



HAL
open science

Annotation sémantique de documents cliniques psychiatriques français fondée sur une ontologie de domaine

Ons Aouina, Jacques Hilbey, Jean Charlet

► To cite this version:

Ons Aouina, Jacques Hilbey, Jean Charlet. Annotation sémantique de documents cliniques psychiatriques français fondée sur une ontologie de domaine. Ingénierie des Connaissances 2023, Jul 2023, Strasbourg, France. hal-04146747

HAL Id: hal-04146747

<https://hal.science/hal-04146747v1>

Submitted on 30 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.

Annotation sémantique de documents cliniques psychiatriques français fondée sur une ontologie de domaine

O. Aouina^{1,2}, J. Hilbey^{1,2}, J. Charlet^{2,3}

¹ Sorbonne Université, Paris, France

²Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, LIMICS, Paris, France

³ Assistance Publique-Hôpitaux de Paris, Paris, France

ons.aouina@etu.sorbonne-universite.fr

Résumé

L'établissement d'une chronologie des profils de patients psychiatriques peut répondre à de nombreuses questions précieuses, telles que la façon dont les événements médicaux importants affectent la progression de la psychose chez les patients. Cependant, la majorité des outils d'extraction d'informations textuelles et d'annotation sémantique, ainsi que les ontologies de domaine, ne sont disponibles qu'en anglais et ne peuvent être facilement étendus à d'autres langues, en raison de différences linguistiques fondamentales. Dans cet article, nous décrivons un système d'annotation sémantique fondé sur une ontologie développée dans le cadre de PsyCARE. Notre système a été comparé avec un autre annotateur sémantique ECMT sur un échantillon de 20 comptes rendus d'hospitalisation français et évalué manuellement par deux annotateurs sur un ensemble de 50 comptes rendus d'hospitalisation, montrant des résultats prometteurs.

Mots-clés

Annotation sémantique, GATE, Ontologie, Psychiatrie, TALN

Abstract

Building a timeline of psychiatric patient profiles can answer many valuable questions, such as how important medical events affect the progression of psychosis in patients. However, the majority of text information extraction and semantic annotation tools, as well as domain ontologies, are only available in English and cannot be easily extended to other languages, due to fundamental linguistic differences. In this paper, we describe a semantic annotation system based on an ontology developed in the framework of PsyCARE. Our system was compared with another semantic annotator called ECMT on a sample of 20 French patient discharge summaries and evaluated by two annotators on a set of 50 summaries, showing promising results.

Keywords

Semantic Annotation, GATE, Ontology, Psychiatry, NLP

1 Introduction

La schizophrénie et la psychose chronique comptent parmi les troubles les plus débilissants chez les adolescents et les jeunes adultes et sont associées à des troubles cognitifs, à une moins bonne réussite professionnelle et à une mauvaise qualité de vie. Des études ont montré que plus la durée de la psychose non traitée est longue, plus les résultats de l'intervention sont mauvais, plus le rétablissement et le fonctionnement général sont mauvais et plus l'affaiblissement social à long terme est important [13].

Ce problème est abordé par le projet RHU PsyCARE¹, qui vise à améliorer l'intervention précoce dans la psychose en fournissant des outils pour faciliter l'accès aux soins et offrir des programmes de traitement personnalisés. Dans ce contexte, l'analyse des Comptes Rendus d'Hospitalisation (CRH) peut nous donner l'opportunité d'étudier de nombreuses questions essentielles, telles que l'impact des événements médicaux importants sur la progression de la psychose chez les patients. Ces résumés fournissent des informations sur l'historique du patient (p. ex. liées au début des symptômes ou au début du traitement ou à des antécédents familiaux).

Cependant, extraire de telles informations pour retracer l'historique de la psychose et développer la chronologie est une affaire complexe qui nécessite des corpus soigneusement annotés. Être en mesure d'extraire automatiquement ces informations peut améliorer les soins médicaux et aiderait également la recherche clinique [1].

Dans ce projet, nous proposons une méthode d'annotation sémantique des CRH fondée sur une ontologie développée dans le cadre de PsyCARE. Notre approche permet non seulement d'extraire les entités médicales d'un texte mais aussi de les transformer en connaissances structurées et formalisées.

La reconnaissance d'entités médicales et l'établissement de liens sont des tâches difficiles dans le traitement du langage naturel en français, en particulier dans notre contexte de textes narratifs non structurés en psychiatrie. Cependant, l'utilisation des ontologies aiderait à concevoir des index de données sémantiques qui exploitent les connaissances mé-

1. <https://psy-care.fr/>

dicales pour améliorer la recherche et la récupération d'informations. Cette proposition s'inscrit, dans le projet PsyCARE, dans une motivation, large, de construire une chronologie complète de la psychose d'un patient à partir de son dossier médical.

2 Contexte

L'annotation automatique de textes consiste à identifier les passages d'un texte (mots simples ou composés) faisant référence à des concepts identifiés dans des terminologies normalisées (p. ex. les concepts UMLS dans le domaine de la santé, les URI de Dbpedia ou d'autres « Linked Open Data » dans d'autres domaines, etc), puis à identifier les modalités (les concepts qui apparaissent précédés d'une négation ou d'un doute), et enfin, idéalement, les liens entre les concepts.

En raison de l'ambiguïté, de la polysémie du langage naturel, et de la richesse des moyens dont il dispose pour exprimer le sens, cette tâche n'est pas aisée. C'est un champ de recherche actif depuis la fin des années 1970 : le Traitement Automatique du Langage naturel (TALN).

Au cours de la dernière décennie, l'expérience accumulée a conduit à l'émergence de plusieurs plateformes d'analyse de texte solides, offrant la possibilité de procéder avec des performances raisonnables à un premier passage d'annotation automatique. Bien qu'il y ait eu de nombreuses recherches sur l'utilisation de techniques d'annotation sémantique, en particulier l'utilisation d'ontologies, pour améliorer la détection et le diagnostic des troubles mentaux dans les textes en langue anglaise, peu d'études ont exploré l'application de ces techniques dans les textes en langue française [3]. Nous citons notamment :

SIFR [1] est un service web accessible à tous permettant à la fois la reconnaissance et la contextualisation de concepts issus de 30 terminologies et ontologies médicales. Le service d'annotation traite les descriptions textuelles, en les associant à des concepts des ontologies biomédicales adéquats, y compris UMLS. Il crée également des annotations en utilisant les connaissances intégrées dans les ontologies, et contextualise les annotations avant de les renvoyer aux utilisateurs dans plusieurs formats. Son installation en local étant très compliquée, il n'a pas été testé.

ECMT (CHU Rouen, France) [17] est un service web disponible sous la forme d'une API web RESTful, qui permet d'annoter automatiquement un texte à l'aide des concepts des principales terminologies de santé disponibles en français. Il est inspiré de l'algorithme CISMef pour la recherche d'information avec le moteur de recherche Doc'CISMeF et F-MTI [18] qui est un indexeur automatique multiterminologique. L'ECMT est adapté à la langue française. Il s'appuie sur l'algorithme du sac de mots et sur la reconnaissance de motifs pour analyser des CRH, des rapports de procédures ou des résultats de laboratoire qui contiennent des données symboliques (présence ou absence), des données numériques et des unités de mesure[4]. En revanche, ce dernier ne donne pas la possibilité de choisir l'ontologie à utiliser pour le processus d'annotation. L'annotateur sup-

porte des fonctions d'expansion sémantique. Nous avons testé la version disponible localement avec sept terminologies afin de garantir le respect de la confidentialité des données.

GATE (Univ. Sheffield, UK) [8] est un logiciel *open source* développé à l'origine à l'Université de Sheffield. Il s'agit d'un environnement de TALN qui permet d'effectuer diverses tâches d'analyse de texte, telles que l'extraction d'information, la classification de documents, la recherche d'information et la traduction automatique. Il y a notamment des développements prometteurs dans le domaine de l'annotation sémantique, l'accent étant mis sur l'évolutivité dans le contexte des données ouvertes liées (Linked Open Data) [12]. Il a également été largement appliqué au domaine de la santé[5]. Sa conception permet de créer des flux de traitement en associant des modules *processing units* dans une architecture ouverte. GATE met donc l'accent sur la réutilisabilité et le partage des ressources. Sa licence (AGPL) permet de développer d'autres applications, y compris commerciales (sous réserve d'accord de licence). Chacune de ces plateformes a ses spécificités et ses domaines d'excellence. Parmi elles, GATE est celle qui a été appliquée au domaine de la santé, avec des expériences réussies d'utilisation dans un contexte opérationnel, par exemple au South London and Maudsley University Hospital [6], ou comme spin-off du projet européen Khresmoi/KConnect (outils d'extraction de données sémantiques distribués par la PME bulgare Ontotext). C'est donc cette plateforme qui a été retenue pour notre travail d'annotation.

3 Méthodes

3.1 Jeu de données sur la psychiatrie

Les documents cliniques utilisés dans ce travail sont issus du projet français PsyCARE. Il s'agit d'une compilation d'environ 8000 CRH couvrant une période de dix ans, ce qui représente un volume d'environ 3 500 000 mots. Ces CRH proviennent du Groupe Hospitalier Universitaire Psychiatrie et Neurosciences de Paris. Ils sont semi-standardisés, en format Word et ont été pseudo-anonymisés au préalable, en remplaçant tous les noms, dates, lieux, etc. De plus, le diagnostic est indiqué à la fin de chaque document en utilisant la CIM-10². Ces documents sont rédigés en français et décrivent l'histoire et le contexte social du patient, les médicaments, les détails de l'admission à l'hôpital et les diagnostics psychiatriques actuels et antérieurs.

3.2 Description de l'ontologie de domaine

L'ontologie utilisée dans notre processus est développée dans le cadre de PsyCARE afin d'intégrer les données et de permettre leur annotation sémantique représentant des domaines tels que les aspects cliniques psychiatriques (signes, symptômes, troubles), les médicaments avec leur code ATC, l'imagerie, la biologie, etc. Comme nous l'avons déjà mentionné, notre objectif final est de modéliser les événements cliniques en psychiatrie. D'où la nécessité d'in-

2. Liste de classification médicale de l'OMS <https://icd.who.int/browse10/2019/en>

clure une représentation temporelle des connaissances médicales [11].

À partir de cette ontologie, nous avons reconstruit un schéma d'annotation, auquel nous avons ajouté une branche décrivant la structure des CRH³ comme indique la figure 1. Cela nous permet de relier les concepts à leur contexte d'occurrence. Par exemple, les médicaments apparaissant dans la section Historique de la maladie ne partagent pas le même contexte que ceux apparaissant dans la section Traitement de sortie.

Il faut noter que la version utilisée et présentée dans l'article n'est pas finalisée, que ce soit pour la clinique ou pour la description des sections.

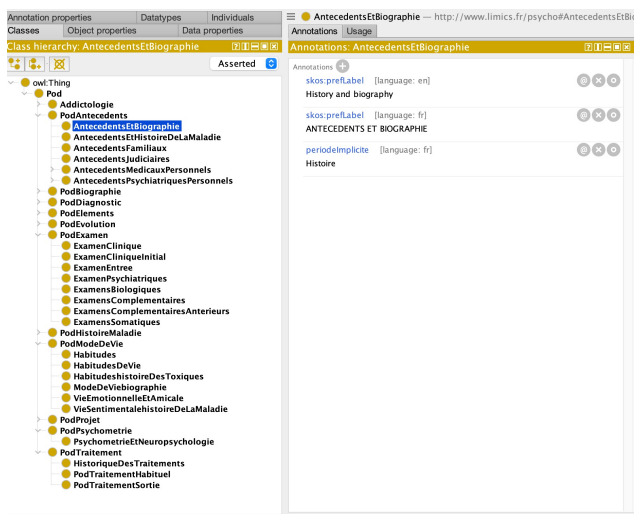


FIGURE 1 – Branche de l'ontologie utilisée pour décrire la structure des CRH des patients.

3.3 Annotation sémantique

GATE fournit différents composants pour l'extraction d'informations sémantiques. Nous avons réutilisé des composants précédemment développés par le projet OnBaSsam [19], qui ont été modifiés et améliorés pour s'adapter aux spécificités des rapports liés à la psychiatrie (p. ex. l'identification des sections et l'extraction des syntagmes nominaux). Notre pipeline est composé de plusieurs ressources de traitement qui s'exécutent séquentiellement sur un document donné, comme le montre la figure 2.

3.3.1 Pré-traitement des comptes rendus médicaux

Tout d'abord, la tokenisation des mots est appliquée sur les CRH, puis le découpage des phrases, suivi du marquage des parties du discours (POS Tagging) et de la lemmatisation, et enfin, l'extraction des syntagmes nominaux ou NP-chunking. Pour ce faire, nous avons utilisé Gate Corpus Pipeline, un composant Gate pour le traitement de corpus, auquel nous avons ajouté les Ressources de Traitement (RT) suivantes : un *tokenizer* français pour la *tokenisation*

3. Cette branche de l'ontologie est disponible à l'adresse suivante https://github.com/AouinaOns/PartOfDocuments_ontology

et le TreeTagger⁴ français pour l'annotation des textes avec des informations sur les parties du discours et les lemmes. Ce dernier a été utilisé avec succès pour annoter des textes français. Enfin, pour récupérer les syntagmes nominaux des CRH, OpenNLP⁵ est utilisé et adapté à la langue française. Cette étape nous permet de construire une liste de phrases nominales à partir de la sortie du TreeTagger. Dans notre solution, nous supposons que les entités telles que les signes et les symptômes, les maladies, les troubles et les événements cliniques sont des syntagmes nominaux [16]. En ce qui concerne le traitement de l'ontologie, nous suivons les mêmes étapes pour la construction de la liste Gazetteer. Cette liste est constituée des concepts de l'ontologie et de leurs étiquettes (PrefLabels et AltLabels), prétraités selon les étapes présentées dans la figure 2, ainsi que des URI.

3.3.2 Détection des sections

La structure discursive d'un document peut être très utile pour améliorer les outils de recherche d'information. L'identification des sections, par exemple « Histoire de la maladie actuelle » ou « Histoire de la famille », est cruciale dans notre contexte car elle est la clé majeure pour identifier le contexte temporel des passages narratifs. Comme mentionné précédemment, les CRH sont organisés selon une structure taxinomique (Cf. sec. 3.2).

Cette structure n'est pas toujours respectée, certaines sections peuvent être manquantes, fusionnées ou dans un ordre différent, et une même section peut avoir plusieurs titres. Par exemple, pour la section « histoire de la famille », on peut trouver « histoire de la famille », « antécédents familiaux psychiatriques », etc. Par conséquent, nous appliquons les règles JAPE [9] et la correspondance floue des chaînes de caractères sur les termes de la nomenclature des noms de sections précédemment construite pour identifier les limites des sections. Une règle habituelle utilise la distance de Levenshtein qui est une métrique représentant le nombre de changements de caractères entre les mots. Cette règle identifie le début d'une section par l'apparition d'un terme disponible dans la terminologie et la fin de la section avant le début du terme suivant.

3.3.3 Reconnaissance d'entités et établissement des liaisons

Pour cette tâche, nous classons les candidats *NP-chunk* obtenus dans la phase de prétraitement en concepts ontologiques en leur attribuant un URI. Pour ce faire, nous développons un système à base de règles combiné à une classification par chaîne de correspondance floue. Nous définissons le seuil d'acceptation des annotations comme étant proportionnel à la longueur du terme du dictionnaire.

Pour extraire les entités temporelles, nous avons utilisé le plugin TIMEX⁶ GATE pour annoter les documents avec des balises TIMEX3 en utilisant la bibliothèque SUTime (Stanford Temporal Tagger). Cela nous a permis d'extraire des dates, des durées, par exemple « depuis environ 1 mois », et des fréquences, par exemple « 2 fois par jour ».

4. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger>

5. <https://github.com/GateNLP/gateplugin-OpenNLP>

6. <https://github.com/pkourdis/gateplugin-SUTime>

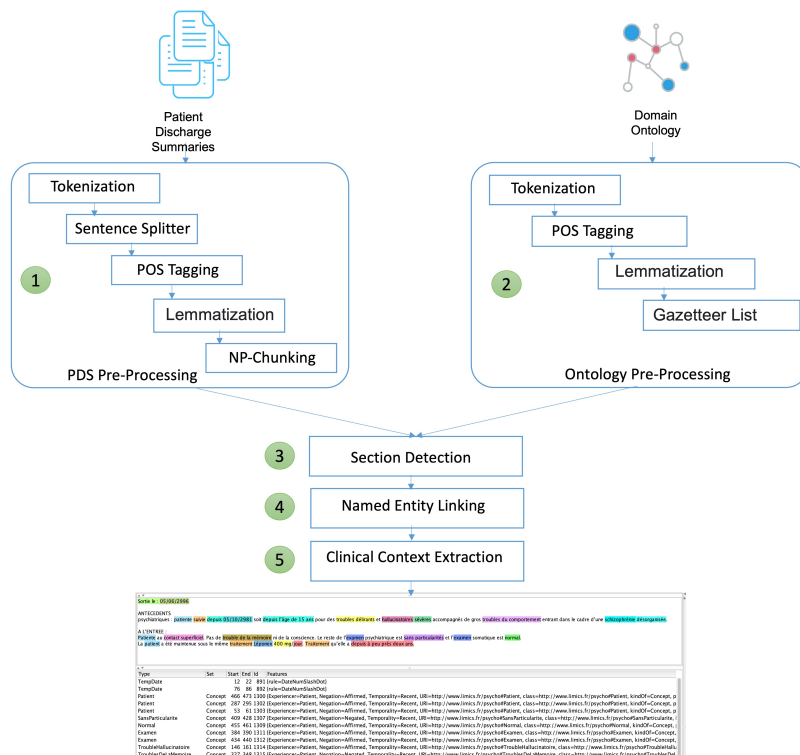


FIGURE 2 – Illustration du processus d’annotation sémantique fondé sur l’ontologie.

3.3.4 Extraction du contexte clinique

En plus d’extraire les entités nommées elles-mêmes, il est nécessaire d’identifier le contexte dans lequel elles apparaissent dans le texte. Les documents sont pré-annotés et mis en correspondance avec les concepts de l’ontologie (Cf. sec. 3.2) (signes, maladies, traitements, symptômes, expressions temporelles, etc.) en utilisant le processus décrit ci-dessus.

Ensuite, l’algorithme français FastContext [14] est appliqué pour identifier le contexte des conditions cliniques annotées dans une phrase. On considère trois contextes : la négation, l’hypothèse et la détermination du sujet, soit le patient, les membres de la famille du patient ou le professionnel de santé.

4 Résultats

4.1 Schéma d’annotation

Nous évaluons d’abord notre approche en analysant manuellement le résultat de la phase d’extraction des entités nommées. Pour ce faire, un schéma d’annotation a été créé pour évaluer les annotations du pipeline. Nous avons utilisé l’outil d’annotation rapide BRAT. Un examen des outils d’annotation Neves et Leser [15] a montré que BRAT était facile à utiliser et pouvait prendre en charge à la fois notre schéma d’annotation et les pré-annotations automatiques. Afin de faciliter l’évaluation manuelle, nos entités extraites ont été regroupées en 10 concepts uniques de premier niveau.

Signe et symptôme. Ce sont des manifestations de troubles mentaux et émotionnels qui peuvent affecter le comportement, la pensée, l’humeur et la perception d’une personne. Ils sont énoncés dans le premier cas par le praticien (signe) et dans le second cas par le patient (symptôme).

Trouble. Cela fait référence à une gamme de conditions de santé mentale qui affecte l’humeur, les émotions, les pensées, le comportement et la perception d’une personne. Les troubles psychiatriques comprennent, par exemple, la dépression, l’anxiété, les troubles bipolaires, les troubles de l’alimentation, les troubles du sommeil et la schizophrénie.

Événement clinique. Les événements cliniques peuvent être des blessures, des accidents, des procédures médicales, des hospitalisations, des interventions chirurgicales, des résultats de tests de laboratoire, des changements dans l’état de santé d’un patient, etc.

Médicament. Ce concept fait référence à une prescription destinée à guérir les maladies ou à soulager les symptômes. On trouve également :

- nom de médicament : Le nom d’un médicament, éventuellement accompagné de sa spécification ;
- dose de médicament : la quantité d’un médicament, y compris le nombre et l’unité.

Situation personnelle. Correspond à tous les aspects de la vie de la personne qui peuvent affecter sa santé et son bien-être. Cela peut inclure des informations sur son environnement familial, social, professionnel, économique, culturel et personnel.

Informations temporelles. Fait référence à des dates, des heures, des durées ou des fréquences dans le texte clinique.

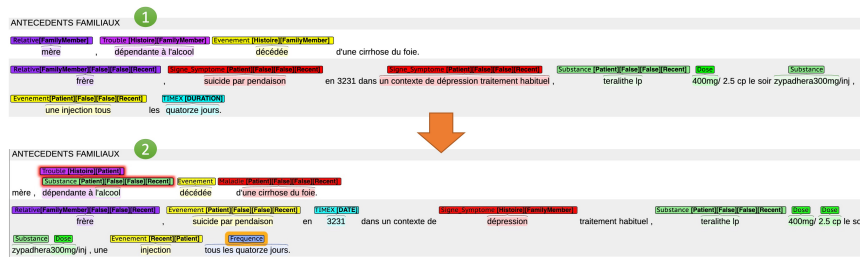


FIGURE 3 – Correction de l’annotation du Pipeline. La partie haute montre la sortie du pipeline et la partie basse, les modifications du correcteur. On constate, par exemple, que l’annotation de « une injection tous les quatorze jours » indique une durée là où il fallait voir une fréquence.

4.2 Évaluation de l’annotation

Nous avons évalué notre approche de deux manières différentes, la première consistant à analyser manuellement la phase de reconnaissance des entités nommées (REN) et d’extraction par rapport à un schéma d’annotation (Cf. sec. 4).

Notre corpus d’évaluation est composé de 50 CRH extraits aléatoirement de l’ensemble de données (4100 phrases, 4213 concepts ontologiques non uniques annotés). Deux personnes ont évalué les annotations et leur contexte, notamment le repérage de la négation, l’hypothétique, la temporalité et la personne impliquée (p. ex. le patient vs un membre de sa famille). La figure 3 montre l’interface de Brat après une correction manuelle des annotations. Nous les avons donc regroupées en 10 concepts uniques de plus haut niveau pour faciliter l’évaluation manuelle. La précision, le rappel et la F-mesure de la phase REN sont présentés dans le tableau 1. La phase NER et l’extraction du contexte des concepts obtiennent une précision globale de 0,9674, un rappel de 0,9780 et une F-mesure de 0,9727.

TABLE 1 – Résultats quantitatifs des évaluations de l’extraction d’entités nommées par les 2 annotateurs avec un accord inter annotateur de 0.88.

	Quantité	Précision	Rappel	F-mes.	
Sign Or Symptom	1747	0.9544	0.9503	0.9524	
Disease	150	0.9826	0.7635	0.8593	
Trouble	459	0.9894	0.9493	0.9690	
Clinical Event	1459	0.9744	0.8140	0.8870	
Personal Situation	188	0.9895	0.8468	0.9126	
Drug	Drug Name	1034	0.8200	0.9805	0.8822
	Drug Dose	650	0.9848	0.9610	0.9727
Temporal Inf.	Date	840	0.9942	0.9709	0.9824
	Duration	529	0.9574	0.9777	0.9674
	Frequency	212	0.9459	0.8373	0.8883

Le deuxième mode d’évaluation consiste à comparer nos performances avec celles d’un autre outil d’annotation sémantique [2], l’ECMT, sur un échantillon de 20 CRH.

Nous avons observé que notre pipeline a extrait le plus grand nombre de concepts : 2384 alors que 1838 ont été extraits avec l’ECMT. Notre pipeline est plus efficace dans l’identification des signes, symptômes et troubles psychiatriques, tels que les délires et le tempérament hyperthyroïdique, ainsi que les informations temporelles, qui sont

des informations courantes dans les CRH. Inversement, l’ECMT détecte mieux les maladies et les événements non psychiatriques, tels que la sinusite maxillaire récurrente et l’appendicectomie.

5 Discussion

L’évaluation initiale des annotations sémantiques montre que nous pouvons identifier correctement les concepts de l’ontologie en utilisant GATE et des algorithmes modifiés pour tenir compte des spécificités des textes en français. Ces résultats encourageants peuvent être expliqués par deux facteurs. Premièrement, la richesse du vocabulaire de l’ontologie, notamment pour la composante signes et symptômes psychiatriques. Deuxièmement, la prise en compte de la structuration des CRH, qui pourrait être une faiblesse pour d’autres types de documents psychiatriques. De plus, cette méthode permet de déterminer les limites correctes des entités nommées détectées.

Cependant, il convient de noter que les expressions nominales obtenues automatiquement (*NP-Chunk*) ne sont pas toujours parfaites et peuvent entraîner des erreurs. Bien que nous ayons consacré des efforts considérables à la correction des annotations du corpus, la taille de l’ensemble des données annotées reste modeste par rapport à d’autres travaux.

Une autre limite réside dans le fait que le pipeline présenté dans cet article repose principalement sur des techniques traditionnelles de traitement du langage. Pour l’améliorer, des règles doivent être ajoutées manuellement, ce qui peut être chronophage et ne permet pas toujours de capturer toutes les nuances des CRH. En réponse aux limites de notre approche, une combinaison d’approches fondées sur les règles et l’apprentissage automatique peut être utilisée pour construire des systèmes d’annotation sémantique plus efficaces [10].

6 Conclusion

L’objectif de ce travail est de reconstruire les données structurées des patients à partir des CRH afin de compléter les données des patients dans le projet PsyCARE. Dans cet article, nous avons présenté une première étape qui consiste à annoter sémantiquement les CRH de psychiatrie française. À cette fin, à partir d’un ensemble non structuré, nous avons

pu réaliser une annotation sémantique en utilisant des plugins GATE et des algorithmes que nous avons modifiés pour les adapter à la structure des CRH.

Les prochaines étapes de notre travail consistent à intégrer dans l'ontologie les événements cliniques et les maladies non-psychiatriques les plus fréquemment présents dans les CRH. Ensuite, il sera nécessaire d'identifier les relations entre les concepts. Par exemple, extraire les relations temporelles entre les événements cliniques et psychiatriques et les entités temporelles. Cela permettra de retracer de manière exhaustive la chronologie des événements dans les rapports cliniques, offrant ainsi une représentation plus complète du parcours thérapeutique des patients. Compte tenu de l'état de l'art, cela sera fait avec des algorithmes d'apprentissage automatique [7, 20].

Remerciements

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du Programme d'Investissements d'Avenir portant la référence PsyCARE ANR-18-RHUS-0014.

Références

- [1] A. Tchechmedjiev, A. Abdaoui, V. Emonet, S. Zevio, C. Jonquet. SIFR annotator : ontology-based semantic annotation of French biomedical text and clinical notes. *BMC Bioinformatics*. 2018.
- [2] A. Redjdal, J. Bouaud J, J. Gligorov, B. Séroussi. Comparison of MetaMap, cTAKES, SIFR, and ECMT to Annotate Breast Cancer Patient Summaries. *Stud Health Technol Inform*. 2022 Jun 6.
- [3] A. Le Glaz, Y. Haralambous Y, DH. Kim-Dufor, P.Lenca et al. Machine Learning and Natural Language Processing in Mental Health : Systematic Review. *J Med Internet Res*. 2021 May 4.
- [4] C. Cabot, Recherche d'information clinomique dans le Dossier Patient Informatisé : modélisation, implantation et évaluation. Thèse de doctorat, 2017.
- [5] G. Gorrell, J. Petrak, and K. Bontcheva. Using Twitter Conventions to Improve #LOD-Based Named Entity Disambiguation. *The Semantic Web. Springer International Publishing*, pages 171–186, Cham, 2015.
- [6] G. Perera, M. Broadbent, F. Callard, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register : current status and recent enhancement of an Electronic Mental Health Record-derived data resource, *BMJ Open* 2016.
- [7] G. Alfattni, N. Peek, and G. Nenadic. Extraction of temporal relations from clinical free text : A systematic review of current approaches. *Journal of biomedical informatics*, 2020.
- [8] H. Cunningham, D. Maynard, K. Bontcheva et al. Text processing with GATE (Version 6). Sheffield : GATE, 2011.
- [9] H. Cunningham, D. Maynard, and V. Tablan. JAPE : a java annotation patterns engine (second edition). department of computer science, university of sheffield, 2000.
- [10] J. Jouffroy, SF. Feldman, I. Lerner, B. Rance, A. Burgun, A. Neuraz. Hybrid Deep Learning for Medication-Related Information Extraction From Clinical Texts in French : MedExt Algorithm Development Study. *JMIR Med Inform*. 2021 Mar 16.
- [11] J. Hilbey, X. Aimé, and J. Charlet. Temporal Medical Knowledge Representation Using Ontologies. *Studies in Health Technology and Informatics* 294 (May 25, 2022) : 337–41.
- [12] L. Derczynski, D. Maynard, G. Rizzo et al. Analysis of named entity recognition and linking for tweets, *Information Processing & Management*, Volume 51, Issue 2, 2015.
- [13] L. Souaiby, R. Gaillard R, MO.Krebs . Durée de psychose non traitée : état des lieux et analyse critique [Duration of untreated psychosis : A state-of-the-art review and critical analysis]. 2016 Aug.
- [14] M. Mirzapour, A. Abdaoui, A. Tchechmedjiev et al. French FastContext : A publicly accessible system for detecting negation, temporality and experienter in French clinical notes, *Journal of Biomedical Informatics*, Volume 117, 2021.
- [15] M. Neves, U. Leser, A survey on annotation tools for the biomedical literature, *Briefings in Bioinformatics*, Volume 15, Issue 2, March 2014.
- [16] S. Zhang, N. Elhadad, Unsupervised Biomedical Named Entity Recognition : Experiments with Clinical and Biological Texts. *Journal of biomedical informatics*. 2013.
- [17] S. Pereira, A. Névéol, G. Kerdelhué, E. Serrot, M. Joubert, S. Darmoni. Using multi-terminology indexing for the assignment of MeSH descriptors to health resources in a French online catalogue. *AMIA symp*. 2008 :586-590.
- [18] S. Darmoni, JP. Leroy, M. Douyère et al. Doc'CIS-MeF : un outil de recherche Internet orienté vers l'enseignement et la formation à distance en médecine.
- [19] S. Cardoso, P. Meneton, X. Aimé, V. Meininger, D. Grabli, G. Guezennec, J. Charlet, Use of a modular ontology and a semantic annotation tool to describe the care pathway of patients with amyotrophic lateral sclerosis in a coordination network. *PLoS One*. 2021 Jan.
- [20] Y. Luo, Ö. Uzuner, P. Szolovits . Bridging semantics and syntax with graph algorithms-state-of-the-art of extracting biomedical relations. *Brief Bioinform*. 2017.