Model - Description du modèle.* http://igm.univ-mlv.fr/~dr/XPOSE2012/HiddenMarkovModel/description.html. Page consultée le 31 juillet 2017.

[89] Shaughnessy (ed.) E.L (2008). *La Chine*. Histoire universelle. Evergreen, Köln.

[90] Service public fédéral. *Santé publique, sécurité de la chaîne alimentaire et environnement.*
https://www.health.belgium.be/fr/terminologie-et-systemes-de-codes-snomed-ct. Page consultée le mercredi 19 juillet 2017.

[91] Strötgen, J., & Gertz, M. (2010, July). Heideltime : High quality rule-based extraction and normalization of temporal expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation* (pp. 321-324). Association for Computational Linguistics.

[92] Strötgen, J., Bögel, T., Zell, J., Armiti, A., Van Canh, T., & Gertz, M. (2014). Extending HeidelTime for Temporal Expressions Referring to Historic Dates. In *LREC* (pp. 2390-2397).

[93] Strötgen, J., & Gertz, M. (2015). A Baseline Temporal Tagger for all Languages. In *EMNLP* (Vol. 15, pp. 541-547).

[94] Styler IV, W. F., Bethard, S., Finan, S., Palmer, M., Pradhan, S., de Groen, P. C., Erickson, B., Miller, T., Lin, C., Savova, G., & Pustejovsky, J. (2014). Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2, 143-154.

[95] Sun, W., Rumshisky, A., & Uzuner, O. (2013a). Temporal reasoning over clinical text : the state of the art. *Journal of the American Medical Informatics Association, 20*(5), 814-819.

[96] Sun, W., Rumshisky, A., & Uzuner, O. (2013b). Annotating temporal information in clinical narratives. *Journal of biomedical informatics, 46*, S5-S12.

[97] Sun, W., Rumshisky, A., & Uzuner, O. (2013cb). Evaluating temporal relations in clinical text : 2012 i2b2 Challenge. *Journal of the American Medical Informatics Association*, 20(5), 806-813.

[98] Sun, W., Rumshisky, A., & Uzuner, O. (2015). Normalization of relative and incomplete temporal expressions in clinical narratives. *Journal of the American Medical Informatics Association, 22*(5), 1001-1008.

[99] Tang, B., Wu, Y., Jiang, M., Chen, Y., Denny, J. C., & Xu, H. (2013). A hybrid system for temporal information extraction from clinical text. *Journal of the American Medical Informatics Association, 20*(5), 828-835.

[100] Tannier, X. (2003). Calcul des relations temporelles du discours. Institut National des Sciences Appliquées. Ecole Doctorale Systèmes. Toulouse.

[101] Tannier, X. (2006). *Traitement automatique du langage naturel pour l'extraction et la recherche d'informations* (Vol. 16, p. 17). Technical report, Ecole Nationale Supérieure des Mines de Saint-Etienne.

[102] Tannier, X., & Muller, P. (2011). Evaluating temporal graphs built from texts via transitive reduction. *Journal of Artificial Intelligence Research, 40*, 375-413.

[103] Tannier, X. (2014). Mémoire d'habilitation à diriger des recherches : Traitement des événements et ciblage d'information.

[104] Tao, C., Wei, W. Q., Solbrig, H. R., Savova, G., & Chute, C. G. (2010). CNTRO : a semantic web ontology for temporal relation inferencing in clinical narratives. In *AMIA annual symposium proceedings* (Vol. 2010, p. 787). American Medical Informatics Association.

[105] Tao, C., Wongsuphasawat, K., Clark, K., Plaisant, C., Shneiderman, B., & Chute, C. G. (2012, January). Towards event sequence representation, reasoning and visualization for EHR data. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium* (pp. 801-806). ACM.

[106] Technolangues.net. *Qu'est-ce qu'une ontologie ?*, 2006. `http://www.technolangue.net/article.php3?id_article=280`, page consultée le 29 juillet 2017.

[107] Tissot, H. C. (2016). Normalisation of imprecise temporal expressions extracted from text.

[108] Tomassone, R., 2001. *Grands repères culturels pour une langue : le français.* Hachette Education, Paris.

[109] Tomassone, R., 2002. *Pour enseigner la grammaire, nouvelle édition.* Delgrave Pédagogie et formation, Paris.

[110] Tourille, J., Ferret, O., Névéol, A., & Tannier, X. (2016). LIMSI-COT at SemEval-2016 Task 12 : Temporal relation identification using a pipeline of classifiers. In *SemEval@ NAACL-HLT* (pp. 1136-1142).

[111] UIMA Unstructured Information Management Architecture, An Apache Project. *Apache UIMA*. `https://uima.apache.org/`. Page consultée le 29 juillet 2017.

[112] UzZaman, N., & Allen, J. F. (2010, July). TRIPS and TRIOS system for TempEval-2 : Extracting temporal information from text. In *Proceedings of the 5th International Workshop on Semantic Evaluation* (pp. 276-283). Association for Computational Linguistics.

[113] UzZaman, N., & Allen, J. F. (2011, June). Temporal evaluation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies : short papers-Volume 2* (pp. 351-356). Association for Computational Linguistics.

[114] UzZaman, N., Llorens, H., Allen, J., Derczynski, L., Verhagen, M., & Pustejovsky, J. (2012). Tempeval-3 : Evaluating events, time expressions, and temporal relations. *arXiv preprint arXiv :1206.5333*.

[115] Velupillai, S., Mowery, D. L., Abdelrahman, S., Christensen, L., & Chapman, W. W. (2015). Towards a generalizable time expression model for temporal reasoning in clinical notes. In *AMIA Annual Symposium Proceedings* (Vol. 2015, p. 1252). American Medical Informatics Association.

[116] Wallach, H. M., *conditional random fields*. 2005. `http://www.inference.org.uk/hmw26/crf/`. Page consultée le 31 juillet 2017.

[117] Weiser, S. (2010). *Repérage et typage d'expressions temporelles pour l'annotation sémantique automatique de pages Web-Application au e-tourisme* (Doctoral dissertation, Université de Nanterre-Paris X).

[118] Wolff Philippe. 1. Le temps et sa mesure au Moyen Âge. In : Annales. Économies, Sociétés, Civilisations. 17e année, N. 6, 1962. pp. 1141-1145 ; doi : 10.3406/ahess.1962.420927. `http://www.persee.fr/doc/ahess_0395-2649_1962_num_17_6_420927`

[119] Yang-Drocourt Z., Jianmin F., Rabut R. (2011), *Méthode de chinois, deuxième niveau*. L'Asiathèque, maison des langues du monde, Paris.

[120] Zhou, L., Melton, G. B., Parsons, S., & Hripcsak, G. (2006). A temporal constraint structure for extracting temporal information from clinical narrative. *Journal of biomedical informatics, 39*(4), 424-439.

[121] Zhou, L., & Hripcsak, G. (2007). Temporal reasoning with medical data—a review with emphasis on medical natural language processing. *Journal of biomedical informatics*, 40(2), 183-202.

[122] Zweigenbaum, P., & Tannier, X. (2013, June). Extraction des relations temporelles entre événements médicaux dans des comptes rendus hospitaliers. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN 2013, article court), Les Sables d'Olonne, France* (Vol. 35).

Septième partie

Annexes

Annexe 1 - Présentation des défis TempEval, SemEval et i2b2 2012

TempEval TempEval se concentre sur l'évaluation de systèmes d'extraction d'expressions temporelles (timexes), d'événements et des relations temporelles. (comme définis dans les critères de TimeML). L'annotation se fait avec TimeML, langage de balisage multilingue pour l'information temporelle. TempEval a élaboré différentes tâches [10] :

1. l'identification des événements ;
2. l'identification des expressions temporelles ;
3. l'identification les relations temporelles, elle-même divisée en :
   (a) identifier les relations temporelles entre événements et expressions temporelles dans une même phrase ;
   (b) identifier les relations temporelles entre un événement et la date de création du document ;
   (c) identifier les relations temporelles entre les événements principaux de phrases consécutives ;
   (d) identifier les relations temporelles entre deux événements dont un est syntaxiquement dominant.

i2b2 2012 Le défi i2b2, vise à

> *developing a scalable informatics framework that will enable clinical researchers to use existing clinical data for discovery research and [...] facilitate the design of*

10. Le site Internet suivant présentent plus en détails ce défi : `http://www.timeml.org/tempeval/`.

targeted therapies[11].

Il y a eu différents défis i2b2, mais la temporalité dans le domaine médical est abordée dans l'i2b2/VA 2012. Cette activité est dédiée à la détection de relations temporelles entre événements et au repérage d'expressions temporelles dans des comptes-rendus hospitaliers en anglais.

11. `https://www.i2b2.org/`, page consultée le 23 mai 2016.

Annexe 2 - Présentation des graphes d'extraction de dates

Nous avons créé un total de 139 graphes pour l'extraction des expressions temporelles. La liste des graphes ci-dessous n'est pas exhaustive car certains graphes ne contiennent que des chiffres, des prépositions propres au mois, années, âge, etc.

1. **DATES_MAIN** : Graphe principal permettant l'accès à tous les autres graphes. Il renvoie vers
   - ***age***
   - ***Date_format_main***
   - ***dates_medicales***
   - ***periode_duree***
2. **age** : graphe principal de l'âge, qu'il donne sous la forme d'années, d'années-mois, d'années-mois-jours, d'années-jours, de mois, de mois-jours et de jours.
   1 **age_vie_min** : donne la durée de vie moyenne ou minimale : *vie moyenne de 2 à 3 ans.*
   2 **age_annee** : donne l'âge en nombre d'années : *Age : 7, âgé de 23 ans.*
   3 **age_mois** : précédé du graphe *age_annee*, il donne le nombre de mois : *âgé de 23 ans et 1 mois.*
   4 **age_jour** : donne l'âge en jours. Il peut être seul, ou précédé d'*age_annee* ou *age_mois* : *âgé de 23 ans 1 mois et 19 jours.*
   5 **age_amj** : donne l'âge sous la forme "x ans x mois x jours" : *23 ans 1 mois 19 jours.*

6 **age_mois_seul** : comme *age_mois*, ce graphe donne l'âge en mois. La différence entre les deux graphes tient des contextes négatifs de gauche : *age_mois_seul* a besoin de plus de conditions qu'*age_mois* : *vers 3 mois, âgée de 18 mois.*

3. **Date_format_main** : Graphe principal des dates :

1 **passe_present_futur** : graphe principal reprenant les diverses expressions dans le passé, le présent et le futur.

1 **passe_main**

1 **annee_passe** : gère les expressions : *dans les années 90*, *dans les années cinquante.*

2 **periode_passe** : gère les autres tournures temporelles passées : (*il y a x temps*, *la semaine dernière*).

2 **present_main**

1 **actuellement** : gère les expressions : *actuellement, à présent.*

2 **periode_present** : gère les autres tournures temporelles du présent : *ce jour, ce matin, en cours de soirée.*

3 **futur_main**

1 **gère les autres tournures temporelles futures : *au long cours, demain, dans un délai de 10 jours***.

2 **jour_mois_ppp** : ou *jour - mois passé-présent-futur* : récupère les jours suivis de mois et éventuellement d'une valeur passée, présente, future : *11 octobre passé, 2 avril, 11 mai prochain, 17 novembre de cette année.*

3 **dmf_annee** : ou début-mi-fin de l'année : reprend les années, précédée d'une formulation représentant le début, la moitié ou la fin de l'année : *mi 2015, fin 2017, 2018.*

4 **dmf_mois** : ou début-mi-fin du mois : reprend les mois, précédée d'une formulation représentant le début, la moitié ou la fin du mois : *mi novembre, fin janvier, mai.*

5 **Format Jour-mois-annee** : gère les dates chiffrées et non chiffrées, séparées (ou pas) par "/ - . :" et les dates de naissance : *11/05/2017, 2 avril 2017, 17-11-2012, 4.11.2011, 22 :01 :2017, née le 11 mai 1993.*

6 **Format Jsem-jour-mois-annee** : gère les dates sous la forme jour de la semaine - jour - mois - année : *dimanche 2 avril 2017.*

7 **Format dates partielles**

1 **Format JMA seuls** :

1 **jour_semaine** : récupère les constructions temporelles qui ne sont formées que du jour de la semaine.

2 **mois_seul** : récupère les constructions temporelles qui ne sont formées que du mois.

2 **Format Jsem-jour** : gère les expressions composées du jour de la semaine et du jour : *lundi 15.*

3 **Format mois-annee** : gère les constructions composées du mois et de l'année : *novembre 2012.*

4 **Format annee** : gère les expressions temporelles composées uniquement de l'année.

8 **Recursion annee** : gère les constructions dans lesquelles plusieurs années se répètent : *en 2012 et 2015, en 2011, 2015 ou 2016 et 2018.*

9 **periode_date** : gère les constructions temporelles qui indiquent une date de début et une date de fin : *du 15 au 18 avril 2016, du 9 décembre 2009 au 18 avril 2017.*

10 **Heure_main** : gère les expressions d'heures : *18h25, 9 heures.*

11 **Format Jsem-jour-mois-annee-heure_main** : gère les dates suivies d'une heure : *jeudi 11 mai 2017 à 18h30.*

4. **dates_medicales** : Graphe principal des dates médicales :

1 **temps_trait** : gère les constructions temporelles médicales concernant la fréquence, la période de traitement et les jours de traitement : *1x/j à midi, 3/sem, 3/j pdt 1 sem puis 1/j, deux aux repas.*

2 **gyneco** : gère les expressions relevant d'une grossesse, d'un accouchement, d'antécédents gynécologiques,... : *terme à 18 SA, âge gestationnel : 17sem 2/7.*

3 **hospitalisation** : gère les constructions temporelles précédant, ou suivant une date d'hospitalisation : *2 heures avant votre intervention, 3 jours après l'examen, J5, H3.*

4 **rdv_postop** : gère les dates des rendez-vous après une opération : *à J2, J5 et J18, à J1, J3, J5 puis toutes les semaines.*

5 **expressions_medicales** : gère les expressions temporelles : *post-opératoire, per-opératoire, la veille de l'intervention.*

5. **periode-duree** : Graphe principal reprenant les autres expressions temporelles non reprises dans les catégories ci-dessus.

   1 **delais** : récupère les délais : *de 2 à 3 jours, une délai d'environ 3 semaines, quelques jours.*

   2 **duree** : récupère les durées (ressemble au graphe *delais*) : *durant de très nombreux mois, depuis une dizaine de jours, à 3 mois 1/2.*

   3 **moment_frequence** : récupère les fréquences temporelles : *mensuellement, au sommeil nocture, pendant la journée, en fin de nuit.*

   4 **periode_vie** : récupère les périodes propres à la vie d'une personne : *dans la jeunesse, pendant l'enfance.*

   5 **saisons** : récupère les saisons.

Annexe 3 - Présentation des graphes de contextualisation des dates

Nous avons créé 74 graphes pour la contextualisation des dates. Nous en présentons ci-dessous les principaux.

1. **timed_procedure** : permet de contextualiser une date avec une *procedure* : *elle a passé ce matin une radiographie.*
2. **timed_substance** : permet de contextualiser une date avec une *substance* : *il fume à raison de 15 cigarettes/j.*
3. **timed_function** : permet de contextualiser une date avec une *function* : *grossesse de 18 sem.*
4. **timed_disorder** : permet de contextualiser une date avec un *disorder* : *atteinte d'une grippe depuis 3 jours.*
5. **timed_finding** : permet de contextualiser une date avec un *finding* : *douleurs au ventre depuis 10 jours.*
6. **timed_drug** : permet de contextualiser une date avec un médicament (*drug*) : *aspirine 2x/j.*
7. **timed_finding_procedure** : permet de contextualiser une date avec un *finding* et une *procedure* : *Il a passé un scanner ce matin suite à une chute.*
8. **timed_disorder_procedure** : permet de contextualiser une date avec un *disorder* et une *procedure* : *Elle a passé ce jour un examen des suite de sa fausse-couche.*
9. **timed_bodypart** : permet de contextualiser une date avec une partie du corps (*bodypart*) : *mal de ventre depuis 1 semaine.*

10. **time** : permet de récupérer les expressions annotées comme temporelles dans le corpus, mais qui pourraient poser problème si extraites dans la première grammaire [12] : *traitement à la sortie, traitement actuel*.

12. Si extraites avant, elles pourraient être contextualisée, ce qui causerait des erreurs.

Annexe 4 - Mode d'emploi de Time extractor

Tous les fichiers nécessaires au bon fonctionnement de time-extractor se trouvent dans le dossier *server*. Ce dossier est ainsi divisé :

- server
  - html
    - js
      - javascript.js
      - jquery-3.2.1.js
    - style
      - logo_earlytracks.png
      - time-extractor.css
    - time-extractor-webpage.html
  - ressources
    - App
    - contextualisation_dates
      - src
        - Dela
          - les dictionnaires Unitex pour la contextualisation de dates
        - Graphs
          - les graphes de contextualisation de dates (cf. Annexe B pour la liste)
    - dates
      - src
        - Dela
          - les dictionnaires Unitex
        - Graphs
          - les graphes d'extraction des dates (cf. Annexe A pour la liste)
    - French
  - services
    - _init_.py
    - Normalization.py
    - TimeExtractor.py
    - Unitex.py
  - utils
    - _init_.py
    - server_utils.py
  - lauch.sh
  - time-extractor-server

TABLE 11.2 – Division du dossier *server*, permettant le lancement du serveur *time-extractor*.

Le dossier *html* comprend le code de la page html, son css et deux fichiers Javascript ; *javascript.js* permet l'interaction entre la page html et le serveur, tandis que *jquery-3.2.1.js* permet l'utilisation de la bibliothèque JQuery.

Le dossier *ressources* comprend les fichiers *.fst2* des graphes Unitex et les fichiers *.bin* des dictionnaires médicaux utilisés dans la contextualisation de dates.

Les dossiers *services* et *utils* comprennent les scripts python gérant l'appel au serveur, les deux appels par Unitex, la normalisation des dates, etc.

Enfin, le fichier *launch.sh* lance le serveur time-extractor (défini dans le fichier *time-extractor-server*).

Tous les fichiers nécessaires au bon fonctionnement de time-extractor sont dans ce dossier, mais il est important de modifier le chemin vers ledit dossier dans le script Unitex.py avant de lancer le serveur ! De plus, vous devez installer Unitex [13] et prendre le dossier *App* de votre dossier Unitex et remplacer celui se trouvant déjà dans le dossier *ressources*.

Une fois les chemins correctement modifiés dans *Unitex.py* et le dossier *App* correctement modifié, il suffit de lancer le serveur [14], puis d'ouvrir la page html *time-extractor-webpage.html*, d'insérer un texte et sa date de rédaction (sous le format AAAA-MM-JJ HH-MM-SS, ou année-mois-jour heures-minutes-secondes en chiffres), un texte dans le deuxième cadre et enfin, d'appuyer sur le bouton *lancer*.

13. Disponible à l'adresse `http://unitexgramlab.org/`.
14. Avec la commande « ./launch.sh » dans une invite de commande.

Annexe 5 - Les sorties des graphes Unitex

Cette annexe reprend toutes les sorties des graphes d'extraction d'expressions temporelles et de contextualisation des paires événements médicaux - timex [15].

Sorties d'extraction

AGE :

{AGE,AGE.Time+vie_moyenne+values=vie+ANS}

{AGE,AGE.Time+âge+format=années+values=ANS}

{AGE,AGE.Time+âge+format=années_mois+values=ANS+$MOIS$}

{AGE,AGE.Time+âge+format=années_mois_jours+values=ANS+$MOIS$+$JOURS$}

{AGE,AGE.Time+âge+format=années_jours+values=ANS+$JOURS$}

{AGE,AGE.Time+âge+format=mois+values=$MOIS$}

{AGE,AGE.Time+âge+format=mois_jours+values=$MOIS$+$JOURS$}

{AGE,AGE.Time+âge+format=jours+values=$JOURS$}

15. Ces sorties sont également disponibles dans le fichier texte *sorties_graphes.txt* dans les annexes électroniques.

PASSE_MAIN :

{$PASSE$,$PASSE$.Time+Date+format=anneePasse+values=$PASSE$}

{$PASSE$,$PASSE$.Time+Date+format=périodePasse+values=$PASSE$}

PRESENT_MAIN :

{$PRESENT$,$PRESENT$.Time+Date+format=actuellement+values=$PRESENT$}

{$PRESENT$,$PRESENT$.Time+Date+format=periodePresent+values=$PRESENT$}

FUTUR_MAIN :

{$FUTUR$,$FUTUR$.Time+Date+format=periodeFutur+values=$FUTUR$}

JOUR_MOIS_PPP :

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM01+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM01+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM01+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM01+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM02+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM02+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM02+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM02+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM03+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM03+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM03+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM03+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM04+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM04+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM04+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM04+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM05+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM05+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM05+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM05+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM06+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM06+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM06+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM06+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM07+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM07+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM07+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM07+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM08+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM08+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM08+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM08+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM09+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM09+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM09+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM09+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM10+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM10+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM10+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM010+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM11+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM11+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM11+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM11+futur}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_undef+values=$JOUR$+MM12+undef}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_passé+values=$JOUR$+MM12+passé}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_présent+values=$JOUR$+MM12+présent}

{$DATE$,$DATE$.Time+Date+format=JJ_MM_futur+values=$JOUR$+MM12+futur}

DMF_ANNEE :

{deb,deb.Time+Date+format=débutAAAA+values=MM01+$ANNEE$}

{mi,mi.Time+Date+format=miAAAA+values=MM06+$ANNEE$}

{fin,fin.Time+Date+format=finAAAA+values=MM12+$ANNEE$}

DMF_MOIS :

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM01}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM01}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM01}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM02}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM02}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM02}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM03}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM03}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM03}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM04}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM04}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM04}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM05}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM05}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM05}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM06}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM06}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM06}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM07}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM07}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM07}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM08}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM08}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM08}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM09}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM09}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM09}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM10}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM10}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM10}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM11}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM11}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM11}

{$DATE$,$DATE$.Time+Date+format=débutMM+values=5+MM12}

{$DATE$,$DATE$.Time+Date+format=miMM+values=15+MM12}

{$DATE$,$DATE$.Time+Date+format=finMM+values=25+MM12}

FORMAT JOUR-MOIS-ANNEE :

{$DATE$,$DATE$.Time+Date+format=mois_AAAA+values=$Mois$+$Annee$}

{$DATE$,$DATE$.Time+Date+format=JJ_mois_AAAA+values=$Jour$+$Mois$+$Annee$}

{$DATE$,$DATE$.Time+Date+format=JJ_mois+values=$Jour$+$Mois$}

{$DATE$,$DATE$.Time+Date+format=JJ-MM-AAAA+values=$Jour$-$Mois$-$Annee$}

{$DATE$,$DATE$.Time+Date+format=JJ-MM-AAAA+values=$Jour$-$Mois$-$Annee$}

{$DATE$,$DATE$.Time+Date+format=JJ-MM-AAAA+values=$Jour$-$Mois$-$Annee$}

{$DATE$,$DATE$.Time+Date+format=JJ-MM-AAAA+values=$Jour$-$Mois$-$Annee$}

{$DATE$,$DATE$.Time+Date_exam+format=JJ-MM-AAAA+values=$Jour$-$Mois$-$Annee$}

$Jour$$Mois$$Annee$,$Jour$$Mois$$Annee$.Time+Date_naissance+format=JJ-MM-AAAA+values=$Jour$-$Mois$-$Annee$}

JOUR_SEMAINE :

{$DATE$,$DATE$.Time+format=JSEM_undef+values=LUNDI+undef}

{$DATE$,$DATE$.Time+format=JSEM_passé+values=LUNDI+passé}

{$DATE$,$DATE$.Time+format=JSEM_présent+values=LUNDI+présent}

{$DATE$,$DATE$.Time+format=JSEM_futur+values=LUNDI+futur}

{$DATE$,$DATE$.Time+format=JSEM_undef+values=MARDI+undef}

{$DATE$,$DATE$.Time+format=JSEM_passé+values=MARDI+passé}

{$DATE$,$DATE$.Time+format=JSEM_présent+values=MARDI+présent}

{$DATE$,$DATE$.Time+format=JSEM_futur+values=MARDI+futur}

{$DATE$,$DATE$.Time+format=JSEM_undef+values=MERCREDI+undef}

{$DATE$,$DATE$.Time+format=JSEM_passé+values=MERCREDI+passé}

{$DATE$,$DATE$.Time+format=JSEM_présent+values=MERCREDI+présent}

{$DATE$,$DATE$.Time+format=JSEM_futur+values=MERCREDI+futur}

{$DATE$,$DATE$.Time+format=JSEM_undef+values=JEUDI+undef}

{$DATE$,$DATE$.Time+format=JSEM_passé+values=JEUDI+passé}

{$DATE$,$DATE$.Time+format=JSEM_présent+values=JEUDI+présent}

{$DATE$,$DATE$.Time+format=JSEM_futur+values=JEUDI+futur}

{$DATE$,$DATE$.Time+format=JSEM_undef+values=VENDREDI+undef}

{$DATE$,$DATE$.Time+format=JSEM_passé+values=VENDREDI+passé}

{$DATE$,$DATE$.Time+format=JSEM_présent+values=VENDREDI+présent}

{$DATE$,$DATE$.Time+format=JSEM_futur+values=VENDREDI+futur}

{$DATE$,$DATE$.Time+format=JSEM_undef+values=SAMEDI+undef}

{$DATE$,$DATE$.Time+format=JSEM_passé+values=SAMEDI+passé}

{$DATE$,$DATE$.Time+format=JSEM_présent+values=SAMEDI+présent}

{$DATE$,$DATE$.Time+format=JSEM_futur+values=SAMEDI+futur}

{$DATE$,$DATE$.Time+format=JSEM_undef+values=DIMANCHE+undef}

{$DATE$,$DATE$.Time+format=JSEM_passé+values=DIMANCHE+passé}

{$DATE$,$DATE$.Time+format=JSEM_présent+values=DIMANCHE+présent}

{$DATE$,$DATE$.Time+format=JSEM_futur+values=DIMANCHE+futur}

MOIS_SEUL :

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM01+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM01+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM01+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM01+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM02+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM02+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM02+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM02+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM03+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM03+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM03+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM03+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM04+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM04+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM04+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM04+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM05+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM05+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM05+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM05+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM06+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM06+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM06+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM06+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM07+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM07+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM07+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM07+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM08+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM08+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM08+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM08+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM09+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM09+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM09+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM09+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM10+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM10+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM10+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM10+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM11+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM11+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM11+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM11+futur}

{$DATE$,$DATE$.Time+Date+format=MM_undef+values=MM12+undef}

{$DATE$,$DATE$.Time+Date+format=MM_passé+values=MM12+passé}

{$DATE$,$DATE$.Time+Date+format=MM_présent+values=MM12+présent}

{$DATE$,$DATE$.Time+Date+format=MM_futur+values=MM12+futur}

FORMAT JSEM_JOUR :

{$DATE$,$DATE$.TIME+DATE+format=JSEM_JJ+values=LUNDI+$JOUR$}

{$DATE$,$DATE$.TIME+DATE+format=JSEM_JJ+values=MARDI+$JOUR$}

{$DATE$,$DATE$.TIME+DATE+format=JSEM_JJ+values=MERCREDI+$JOUR$}

{$DATE$,$DATE$.TIME+DATE+format=JSEM_JJ+values=JEUDI+$JOUR$}

{$DATE$,$DATE$.TIME+DATE+format=JSEM_JJ+values=VENDREDI+$JOUR$}

{$DATE$,$DATE$.TIME+DATE+format=JSEM_JJ+values=SAMEDI+$JOUR$}

{$DATE$,$DATE$.TIME+DATE+format=JSEM_JJ+values=DIMANCHE+$JOUR$}

FORMAT MOIS-ANNEE :

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM01+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM02+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM03+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM04+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM05+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM06+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM07+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM08+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM09+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM10+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM11+$ANNEE$}

{$DATE$,$DATE$.Time+Date+format=prep_MM_AAAA+values=$prep$+MM12+$ANNEE$}

FORMAT ANNEE :

{$DATE$,$DATE$.Time+Date+format=prep_AAAA+values=$prep$+$ANNEE$}

FORMAT LUNDI-JOUR-MOIS-ANNEE :

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM01_AAAA+values=LUNDI+$JOUR$
+MM01+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM02_AAAA+values=LUNDI+$JOUR$
+MM02+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM03_AAAA+values=LUNDI+$JOUR$
+MM03+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM04_AAAA+values=LUNDI+$JOUR$
+MM04+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM05_AAAA+values=LUNDI+$JOUR$
+MM05+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM06_AAAA+values=LUNDI+$JOUR$
+MM06+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM07_AAAA+values=LUNDI+$JOUR$
+MM07+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM08_AAAA+values=LUNDI+$JOUR$
+MM08+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM09_AAAA+values=LUNDI+$JOUR$
+MM09+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM10_AAAA+values=LUNDI+$JOUR$
+MM10+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM11_AAAA+values=LUNDI+$JOUR$
+MM11+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=LUNDI_JJ_MM12_AAAA+values=LUNDI+$JOUR$
+MM12+$ANNEE$}

FORMAT MARDI-JOUR-MOIS-ANNEE :

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM01_AAAA+values=MARDI+$JOUR$
+MM01+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM02_AAAA+values=MARDI+$JOUR$
+MM02+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM03_AAAA+values=MARDI+$JOUR$
+MM03+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM04_AAAA+values=MARDI+$JOUR$
+MM04+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM05_AAAA+values=MARDI+$JOUR$
+MM05+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM06_AAAA+values=MARDI+$JOUR$
+MM06+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM07_AAAA+values=MARDI+$JOUR$
+MM07+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM08_AAAA+values=MARDI+$JOUR$
+MM08+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM09_AAAA+values=MARDI+$JOUR$
+MM09+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM10_AAAA+values=MARDI+$JOUR$
+MM10+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM11_AAAA+values=MARDI+$JOUR$
+MM11+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MARDI_JJ_MM12_AAAA+values=MARDI+$JOUR$
+MM12+$ANNEE$}

FORMAT MERCREDI-JOUR-MOIS-ANNEE :

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM01_AAAA+values=MERCREDI+
$JOUR$+MM01+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM02_AAAA+values=MERCREDI+
$JOUR$+MM02+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM03_AAAA+values=MERCREDI+
$JOUR$+MM03+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM04_AAAA+values=MERCREDI+
$JOUR$+MM04+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM05_AAAA+values=MERCREDI+
$JOUR$+MM05+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM06_AAAA+values=MERCREDI+
$JOUR$+MM06+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM07_AAAA+values=MERCREDI+
$JOUR$+MM07+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM08_AAAA+values=MERCREDI+
$JOUR$+MM08+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM09_AAAA+values=MERCREDI+
$JOUR$+MM09+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM10_AAAA+values=MERCREDI+
$JOUR$+MM10+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM11_AAAA+values=MERCREDI+$JOUR$+MM11+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=MERCREDI_JJ_MM12_AAAA+values=MERCREDI+$JOUR$+MM12+$ANNEE$}

FORMAT JEUDI-JOUR-MOIS-ANNEE :

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM01_AAAA+values=JEUDI+$JOUR$+MM01+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM02_AAAA+values=JEUDI+$JOUR$+MM02+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM03_AAAA+values=JEUDI+$JOUR$+MM03+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM04_AAAA+values=JEUDI+$JOUR$+MM04+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM05_AAAA+values=JEUDI+$JOUR$+MM05+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM06_AAAA+values=JEUDI+$JOUR$+MM06+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM07_AAAA+values=JEUDI+$JOUR$+MM07+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM08_AAAA+values=JEUDI+$JOUR$+MM08+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM09_AAAA+values=JEUDI+$JOUR$+MM09+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM10_AAAA+values=JEUDI+$JOUR$+MM10+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM11_AAAA+values=JEUDI+$JOUR$+MM11+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=JEUDI_JJ_MM12_AAAA+values=JEUDI+$JOUR$+MM12+$ANNEE$}

FORMAT VENDREDI-JOUR-MOIS-ANNEE :

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM01_AAAA+values=VENDREDI+$JOUR$+MM01+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM02_AAAA+values=VENDREDI+$JOUR$+MM02+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM03_AAAA+values=VENDREDI+$JOUR$+MM03+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM04_AAAA+values=VENDREDI+$JOUR$+MM04+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM05_AAAA+values=VENDREDI+$JOUR$+MM05+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM06_AAAA+values=VENDREDI+$JOUR$+MM06+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM07_AAAA+values=VENDREDI+$JOUR$+MM07+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM08_AAAA+values=VENDREDI+$JOUR$+MM08+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM09_AAAA+values=VENDREDI+$JOUR$+MM09+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM10_AAAA+values=VENDREDI+$JOUR$+MM10+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM11_AAAA+values=VENDREDI+$JOUR$+MM11+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=VENDREDI_JJ_MM12_AAAA+values=VENDREDI+$JOUR$+MM12+$ANNEE$}

FORMAT SAMEDI-JOUR-MOIS-ANNEE :

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM01_AAAA+values=SAMEDI+$JOUR$+MM01+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM02_AAAA+values=SAMEDI+$JOUR$+MM02+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM03_AAAA+values=SAMEDI+$JOUR$+MM03+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM04_AAAA+values=SAMEDI+$JOUR$+MM04+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM05_AAAA+values=SAMEDI+$JOUR$+MM05+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM06_AAAA+values=SAMEDI+$JOUR$+MM06+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM07_AAAA+values=SAMEDI+$JOUR$+MM07+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM08_AAAA+values=SAMEDI+$JOUR$+MM08+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM09_AAAA+values=SAMEDI+$JOUR$+MM09+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM10_AAAA+values=SAMEDI+$JOUR$+MM10+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM11_AAAA+values=SAMEDI+$JOUR$+MM11+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=SAMEDI_JJ_MM12_AAAA+values=SAMEDI+$JOUR$
+MM12+$ANNEE$}

FORMAT DIMANCHE-JOUR-MOIS-ANNEE :

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM01_AAAA+values=DIMANCHE+
$JOUR$+MM01+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM02_AAAA+values=DIMANCHE+
$JOUR$+MM02+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM03_AAAA+values=DIMANCHE+
$JOUR$+MM03+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM04_AAAA+values=DIMANCHE+
$JOUR$+MM04+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM05_AAAA+values=DIMANCHE+
$JOUR$+MM05+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM06_AAAA+values=DIMANCHE+
$JOUR$+MM06+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM07_AAAA+values=DIMANCHE+
$JOUR$+MM07+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM08_AAAA+values=DIMANCHE+
$JOUR$+MM08+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM09_AAAA+values=DIMANCHE+
$JOUR$+MM09+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM10_AAAA+values=DIMANCHE+
$JOUR$+MM10+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM11_AAAA+values=DIMANCHE+
$JOUR$+MM11+$ANNEE$}

{$DATE$,$DATE$.Time+DATE+format=DIMANCHE_JJ_MM12_AAAA+values=DIMANCHE+
$JOUR$+MM12+$ANNEE$}

RECURSION ANNEE :

{$DATE$,$DATE$.Time+Date+format=recursion_AAAA+values=$prep$+$ANNEE1$+$conj$+
$ANNEE2$}

{$DATE$,$DATE$.Time+Date+format=recursion_3xAAAA+values=$prep$+$ANNEE1$+$conj$+
$ANNEE2$+$conj2$+$ANNEE3$}

{$DATE$,$DATE$.Time+Date+format=recursion_4xAAAA+values=$prep$+$ANNEE1$+$conj$+
$ANNEE2$+$conj2$+$ANNEE3$+$conj3$+$ANNEE4$}

DATE_PERIODE :

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=JJauJJ-MM+values=$Jour1$
+au+$Jour2$-$Mois$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=JJ-MMauJJ-MM+values=
$Jour1$-$Mois1$+au+$Jour2$-$Mois2$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=JJ-MMauJJ-MM-AAAA+values=
$Jour1$-$Mois1$+au+$Jour2$-$Mois2$-$Annee$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=JJ-MM-AAAAauJJ-MM-AAAA
+values=$Jour1$-$Mois1$-$Annee1$+au+$Jour2$-$Mois2$-$Annee2$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=JJauJJ-MM-AA+values=
$Jour1$+au+$Jour2$-$Mois$-$Annee$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=LETTREJJauJJ-MM+values=
$Jour1$+au+$Jour2$-$Mois$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=LETTREJJ-MMauJJ-MM+values=
$Jour1$-$Mois1$+au+$Jour2$-$Mois2$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=LETTREJJ-MMauJJ-MM-AAAA
+values=$Jour1$-$Mois1$+au+$Jour2$-$Mois2$-$Annee$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=LETTREJJ-MM-AAAAau
JJ-MM-AAAA+values=$Jour1$-$Mois1$-$Annee1$+au+$Jour2$-$Mois2$-$Annee2$}

{$DATE_PERIODE$,$DATE_PERIODE$.Time+Date-période+format=LETTREJJauJJ-MM-AA+values=
$Jour1$+au+$Jour2$-$Mois$-$Annee$}

HEURE_MAIN :

{$HEURE$,$HEURE$.Time+Heure+format=hh+values=h}

{$HEURE$,$HEURE$.Time+Heure+format=hh_mm+values=h+m}

{$HEURE$,$HEURE$.Time+Heure+format=hh :mm+values=h :m}

{H_EXAM,H_EXAM.Time+Heure+format=hhmmss+values=h :m :s}

{$nieme_heure$,$nieme_heure$.Time+nièmeMoment+format=undef+values=$nieme_heure$}

{$DUREE_MIN$,$DUREE_MIN$.Time+Durée_minutes+format=mm'-minutes+values=m}

{$HEURE1$ $HEURE2$,$HEURE1$ $HEURE2$.Time+HeureRécursion+format= de_hh_mm_à_hh_mm
+values=de+$h1$:$m1$+à+$h2$:$m2$}

FORMAT LUNDI-JOUR-MOIS-ANNEE-HEURE :

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM01_AAAA_
à_HHhMMm+values=LUNDI+$JOUR$+MM01+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM02_AAAA_
à_HHhMMm+values=LUNDI+$JOUR$+MM02+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM03_AAAA_
à_HHhMMm+values=LUNDI+$JOUR$+MM03+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM04_AAAA_
à_HHhMMm+values=LUNDI+$JOUR$+MM04+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM05_AAAA_à_HHhMMm+values=LUNDI+$JOUR$+MM05+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM06_AAAA_à_HHhMMm+values=LUNDI+$JOUR$+MM06+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM07_AAAA_à_HHhMMm+values=LUNDI+$JOUR$+MM07+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM08_AAAA_à_HHhMMm+values=LUNDI+$JOUR$+MM08+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM09_AAAA_à_HHhMMm+values=LUNDI+$JOUR$+MM09+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM10_AAAA_à_HHhMMm+values=LUNDI+$JOUR$+MM10+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM11_AAAA_à_HHhMMm+values=LUNDI+$JOUR$+MM11+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=LUNDI_JJ_MM12_AAAA_à_HHhMMm+values=LUNDI+$JOUR$+MM12+$ANNEE$+h :m}

FORMAT MARDI-JOUR-MOIS-ANNEE-HEURE :

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM01_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM01+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM02_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM02+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM03_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM03+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM04_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM04+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM05_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM05+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM06_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM06+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM07_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM07+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM08_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM08+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM09_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM09+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM10_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM10+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM11_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM11+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MARDI_JJ_MM12_AAAA_à_HHhMMm+values=MARDI+$JOUR$+MM12+$ANNEE$+h :m}

FORMAT MERCREDI-JOUR-MOIS-ANNEE-HEURE :

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM01_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM01+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM02_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM02+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM03_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM03+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM04_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM04+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM05_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM05+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM06_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM06+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM07_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM07+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM08_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM08+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM09_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM09+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM10_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM10+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM11_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM11+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=MERCREDI_JJ_MM12_AAAA_à_HHhMMm+values=MERCREDI+$JOUR$+MM12+$ANNEE$+h :m}

FORMAT JEUDI-JOUR-MOIS-ANNEE-HEURE :

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM01_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM01+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM02_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM02+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM03_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM03+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM04_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM04+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM05_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM05+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM06_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM06+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM07_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM07+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM08_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM08+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM09_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM09+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM10_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM10+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM11_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM11+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=JEUDI_JJ_MM12_AAAA_à_HHhMMm+values=JEUDI+$JOUR$+MM12+$ANNEE$+h :m}

FORMAT VENDREDI-JOUR-MOIS-ANNEE-HEURE :

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM01_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM01+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM02_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM02+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM03_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM03+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM04_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM04+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM05_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM05+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM06_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM06+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM07_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM07+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM08_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM08+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM09_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM09+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM10_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM10+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM11_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM11+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=VENDREDI_JJ_MM12_AAAA_à_HHhMMm+values=VENDREDI+$JOUR$+MM12+$ANNEE$+h :m}

FORMAT SAMEDI-JOUR-MOIS-ANNEE-HEURE :

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM01_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM01+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM02_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM02+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM03_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM03+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM04_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM04+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM05_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM05+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM06_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM06+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM07_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM07+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM08_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM08+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM09_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM09+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM10_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM10+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM11_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM11+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=SAMEDI_JJ_MM12_AAAA_à_HHhMMm+values=SAMEDI+$JOUR$+MM12+$ANNEE$+h :m}

FORMAT DIMANCHE-JOUR-MOIS-ANNEE-HEURE :

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM01_AAAA_à_HHhMMm+values=DIMANCHE+$JOUR$+MM01+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM02_AAAA_à_HHhMMm+values=DIMANCHE+$JOUR$+MM02+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM03_AAAA_à_HHhMMm+values=DIMANCHE+$JOUR$+MM03+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM04_AAAA_à_HHhMMm+values=DIMANCHE+$JOUR$+MM04+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM05_AAAA_à_HHhMMm+values=DIMANCHE+$JOUR$+MM05+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM06_AAAA_à_HHhMMm+values=DIMANCHE+$JOUR$+MM06+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM07_AAAA_
à_HHhMMm+values=DIMANCHE+$JOUR$+MM07+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM08_AAAA_
à_HHhMMm+values=DIMANCHE+$JOUR$+MM08+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM09_AAAA_
à_HHhMMm+values=DIMANCHE+$JOUR$+MM09+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM10_AAAA_
à_HHhMMm+values=DIMANCHE+$JOUR$+MM10+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM11_AAAA_
à_HHhMMm+values=DIMANCHE+$JOUR$+MM11+$ANNEE$+h :m}

{$DATE$ $HEURE$,$DATE$ $HEURE$.TIME+DATE-HEURE+format=DIMANCHE_JJ_MM12_AAAA_
à_HHhMMm+values=DIMANCHE+$JOUR$+MM12+$ANNEE$+h :m}

TEMPS_TRAIT :

{$TRAITEMENT$,$TRAITEMENT$.Time+Traitement+format=période+values=
$PERIODE_TRAITEMENT$}

{$TRAITEMENT$,$TRAITEMENT$.Time+Traitement+format=période_fréquence+values=
$FREQ_TRAITEMENT$+$PERIODE_TRAITEMENT$}

{$TRAITEMENT$,$TRAITEMENT$.Time+Traitement+format=fréquence+values=
$FREQ_TRAITEMENT$}

{$JOUR_TRAITEMENT$,$JOUR_TRAITEMENT$.TIME+Traitement+format=jour+values=
$JOUR_TRAITEMENT$}

{$H_TRAITEMENT$,$H_TRAITEMENT$.TIME+Traitement+format=heure+values=$H_TRAITEMENT$}

{$JOUR_TRAITEMENT$,$JOUR_TRAITEMENT$.TIME+Traitement+format=jour-jour+values=de-a}

{$H_TRAITEMENT$,$H_TRAITEMENT$.TIME+Traitement+format=heure-heure+values=de-a}

{$TRAITEMENT$,$TRAITEMENT$.Time+Traitement+format=période_fréquence+values=pre
/$post$+$PERIODE_TRAITEMENT$}

{$TRAITEMENT$,$TRAITEMENT$.Time+Traitement+format=période_fréquence+values=$predemi$1
/2$postdemi$+$PERIODE_TRAITEMENT$}

{$TRAITEMENT$,$TRAITEMENT$.Time+Traitement+format=fréquence+values=pre
/$post$}

{$TRAITEMENT$,$TRAITEMENT$.Time+Traitement+format=fréquence+values=$predemi$1
/2$postdemi$}

GYNECO :

{$AGE_GESTATIONNEL$,$AGE_GESTATIONNEL$.Time+Gynécologie+format=
semaineGesta_jourGesta+values=$semaines_gestationnel$+$jours_gestationnel$}

{$AGE_GESTATIONNEL$,$AGE_GESTATIONNEL$.Time+Gynécologie+format=
semaineGesta+values=$semaines_gestationnel$}

{$AGE_GESTATIONNEL$,$AGE_GESTATIONNEL$.Time+Gynécologie+format=
semaineGesta+values=$semaines_gestationnel$}

{$POST_GROSSESSE$,$POST_GROSSESSE$.Time+Gynécologie+format=
postGrossesse+values=$POST_GROSSESSE$}

{$ANTECEDENT_GYNECO$,$ANTECEDENT_GYNECO$.Time+Gynécologie+format=
antécédentsGynécologiques+values=$ANTECEDENT_GYNECO$}

{$TERME$,$TERME$.Time+Gynécologie+format=terme-semaineGesta+jourGesta+values=
$semaines_gestationnel$+$jours_gestationnel$}

{$TERME$,$TERME$.Time+Gynécologie+format=terme-semaineGesta+values=$semaines_gestationnel$}

HOSPITALLISATION :

{$HOSPITALISATION$,$HOSPITALISATION$.Time+Hospitalisation+format=PREMED_PASSE
+values=$DATE_HOSPI$+PRE_MED}

{$HOSPITALISATION$,$HOSPITALISATION$.Time+Hospitalisation+format=PRESMED_PASSE
+values=$DATE_HOSPI$+$PRES_MED$}

{$HOSPITALISATION$,$HOSPITALISATION$.Time+Hospitalisation+format=POSTMED_PASSE
+values=$DATE_HOSPI$+$POST_MED$}

RDV_POSTOP :

{RDV_POSTOP,RDV_POSTOP.Time+RDVpost-op+format=jour+values=$JOUR$}

{RDV_POSTOP,RDV_POSTOP.Time+RDVpost-op+format=jour-semaine+values=
$JOUR$-$SEMAINE$}

{RDV_POSTOP,RDV_POSTOP.Time+RDVpost-op+format=jour-semaine-mois+values=
$JOUR$-$SEMAINE$-$MOIS$}

{RDV_POSTOP,RDV_POSTOP.Time+RDVpost-op+format=jour-semaine-mois-année+values=
$JOUR$-$SEMAINE$-$MOIS$-$ANNEE$}

EXPRESSIONS_MEDICALES :

{$expr_med$,$expr_med$.Time+expressionMedicale+values=$expr_med$}

Sorties de contextualisation

{<EVENT type = "TIMED_PROCEDURE"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$
<ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY></EVENT>

{<EVENT type = "TIMED_PROCEDURE"> <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY>
$word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "TIMED_PROCEDURE"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$
<ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY></EVENT>

{<EVENT type = "TIMED_SUBSTANCE"> <ENTITY type = "TIME">$TIME$</ENTITY> <ENTITY type = "SUBSTANCE">$SUBSTANCE$</ENTITY></EVENT>

{<EVENT type = "TIMED_SUBSTANCE"> <ENTITY type = "SUBSTANCE">$SUBSTANCE$</ENTITY> <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "TIMED_FUNCTION"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "FUNCTION">$FUNCTION$</ENTITY></EVENT>

{<EVENT type = "TIMED_FUNCTION"> <ENTITY type = "FUNCTION">$FUNCTION$</ENTITY> $word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "TIMED_FUNCTION"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "FUNCTION">$FUNCTION$</ENTITY></EVENT>

{<EVENT type = "TIMED_CLINICALFINDING-DISORDER"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "CLINICAL DISORDER">$DISORDER$</ENTITY></EVENT>

{<EVENT type = "TIMED_CLINICALFINDING-DISORDER"> <ENTITY type = "CLINICAL DISORDER">$DISORDER$</ENTITY> $word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "TIMED_CLINICALFINDING-DISORDER"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "CLINICAL DISORDER">$DISORDER$</ENTITY></EVENT>

{<EVENT type = "TIMED_CLINICALFINDING-FINDING"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "CLINICAL FINDING">$FINDING$</ENTITY></EVENT>

{<EVENT type = "TIMED_CLINICALFINDING-FINDING"> <ENTITY type = "CLINICAL FINDING">$FINDING$</ENTITY> $word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "TIMED_CLINICALFINDING-FINDING"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "CLINICAL FINDING">$FINDING$</ENTITY></EVENT>

{<EVENT type = "DRUG"> <ENTITY type = "DRUG_NAME">$DRUG$</ENTITY> <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "DRUG"> <ENTITY type = "DRUG_NAME">$DRUG$</ENTITY> <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "DRUG"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "DRUG">$DRUG$</ENTITY></EVENT>

{<EVENT type = "TIMED_FINDING_PROCEDURE"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "CLINICAL FINDING">$FINDING$</ENTITY> $word$ <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY></EVENT>

{<EVENT type = "TIMED_FINDING_PROCEDURE"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY> $word$ <ENTITY type = "CLINICAL FINDING">$FINDING$</ENTITY></EVENT>

{<EVENT type = "TIMED_FINDING_PROCEDURE"> <ENTITY type = "CLINICAL FINDING">$FINDING$</ENTITY> $word$ <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY> $word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "TIMED_FINDING_PROCEDURE"> <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY> $word$ <ENTITY type = "CLINICAL FINDING">$FINDING$</ENTITY> $word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "TIMED_DISORDER_PROCEDURE"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "CLINICAL DISORDER">$DISORDER$</ENTITY> $word$ <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY></EVENT>

{<EVENT type = "TIMED_DISORDER_PROCEDURE"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY> $word$ <ENTITY type = "CLINICAL DISORDER">$DISORDER$</ENTITY></EVENT>

{<EVENT type = "TIMED_DISORDER_PROCEDURE"> <ENTITY type = "CLINICAL DISORDER">$DISORDER$</ENTITY> $word$ <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY> $word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "TIMED_DISORDER_PROCEDURE"> <ENTITY type = "PROCEDURE">$PROCEDURE$</ENTITY> $word$ <ENTITY type = "CLINICAL DISORDER">$DISORDER$</ENTITY> $word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "BODYPART"> <ENTITY type = "BODYPART">$BODYPART$</ENTITY> $word$ <ENTITY type = "TIME">$TIME$</ENTITY></EVENT>

{<EVENT type = "BODYPART"> <ENTITY type = "TIME">$TIME$</ENTITY> $word$ <ENTITY type = "BODYPART">$BODYPART$</ENTITY></EVENT>

{<EVENT type = "TIME"> <ENTITY type = "DRUG">$TIME$</ENTITY></EVENT>

Annexe 6 - Présentation des annexes électroniques

Dans la clé USB ci-jointe, nous proposons une série d'annexes électroniques. Ne seront pas inclus les dictionnaires Unitex utilisés pour la contextualisation des événements médicaux, excepté un petit dictionnaire que nous avons réalisé nous-même comme exemple et pour entraînement.

Dans le dossier ***server*** (dont la division est explicitée à la table 80 du présent document), nous reprenons tous les scripts nécessaires au bon fonctionnement de *time-extractor*. Les graphes Unitex peuvent également être retrouvés dans le dossier *ressources*.

Dans le dossier ***graphes***, se trouvent les graphes de contextualisation (dossier ***contextualisation***), les graphes d'extraction (dossier ***extraction***) et les images des graphes Unitex (dossier ***images***).

Dans le dossier ***autres scripts***, il y a :

- *concordancier.py* : permet de trouver un fichier texte et le fichier *.annotations* correspondant.
- *random_files_selection.py* : prend aléatoirement x fichiers dans un dossier a et les insère dans un dossier b.
- *sous_annot.py* : modifie les annotations des fichiers pour la deuxième couche d'annotations.

Dans le dossier ***images_interface***, nous reprenons les figures 6.1 et 9.1 à 9.3.

sorties_graphes.txt : le fichier de sorties des graphes.

memoire_neuzy_2017.pdf : le fichier pdf du présent mémoire.

Place Blaise Pascal, 1 bte L3.03.11, 1348 Louvain-la-Neuve, Belgique **www.uclouvain.be/fial**