HAL open science

Positionnement temporel indépendant des évènements : application à des textes cliniques en français

Nesrine Bannour, Xavier Tannier, Bastien Rance, Aurélie Névéol

To cite this version:

Nesrine Bannour, Xavier Tannier, Bastien Rance, Aurélie Névéol. Positionnement temporel indépendant des évènements : application à des textes cliniques en français. 18e Conférence en Recherche d'Information et Applications – 16e Rencontres Jeunes Chercheurs en RI – 30e Conférence sur le Traitement Automatique des Langues Naturelles – 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues, Jun 2023, Paris, France. pp.1-14. hal-04130201

HAL Id: hal-04130201

https://hal.science/hal-04130201

Submitted on 20 Jun 2023

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.

Distributed under a Creative Commons Attribution 4.0 International License

Actes de CORIA-TALN 2023
Paris, France, 5 au 9 juin 2023

Positionnement temporel indépendant des évènements : application à des textes cliniques en français

Nesrine Bannour[1] Xavier Tannier[2] Bastien Rance[3,4,5] Aurélie Névéol[1]
(1) Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400, Orsay, France
(2) Sorbonne Université, Inserm, Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, Paris, 75006, France
(3) Inserm, Centre de Recherche des Cordeliers, UMRS 1138, Université Paris Cité, Sorbonne Paris Cité, Paris 75006, France
(4) Assitance Publique - Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Paris 75015, France
(5) HeKA, Inria Paris, Paris 75006, France
nesine.bannour@lisn.upsaclay.fr, xavier.tannier@sorbonne-universite.fr, bastien.rance@aphp.fr, aurelie.neveol@lisn.upsaclay.fr

RÉSUMÉ

L'extraction de relations temporelles consiste à identifier et classifier la relation entre deux mentions. Néanmoins, la définition des mentions temporelles dépend largement du type du texte et du domaine d'application. En particulier, le texte clinique est complexe car il décrit des évènements se produisant à des moments différents et contient des informations redondantes et diverses expressions temporelles spécifiques au domaine. Dans cet article, nous proposons une nouvelle représentation des relations temporelles, qui est indépendante du domaine et de l'objectif de la tâche d'extraction. Nous nous intéressons à extraire la relation entre chaque portion du texte et la date de création du document. Nous formulons l'extraction de relations temporelles comme une tâche d'étiquetage de séquences. Une macro F-mesure de 0,8 est obtenue par un modèle neuronal entraîné sur des textes cliniques, écrits en français. Nous évaluons notre représentation temporelle par le positionnement temporel des évènements de toxicité des chimiothérapies.

ABSTRACT

Event-independent temporal positioning : application to French clinical texts.

The extraction of temporal relations entails identifying and classifying the relation between two mentions. However, the definition of temporal mentions is strongly dependent on the type of text and the application domain. Clinical text, in particular, is complex, describing events that occurred at different times, containing redundant information, and a variety of domain-specific temporal expressions. In this paper, we propose a new representation of temporal relations, which is independent of the domain and the goal of the extraction task. We are interested in extracting the relation between each text portion and the document creation time. Temporal relation extraction is cast as a sequence labeling task. An overall macro F-measure of 0.8 is obtained by a neural model trained on clinical texts written in French. We evaluate our temporal representation by the temporal positioning of chemotherapy toxicity events.

MOTS-CLÉS : Texte clinique, Extraction d'informations temporelles, étiquetage de séquences.

KEYWORDS: Clinical Text, Temporal Information Extraction, Sequence Labeling.

Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), volume 2 : travaux de recherche originaux – articles courts, pages 1–14.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

1 Introduction

L'extraction d'informations temporelles est une tâche cruciale pour la compréhension de textes en langage naturel. Cette tâche a été utilisée dans de nombreuses applications de traitement du langage naturel, dont la reconstitution de récits (Do *et al.*, 2012; Ning *et al.*, 2017; Han *et al.*, 2019) et le traitement de textes cliniques (Tourille *et al.*, 2017; Moharasan & Ho, 2019; Lin *et al.*, 2020). Une attention particulière a été portée à l'extraction d'informations temporelles dans les textes cliniques par le biais des campagnes d'évaluation i2b2-2012 (Sun *et al.*, 2013) et Clinical TempEval (Bethard *et al.*, 2015, 2016, 2017). Les informations temporelles contenues dans les textes cliniques des Dossiers Patients Informatisés (DPIs) permettent de mieux comprendre divers évènements cliniques tels que la progression de la maladie et les effets longitudinaux des médicaments (Lin *et al.*, 2016). Le corpus de la campagne i2b2-2012 a été annoté en se basant sur les schémas d'annotations TimeML (Pustejovsky *et al.*, 2003) et THYME-TimeML (Styler IV *et al.*, 2014). THYME-TimeML a été développé pour annoter l'évolution temporelle des évènements cliniques dans le corpus THYME. Ce schéma a été utilisé pour l'annotation du corpus de la campagne Clinical TempEval.

L'extraction d'informations temporelles a d'abord été représentée comme l'extraction des relations TLINKs, reliant les évènements (EVENT) à un ancrage temporel représenté par des expressions temporelles (TIMEX). Pour le domaine clinique, des relations DocTimeRel entre des évènements médicaux et la date du création du document (Document Creation Time, DCT) ont été introduites (Pustejovsky & Stubbs, 2011; Styler IV *et al.*, 2014). Plusieurs défis sont associés à la représentation des informations temporelles cliniques (Najafabadipour *et al.*, 2020). En effet, les expressions temporelles sont diverses et incluent des dates spécifiques au domaine, non standards et abrégées. Le texte clinique étant généralement non-standard, il est également difficile d'associer des évènements à des expressions temporelles. Dans certains cas, le moment associé à un évènement n'est même pas explicitement mentionné. De plus, le texte narratif va et vient dans le temps en décrivant des évènements survenus à des moments différents, et peut contenir des informations redondantes provenant de rapports cliniques antérieurs, ce qui rend difficile l'identification de l'ordre chronologique des évènements.

Plusieurs études se sont intéressées à extraire les relations DocTimeRel (Tourille *et al.*, 2017; Viani *et al.*, 2019; Lin *et al.*, 2020; Alfattni *et al.*, 2021) et les relations TLINKs (Alfattni *et al.*, 2020; Tourille *et al.*, 2016; Liu *et al.*, 2019; Lin *et al.*, 2020). Néanmoins, les performances de la plupart des systèmes proposés sont loin d'être suffisantes pour des applications pratiques (Gumiel *et al.*, 2021). D'une part, outre l'aspect long et coûteux du processus d'annotation, l'annotation des relations temporelles cliniques est beaucoup plus difficile car cela nécessite une expertise médicale. D'autre part, la définition du schéma d'annotation dépend principalement du type de texte clinique et de l'objectif de la tâche d'extraction. La plupart des travaux de recherche portent aussi sur des corpus écrits en anglais. Peu d'études utilisent les corpus en français (Tourille *et al.*, 2016, 2017).

Dans cet article, nous nous intéressons à extraire le positionnement temporel des mentions en fonction de la DCT. Contrairement à l'extraction des relations DocTimeRel, nous ne commençons pas par l'extraction des évènements cliniques pour ensuite les classifier selon la DCT. Au lieu de cela, nous souhaitons l'extraction de la relation temporelle de chaque portion du texte avec la DCT, indépendemment des évènements. Nous pouvons procéder par la suite à l'extraction des évènements et chaque évènement aura alors la même relation temporelle que la portion du texte qui le contient. Voici les principales contributions de cet article :

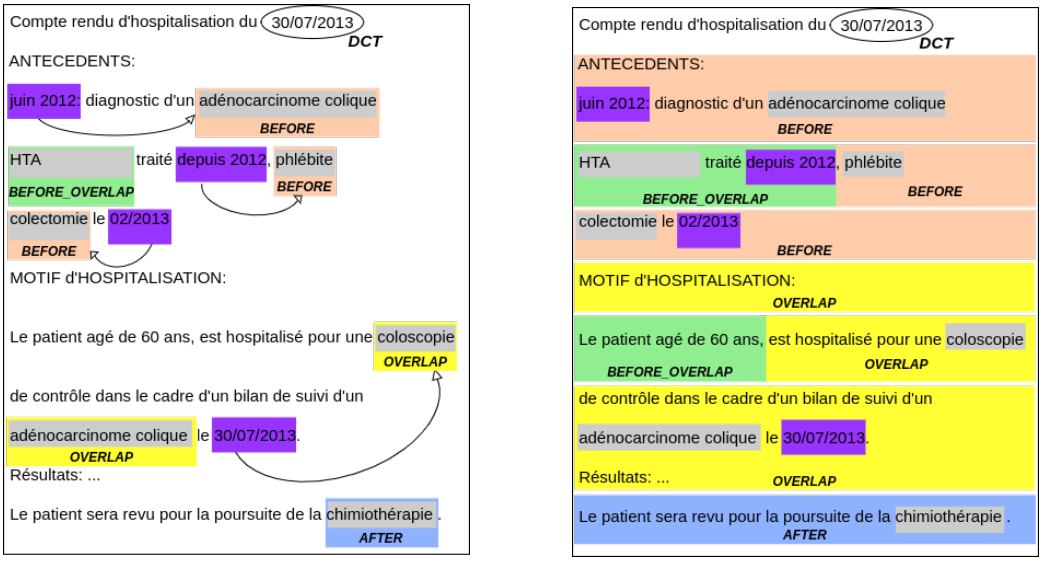
— Nous introduisons une nouvelle représentation des relations temporelles qui nous permet d'identifier des portions de textes homogènes du point de vue temporel et de caractériser ce

positionnement temporel, indépendamment du domaine et de la tâche de l'extraction.
— Pour évaluer notre représentation, nous annotons un corpus de textes cliniques, écrits en français en utilisant le schéma d'annotation THYME-TimeML et nous modélisons la tâche d'extraction de ces relations temporelles en tant qu'une tâche de classification de séquences. Le modèle de classification est comparé à un modèle *Baseline* à base de règles.
— Pour valider l'efficacité de notre représentation des relations temporelles, nous appliquons notre modèle de positionnement temporel sur un autre corpus clinique, nous identifions les évènements de toxicité des chimiothérapies dans ce corpus et nous inférons ensuite la relation temporelle de chaque évènement par rapport à la DCT.

2 Méthodologie

Représentation des relations temporelles. Comme illustré dans la Figure 1a, les relations temporelles sont souvent représentées par les relations DocTimeRel et les relations TLINKs. L'extraction des DocTimeRel revient à l'identification des évènements et à la classification de leur relation temporelle avec la date de création du document. Chaque évènement sera associé à une catégorie selon le schéma THYME-TimeML (Figure 1) : *Before (orange)*, *Before_Overlap (vert)*, *Overlap (jaune)* et *After (bleu)*. Cependant, la définition des événements étant très spécifique au domaine, la tâche d'extraction des DocTimeRel diffère selon le domaine et aucune généralisation n'est possible. Des défis supplémentaires sont également rencontrés dans la définition des évènements cliniques dus à la complexité et à la variété des terminologies médicales présentes dans le texte clinique. L'extraction des relations TLINKs revient dans une première étape à extraire les paires possibles des évènements et des expressions temporelles. La stratégie la plus adoptée consiste à choisir les paires dans la même phrase et d'extraire les relations temporelles intra-phrastiques. Or, les spécificités des textes cliniques présentent des difficultés pour l'identification des frontières de phrase comme l'usage des termes contenant des marques de ponctuation et l'oubli des marqueurs de début et de fin de phrases. Par ailleurs, si l'évènement et l'expression temporelle sont dans des phrases différentes, d'autres stratégies doivent être adoptées pour résoudre les dépendances à longue distance. De ce fait, nous introduisons une nouvelle représentation des relations temporelles qui est indépendante des évènements. Comme illustré dans la Figure 1b, les portions de texte homogènes du point de vue temporel sont extraites et associées à une catégorie du schéma d'annotation THYME-TimeML qui reflète la relation avec la DCT. Les évènements auront par la suite la même catégorie que la portion qui les inclut. Ainsi, nous n'avons pas à gérer les problèmes de frontières de phrases ni la problématique de dépendance longue. Certes, cette représentation est moins fine que la représentation classique des informations temporelles, mais il s'agit d'une représentation qui est totalement indépendante du type des mentions à extraire et donc du domaine d'application.

Extraction et classification des relations temporelles. Nous modélisons la tâche d'extraction des relations temporelles comme une tâche de classification supervisée de séquences. Notre but est de classifier chaque portion du texte en une catégorie pré-définie. Pour ceci, nous entraînons un modèle de classification de tokens, en utilisant le modèle français CamemBERT (Martin *et al.*, 2020) de la librairie *transformers* de HuggingFace (Wolf *et al.*, 2020). Nous classifions chaque token en nous appuyant sur le format BIO (Beginning-Inside-Outside). De cette manière, le modèle sera capable de détecter les tokens qui marquent le changement temporel dans le texte. Les poids du modèle ont été optimisés avec Adam sans *weight decay* pendant 20 époques.



(a) Représentation classique des informations temporelles

(b) Positionnement temporel indépendant des évènements

FIGURE 1 – Représentation des informations temporelles. La DCT est entourée, les expressions temporelles sont représentées en violet, les évènements sont représentés en gris et encadrés par leurs relations DocTimeRel et les TLINKs sont représentées par les flèches. La Figure 1a correspond à la représentation classique des DocTimeRel entre la DCT et les évènements, et des TLINKs entre les évènements et les expressions temporelles. La Figure 1b illustre notre représentation du positionnement des portions du texte par rapport à la DCT, indépendemment des évènements.

Extraction des évènements de toxicité. Pour une première pré-annotation et identification des évènements de toxicité des chimiothérapies, nous avons utilisé l'algorithme QuickUMLS (Soldaini, 2016) en utilisant un dictionnaire de toxicité (Rogier *et al.*, 2021), contenant des termes de toxicité en français issus de différentes terminologies.

3 Expérimentations

Données d'évaluation [1]**.** Pour l'extraction des relations temporelles, nous avons sélectionné au hasard des comptes rendus hospitaliers, opératoires et de consultation de patients atteints de cancer du côlon issus d'un entrepôt de données de santé de l'Hôpital Européen Georges Pompidou (HEGP) (Jannot *et al.*, 2017). Nous avons annoté 180 documents pour l'entraînement et la validation de notre modèle et 42 documents pour l'évaluation. Nous utilisons les catégories du schéma THYME-TimeML ainsi que deux autres catégories *TemporalReference* et *End_Scope*. La catégorie *TemporalReference* sert à identifier le début d'un compte rendu associé à une nouvelle date de création de document, ce qui est nécessaire pour les cas où différents comptes rendus sont inscrits dans le même document. *End_Scope* marque la fin d'une portion de texte si la prochaine portion est une en-tête ou une signature. Cela nous permet uniquement d'exclure ces zones dans le pré-traitement. Trois annotateurs ont annoté un échantillon de 9 documents. Ces accords inter-annotateurs en termes de macro F-mesure sont obtenus : 0,62, 0,73 et 0,69. Pour la classification, nous aurons donc ces cinq catégories : *TemporalReference*,

1. L'accès aux données cliniques a été autorisé par le conseil scientifique et éthique de l'AP-HP (CSE21-15_TALONCO).

Before, *Before_Overlap*, *Overlap* et *After*. La catégorie temporelle par défaut de *TemporalReference* est *Overlap*. Le guide d'annotation détaillé est fourni en Annexe A. Pour la validation de l'efficacité de notre approche, nous utilisons un deuxième corpus qui contient des informations de toxicité de chimiothérapies de patients atteints du cancer du côlon et du poumon (Jannot *et al.*, 2017). Une validation manuelle des annotations des évènements de toxicité a été réalisée par un expert sur 25 documents cliniques, dont 5 documents qui appartiennent au corpus de test de notre modèle de positionnement temporel. L'annotation a été effectuée avec l'outil BRAT (Stenetorp *et al.*, 2012).

Métriques d'évaluation. Dans notre étude, nous sommes intéressés par la détection du changement temporel entre des grandes portions du texte. Dans ce cas, la segmentation en phrases et en tokens n'a donc plus de sens et nous évaluons la performance de notre système de classification temporelle au niveau des caractères en mesurant la macro précision, le macro rappel et la macro F-mesure. Nous calculons les intervalles de confiance à 95 % de nos résultats de classification en utilisant la technique d'*empirical bootstrap* (Dekking *et al.*, 2005, p.275). Pour cela, nous avons échantillonné notre corpus de test avec remise 1000 fois. Les métriques seront calculées pour chaque sous-échantillon. Pour mesurer l'empreinte carbone de l'entraînement et l'évaluation de notre modèle, nous utilisons l'outil Carbontracker (Anthony *et al.*, 2020).

Modèle *Baseline*. Pour évaluer la performance de notre modèle d'extraction de relations temporelles, nous avons développé un modèle à base de règles pour le positionnement temporel des portions de texte. Comme pour l'annotation, nous nous sommes appuyés sur des mots-clés qui sont souvent utilisés pour définir certaines sections médicales, en particulier dans les comptes-rendus opératoires et hospitaliers telles que *"Antécédents", "Indication", "Gestes réalisés", "Traitement de sortie"*, etc. Ces mots-clés sont généralement utiles pour l'annotation temporelle, bien qu'ils soient insuffisants pour couvrir tous les types de comptes rendus. Le modèle *Baseline* sera évalué sur notre corpus de test.

4 Résultats et discussion

Extraction des relations temporelles cliniques. La Table 1 présente une comparaison de notre modèle de positionnement temporel avec le modèle *Baseline* à base de règles. Les meilleurs résultats sont obtenus avec notre modèle avec une macro F-mesure de 0,82, ce qui est supérieur aux accords inter-annotateurs. Les résultats sont beaucoup plus bas avec le modèle *Baseline* avec une F-mesure de 0,39. Malgré l'utilité des mots-clés introduisant les sections médicales pour la catégorisation temporelle, nous ne disposons pas d'une liste exhaustive. De plus, les sections varient énormément selon l'hôpital et même selon le service hospitalier et aucune catégorisation en sections n'est présente dans les comptes rendus du type courrier. Nous pouvons avoir des changements temporels dans une même section médicale et même dans une phrase (Figure 1b). Une approche liée uniquement à la macro-structure du texte ne permet donc pas de faire une analyse temporelle sur le document. De ce fait, ce type de changement ne sera donc pas identifié par le modèle *Baseline*, ce qui peut expliquer les mauvais résultats de ce modèle. L'émission de CO_2 résultant de l'entraînement et du test de notre modèle est estimée à 167 g, qui est l'équivalent de 1,55 km parcourus en voiture. La Table 2 présente les résultats par catégorie de notre modèle. La Table 3 représente le nombre et le pourcentage de portions de texte pour chaque catégorie pour les corpus d'entraînement et de test. Les catégories les plus prévalentes sont les mieux prédites. Ainsi, une F-mesure de 0,88 est obtenue pour la catégorie *Overlap*, représentant 35,5 % du corpus d'entraînement et de 0,86 pour la catégorie *Before*, représentant 22,8 % du corpus d'entraînement. Pour les catégories moins représentées telles que

	Précision	**Rappel**	**F-Mesure**	**Équivalent CO_2 (g.)**
Modèle *Baseline*	0,43 [0,37-0,52]	0,59 [0,51-0,66]	0,39 [0,31-0,46]	-
Notre modèle	**0,82 [0,78-0,85]**	**0,79 [0,75-0,85]**	**0,80 [0,76-0,85]**	167

TABLE 1 – Performance globale des modèles étudiés sur le corpus du test

TemporalReference et *After*, les résultats sont moins bons (F-mesure de 0,75 et 0,79 respectivement). Les portions de texte ayant la catégorie *Before_Overlap* sont souvent des phrases incluses dans des portions de la catégorie *Before* avec une indication temporelle qui indique la continuité dans le temps (Figure 2, Annexe A). Ce changement temporel est rarement prédit correctement et malgré la couverture de la la catégorie *Before_Overlap* (18,5 % dans le corpus d'entrainement) la performance est moins élevée (0,74 de F-mesure).

	P	**R**	**F**
TemporalReference	0,78	0,71	0,75
Before	0,91	0,82	0,86
Before_Overlap	0,75	0,73	0,74
Overlap	0,86	0,90	0,88
After	0,78	0,81	0,79
Macro-moyenne	**0,82**	**0,79**	**0,80**

TABLE 2 – Performance du positionnement temporel des portions du texte sur le corpus de test

	# portions de texte (test)	**# portions de texte (train)**
TemporalReference	46 (8,4%)	207 (11,2%)
Before	121 (22,2%)	423 (22,8%)
Before_Overlap	118 (21,6%)	343 (18,5%)
Overlap	191 (35%)	658 (35,5%)
After	70 (12,8%)	223 (12%)
Total	546	1854

TABLE 3 – Nombre de portions de texte pour chaque catégorie dans le corpus de test et d'entraînement

Positionnement temporel des évènements de toxicité des chimiothérapies. Afin d'évaluer l'efficacité de notre représentation des relations temporelles, nous utilisons notre modèle de positionnement temporel sur 25 documents cliniques contenant des évènements de toxicité. Ces évènements ont été identifiés en se basant sur un dictionnaire de toxicité et validés par un expert. La Table 4 présente les résultats du positionnement temporel des portions du texte, indépendamment des évènements ainsi que les résultats du positionnement temporel des évènements de toxicité sur 5 documents appartenant à la fois au corpus de test de la temporalité et de la toxicité. Nous avons une bonne performance de notre modèle pour le positionnement temporel des portions du texte avec une F-mesure de 0,84, indépendamment des évènements. Pour le positionnement temporel des évènements de toxicité, une F-mesure de 0,7 est obtenue. En moyenne, un positionnement de 10 évènements de toxicité par document est effectué sur cet échantillon de 5 documents. Nous observons une distribution similaire pour les 20 autres documents annotés en toxicité et nous estimons alors que les résultats seront similaires pour ces autres documents.

5 Conclusion

Nous proposons une nouvelle représentation des relations temporelles dans les textes, qui est indépendante des évènements et donc du domaine d'application. La tâche de l'extraction est formulée

	Précision	Rappel	F-Mesure
Positionnement des portions du texte	0,83 [0,76-0,88]	0,85 [0,74-0,97]	0,84 [0,74-0,92]
Positionnement des évènements de toxicité	0,87 [0,65-1]	0,70 [0,38-1]	0,70 [0,44-1]

TABLE 4 – Performance du positionnement temporel des portions du texte et des évènements de toxicité sur 5 documents du corpus de test

comme une tâche de classification des portions de texte homogènes du point de vue temporel en des catégories temporelles. Nous obtenons de bonnes performances de positionnement temporel des portions de texte cliniques écrits en français. Notre représentation temporelle nous a permis d'inférer le positionnement temporel des évènements de toxicité des chimiothérapies et semble être une bonne méthode pour le positionnement temporel de tout type d'évènement indépendamment du domaine.

Remerciements

Nous remercions le conseil scientifique et éthique de l'entrepôt de données de santé de l'AP-HP et l'Hôpital Européen Georges Pompidou qui nous ont permis d'avoir accès aux corpus utilisés dans ce travail. Nesrine Bannour a bénéficié d'un financement de l'ITMO Cancer Aviesan. Bastien Rance est soutenu par le programme SIRIC CARPEM.

Références

ALFATTNI G., PEEK N. & NENADIC G. (2020). Extraction of temporal relations from clinical free text : A systematic review of current approaches. *Journal of Biomedical Informatics*, **108**, 103488.

ALFATTNI G., PEEK N. & NENADIC G. (2021). Attention-based bidirectional long short-term memory networks for extracting temporal relationships from clinical discharge summaries. *Journal of Biomedical Informatics*, **123**, 103915.

ANTHONY L. F. W., KANDING B. & SELVAN R. (2020). Carbontracker : Tracking and predicting the carbon footprint of training deep learning models. In *ICML Workshop on "Challenges in Deploying and monitoring Machine Learning Systems"*.

BETHARD S., DERCZYNSKI L., SAVOVA G., PUSTEJOVSKY J. & VERHAGEN M. (2015). SemEval-2015 task 6 : Clinical TempEval. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, p. 806–814, Denver, Colorado : Association for Computational Linguistics. DOI : 10.18653/v1/S15-2136.

BETHARD S., SAVOVA G., CHEN W.-T., DERCZYNSKI L., PUSTEJOVSKY J. & VERHAGEN M. (2016). SemEval-2016 task 12 : Clinical TempEval. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, p. 1052–1062, San Diego, California : Association for Computational Linguistics. DOI : 10.18653/v1/S16-1165.

BETHARD S., SAVOVA G., PALMER M. & PUSTEJOVSKY J. (2017). SemEval-2017 task 12 : Clinical TempEval. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, p. 565–572, Vancouver, Canada : Association for Computational Linguistics. DOI : 10.18653/v1/S17-2093.

DEKKING F. M., KRAAIKAMP C., LOPUHAÄ H. P. & MEESTER L. E. (2005). *A Modern Introduction to Probability and Statistics : Understanding why and how*, volume 488. Springer.

DIAS G., Éd. (2015). *Actes de TALN 2015 (Traitement automatique des langues naturelles)*, Caen. ATALA, HULTECH.

DO Q., LU W. & ROTH D. (2012). Joint inference for event timeline construction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, p. 677–687, Jeju Island, Korea : Association for Computational Linguistics.

GUMIEL Y. B., SILVA E OLIVEIRA L. E., CLAVEAU V., GRABAR N., PARAISO E. C., MORO C. & CARVALHO D. R. (2021). Temporal relation extraction in clinical texts : A systematic review. *ACM Computing Surveys (CSUR)*, **54**(7), 1–36.

HAN R., NING Q. & PENG N. (2019). Joint event and temporal relation extraction with shared representations and structured prediction. In *Conference on Empirical Methods in Natural Language Processing*.

JANNOT A.-S., ZAPLETAL E., AVILLACH P., MAMZER M.-F., BURGUN A. & DEGOULET P. (2017). The georges pompidou university hospital clinical data warehouse : A 8-years follow-up experience. *International Journal of Medical Informatics*, **102**, 21–28. DOI : https ://doi.org/10.1016/j.ijmedinf.2017.02.006.

LIN C., DLIGACH D., MILLER T. A., BETHARD S. & SAVOVA G. K. (2016). Multilayered temporal modeling for the clinical domain. *Journal of the American Medical Informatics Association*, **23**(2), 387–395.

LIN C., MILLER T., DLIGACH D., SADEQUE F., BETHARD S. & SAVOVA G. (2020). A BERT-based one-pass multi-task model for clinical temporal relation extraction. In *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing*, p. 70–75, Online : Association for Computational Linguistics. DOI : 10.18653/v1/2020.bionlp-1.7.

LIU S., WANG L., CHAUDHARY V. & LIU H. (2019). Attention neural model for temporal relation extraction. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, p. 134–139.

MARTIN L., MULLER B., SUÁREZ P. J. O., DUPONT Y., ROMARY L., DE LA CLERGERIE E. V., SEDDAH D. & SAGOT B. (2020). Camembert : a tasty french language model. *ArXiv*, **abs/1911.03894**.

MOHARASAN G. & HO T.-B. (2019). Extraction of temporal information from clinical narratives. *Journal of Healthcare Informatics Research*, **3**(2), 220–244.

NAJAFABADIPOUR M., ZANIN M., GONZÁLEZ A. R., TORRENTE M., GARCÍA B. N., BERMUDEZ J. L. C., PROVENCIO M. & RUIZ E. M. (2020). Reconstructing the patient's natural history from electronic health records. *Artificial intelligence in medicine*, **105**, 101860.

NING Q., FENG Z. & ROTH D. (2017). A structured learning approach to temporal relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, p. 1027–1037, Copenhagen, Denmark : Association for Computational Linguistics. DOI : 10.18653/v1/D17-1108.

PUSTEJOVSKY J., CASTANO J. M., INGRIA R., SAURI R., GAIZAUSKAS R. J., SETZER A., KATZ G. & RADEV D. R. (2003). Timeml : Robust specification of event and temporal expressions in text. *New directions in question answering*, **3**, 28–34.

PUSTEJOVSKY J. & STUBBS A. (2011). Increasing informativeness in temporal annotation. In *Proceedings of the 5th Linguistic Annotation Workshop*, p. 152–160, Portland, Oregon, USA : Association for Computational Linguistics.

ROGIER A., COULET A. & RANCE B. (2021). Using an ontological representation of chemotherapy toxicities for guiding information extraction and integration from EHRs. In *Medinfo 2021 - 18th World Congress on Medical and Health Informatics*, Virtual conference, Australia. HAL : hal-03364585.

SOLDAINI L. (2016). QuickUMLS : a fast, unsupervised approach for medical concept extraction.

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 Session at EACL 2012*, Avignon, France : Association for Computational Linguistics.

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. DOI : 10.1162/tacl_a_00172.

SUN W., RUMSHISKY A. & UZUNER Ö. (2013). Evaluating temporal relations in clinical text : 2012 i2b2 challenge. *Journal of the American Medical Informatics Association : JAMIA*, **20 5**, 806–13.

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 *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, p. 1136–1142.

TOURILLE J., FERRET O., TANNIER X. & NÉVÉOL A. (2017). Temporal information extraction from clinical text. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics : Volume 2, Short Papers*, p. 739–745, Valencia, Spain : Association for Computational Linguistics.

VIANI N., MILLER T. A., NAPOLITANO C., PRIORI S. G., SAVOVA G. K., BELLAZZI R. & SACCHI L. (2019). Supervised methods to extract clinical events from cardiology reports in italian. *Journal of biomedical informatics*, **95**, 103219.

WOLF T., DEBUT L., SANH V., CHAUMOND J., DELANGUE C., MOI A., CISTAC P., RAULT T., LOUF R., FUNTOWICZ M., DAVISON J., SHLEIFER S., VON PLATEN P., MA C., JERNITE Y., PLU J., XU C., SCAO T. L., GUGGER S., DRAME M., LHOEST Q. & RUSH A. M. (2020). Transformers : State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing : System Demonstrations*, p. 38–45, Online : Association for Computational Linguistics.

A Temporal Annotation scheme for our clinical corpus

A.1 Definitions of temporal categories

To annotate the temporal information in a clinical report, we define a temporal annotation scheme based on the Document Creation Time (DCT) and the possible categories of the Document creation Time Relation (DocTimeRel). The DCT might be the current medical visit date, usually stated in the document heading. It might also be the length of time spent in the hospital. The DCT does not need to be annotated.

A.1.1 Document creation Time Relation

Document creation Time Relation is the relation between events and Document Creation Time. We consider these four possible categories for this time relation : Before, Before_Overlap, Overlap, and After. We annotate only the first word of each temporal portion. We consider that the start of a temporal portion denotes the end of the previous one.

A.1.2 Before

The Before category is used to annotate narrative portions referring to what occurred before the Document Creation Time.

Examples

- Antécédents, antécédents médicaux, antécédents chirurgicaux, Antécédents familiaux, Histoire de la maladie, Rappel clinique, Rappel sur la pathologie → All terms referring to the medical history section.
- **Except :** Maladie traitée depuis le → Before_Overlap since we have a temporal indication that the procedure/disease is still ongoing for the patient (cf. Figure 2).

A.1.3 Before_Overlap

The Before_Overlap category is used to annotate narrative portions that started before the document creation time and are still ongoing at that time.

Examples

- Comorbidités, Mode de vie, Autonomie, traitement habituel, traitement à l'entrée, Allergies, Traitements concomitants, Facteurs de risque, Indication, Indication opératoire, décision d'une intervention
- Patient de 70 ans
- HTA traitée depuis, dans le cadre d'un suivi d'un cancer → The patient is still suffering from the disease.
- METASTASES HEPATIQUES D'UN ADENOCARCINOME → The disease's name as a title in operative reports, which is generally capitalized (cf. Figure 3).

A.1.4 Overlap

The Overlap category is used to annotate narrative portions that happen at the same time as the document creation time.

Examples

- Examen pratique, Tolérance intercure, Au total, Conclusion, Gestes opératoires, Gestes réalisés, Motif d'hospitalisation, Biologie, Biologie de sortie, INTERVENTION, constantes à l'arrivée, Date d'hospitalisation, Date d'entrée, Date de l'intervention, Motif

— Examens complémentaires, Examens paracliniques → Sometimes, some complementary exams are conducted before the document creation time but because they are done for the purpose of the hospital stay, we annotate them as Overlap (cf. Figure 2).
— Je vois ce jour, Je revois en consultation

A.1.5 After

The After category is used to annotate narrative portions referring to what occurs after the document creation time.

Examples
— Traitement de sortie, Prochains rendez-vous, Rendez-vous à venir, Prescription de médicaments, Date de la prochaine cure, Ordonnance de sortie, Prochains examens
— Je reverrai ce patient, je prévois une coloscopie
— La pièce est envoyée pour un examen histologique

A.2 Other categories

A.2.1 TemporalReference

Because several medical reports might be written in the same document, the TemporalReference category specifies the beginning of a new clinical report. Because several medical reports might be written in the same document, the TemporalReference category specifies the beginning of a new clinical report. Each clinical report will then have its own Document Creation Time, and the annotations will be based on this DCT. The TemporalReference category's default Document Time Relation is assumed to be Overlap and does not need to be annotated.

Examples
— Compte-rendu opératoire, Compte-rendu d'hospitalisation, Paris, le 14 octobre 2018

A.2.2 End_Scope

We do not consider heading and signature information in our annotation. Therefore, we use the category End_Scope to mark the ending of a narrative portion if the next narrative portion is a heading or a signature. This way, we avoid annotating the contact information for the health care unit, which may be repeated in several clinical reports. Despite the fact that the clinical documents are de-identified, we avoid annotating specific patient information. In cases other than headings or signatures, the end of a temporal portion is implicitly considered the start of a new temporal portion.

A.3 Examples of annotations made in accordance with the above scheme and guidelines

Annotations of the first example (cf. Figure 2)

— From *Compte* to *d'hospitalisation* as TemporalReference
— From *Hospitalisé* to 30/07/2013 as Overlap
— From *Motif* to *d'HOSPITALISATION :* as Overlap, note that we don't annotate the temporal portion after the End_Scope category containing contact information of doctors
— From *HISTOIRE* to *ANTECEDENTS* as Before
— From *HTA* to *2012,* as Before_Overlap since we have a temporal indication that the disease is still ongoing for the patient
— From *phlébite* to *07/2012* as Before since it's part of the medical patient history
— From *ALLERGIES* to *Autonome* as Before_Overlap
— From *Examens* to *et* as Overlap despite the fact that the medical exams are conducted before the date of hospital admission
— From *sera* to *10/09/2013* as After. The signature of the document after the End_Scope category is not annotated

FIGURE 2 – A first example of hospital report annotations

Annotations of the second example (cf. Figure 3)

— From *COMPTE* to *OPERATOIRE* as Temporal Reference
— *ADENOCARCINOME* as Before_Overlap
— *COLECTOMIE* as Overlap
— From *Rappel* to *clinique :* as Before
— From *Indication* to *opératoire.* as Before_Overlap
— From *Gestes* to *réalisés :* as Overlap
— From *La* to *histologique.* as After

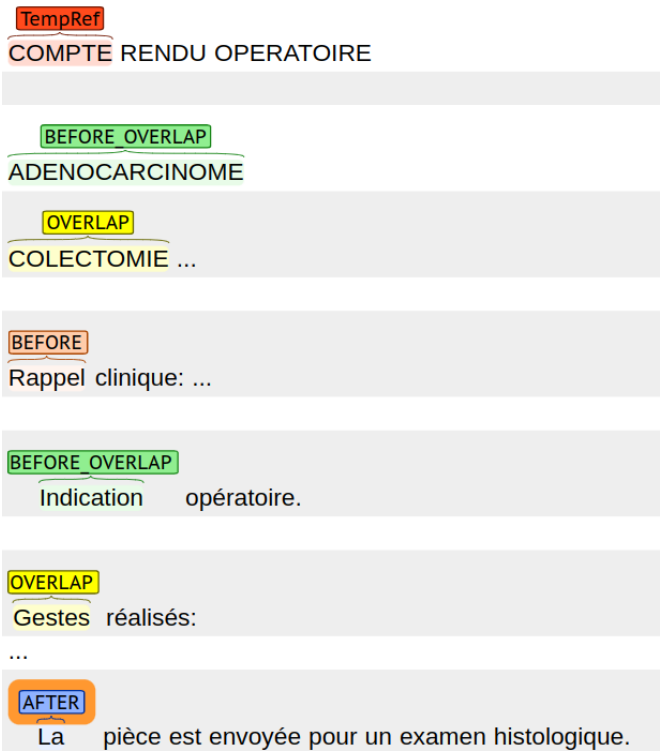


FIGURE 3 – A second example of annotating an operative report

Annotations of the third example (cf. Figure 4)

— From *Paris* to *2014,* as TemporalReference
— From *Je* to *jour* as Overlap
— From *Monsieur* to *comme* as Before_Overlap for the patient's age and since it is stated that the purpose of the medical visit is a disease follow-up
— From *antécédent* to *Rappel :* as Before
— From *Examen* to *pratique :* as Overlap
— From *A* to *mois* as After
— From *Dossier* to *staff* as TemporalReference, it's a new clinical report
— From *Dernières* to *2014 :* as Before, based on the document creation time of the second clinical report.
— From *Décisions* to *staff :* as Overlap
— From *Le* to *consultation.* as After

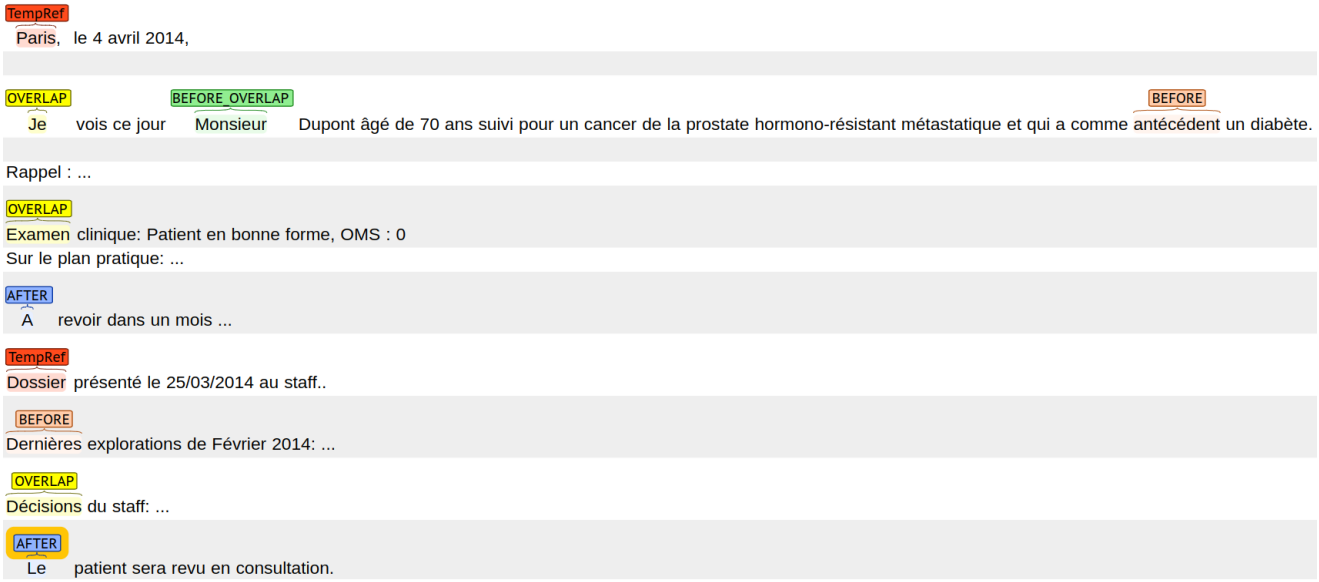


FIGURE 4 – A third example of annotating a clinical document containing two clinical reports