



HAL
open science

Analyse d'expressions temporelles dans les dossiers électroniques patients

Mike Donald Tapi Nzali, Aurélie Névéol, Xavier Tannier

► **To cite this version:**

Mike Donald Tapi Nzali, Aurélie Névéol, Xavier Tannier. Analyse d'expressions temporelles dans les dossiers électroniques patients. TALN: Traitement Automatique des Langues Naturelles, Jun 2015, Caen, France. hal-01222638

HAL Id: hal-01222638

<https://hal.science/hal-01222638>

Submitted on 30 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse d'expressions temporelles dans les dossiers électroniques patients

Mike Donald Tapi Nzali¹ * Aurélie Névéal¹ Xavier Tannier^{2,1}

(1) LIMSI-CNRS, Campus Universitaire d'Orsay, bât 508, 91405 ORSAY, FRANCE

(2) Université Paris-Sud, 91403 ORSAY, FRANCE

prenom.nom@limsi.fr

Résumé. Les références à des phénomènes du monde réel et à leur caractérisation temporelle se retrouvent dans beaucoup de types de discours en langue naturelle. Ainsi, l'analyse temporelle apparaît comme un élément important en traitement automatique de la langue. Cet article présente une analyse de textes en domaine de spécialité du point de vue temporel. En nous appuyant sur un corpus de documents issus de plusieurs dossiers électroniques patient désidentifiés, nous décrivons la construction d'une ressource annotée en expressions temporelles selon la norme TimeML. Par la suite, nous utilisons cette ressource pour évaluer plusieurs méthodes d'extraction automatique d'expressions temporelles adaptées au domaine médical. Notre meilleur système statistique offre une performance de 0,91 de F-mesure, surpassant pour l'identification le système état de l'art HeidelTime. La comparaison de notre corpus de travail avec le corpus journalistique FR-Timebank permet également de caractériser les différences d'utilisation des expressions temporelles dans deux domaines de spécialité.

Abstract.

An analysis of temporal expressions in Electronic Health Records in French

References to phenomena occurring in the world and their temporal characterization can be found in a variety of natural language utterances. For this reason, temporal analysis is a key issue in natural language processing. This article presents a temporal analysis of specialized documents. We use a corpus of documents contained in several de-identified Electronic Health Records to develop an annotated resource of temporal expressions relying on the TimeML standard. We then use this corpus to evaluate several methods for the automatic extraction of temporal expressions. Our best statistical model yields 0.91 F-measure, which provides significant improvement on extraction, over the state-of-the-art system HeidelTime. We also compare our medical corpus to FR-Timebank in order to characterize the uses of temporal expressions in two different subdomains

Mots-clés : Extraction d'Information ; Analyse Temporelle ; Développement d'un Corpus Annoté.

Keywords: Information Extraction, Temporal Analysis, Development of Annotated Corpus.

1 Introduction

1.1 Contexte et motivation

Des informations importantes sur des questions de santé publique se trouvent dans les dossiers électroniques patient rédigés en langue naturelle. Ces dossiers patient constituent un corpus considérable qui croît au rythme de plusieurs centaines de milliers de documents chaque année pour un seul établissement hospitalier. L'analyse rétrospective des parcours de santé permet aux professionnels de santé d'avoir une vue synthétique sur la prise en charge d'une pathologie au sein de leur établissement. Elle permet de comparer les prises en charge effectives avec les recommandations de bonne pratique afin d'évaluer la qualité des parcours de soin, d'identifier des éléments du parcours demandant une prise en charge spécifique, d'améliorer la démarche diagnostique ou thérapeutique. Le but à long terme de ce travail est de développer des méthodes de traitement automatique de la langue biomédicale permettant de faciliter la synthèse d'informations sur les parcours de soin dans le cadre d'analyses rétrospectives. Notre travail porte sur le contenu textuel des dossiers patients, notamment les compte-rendus de séjours et compte-rendus d'actes.

*. Nouvelle affiliation: Université de Montpellier, LIRMM, I3M - Montpellier, France - mike-donald.tapi-nzali@lirmm.fr

Le potentiel des méthodes de traitement automatique de la langue biomédicale (bioNLP) pour exploiter ces documents a été démontré (Demner-Fushman *et al.*, 2009) et fait l’objet de nombreux travaux aussi bien sur l’anglais (par exemple, Friedman *et al.* (1994); Savova *et al.* (2010)) que sur le français (par exemple, Deléger *et al.* (2010); Grouin *et al.* (2011)). Le dossier électronique patient contient plusieurs types de documents (comptes-rendus de séjours, comptes-rendus d’actes, correspondance entre professionnels de santé, ordonnances...) permettant de retracer le parcours du patient dans l’hôpital depuis la première admission, les examens diagnostiques, la prise en charge et le suivi thérapeutique. Un même événement de l’historique médical du patient peut ainsi être mentionné à plusieurs reprises dans divers documents. Il est donc nécessaire de détecter les mentions d’événements ainsi que les mentions d’expressions temporelles afin de pouvoir les relier entre eux et d’agréger les informations associées.

1.2 Travaux antérieurs en analyse temporelle

L’analyse des expressions temporelles dans les textes est une problématique du traitement automatique des langues qui a connu un intérêt grandissant ces dernières années. Les efforts ont d’abord et principalement porté sur les textes journalistiques en langue anglaise, grâce à la création du corpus TimeBank (Pustejovsky *et al.*, 2003), utilisé dans les campagnes d’évaluation TempEval, dont la première édition a eu lieu en 2007 (Verhagen *et al.*, 2007). Plus récemment, ces travaux ont été étendus à d’autres langues (Li *et al.*, 2014; Strötgen *et al.*, 2014a), dont le français (Moriceau & Tannier, 2014), ainsi qu’à d’autres domaines (SMS, textes historiques, résumés d’essais cliniques), montrant des différences intéressantes dans l’extraction et la normalisation des expressions temporelles (Strötgen & Gertz, 2012), ces différences variant d’ailleurs selon la langue étudiée (Strötgen *et al.*, 2014b). Ces constats soulignent le besoin de considérer l’analyse temporelle à la fois sous l’angle du domaine et de la langue considérée.

Le domaine clinique, quant à lui, a été pris en compte à partir de la campagne d’évaluation *i2b2* 2012 (Sun *et al.*, 2013b,a), dans une tâche dédiée à l’extraction de relations temporelles dans des notes cliniques en anglais. Par la suite, d’autres travaux dans ce domaine ont été menés pour les langues anglaise (Jindal & Roth, 2013) et suédoise (Velupillai, 2014), donnant lieu notamment à la création d’un guide détaillé pour la création d’annotations temporelles, ainsi qu’à une étude des spécificités du domaine clinique (Styler IV *et al.*, 2014).

La tâche principale de l’analyse temporelle consiste en l’extraction et la normalisation des expressions temporelles. La normalisation est l’opération de transformation d’une expression (par exemple, “hier” ou “le 1^{er} janvier 2015”) en une représentation formatée et entièrement spécifiée (en particulier, indiquant la valeur absolue des dates relatives ; par exemple, selon le contexte, l’expression “hier” pourrait être normalisée en une date telle que “2015-01-01”). Plusieurs outils réalisant une telle annotation temporelle ont été créés ces dernières années, parmi lesquels on peut citer SUTime (Chang & Manning, 2012), TIMEN (Llorens *et al.*, 2012) pour la langue anglaise, et XIP (Hagège & Tannier, 2008; Bittar & Hagège, 2012) et HeidelTime (Strötgen & Gertz, 2013) pour des adaptations multilingues.

Le travail que nous décrivons ici s’appuie en partie sur l’outil HeidelTime, que nous décrivons donc avec plus de précisions ci-dessous.

1.3 HeidelTime

HeidelTime est un système libre, à base de règles, d’étiquetage d’expressions temporelles, qui a déjà été décliné dans plusieurs langues et, pour l’anglais, dans plusieurs domaines de spécialités (journalistique, scientifique) (Strötgen & Gertz, 2013). Il a notamment obtenu les meilleurs résultats pour l’extraction et la normalisation des expressions temporelles pour l’anglais, dans le contexte des campagnes TempEval-2 et TempEval-3 (Verhagen *et al.*, 2010; UzZaman *et al.*, 2013). HeidelTime est disponible pour 11 langues¹.

HeidelTime produit des annotations dans le format ISO-TimeML (Pustejovsky *et al.*, 2010), norme devenue un standard en la matière, que nous avons également respectée dans les travaux décrits ici. En particulier, cette norme fait la distinction entre quatre catégories d’expressions temporelles : les *dates*, les *heures*, les *durées* et les *fréquences*.

1. <http://code.google.com/p/heideltime/>

1.4 Objectif et contribution

Dans cet article, nous nous intéressons à la phase d'extraction des expressions temporelles dans des dossiers électroniques patient en français. Nous présentons un nouveau corpus clinique annoté en expressions temporelles normalisées (section 2). Ensuite, nous utilisons cette ressource pour évaluer plusieurs méthodes d'extraction automatique d'expressions temporelles adaptées au domaine médical (sections 3 et 4).

La principale contribution scientifique de ce travail est le développement d'une importante ressource annotée pour le domaine biomédical, ainsi que d'outils d'analyse temporelle nous permettant de traiter des textes du domaine et de caractériser le domaine de spécialité de manière globale. Par ailleurs, notre travail offre une contribution méthodologique en ce qui concerne l'adaptation d'outils d'analyse temporelle à un nouveau domaine de spécialité, notamment en quantifiant la charge de travail nécessaire à l'adaptation pour des méthodes symboliques et statistiques, et les résultats que l'on peut en attendre.

2 Développement du corpus clinique annoté pour les expressions temporelles

Pour constituer le corpus de travail, nous avons sélectionné trois dossiers patient complets, issus de patients du service de Néphrologie ayant subi une greffe de rein. Le choix de dossiers complets est guidé par la perspective à long terme d'étudier les parcours de soin, qui nécessite de pouvoir reconstituer l'historique médical complet d'un patient. Le contexte de transplantation rénale est suffisamment complexe pour fournir des exemples de parcours de soin riches du point de vue chronologique.

2.1 Protocole de développement du corpus

Deux dossiers ont été utilisés comme corpus d'entraînement pour les expériences décrites en section 3, et le troisième comme corpus de test. L'ensemble des 361 documents ont été désidentifiés selon le protocole établi par Grouin & Névéol (2014) : douze types d'informations identifiantes sont marqués par deux annotateurs indépendants. Après adjudication, les informations identifiantes sont remplacées automatiquement par des substituts plausibles en veillant à conserver la cohérence et l'aspect naturel des textes.

Le corpus a ensuite été pré-annoté automatiquement à l'aide de l'outil HeidelbergTime, qui a fourni des annotations automatiques pour les expressions temporelles de type *Date*, *Durée*, *Fréquence* et *Heure*. Dans un deuxième temps, ces annotations ont été révisées manuellement par trois annotateurs, qui ont validé les annotations correctement proposées par HeidelbergTime, corrigé les annotations erronées et ajouté les annotations manquantes. Nous avons également annoté les "signaux" (*Signal*), qui selon la norme TimeML sont des prépositions et conjonctions temporelles, ou des caractères spéciaux qui explicitent les relations temporelles entre entités. Nous anticipons une utilité de ces annotations dans une phase ultérieure du travail, à savoir la détection de relations entre événements et expressions temporelles.

Un premier jeu de 20 documents (issus du corpus d'entraînement), utilisé comme jeu d'entraînement pour l'annotation, a été annoté indépendamment par les trois annotateurs (les auteurs de cet article). Les désaccords ont ensuite été résolus lors d'une réunion de consensus. Le reste du corpus a ensuite été divisé de façon à ce que chaque document soit annoté indépendamment par deux annotateurs, les désaccords étant ensuite résolus dans une réunion de consensus. Dans cette deuxième phase, la pré-annotation utilisée a bénéficié de la mise en place de quelques règles spécifiques au domaine clinique dans l'outil HeidelbergTime comme indiqué en section 3.1.

La Figure 1 présente un extrait du corpus annoté selon la norme TimeML.

Pour réaliser l'annotation, nous avons utilisé l'outil libre BRAT rapid annotation tool (BRAT (Stenetorp *et al.*, 2012)). À la différence d'HeidelbergTime, outil dédié à l'annotation automatique d'expressions temporelles, BRAT est un outil générique qui permet de visualiser des annotations existantes et de les modifier manuellement. Selon une revue récente d'outils d'annotations utilisés dans le domaine biomédical, BRAT présente l'avantage d'être facile à installer et à prendre en main et de permettre l'utilisation de pré-annotations (Neves & Leser, 2012) – dans notre travail, les pré-annotations utilisées sont issues du traitement du corpus par HeidelbergTime.

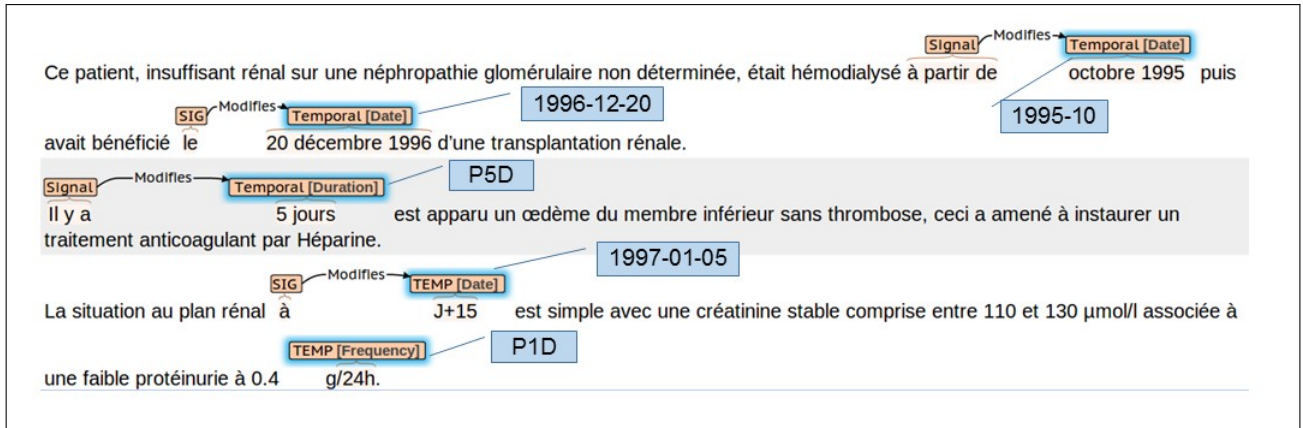


FIGURE 1 – Extrait du corpus clinique annoté en expressions temporelles. Pour cet exemple, nous montrons également la normalisation des expressions qui fait partie du travail en cours (cf section 5). Toutes les informations identifiantes ainsi que les dates ont été remplacées par des substituts plausibles.

2.2 Détails quantitatifs concernant le corpus

La Table 1 présente la distribution des expressions temporelles annotées dans notre corpus de travail, comparée à la distribution des expressions temporelles dans le corpus journalistique FR-TimeBank (Bittar *et al.*, 2011).

Par ailleurs, l'accord inter-annotateur (mesuré en termes de F-mesure) sur le corpus clinique est supérieur à 0,90 pour l'ensemble des paires d'annotateurs. Le temps d'annotation moyen observé est de 25 minutes pour 100 annotations (pour un annotateur), plus 9 minutes consacrées à l'adjudication (lors des réunions de consensus auxquelles participent deux annotateurs). Cela nous permet d'estimer l'annotation en double du corpus complet à environ 45 heures de travail.

	FR-TimeBank		Corpus Clinique	
	#	%	#	%
Date	227	53,41	2594	65,14
Durée	52	12,24	343	8,61
Fréquence	16	3,76	994	24,96
Heure	130	30,59	51	1,28

TABLE 1 – Distribution des expressions temporelles dans deux domaines de spécialité

La comparaison de notre corpus clinique avec le corpus journalistique FR-TimeBank montre qu'il existe une différence de taille considérable entre les deux ressources : 109 documents comprenant un total de 16 208 tokens (soit en moyenne, 148 tokens par document) annotés avec 425 expressions temporelles pour FR-TimeBank et 361 documents comprenant un total de 141 811 tokens (soit en moyenne, 393 tokens par document) annotés avec 3 982 expressions temporelles pour le corpus clinique. Il faut remarquer que la méthode de sélection des documents et la finalité envisagée des deux corpus diffèrent également. Le corpus FR-TimeBank a été construit pour être représentatif du genre journalistique et regroupe sept types d'articles. Le corpus clinique n'a pas cherché à représenter de diversité thématique, mais s'est plutôt attaché à sélectionner des dossiers patients complets afin de permettre à terme une étude des parcours de soin. Par ailleurs, le corpus FR-TimeBank est actuellement beaucoup plus riche que le corpus clinique car il contient des annotations des événements et des relations temporelles en plus des expressions temporelles normalisées.

En terme de distribution des expressions temporelles, on constate d'après la table 1 que les dates et les durées sont présentes dans les deux corpus de manière équivalente. Cependant, la distribution des heures et des fréquences révèle des différences considérables entre les deux domaines de spécialité en présence. En effet, les fréquences sont très présentes dans le corpus clinique (24,96 % vs. 3,76 %) alors que les heures sont prépondérantes dans le corpus journalistique (30,59 % vs. 1,28 %). En pratique, cela s'explique par la nécessité de décrire un événement de l'actualité avec une précision incluant l'horaire de survenue, alors que pour les événements médicaux, une granularité de l'ordre de la journée semble suffisante. En revanche, la prescription de médicaments et de traitements implique d'indiquer la fréquence de prise par le patient, alors que les événements de l'actualité ont lieu ponctuellement.

3 Extraction automatique des expressions temporelles

Nous avons mis en œuvre trois approches pour l'adaptation de l'extraction d'expressions temporelles dans notre corpus de dossiers cliniques patients. La première est un enrichissement manuel des règles de l'outil HeidelbergTime ; la deuxième est une méthode supervisée de prédiction de séquences à base de CRF (champs aléatoires conditionnels) ; enfin, la troisième est une approche hybride utilisant les deux premières.

3.1 Approche symbolique

L'adaptation d'un système symbolique d'extraction d'information temporelle à un nouveau domaine nécessite un réglage de l'outil pour tenir compte de phénomènes nouveaux ou différents (Strötgen & Gertz, 2012). Nous avons ainsi enrichi les règles francophones d'HeidelbergTime pour le traitement d'un certain nombre d'expressions temporelles spécifiques au domaine clinique, telles que *J+1* (qui désigne par exemple le lendemain d'une opération chirurgicale), *J11* (qui désigne par exemple le onzième jour après le début d'un protocole de traitement), des expressions particulières de fréquences ou de durée (les abréviations étant bien plus usitées que dans le domaine journalistique – par exemple, *5 fois/j*). Au total, 14 règles d'extraction et de normalisation ont été ajoutées, et aucune modifiée ou supprimée². Nous estimons le temps de travail pour développer les règles permettant d'adapter HeidelbergTime au domaine biomédical à environ 8 heures.

3.2 Approche supervisée (CRF)

Les champs aléatoires conditionnels (CRF (Lafferty *et al.*, 2001)) linéaires se sont imposés récemment comme l'une des approches les plus efficaces et robustes pour l'étiquetage supervisé de séquences, appliquée avec succès pour des tâches telles que l'étiquetage morphosyntaxique (Lafferty *et al.*, 2001), l'extraction d'entités nommées (McCallum & Li, 2003), l'extraction d'informations structurées (Pinto *et al.*, 2003).

Les CRF sont des modèles probabilistes graphiques non dirigés conçus pour définir une distributions de probabilités conditionnelles sur des séquences d'étiquettes, étant données des séquences observées. Cette nature conditionnelle démarque les CRF des modèles qui nécessitent une hypothèse d'indépendance des variables, tels que les modèles de Markov cachés (HMM). En pratique, une qualité des modèles CRF est leur robustesse sur des ensembles de données de petite taille. Une introduction détaillée aux CRF peut être trouvée dans Sutton & McCallum (2006)

Nous avons créé un modèle CRF pour la détection des expressions temporelles dans les textes, ainsi que pour leur typage (date, durée, fréquence, heure).

3.2.1 Modèle CRF

Les traits utilisés pour l'apprentissage sont d'ordres morphologique, syntaxique et sémantique.

Traits morphologiques

- Capitalisation du token ;
- Longueur du token ;
- Présence d'un chiffre dans le token ;
- Présence de ponctuation dans le token ;

Traits syntaxiques : Étiquettes morphosyntaxiques. Nous avons utilisé TreeTagger (Schmid, 1994) mais nous avons modifié l'étape de segmentation des mots. En effet, pour tenir compte en particulier du grand nombre d'abréviations présentes dans les documents, et de l'absence fréquente d'espaces entre ces abréviations (*5 fois/j*) ou dans les dates (*21/05/05*), les marques de ponctuation ou les chiffres insérés dans les tokens sont considérés comme des séparateurs de mots.

Traits sémantiques :

- Le token lui-même ;

2. Pour un détail sur le fonctionnement des règles, voir Strötgen & Gertz (2013) ou Moriceau & Tannier (2014) pour le français.

- La présence du token dans des listes de déclencheurs construits au préalable (indices temporelles comme les noms de mois ou de jours de la semaine, ainsi qu’unités de mesures) ;
- Identifiant du cluster de Brown associé au token (Liang, 2005).

3.2.2 Représentation des données

Les données sont représentées selon le format tabulaire BIO standard pour les CRF. Chaque token est étiqueté par les traits décrits ci-dessus ainsi que par une classe de la forme *B-Type* (début d’un segment de classe *Type*), *I-Type* (intérieur d’un segment) ou *O* (extérieur à toutes les classes étudiées), où *Type* prendra les valeurs des catégories d’expressions temporelles considérées (*Date*, *Duration*, *Set*, *Time*).

La figure 2 illustre ce format, qui est à la fois l’entrée et la sortie du modèle CRF utilisé.

Token	POSTreeTagger	Length	IsCapitalized	IsPunctuation	IOB tags
à	PRP	1	O	NO_PUNCT	○
partir	VER	6	mm	NO_PUNCT	○
de	PRP	2	mm	NO_PUNCT	B-Signal
octobre	NOM	4	Mm	NO_PUNCT	B-Date
1995	NUM	4	O	NO_PUNCT	I-Date
,	PUN	1	O	PUNCT	○
puis	ADV	4	mm	NO_PUNCT	○
avait	VER	5	mm	NO_PUNCT	○
[...]					
À	PRP	1	MM	NO_PUNCT	B-Signal
J+1	NAM	3	Mm	NO_PUNCT	B-Date

FIGURE 2 – Exemple de fichier tabulaire au format BIO.

3.2.3 Mise en œuvre expérimentale

Nous avons utilisé l’outil Wapiti (Lavergne *et al.*, 2010) pour la mise en œuvre du modèle CRF, avec l’algorithme RPROP (*resilient backpropagation*), qui donne généralement de meilleurs résultats sur ce type de tâches (Grouin, 2013, p. 158). La valeur de régularisation $L1$ a été déterminée à 1,8 par validation croisée 10 fois sur le corpus d’apprentissage. Les résultats présentés à la section 4 sont ceux calculés sur le corpus de test à partir du modèle entraîné finalement sur l’ensemble du corpus d’apprentissage.

3.3 Approche hybride

L’approche hybride consiste à utiliser à la fois les résultats de la méthode symbolique (HeidelTime) et de la méthode statistique. Notre choix a été d’utiliser simplement la sortie d’HeidelTime comme trait supplémentaire dans notre modèle CRF, en utilisant le format BIO décrit ci-dessus.

3.4 Évaluation

Pour chacune des trois méthodes, nous avons utilisé deux dossiers (soit 246 documents) comme corpus d’entraînement, et un dossier (soit 115 documents) comme corpus de test.

Les performances de l’extraction d’expressions temporelles ont été mesurées en terme de précision, rappel et F-mesure pour chaque type d’expression temporelles. Nous avons également calculé la micro-moyenne sur les catégories afin d’obtenir une mesure de performance globale.

4 Résultats

Dans cette section, nous présentons le résultat de nos expérimentations sur l'extraction d'expressions temporelles dans le corpus clinique, puis nous discutons de la question de l'adaptation de méthodes d'extraction d'expressions temporelles en français à un nouveau domaine de spécialité. En raison de la nature sensible du corpus, nous précisons qu'il n'est actuellement pas possible de le diffuser librement. En revanche, le jeu de règles spécifiques ajouté dans HeidelbergTime sera disponible dans la prochaine mise à jour de la distribution du logiciel.

4.1 Extraction automatique d'expressions temporelles

La Table 2 montre les performances globales (micro-moyenne) des différents systèmes évalués sur le corpus d'entraînement (en validation croisée pour les modèles statistiques) et de test.

	Corpus d'entraînement			Corpus de test		
	Précision	Rappel	F-Mesure	Précision	Rappel	F-Mesure
HeidelTime	0,5167	0,6433	0,5731	0,5291	0,6283	0,5744
HeidelTime avec ajout de règles	0,8280	0,8504	0,8391	0,7666	0,7941	0,7801
CRF	0,9458	0,9745	0,9599	0,8068	0,9231	0,8610
CRF avec traits HeidelbergTime	0,9470	0,9789	0,9627	0,8837	0,9403	0,9111

TABLE 2 – Performances globales de l'extraction d'expressions temporelles

Nous constatons que les performances des différentes méthodes augmentent avec la quantité d'information temporelle et d'information spécifique au domaine fournie à l'outil : la méthode symbolique HeidelbergTime standard n'inclut aucune information spécifique au domaine clinique et offre les moins bonnes performances avec une F-mesure de 0,57. Ensuite, HeidelbergTime enrichi de règles spécifiques au domaine biomédical obtenues en analysant le corpus d'entraînement, ainsi que le modèle CRF simple offre de bonnes performances avec des F-mesures de 0,78 et 0,86, respectivement. Les meilleures performances sont obtenues avec le modèle hybride qui utilise des informations spécifiques au domaine *via* le corpus d'entraînement, et des informations temporelles *via* la sortie d'HeidelbergTime. Il est intéressant de remarquer que les performances de cet outil sont comparables à l'accord inter-annotateur pour l'annotation d'expressions temporelles, ce qui indique que la limite supérieure de performance que l'on peut attendre d'un tel outil est atteinte.

Par ailleurs, les performances obtenues sur le corpus d'entraînement se transfèrent avec peu de perte sur le corpus de test, ce qui indique d'une part qu'il n'y a pas de surentraînement des méthodes statistiques, et d'autre part que les corpus d'entraînement et de test sont relativement équilibrés, bien que cet aspect n'ait pas été formellement contrôlé au moment de la construction du corpus (car nous avons privilégié la cohésion des parcours de soin en incluant des dossiers complets).

Les Tables 3 et 4 présentent le détail des résultats de HeidelbergTime (avec les règles spécifiques au domaine clinique) et du modèle CRF hybride sur chaque type d'expression temporelle dans le corpus de test.

	Précision	Rappel	F-mesure
Date	0,9144	0,9334	0,9238
Durée	0,8182	0,8090	0,8136
Fréquence	0,4242	0,8209	0,5593
Heure	0,4688	0,0798	0,1364
Global	0.7666	0.7941	0.7801

TABLE 3 – Performance de l'extraction d'expressions temporelles sur le corpus de test avec l'outil HeidelbergTime adapté pour le domaine clinique

La table 4 indique que les performances du modèle statistique sont inférieures pour la catégorie *Heure* par rapport aux autres catégories. Il est probable que cette différence s'explique par le faible nombre d'expressions temporelles de ce type dans le corpus (seulement 1,28 %). On constate par ailleurs que les performances obtenues par l'outil HeidelbergTime sur les expressions temporelles de type *Date* et *Durée* sont comparables à celles du modèle statistique hybride (0,92 vs. 0,95 pour les dates, 0,81 vs. 0,83 pour les durées). En revanche, pour les *Heure* et les *Fréquence*, qui semblent concentrer l'essentiel des différences d'utilisation des expressions temporelles entre le domaine journalistique et le domaine clinique,

	Précision	Rappel	F-mesure
Date	0,9262	0,9774	0,9511
Durée	0,7500	0,9296	0,8302
Fréquence	0,8535	0,8601	0,8568
Heure	0,3750	0,8571	0,5217
Global	0,8837	0,9403	0,9111

TABLE 4 – Performance de l’extraction d’expressions temporelles sur le corpus de test avec le modèle statistique hybride

les performances d’HeidelTime sont d’au moins trente points inférieures à celle du modèle statistique. Il faut également remarquer que l’accord inter-annotateur est moins élevé sur la catégorie *Heure* que sur les autres. Outre le fait que cette catégorie est relativement peu présente dans le corpus clinique, certaines expressions la représentant sont ambiguës par rapport à la catégorie *Fréquence*. Par exemple, l’expression “le soir” sera considérée comme une *Fréquence* dans (1) mais comme une *Heure* dans (2).

(1) Outre le SKENAN et l’ACTISKENAN, le malade prend du LAROXYL 10 gouttes **le soir**.

(2) Le patient a été admis aux Urgences **le soir du 7 juillet**.

4.2 Adaptation à un nouveau domaine de spécialité

Nos expériences montrent qu’il n’est pas possible d’utiliser directement des outils d’extraction d’expressions temporelles conçus pour le domaine journalistique à un nouveau domaine de spécialité. En cela, notre travail sur le français rejoint les conclusions de Strötgen *et al.* (2014b) sur l’anglais.

Cependant, nous montrons qu’il est possible d’adapter l’extraction d’expression temporelles à un nouveau domaine de spécialité grâce à un effort modéré (charge de travail estimée à 8 heures). En effet, l’ajout de 14 règles à l’outil HeidelTime (qui en comporte déjà 154) suffit à augmenter ses performances sur le domaine clinique de 20 points de F-mesure, et d’obtenir des résultats tout à fait honorables (F-mesure de 0,78). Notons cependant que le corpus n’est représentatif que d’une spécialité médicale et d’un hôpital ; même si les règles ajoutées ne sont pas spécifiques à ces paramètres, il est possible que d’autres pratiques médicales fassent émerger des expressions temporelles de nature différente.

La mise en œuvre de méthodes statistiques demande un effort supplémentaire plus conséquent avec le développement d’un corpus annoté. Il faut néanmoins souligner que les traits utilisés par nos modèles CRF sont entièrement génériques et ne comportent aucune information spécifique au domaine biomédical qui ne soit pas transposable directement à un autre domaine (typiquement, les clusters de Brown peuvent être obtenus à partir d’un corpus de tout domaine de spécialité). Ainsi, nous pouvons faire l’hypothèse que, moyennant la disponibilité d’un corpus d’entraînement, notre modèle statistique est transposable à un autre domaine.

5 Conclusion et perspectives

Ce travail a permis de développer une nouvelle ressource annotée en expressions temporelles pour un domaine de spécialité, le domaine clinique. À partir de cette ressource, nous avons construit un modèle hybride de reconnaissance des expressions temporelles qui offre de très bonnes performances, comparables à l’accord-inter annotateur.

Une perspective à court terme de ce travail est d’étendre l’annotation et la reconnaissance des expressions temporelles à la normalisation des expressions, selon la norme TimeML. L’enrichissement du corpus avec la normalisation des expressions est en cours. Nous souhaitons ensuite tester la généralisabilité des méthodes à d’autres sous-domaines, et nous intéresser à la détection des événements et des relations entre événements et expressions temporelles.

Remerciements

Nous remercions le Service d’Informatique Biomédicale (SIBM) ainsi que l’équipe CISMef du CHU de Rouen qui nous ont permis d’utiliser le corpus LERUDI pour cette étude. Ce travail a bénéficié d’une aide de l’Agence Nationale de la

Recherche portant la référence CABeRneT³ ANR-13-JS02-0009-01.

Références

- BITTAR A., AMSILI P., DENIS P. & DANLOS L. (2011). French timebank : An iso-timeml annotated reference corpus. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies : Short Papers - Volume 2*, HLT '11, p. 130–134, Stroudsburg, PA, USA : Association for Computational Linguistics.
- 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*.
- CHANG A. X. & MANNING C. (2012). SUTime : A library for recognizing and normalizing time expressions. In (LREC2012, 2012).
- DELÉGER L., GROUIN C. & ZWEIGENBAUM P. (2010). Extracting medication information from french clinical texts. In *Stud Health Technol Inform*, volume 160, p. 949–953.
- DEMNER-FUSHMAN D., CHAPMAN W. W. & MCDONALD C. J. (2009). What can natural language processing do for clinical decision support ? *J Biomed Inform*, **42**, 760–772.
- FRIEDMAN C., ALDERSON P., AUSTIN J., CIMINO J. & JOHNSON S. (1994). A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc*, **1**, 161–174.
- GROUIN C. (2013). *Anonymisation de documents cliniques : performances et limites des méthodes symbolique et par apprentissage statistique*. PhD thesis, Université Pierre et Marie Curie (Paris VI).
- GROUIN C., DELÉGER L., ROSIER A., TEMAL L., DAMERON O., VAN HILLE P., BURGUN A. & ZWEIGENBAUM P. (2011). Automatic computation of cha2ds2-vasc score : information extraction from clinical texts for thromboembolism risk assessment. In *AMIA Annu Symp Proc*, p. 501–510.
- GROUIN C. & NÉVÉOL A. (2014). De-identification of clinical notes in french : towards a protocol for reference corpus development. In *J Biomed Inform*.
- HAGÈGE C. & TANNIER X. (2008). XTM : A Robust Temporal Text Processor. In *Computational Linguistics and Intelligent Text Processing, proceedings of 9th International Conference CICLing 2008*, p. 231–240, Haifa, Israel : Springer Berlin / Heidelberg.
- JINDAL P. & ROTH D. (2013). Extraction of events and temporal expressions from clinical narratives. *Journal of Biomedical Informatics (JBI)*.
- LAFFERTY J. D., MCCALLUM A. & PEREIRA F. C. N. (2001). Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, p. 282–289, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.
- LAVERGNE T., CAPPÉ O. & YVON F. C. (2010). Practical very large scale crfs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, p. 504–513, Uppsala, Sweden.
- LI H., STRÖTGEN J., ZELL J. & GERTZ M. (2014). Chinese temporal tagging with heideltime. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2 : Short Papers*, p. 133–137 : Association for Computational Linguistics.
- LIANG P. (2005). *Semi-supervised learning for natural language*. PhD thesis.
- LLORENS H., DERCZYNSKI L., GAIZAUSKAS R. & SAQUETE E. (2012). TIMEN : An Open Temporal Expression Normalisation Resource. In (LREC2012, 2012).
- LREC2012 (2012). *Proceedings of the Eighth International Language Resources and Evaluation (LREC'2012)*, Istanbul, Turkey.
- MCCALLUM A. & LI W. (2003). Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4, CONLL '03*, p. 188–191, Stroudsburg, PA, USA : Association for Computational Linguistics.
- MORICEAU V. & TANNIER X. (2014). French Resources for Extraction and Normalization of Temporal Expressions with HeidelTime. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavík, Iceland.

3. CABeRneT : Compréhension Automatique de Textes Biomédicaux pour la Recherche Translationnelle

- NEVES M. & LESER U. (2012). A survey on annotation tools for the biomedical literature. *Briefings in Bioinformatics*.
- PINTO D., MCCALLUM A., WEI X. & CROFT W. B. (2003). Table extraction using conditional random fields. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, p. 235–242, New York, NY, USA : ACM.
- PUSTEJOVSKY J., HANKS P., SAUR R., SEE A., GAIZAUSKAS R., SETZER A., RADEV D., SUNDHEIM B., DAY D., FERRO L. & LAZO M. (2003). The timebank corpus. *Corpus Linguistics*, p. 647–656.
- PUSTEJOVSKY J., LEE K., BUNT H. & ROMARY L. (2010). ISO-TimeML : An International Standard for Semantic Annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation, (LREC'10)*, p. 394–7, La Valette, Malta.
- SAVOVA G., MASANZ J., OGREN P., ZHENG J., SOHN S., KIPPER-SCHULER K. & CHUTE C. (2010). Mayo clinical text analysis and knowledge extraction system (ctakes) : architecture, component evaluation and applications. *J Am Med Inform Assoc*, **17**, 507–513.
- SCHMID H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *International Conference on New Methods in Language Processing*.
- STENETORP P., PYYSALO S., TOPIĆ G., OHTA T., ANANIADOU S. & TSUJII J. (2012). Brat : A web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, p. 102–107, Stroudsburg, PA, USA : Association for Computational Linguistics.
- STRÖTGEN J., ARMITI A., VAN CANH T., ZELL J. & GERTZ M. (2014a). Time for more languages : Temporal tagging of arabic, italian, spanish, and vietnamese. *ACM Transactions on Asian Language Information Processing (TALIP)*, **13**(1), 1–21.
- STRÖTGEN J., BÖGEL T., ZELL J., ARMITI A., CANH T. V. & GERTZ M. (2014b). Extending heideltime for temporal expressions referring to historic dates. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, p. 2390–2397 : European Language Resources Association (ELRA).
- STRÖTGEN J. & GERTZ M. (2012). Temporal tagging on different domains : Challenges, strategies, and gold standards. In N. C. C. CHAIR), K. CHOUKRI, T. DECLERCK, M. U. DOGAN, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Eds., *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey : European Language Resources Association (ELRA).
- STRÖTGEN J. & GERTZ M. (2013). Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation*, **47**(2), 269–298.
- 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*, p. 143–154.
- SUN W., RUMSHISKY A. & ÖZLEM UZUNER (2013a). Annotating temporal information in clinical narratives. *J Biomed Inform*, **46**, Suppl :S5–12.
- SUN W., RUMSHISKY A. & ÖZLEM UZUNER (2013b). Evaluating temporal relations in clinical text : 2012 i2b2 challenge. *J Am Med Inform Assoc.*, **20**, 806–813.
- SUTTON C. & MCCALLUM A. (2006). An Introduction to Conditional Random Fields for Relational Learning. In L. GETOOT & B. TASKAR, Eds., *Introduction to Statistical Relational Learning*. MIT Press. <http://people.cs.umass.edu/mccallum/papers/crf-tutorial.pdf>.
- UZZAMAN N., LLORENS H., DERZYNSKI L., VERHAGEN M., ALLEN J. & PUSTEJOVSKY J. (2013). Semeval-2013 task 1 : Tempeval-3 : Evaluating time expressions, events, and temporal relations. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)*, p. 1–9 : ACL.
- VELUPILLAI S. (2014). Temporal expressions in swedish medical text – a pilot study. In *Proceedings of BioNLP 2014*, p. 88–92, Baltimore, Maryland : Association for Computational Linguistics.
- VERHAGEN M., GAIZAUSKAS R., SCHILDER F., HEPPLER M. & PUSTEJOVSKY J. (2007). Semeval-2007 task 15 : Tempeval temporal relation identification. In *SemEval-2007 : 4th International Workshop on Semantic Evaluations*.
- VERHAGEN M., SAURI R., CASELLI T. & PUSTEJOVSKY J. (2010). SemEval-2010 - 13 : TempEval-2. In *Proceedings of SemEval workshop at ACL*, Uppsala, Sweden.