Analyse automatique de négations pour la radiologie et autres textes cliniques en français par modèles de langage

S Sadoune, Antoine Fraboulet, A Richard, F Talbot, L. Boussel, Hugues Berry

▶ **To cite this version:**

S Sadoune, Antoine Fraboulet, A Richard, F Talbot, L. Boussel, et al.. Analyse automatique de négations pour la radiologie et autres textes cliniques en français par modèles de langage. 2023. hal-04165319

HAL Id: hal-04165319
https://inria.hal.science/hal-04165319

Preprint submitted on 18 Jul 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

Analyse automatique de négations pour la radiologie et autres textes cliniques en français par modèles de langage

S. Sadoune[1], A. Fraboulet[1], A. Richard[2], F. Talbot[2], L. Boussel[3], H. Berry[4]

[1] Inria, Centre de Recherche de Lyon, F-69603 Villeurbanne
[2] DSN Bron, Hospices Civils de Lyon, F-69672 Bron
[3] CREATIS UMR 5220, INSA-Lyon, Univ. Claude Bernard Lyon 1, CNRS, Inserm, F-69621 Lyon
[4] AIstroSight, Inria, Hospices Civils de Lyon, Univ. Claude Bernard Lyon 1, F-69603 Villeurbanne

31 mai 2023

Résumé

Nous présentons une approche d'apprentissage supervisé basée sur un modèle de langage de type BERT pour la reconnaissance automatique et la détermination de la portée de négations dans des comptes rendus médicaux en français. Sur un corpus de comptes rendus de scanners thoraciques annoté par nos soins, notre méthode obtient un score F1 global de 96.9%. Nous appliquons ensuite cette approche à des comptes rendus biomédicaux non radiologiques annotés de façon hétérogène. Dans ces conditions, notre méthode montre un score F1 global de 94.1%, confirmant ainsi son potentiel pour l'usage clinique.

Mots-clés

Analyse des négations, Modèles de langage, Comptes rendus médicaux, Reconnaissance d'entités nommées, CamemBERT

Abstract

We present a supervised learning approach based on a BERT language model for automatic recognition and determination of the scope of negations in medical reports in French. On a corpus of reports of chest scanners annotated by us, our method obtains an overall F1 score of 96.9%. We then apply this approach to non-radiological biomedical reports where negations have been annotated in a heterogeneous manner. Under these conditions, our method shows an overall F1 score of 94.1%, confirming its potential for clinical use.

Keywords

Negation Analysis, Language Models, Medical Reports, Named Entity Recognition, CamemBERT

1 Introduction

L'annotation des textes cliniques permet d'extraire automatiquement des informations pertinentes à partir des données textuelles brutes formulées par des cliniciens. Dans des comptes rendus textuels de scanners thoraciques, elle permet par exemple de détecter automatiquement si le patient ou la patiente souffre d'une embolie pulmonaire, d'une bronchopathie ou d'aucune pathologie thoracique [3]. Cependant la nature non structurée de ces informations implique des difficultés supplémentaires à l'analyse. Prenons par exemple deux bouts de phrases extraites de compte rendus d'imagerie :*"Absence de lésion osseuse évolutive."* et *"Aorte athéromateuse, sans dilatation anévrismale."*. Dans la première, une méthode se limitant à détecter des termes synonymes de pathologie (ici, "lésion", par exemple) échouerait puisque, justement, la négation "absence de" réfute cette pathologie. Le seconde phrase montre un exemple où la négation est utilisée pour préciser la nature de la pathologie et apporte donc une information supplémentaire importante. Afin de capturer l'information niée dans les textes cliniques, une approche séquentielle consiste à d'abord identifier les indices qui déclenchent la négation dans la phrase (par exemple les termes : "pas", "sans", "absence"...), puis de décider quel est l'effet de cet indice et si cet effet s'étend sur toute la phrase (1er exemple plus haut) ou seulement sur une partie de la phrase (2nd exemple plus haut). La portée de la négation est ainsi définie par l'ensemble des mots consécutifs dans la phrase qui font l'objet de la négation. Dans cet article, nous proposons un pipeline de reconnaissance de négations (présence + portée), basé sur les modèles de langage adaptés aux comptes rendus de radiologie. Dans un second temps, nous testons la capacité à généraliser de notre approche en l'appliquant à une variété de notes et comptes rendus cliniques rédigés en français et annotés de façon hétérogène.

2 Travaux existants

Deux familles d'approches méthodologiques sont utilisées pour la détection automatique de la détection : les systèmes à bases de règles, ou systèmes experts et la classification par apprentissage automatique supervisé.

2.1 Systèmes experts

Parmi les nombreux systèmes experts pour la détection des négations (cf e.g., [1] pour une revue en français), quelques uns s'adressent spécifiquement aux comptes rendus médicaux rédigés en français. C'est le cas par exemple, de Med-

kit [1] ou encore EDS-NLP [2], dévelopée par l'Assistance Publique–Hôpitaux de Paris (AP-HP). Les limitations des systèmes experts sont en partie dues à la présence d'abréviations et acronymes, dont la gestion est difficile pour un système expert. Mais surtout, les termes utilisés comme indice de négation sont fréquemment polysémiques et la structure de la portée parfois complexe. Par exemple, dans la phrase : *"Un pas vers une autre approche du diagnostic des masses pulmonaires périphériques."*, le terme "pas" serait détecté à tord par un système de règles comme un indice de négation, dont il ne pourrait pas gérer le contexte.

2.2 Apprentissage automatique supervisé

Dans les approches basées sur l'apprentissage automatique, la détection de négation est considérée comme une tâche d'étiquetage de séquences, qui attribue une étiquette pour les indices de négation et leur portée. A notre connaissance, le premier système de détection de négation par l'apprentissage automatique dans les textes biomédicaux en français a été introduit dans les travaux de Daloux et al. [1]. Dans le cadre de ces travaux, les auteurs ont exploré l'utilisation de plusieurs familles d'approches d'apprentissage différentes (réseaux de neurones récurrents, LSTM, étiquetages par champs aléatoires conditionnels, plongements de mots). Pour l'apprentissage, ces auteurs ont assemblé deux corpus biomédicaux annotés avec les informations de négation : le corpus d'essais clinique ESSAI [1] et le corpus de cas cliniques CAS [2]. Ces deux corpus sont disponibles en libre accès, sur demande, à des fins de recherche. Le tableau 1 montre les statistiques relatives à ces deux corpus. Plus récemment, une étude a proposé d'affiner le modèle de langage CamemBERT en l'adaptant à un contexte de radiologie en langue française [10]. L'approche du présent article est comparable à cette dernière.

	ESSAI	CAS	RADIO
Nombres des phrases	7 248	3 790	11252
Nombres des phrases négatives	980	838	2665
Nombes des indices de négation	56	74	13

TABLE 1 – Statistiques relatives aux corpus CAS, ESSAI et RADIO.

3 Détection de négation en radiologie

3.1 Création d'un corpus

Le développement de systèmes d'apprentissage automatique nécessite des corpus annotés. Pour adapter notre approche au contexte radiologique, nous avons assemblé un corpus de référence annoté avec les informations de négation (indice, portée). Nous disposons pour cela d'un corpus initial composé de plus de 30 000 phrases anonymisées issues de comptes rendus de CT-scans thoraciques fournies par le service de radiologie des Hospices Civils de Lyon (HCL). Les phrases du corpus ont été annotées par des radiologues selon 23 pathologies thoraciques (anomalie pleurale, anomalie cardiaque...). Les 11 252 phrases spécifiques à un examen thoracique ont été conservées avec leur annotation pour former le corpus "RADIO". Elles comprennent 2 665 phrases marquées par la négation par 13 indices de négation différents (cf tableau 1). Les trois indices de négation les plus fréquents sont : *"sans"* (1429 occurrences), *"non"* (676) et *"pas"* (568). Les trois indices de négation les moins fréquents sont : *"exclure"* (1), *"jamais"* (1) et *"n'...plus"* (2).

3.2 Protocole d'annotation

Nous avons annoté la présence des informations de négation dans les phrases, une étape effectuée par un unique annotateur. Deux types de négation ont été distinguées :

- La négation partielle [8] : elle porte sur une partie ou un élément de la phrase. Exemple : "Vésicule biliaire non [distendue], alithiasique, à parois fines."
- La négation totale [8] : Elle porte sur la phrase entière. Exemple : "[La dérivation cardiopulmonaire] n' [est] pas [opacifiée]."

À partir de ce descriptif du domaine de la négation linguistique [5] (pour repérer les portées selon le type de négation), le processus d'annotation a été réalisé selon cinq étapes :

1. Distinguer les phrases portant une négation totale de celles portant une négation partielle.
2. Repérer tous les indices (pas, absence, aucun...) qui déclenchent la négation.
3. Annoter la séquence des tokens impactés par chaque indice de négation.
4. Utiliser le format d'annotation de séquences BIO (Begin, Inside, Outside) pour créer la reconnaissance d'entités nommées (NER) correspondante.
5. Utiliser TreeTagger [3], un outil automatique d'annotation morphosyntaxique, pour lemmatiser les phrases du corpus.

Dans le tableau 2, nous présentons un exemple d'une phrase lemmatisée et annotée avec l'indice de négation et sa portée en format d'annotation BIO.

3.3 Modèles de langage pour la radiologie

Nous traitons la tâche de détection de la négation comme une tâche de reconnaissance de deux entités nommées (indice, portée). Pour cela nous utilisons le modèle de langage pré-entraîné CamemBERT-base comme modèle de langage initial [4], et l'ajustons de façon non supervisée à l'aide d'un Masked Language Model [7](MLM). Pour cela, des mots entiers ont été masqués et chaque phrase de l'ensemble de données d'apprentissage a été tokenisée à l'aide de la version 'CamemBERT-base' du tokenizer CamemBERT. Le Masked-Language-Model de CamemBERT a ensuite été entraîné pendant 30 époques sur un corpus de

1. https://github.com/TeamHeka/medkit
2. https://github.com/aphp/edsnlp
3. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger

Words	Words-lim	Labels
Malheureusement	malheureusement	O
le	le	B-PORTÉE-NEG
temps	temps	I-PORTÉE-NEG
artériel	artériel	I-PORTÉE-NEG
n’	n’	B-INDICE-NEG
a	avoir	I-PORTÉE-NEG
pas	pas	I-INDICE-NEG
été	être	I-PORTÉE-NEG
réalisé	réaliser	I-PORTÉE-NEG

TABLE 2 – Extrait d’une phrase de corpus RADIO limmatisé et annoté en format BIO.

3 3875 phrases anonymisées et lemmatisées. Un mélange aléatoire et un fractionnement en 95 :5 ont été effectués pour créer un ensemble d’entraînement et un ensemble de validation. L’apprentissage a été effectué avec une taille de batch de 8 et un optimiseur Adam avec un learning rate de 2.10^{-5}.

3.3.1 Résultats

L’ajustement du modèle nous a permis de diminuer la perplexité du CamemBERT initial à 1.41 avec une valeur de loss égale à 0.45 sur l’ensemble de validation. Le tableau 3 montre la probabilité obtenue sur les prédictions des mots masqués en comparant le modèle CamemBERT-base et le modèle affiné pour la même phrase :*"absence de <mask>"*. Ceci illustre la forte amélioration de la pertinence et de la certitude du mot proposé avec la version affinée par rapport au modèle CamemBERT-base pré-entraîné.

Avant le réglage fin		**Aprés le réglage fin**	
Mots	Probabilité	Mots	Probabilité
Commentaire	2.32%	Fracture	59.97%
Commentaires	1.75%	Collection	15.42%
Photos	1.29%	Condensation	13.87%

TABLE 3 – Probabilité de détection des mots masqués avant et après l’affinage du modèle CamemBERT

3.4 Détection de négation

La deuxième étape de notre pipeline consiste à affiner le modèle CamemBERT ajusté sur la tâche de reconnaissance de la négation. Nous abordons cette tâche comme une tâche de classification de tokens. Nous formons ainsi un modèle NER pour identifier deux entités (indice, portée) dans les phrases du corpus RADIO présenté ci-dessus. Cette tâche a été implémentée à l’aide d’HuggingFace Transformers [9], en passant par un format de données NER HuggingFace personnalisé ([4]) pour l’entraînement. Pour alimenter la liste des tokens dans notre modèle, les tokens doivent être convertis en identifiants. Pour cela, le tokenizer de CamemBERT-base a été augmenté d’une liste de mots ayant subit une élision (n’, l’,...) afin de gérer les espaces blancs lors de la tokenisation des apostrophes. Les potentiels décalages entre les annotations résultant de ce processus ont été corrigés par une fonction permettant d’aligner et associer chaque token à son label [5].
Une fois les données préparées, alignées et tokenisées, un modèle de langage basé sur les transformateurs est formé puis évalué. Le classifieur de tokens (AutoModelForTokenClassification) est formé de manière supervisée sur la base du modèle de langage affiné, en utilisant les mêmes hyperparamètres que la tâche MLN 3.3 ci-dessus. Pour évaluer les performances du modèle, nous avons utilisé les mesures suivantes : Précision (P), Rappel (R) et F1-score (F1). Le modèle a été validé sur 1 126 phrases dont 252 négatives.

3.4.1 Résultats

Le tableau 4 présente les résultats obtenus sur l’ensemble de validation en comparant le modèle CamemBERT-base et le modèle ajusté, nommé CamemBERT-RADIO, sur la tâche de détection des indices de négation et d’identification de leur portée. Ce tableau montre que les meilleures performances sont obtenues avec le modèle adapté à la radiologie, CamemBERT-RADIO. Les deux modèles arrivent à détecter la quasi-totalité des indices de négation, avec un F1-score très élevé (99.10%). Les autres mesures sont elles-aussi très élevées. En revanche, le tableau montre que l’identification de la portée de la négation est une tâche plus difficile à effectuer que la simple détection : en particulier les F1-scores diminuent à 94.72% pour le modèle CamemBERT-RADIO et 92.99% pour le modèle CamemBERT-base.Une diminution comparable est obtenue pour les autres mesures. En général, la portée de négation se compose de plusieurs tokens de la séquence. Par conséquent, prédire correctement tous les tokens annotés reste une tâche difficile avec peu de données annotées. Cependant les résultats montrent que l’ajustement du modèle initial (CamemBERT-base) sur les données du corpus RADIO (CamemBERT-RADIO) permet une amélioration significative des performances, avec une augmentation des mesures de 1 à 2 %.

3.5 Visualisation et analyse d’erreurs

La visualisation directe des deux entités (indice, portée) permet d’illustrer les principales sources d’erreur. Sur les exemples montrés Fig. 1, les tokens incorrectement prédits dans trois phrases lématisées de l’ensemble de validation sont soulignés. Une des erreurs fréquemment commises apparaît dans la phrase 1, où un tokenizer inadapté à un domaine spécifique (radiologie) produit des erreurs et crée une portée discontinue lors de la tokenization d’un mot hors vocabulaire (“sténose”, ici). Dans les deux phrases qui suivent, le modèle échoue à détecter les limites (fin, début) de la portée de négation. Dans la phrase 2, le modèle a incorrectement inclus *ce jour* dans la portée tandis que dans la phrase 3 le modèle a exclu *acquisition*. Ce type d’erreur se produit généralement dans les phrases niées où la portée est discontinue.

4. https://huggingface.co/datasets/conllpp

5. https://huggingface.co/course/chapter7/2

Modèles	INDICE-NEG			PORTÉE-NEG		
	F1(%)	R(%)	P(%)	F1(%)	R(%)	P(%)
CamemBERT-base	99.10	98.81	99.40	92.99	92.73	93.58
CamemBERT-RADIO	99.10	98.81	99.40	**94.72**	**94.85**	**94.59**

TABLE 4 – Précision, Rappel et F1-score pour les deux modèles de détection de négation en Radiologie.

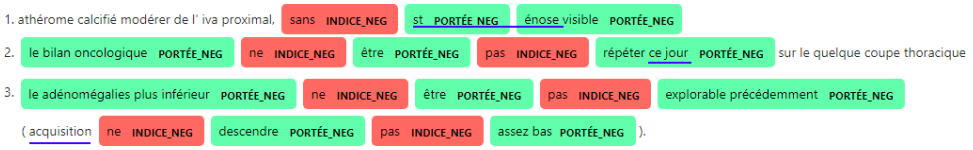

FIGURE 1 – *Visualisation de quelques erreurs de prédictions*

4 Généralisation de l'approche

Les bonnes performances de notre modèle CamemBERT-RADIO pour l'analyse de la négation sur le corpus de CT scans thoraciques sont encourageantes (tableau 4), mais les conditions dans lesquelles elles ont été obtenues restent éloignées d'une application en routine clinique. Les phrases négatives ont été annotées par un seul annotateur, avec des règles d'annotation très homogènes. Dans un contexte clinique, la règle définissant l'identification d'une négation et surtout sa portée varie probablement bien davantage, d'un expert à l'autre. D'autre part, le corpus utilisé pour l'ajustement du modèle est limité aux comptes rendus de CT-scans thoraciques. Il serait plus utile de disposer d'un outil d'analyse de négation efficace dans un contexte plus large de comptes rendus médicaux au delà des scanners thoraciques. Pour aborder ces questions nous avons fait appel à deux corpus de comptes rendus biomédicaux, ESSAI et CAS (cf sect. 2.2), portant sur des essais et des cas cliniques, respectivement, et annotés par d'autres experts, selon des règles non-homogènes. Le tableau 5 résume les performances obtenues pour quatre scénarios envisagés.

CamemBERT-RADIO : Le premier scénario consiste à tester la performance du modèle CamemBERT-RADIO précédent sur les corpus CAS et ESSAI fusionnés. Le tableau 5 montre que CamemBERT-RADIO échoue à détecter plusieurs indices de négation d'ESSAI et CAS (F1-score de 87.96%). En conséquence ses performances pour la détection de la portée des négations sont elles-aussi très limitées (F1-score de 68.67%). Notons que ces résultats étaient attendus dans la mesure où les corpus ESSAI et CAS contiennent bien plus d'indices de négation que RADIO (tableau 1), si bien que de nombreux indices de négation d'ESSAI et CAS n'apparaissent pas dans le corpus d'apprentissage RADIO.

CamemBERT-ESSAI-CAS : Ce second scénario consiste à effectuer le processus inverse du premier, i.e. à ajuster le modèle initial CamemBERT-base sur la fusion des corpus ESSAI et CAS et tester le modèle ainsi ajusté sur le corpus RADIO. Le modèle ainsi obtenu montre un score F1 relativement élevé (91.34%). La performance du modèle pour la détection des portées de négation reste relativement plus faible (75.83%), mais cette valeur est à contre-balancer par le fait que les deux corpus (ESSAI-CAS d'un côté et RADIO de l'autre) concernent des domaines médicaux très différents et ont été annotés de façon très hétérogène. Ces résultats permettent donc d'envisager l'utilisation de notre approche dans un contexte clinique réaliste.

CamemBERT-RADIO-ESSAI : L'objectif ici est de tester les conclusions précédentes, en ajustant le modèle Camembert-RADIO sur le corpus ESSAI (corpus d'apprentissage fusionné RADIO+ESSAI), puis en testant la performance du modèle obtenu sur le corpus CAS. Les résultats obtenus (tableau 5) montrent que l'intégration du corpus CAS avec la base d'apprentissage RADIO a fourni davantage d'exemples et d'informations sur la négation, ce qui a permis d'augmenter le score F1 pour la détection des indices (94.53%). Cette augmentation de la cohérence et de la compatibilité entre les textes et les annotations a aussi permis d'améliorer les résultats pour la détermination de la portée, avec une amélioration par rapport à l'expérience précédente qui se situe entre 4 et 9%. Ces résultats soulignent l'importance de la cohérence entre les données utilisées pour former le modèle et celles utilisées pour le tester.

CamemBERT-RADIO-ESSAI-CAS : Le dernier scénario consiste à maximiser la quantité d'informations et de critères d'annotation du corpus. Pour cela, les trois corpus annotés (RADIO, CAS, et ESSAI) on été fusionnés, puis additionnés de 111 phrases extraites du corpus médical QUAERO [6], annotées par nos soins. Le résultat est un corpus de 22 400 phrases, dont plus de 4 500 négatives (corpus CamemBERT-RADIO-ESSAI-CAS du tableau 5). 1 120 phrases ont été extraites aléatoirement de ce corpus pour former l'ensemble de test (dont 450 phrases négatives). Pour l'adaptation au domaine, le modèle CamemBERT-RADIO a été ajusté avec la fusion des corpus CAS et ESSAI. Les scores obtenus sont bien plus élevés que les scénarios précédents, avec un score F1 de 97.20 % pour la détection des indices de négation. Ce modèle est donc bien adapté aux différents indices de négation qui apparaissent dans les différents corpus. Pour la détection de la portée, ce modèle a obtenu un score F1 certes plus faible

MODEL	CORPUS TRAIN	CORPUS TEST	INDICE-NEG			PORTÉE-NEG		
			F1(%)	R(%)	P(%)	F1(%)	R(%)	P(%)
CamemBERT-RADIO	RADIO	ESSAI + CAS	87.96	87.03	87.96	68.76	71.19	66.50
CamemBERT-ESSAI-CAS	ESSAI + CAS	RADIO	91.34	90.23	.92.49	75.83	76.32	75.35
CamemBERT-RADIO-ESSAI	RADIO+ESSAI	CAS	94.53	94.97	94.08	82.18	80.21	84.15
CamemBERT-RADIO-ESSAI-CAS	RADIO+ESSAI+CAS	10% TRAIN	**97.20**	**96.61**	**97.81**	**91.04**	**90.90**	**91.17**

TABLE 5 – Précision, Rappel et F1-score pour les 4 scénarios envisagés.

mais qui reste très élevé (91.04%). Nous en concluons donc que ce modèle est performant pour l'analyse de la négation (identification et portée) dans des comptes rendus médicaux d'origines diverses et annotées de façon possiblement non-homogène.

Conclusion

Dans cet article, nous proposons l'utilisation de modèles de langages adaptés au domaine médical pour la reconnaissance des informations niées dans les textes cliniques rédigés en français. Pour ce faire, nous utilisons deux corpus biomédicaux publiques ayant été annotés avec des informations liées à la négation et un corpus privé de textes radiologiques (CT scan thoraciques) issu des Hospices Civils de Lyon, dont les négations ont été annotées par nous-mêmes. Après adaptation du modèle de langage CamemBERT-base au corpus radiologique, le modèle a été affiné pour la tâche de la détection des indices de négation mais aussi de leur portée, toujours en radiologie. Nos résultats confirment que l'adaptation du domaine aide à améliorer les performances du modèle pour la reconnaissance de la négation et sa portée (scores F1 respectifs de l'ordre de 99% et 95%). A l'aide de plusieurs scénarios de combinaison des corpus disponibles, nous montrons que l'augmentation de la taille du jeu de données annoté, de la cohérence entre les textes cliniques provenant de divers domaines médicaux ainsi que celle des critères d'annotation impactent positivement les performances de notre système de détection de la négation. Cette étude permet de proposer deux modèles de langage pour le traitement automatique des négations et leur portée : un premier modèle spécifique et bien adapté à la radiologie thoracique, et un second modèle plus généraliste, entraîné sur plus de 4 000 phrases négatives annotées et qui montre une bonne performance sur des comptes rendus issus de différents domaines médicaux (radiologie, essais cliniques, cas cliniques etc). En particulier, un des avantages importants de notre approche est de traiter correctement la polysémie des termes de négation. Par exemple, dans la phrase *"Un pas vers une autre approche du diagnostic des masses pulmonaires périphériques."* évoqué dans la section 2.1 plus haut, notre approche ne considère pas le terme "pas" comme un indice de négation.

Références

[1] C. Dalloux and V. Claveau. Détection de la négation : corpus français et apprentissage supervisé. *Revue des Sciences et Technologies de l'Information*, pages 83–103, 2019.

[2] N. Grabar, C. Grouin, T. Hamon, and V. Claveau. Corpus annoté de cas cliniques en français. In *TALN 2019-26e Conference on Traitement Automatique des Langues Naturelles*, pages 1–14, 2019.

[3] E. Jupin-Delevaux, A. Djahnine, F. Talbot, A. Richard, S. Gouttard, A. Mansuy, P. Douek, S. Si-Mohamed, and L. Boussel. BERT-based natural language processing analysis of french ct reports : Application to the measurement of the positivity rate for pulmonary embolism. *Research in Diagnostic and Interventional Imaging*, 6 :100027, 2023.

[4] L. Martin, B. Muller, P. J. Ortiz Suárez, Y. Dupont, L. Romary, É. de la Clergerie, D. Seddah, and B. Sagot. CamemBERT : a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, 2020.

[5] C. Muller. *La négation en français : syntaxe, sémantique et éléments de comparaison avec les autres langues romanes*. Librairie Droz, 1991.

[6] A. Névéol, C. Grouin, J. Leixa, S. Rosset, and P. Zweigenbaum. The Quaero french medical corpus : a resource for medical entity recognition and normalization. In *Fourth Workshop on Building and Evaluating Ressources for Health and Biomedical Text Processing - BioTxtM2014*, pages 24–30, 2014.

[7] D. Nozza, F. Bianchi, and D. Hovy. What the [mask] ? making sense of language-specific bert models. *arXiv :2003.02912*, 2020.

[8] C. Touratier. La portée de la négation ? *Revue de Linguistique Latine du Centre Alfred Ernout (De Lingua Latina)*, 1, 2008.

[9] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al. Transformers : State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing : system demonstrations*, pages 38–45, 2020.

[10] A. Yan, J. McAuley, X. Lu, J. Du, E. Y. Chang, A. Gentili, and C.-N. Hsu. RadBERT : Adapting transformer-based language models to radiology. *Radiology : Artificial Intelligence*, 4(4) :e210258, 2022.