

Détection de métaphores dans les documents médicaux

Coralie Pottiez¹ Thierry Hamon^{2,3} Natalia Grabar¹

(1) CNRS, Univ. Lille, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France

(2) Université Paris-Saclay, CNRS, LISN, 91405 Orsay, France

(3) Université Sorbonne Paris Nord, France

coralie.pottiez@orange.fr, thierry.hamon@lisn.fr,

natalia.grabar@univ-lille.fr

RÉSUMÉ

La métaphore est une figure de style, qui permet de transférer le sens d'un terme source vers un terme cible, comme dans *LE TEMPS C'EST DE L'ARGENT*. De cette manière, la métaphore identifie des similarités cachées entre deux idées. La métaphore peut jouer plusieurs rôles dans la langue, comme l'embellir, structurer la pensée ou expliquer des notions complexes. Nous nous intéressons à la métaphore utilisée dans le domaine médical. Nous proposons d'abord une typologie de métaphores et un corpus de cas cliniques annoté avec des emplois métaphoriques. Nous effectuons également des expériences de détection automatique des métaphores avec un giga-modèle génératif. Plusieurs types de *prompts* sont testés. Les meilleurs résultats atteignent 67,50 de F-mesure, avec le rappel allant jusqu'à 74 % avec certains *prompts*. Le typage de métaphores montre que 45,51 % de métaphores sont typés correctement.

ABSTRACT

Detection of metaphors in medical documents.

Metaphor is a figure of speech, which permits to transfer the meaning of a given source term towards a target term, like in *TIME IS MONEY*. In this way, metaphor identifies hidden similarities between two different ideas. Metaphor can play several roles in the language, like embellish it, structure the thought or explain complex notions. We address the metaphor occurring in the medical domain. We first propose a typology of metaphors and a metaphor-annotated corpus with clinical cases. Then, we run the experiments in automatic detection of metaphors with a generative LLM. Several types of prompts are tested. The best results reach 67,50 F-measure, with a recall going up to 74% with some prompts. The assignment of types to metaphors shows that 45,51% of metaphors are characterized correctly.

MOTS-CLÉS : Métaphore, typologie, domaine médical, détection, giga-modèle de langue.

KEYWORDS: Metaphor, typology, medical domain, detection, LLM.

1 Introduction

La métaphore est une figure de style, qui permet de transférer le sens d'un terme ou domaine source vers un terme ou domaine cible en se basant sur l'analogie, comme dans *notre passé nous poursuit*, où *le passé* reçoit des caractéristiques humaines. Les métaphores sont assez répandues dans la langue : les observations en corpus montrent par exemple que jusqu'à 13,6 % de mots sont utilisés

métaphoriquement (Krennmayr & Steen, 2017). La presse et la politique semblent être les plus gourmands en métaphores (Shutova, 2017). La métaphore peut jouer plusieurs rôles dans la langue. Selon Aristote, elle est utilisée pour embellir la communication en la rendant plus imagée (Veyne & Tamba, 1979). Elle joue également un rôle dans la structuration de notre pensée. Lakoff & Johson (1980) proposent ainsi que la métaphore influence notre perception et notre compréhension du monde. Par exemple, la métaphore LE TEMPS C'EST DE L'ARGENT permet de percevoir le temps comme une ressource précieuse à gérer, économiser et dépenser. Les auteurs indiquent que cette métaphore se réalise avec des expressions métaphoriques comme : *gagner du temps, perdre du temps, investir du temps*. Dans la communication quotidienne, la métaphore permet aussi de rendre les idées abstraites plus concrètes et de transmettre les concepts complexes en les comparant à des expériences réelles (Jamet, 2008; Rollo, 2015).

Dans le domaine médical, la métaphore est également fréquente. Elle aide les professionnels de santé à expliquer des conditions médicales et à décrire des traitements de manière accessible aux patients et aux étudiants en médecine (Charteris-Black, 2004; Limaïem *et al.*, 2021). Elle permet de créer une image claire et d'améliorer la communication entre les médecins et les patients (Marquer, 2019). Par exemple, lorsqu'un patient est diagnostiqué avec un cancer, la métaphore de GUERRE est mobilisée : les médecins parlent souvent d'une *bataille*, ils évoquent des *stratégies* pour *attaquer* les cellules cancéreuses et disent au patient de *ne pas baisser les bras*. Cette métaphore aide les patients à visualiser leur traitement comme une série d'actions coordonnées pour vaincre un *ennemi*.

L'objectif de notre travail consiste à détecter automatiquement les expressions métaphoriques dans les textes médicaux de type clinique. Nous commençons par une revue de la littérature sur différentes questions liées (section 2). Ensuite, nous présentons le corpus exploité (section 3.1) et la méthodologie utilisée pour constituer et annoter le corpus des textes médicaux (section 3.2). Nous présentons l'approche pour la détection automatique des expressions métaphoriques (section 3.3). Les résultats sont présentés dans la section 4. Une conclusion et les travaux futurs sont présentés dans la section 5.

2 Travaux existants

Nous présentons quelques typologies des métaphores, des méthodes d'annotation et des corpus existants annotés avec des emplois métaphoriques.

2.1 Typologies des métaphores

Il existe plusieurs typologies de métaphores, qui abordent cette notion de différents points de vue.

Dans les manuels scolaires, trois types de métaphores sont distingués, selon la manière d'exprimer et d'explicitement leur construction : les métaphores annoncées (métaphores *in praesentia* ou métaphores explicites, qui établissent explicitement le rapport entre le comparant et le comparé, comme dans *Le cancer est un ennemi qui s'attaque à notre corps*), les métaphores directes (métaphores *in absentia* ou métaphores contextuelles, qui comparent deux entités de manière implicite, comme dans *la maladie a envahi son corps*, où l'image de *maladie* n'est pas directement associée avec celle d'*ennemi*) et les métaphores filées, dans l'exemple en (1), où la métaphore se développe sur plusieurs phrases et utilise ici des images liées à la guerre pour décrire la progression de la maladie et le combat mené.

- (1) *Le cancer est un ennemi invisible. Il s'infiltré silencieusement et attaque les cellules. Chaque nouvelle métastase est une nouvelle bataille. La chimiothérapie est notre armée, les médicaments nos soldats et chaque traitement est une attaque planifiée pour repousser l'envahisseur.*

Une autre typologie focalise sur l'évolution et le figement des métaphores et différencie les métaphores mortes (déjà lexicalisées et devenues très courantes dans la langue, comme *pression artérielle* ou *rythme cardiaque*) et les métaphores vives (qui conservent leur caractère imaginaire, comme *le cancer est un ennemi invisible*) (Ricoeur, 1975).

La typologie des métaphores conceptuelles distingue d'abord trois catégories (métaphores structurales, d'orientation et ontologiques) (Lakoff & Johson, 1980) et ensuite deux catégories (métaphores structurales et ontologiques) (Lakoff & Johnson, 2003). Les métaphores structurales organisent un domaine cible avec des termes d'un domaine source et permettent d'exprimer un concept grâce à un autre concept, comme dans cette image de GUERRE : *Combat contre les allergies alimentaires, bombardement des cellules, agressivité de la maladie, victimes de la cancérose* (Rollo, 2015). Les métaphores ontologiques concernent la manière de conceptualiser une expérience. Elles permettent de définir des expériences abstraites, comme dans ces exemples : *Cellules NK "Natural Killers", petite colonie de bactéries, croissance / mort des cellules cancéreuses* (Rollo, 2015). Les métaphores d'orientation organisent des concepts en fonction de l'espace, en donnant une direction haut-bas, devant-derrrière, etc. Par exemple, le concept de santé est souvent orienté vers le haut *être au sommet de sa forme*, tandis que la maladie est orientée vers le bas *être au plus bas* (Lakoff & Johson, 1980).

Une autre typologie s'est intéressée spécifiquement aux métaphores du domaine médical (Salager-Meyer, 1990; Divasson & León, 2006). La différenciation est effectuée par rapport aux notions médicales (terme source) et aux domaines cibles (terme cible), avec deux grandes catégories :

1. Les *métaphores morphologiques ou structurales* concernent la forme et la structure des objets médicaux, avec la distinction de métaphores architecturales (*cloison membranaire, pont musculaire*), géomorphique (*plateaux vertébraux, fosse nasale*), phytomorphique (*bourgeon génital, rameau supérieur*), anatomique (*bras de la capsule, corps vertébraux*), zoomorphique (*bruit de galop, fourmillement*). Notons que la métaphore structurale se trouve à la base de la terminologie médicale (Skoda, 1988) ;
2. Les *métaphores physiologiques ou fonctionnelles* font référence aux processus, fonctions et relations biologiques, et conçoivent des images plus dynamiques, comme dans *vérouillage du genou* ou *abdomen ballonné*.

2.2 Annotation des métaphores

La formation d'une métaphore est décrite de manière assez claire : il s'agit du transfert de sens d'un terme ou domaine source vers un terme ou domaine cible. Cependant, l'annotation des métaphores dans la langue présente des difficultés liées à la variabilité contextuelle, à l'interprétation des métaphores (Steen *et al.*, 2010b), à l'ambiguïté qu'elle peut introduire (Semino *et al.*, 2016) et à l'ancienneté de la métaphore. Deux méthodes principales ont été proposées pour effectuer cette annotation.

MIP (Metaphor Identification Procedure) (Group, 2007) repose sur plusieurs étapes. Elle commence par la lecture du texte. Ensuite, pour chaque unité lexicale, il faut effectuer plusieurs étapes d'analyse : déterminer le sens contextuel, déterminer le sens premier, et comparer les deux sens. Si

une unité lexicale peut être comprise par rapport à un autre domaine que son sens premier, son emploi est alors considéré comme métaphorique. Dans la méthode MIP, le sens de base est supposé être plus concret et souvent plus ancien que le sens contextuel, ce qui peut être discutable, car le sens de base peut également être contemporain et dominant.

MIPVU (Metaphor Identification Procedure VU University Amsterdam) (Steen *et al.*, 2010a) est une extension de la méthode MIP. Elle apporte des améliorations en termes de systématique et de reproductibilité des annotations. Après les étapes de la méthode MIP, d'autres étapes sont ajoutées, qui permettent de typer directement les emplois métaphoriques : (1) si un mot est utilisé de manière indirecte et si cette utilisation peut être expliquée par un transfert interdomaine, le mot est annoté comme une métaphore ; (2) si un mot est utilisé directement avec un transfert interdomaine à partir d'un référent ou d'un sujet plus basique, ce mot est annoté comme une métaphore directe ; (3) en cas d'anaphores, il faut restituer le référent et, s'il y a une métaphore, la marquer comme une métaphore implicite ; (4) si possible, marquer les mots qui servent d'indicateurs du transfert interdomaine. Notons que, dans cette méthode d'annotation, le sens le plus ancien n'est pas forcément le sens le plus littéral.

2.3 Corpus annotés avec des emplois métaphoriques

Parmi les corpus annotés avec des emplois métaphoriques librement disponibles, nous pouvons mentionner trois corpus très utilisés. Le corpus **VUA (VU Amsterdam Metaphor Corpus)** (Krennmayr & Steen, 2017) est sans doute le plus gros des corpus disponibles, avec environ 200 000 occurrences. Il est annoté selon la méthode MIPVU. Il est très utilisé pour l'étude des métaphores en anglais (Badryzlova *et al.*, 2013; Dunn, 2013). Il contient des documents de diverses sources : textes académiques, conversations, fiction et articles de presse. Le corpus **METCLIL (Metaphor in English as a Medium of Instruction seminars)** (Alejo *et al.*, 2021) est dédié aux métaphores dans les séminaires académiques, où l'anglais est utilisé comme langue d'enseignement. Il montre comment les enseignants et les étudiants utilisent les métaphores pour faciliter la compréhension de concepts complexes dans un contexte académique. Le **CoMeta Corpus** est un corpus annoté avec les métaphores en espagnol. Il contient des textes provenant de diverses sources (corpus linguistiques, comme AnCora et GSD, et discours politiques). Le corpus est annoté selon la méthode MIPVU. CoMeta a été utilisé pour des tâches de détection de métaphores à la fois multilingues et cross-langues. Par exemple, en combinant les données de CoMeta en espagnol avec celles du corpus VUA en anglais, des expériences ont été menées pour effectuer la détection de métaphores cross-langues avec des approches supervisées (Sanchez-Bayona & Agerri, 2022).

2.4 Identification automatique des emplois métaphoriques

Il existe plusieurs tâches liées aux métaphores, dont certaines sont motivées par la compétition FigLang : identification des métaphores (Leong *et al.*, 2018), identification des métaphores selon les parties de discours (Leong *et al.*, 2020), compréhension et explication des métaphores grâce à l'inférence textuelle (Ghosh *et al.*, 2022), interprétation des métaphores (Veale & Hao, 2008; Mao *et al.*, 2019) et leur désabiguïsation (Li *et al.*, 2024). Les travaux sur l'identification des métaphores sont les plus nombreux. Comme dans d'autres tâches de TAL, les méthodes utilisées pour cette identification ont évolué au fil du temps (Shutova, 2015) allant de méthodes basées sur les règles et les ressources (Birke & Sarkar, 2006; Shutova *et al.*, 2010), en passant par l'apprentissage non supervisé (Heintz *et al.*, 2013) et supervisé (Gedigian *et al.*, 2006; Mohler *et al.*, 2013), pour déboucher

actuellement sur les Transformers et les giga-modèles de langues (Dankers *et al.*, 2020; Li *et al.*, 2023). Les chercheurs considèrent qu’il s’agit d’une tâche difficile, qui présente de nombreux défis (Ptiček & Dobša, 2023), où les méthodes par apprentissage atteignent autour de 70 % de précision et 65 % de rappel (Shutova, 2015), tandis qu’avec les modèles de langue, la F-mesure augmente et se trouve entre 0,56 (Do Dinh & Gurevych, 2016) et 0,79-0,80 (Li *et al.*, 2023; Zhang & Liu, 2022).

La métaphore reste également un procédé productif dans les écrits de santé. Les chercheurs ont par exemple étudié la métaphore dans la création des termes de différentes spécialités médicales (Rollo, 2015; Gingras Harvey, 2014) et dans les structures verbales (Vandaele *et al.*, 2006). De tels travaux sont effectués sur les textes de la littérature scientifique et se concentrent sur l’annotation manuelle.

3 Approche

Nous présentons différents aspects de notre approche : le corpus, la typologie des métaphores et l’annotation manuelle du corpus, la détection automatique des métaphores et son évaluation.

3.1 Données

À notre connaissance, il n’existe pas de corpus annotés en métaphores en français librement disponibles, en particulier dans le domaine médical. Nous utilisons des textes de type clinique, relatifs aux dossiers médicaux des patients, provenant du corpus de cas cliniques CAS (Grabar *et al.*, 2019). Les cas cliniques décrivent des situations médicales réelles et désidentifiées ou bien des situations fictives. Nous effectuons le travail avec 30 cas cliniques.

3.2 Typologie en annotation manuelle des métaphores

Comme nous travaillons sur des textes médicaux, nous nous inspirons de la typologie existante (Salager-Meyer, 1990), que nous adaptons à nos objectifs. La figure 1 présente la typologie que nous établissons, avec plusieurs catégories. Les métaphores médicales sont directement liées aux notions médicales. Elles sont couramment utilisées pour décrire des phénomènes biologiques, des traitements, des diagnostics, etc. Nous distinguons huit types de métaphores, selon les domaines sources des termes :

- La métaphore DE GUERRE positionne les notions médicales, comme les maladies et les traitements, dans le contexte de combat. Elle évoque des notions d’agression, de défense et de bataille et souligne la nature combative des situations médicales, comme le terme *envahissement* dans : *La compression urétérale pelvienne était secondaire à un envahissement ganglionnaire iliaque massif visible sur le scanner abdomino-pelvien.*
- La métaphore de VOYAGE compare les processus médicaux à un voyage ou à un déplacement. Elle peut illustrer par exemple les parcours et les étapes des traitements médicaux, suggérant ainsi une progression ou un chemin à parcourir, comme le terme *étape* dans : *Dans une première étape, le diagnostic d’abcès prostatique avait été posé, mais un drainage par prostatotomie n’avait apporté aucune amélioration.*
- La métaphore de la MÉCANIQUE compare le corps humain à une machine ou à un dispositif mécanique. Elle illustre aussi comment ces systèmes peuvent parfois tomber en panne ou

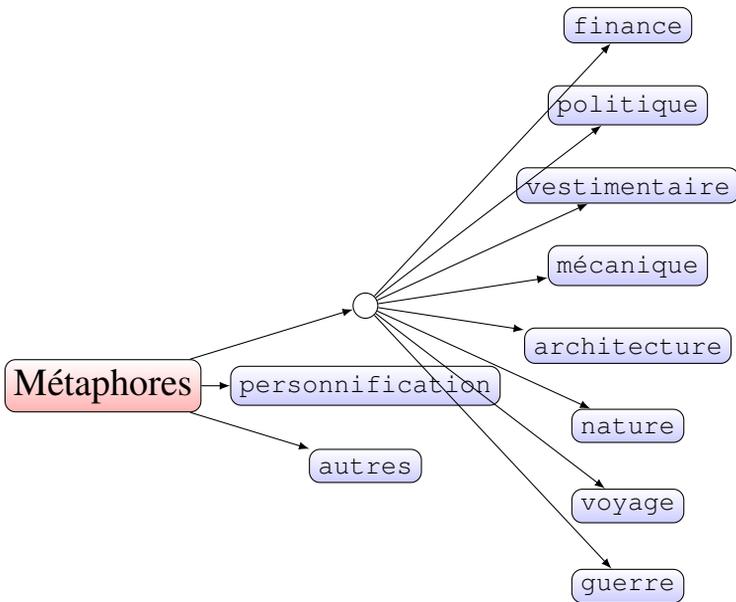


FIGURE 1 – Typologie des métaphores considérées dans notre travail

nécessiter des réparation, comme *calibre* dans : *une vessie d'aspect normal et un urètre de bon calibre sur le cliché permictionnel*.

- La métaphore de l'ARCHITECTURE compare les structures anatomiques ou les processus médicaux à des éléments de construction ou d'architecture. Elle évoque la solidité, la structure et l'assemblage des parties du corps, comme le terme *structure* dans : *Une échographie a permis de mettre en évidence une importante dilatation urétéro-pyélocalycielle bilatérale en amont d'une énorme structure hyperéchogène vésicale générant un cône d'ombre postérieure*.
- La métaphore de NATURE utilise des termes liés à l'environnement pour décrire des processus médicaux. Elle évoque par exemple des notions de croissance et de culture, comme le terme *bourgeonnante* dans : *A l'examen clinique nous retrouvons une lésion tumorale volumineuse, indolore, exophytique, bourgeonnante, irrégulière du pénis latéro-postérieure supérieure*.
- La métaphore de la FINANCE compare les processus médicaux avec des concepts financiers ou économiques, comme le terme *bilan* dans : *Le bilan d'extension était sans anomalies*.
- La métaphore POLITIQUE associe le processus de soins à un protocole et la démarche médicale à la gestion, comme le terme *administration* dans : *l'équipe traitante entreprend l'administration de clonidine par voie orale à une dose de 0,05 mg deux fois par jour*.
- La métaphore VESTIMENTAIRE fait le parallèle entre les éléments de vêtements et les processus ou procédés médicaux, comme le terme *paletot* dans : *Le traitement a consisté en une réduction urétrale avec urétroplastie, suture en paletot et reconstruction du phallus*.

À côté de cela, nous distinguons la *personnification*, qui apparaît lorsqu'un mot ou une expression attribue des caractéristiques humaines à des objets, des animaux ou des idées abstraites, comme dans cet exemple : *L'échographie montrait une masse tissulaire hétérogène du rein droit faisant 12x7 cm, bien encapsulée, et sans envahissement de la veine rénale*. Une dernière catégorie regroupe les métaphores AUTRES, que nous n'avons pas pu classer dans les catégories établies, comme *régime* dans : *Son introduction au régime thérapeutique causera assurément une augmentation des*

Nous pouvons donc voir que certains types de métaphores sont communs avec la typologie de **Salager-Meyer (1990)**, comme les métaphores architecturales, géomorphiques/phytomorphique et nature. En même temps, en nous basant sur la spécificité du corpus, nous avons défini aussi des types de métaphores spécifiques : de guerre, de voyage ou vestimentaire.

Nous avons choisi la méthode MIP car notre objectif consiste d'abord à identifier les métaphores, puis à les catégoriser selon notre propre typologie spécifique aux textes médicaux, tandis que la méthode MIPVU intègre une typologie prédéfinie de métaphores. Notre méthode d'annotation permet d'identifier les expressions métaphoriques en comparant le sens contextuel et le sens premier des mots. Cette comparaison est effectuée en utilisant des dictionnaires de référence de français, tels que le Trésor de la Langue Française informatisé (TLFi)¹, le Larousse² et le Robert³. Ces dictionnaires permettent de rechercher le sens premier des mots et de le comparer avec le sens contextuel observé dans les textes. Cette étape importante pour déterminer si une unité lexicale est utilisée de manière métaphorique.

Deux annotateurs sont impliqués dans le processus d'annotation. L'annotation est effectuée avec Brat (**Stenetorp et al., 2012**). En plus de l'annotation des types de métaphores, les annotateurs sont demandés d'indiquer une justification pour les métaphores (comparer le sens contextuel et le sens premier) et, pour les personnifications, de décrire comment des caractéristiques humaines sont attribuées aux objets, processus ou autres. Une guide d'annotation muni d'exemples aide à systématiser le processus d'annotation et à réduire les ambiguïtés. Cependant, la décision quand à la présence d'un emploi métaphorique d'une unité lexicale est parfois difficile à effectuer. Dans l'exemple (2), le terme *colonne* peut être considéré comme ambigu quant à la présence de la métaphore. D'une part, *colonne* est un terme architectural et son utilisation dans un texte médical témoigne de sa métaphorisation. D'autre part, le terme *colonne vertébrale* est très lexicalisé dans la langue, car sa métaphorisation remonte au grec ancien (**Skoda, 1988**). Après discussion lors du consensus, nous décidons de marquer *colonne* comme métaphorique. Le consensus a permis donc de résoudre certains cas ambigus, mais aussi d'affiner la typologie de métaphores. Ainsi, la consultation des dictionnaires de référence et la discussion entre les annotateurs lors du consensus sont les principaux moyens pour mieux traiter des cas ambigus et difficiles.

- (2) *La colonne vertébrale présentait une scoliose à convexité droite avec une masse douloureuse latéro-vertébrale droite à projection lombaire sans signes neurologiques associés.*

Nous avons annoté 30 documents (17 100 occurrences de mots). Comme le montre le tableau 1, nous annotons 430 emplois métaphoriques et observons tous les types de métaphores. Ainsi, la personnification (182 occ.), les métaphores de voyage (58) et politiques (53) sont les plus fréquentes. Viennent ensuite les métaphores d'architecture (38), de nature (31) et de guerre (21). Les autres types de métaphores sont beaucoup moins représentées. Les métaphores *autres* contient 34 occurrences, ce qui indique que cette typologie devrait être affinée. L'accord inter-annotateur, effectué sur l'ensemble des annotations avec la dernière version de la typologie, obtient un kappa de Cohen (**Cohen, 1960**) de 0,495, ce qui montre la difficulté et la subjectivité de la tâche.

1. <https://www.cnrtl.fr/definition/>

2. <https://www.larousse.fr/>

3. <https://www.lerobert.com/>

Type de métaphores	Nombre	%
Personnification	182	42,3
Voyage	58	13,5
Politique	53	12,3
Architecture	38	8,9
Nature	31	7,2
Guerre	21	4,9
Finance	8	1,9
Vestimentaire	4	0,9
Mécanique	1	0,2
Autres	34	7,9
Total	430	100

TABLE 1 – Résultat d’annotation du corpus de cas cliniques en métaphores.

3.3 Détection automatique des métaphores

Nous effectuons deux tâches automatiques : (1) détection de métaphores et (2) typage des métaphores. Ces tâches sont effectuées avec le giga-modèle génératif ChatGPT 4o d’OpenAI. Le *prompting* est effectué en ligne, sans installation locale ni paramétrage spécifique. Les interactions sont limitées aux prompts sur les métaphores. Il a été noté que, si ce type de modèles permet d’effectuer efficacement différentes tâches sur les données textuelles, la qualité des *prompts* (invites ou requêtes) a un impact direct sur la qualité des résultats (Wei *et al.*, 2022b; Wang *et al.*, 2023; White *et al.*, 2023; Kong *et al.*, 2024), ce qui pousse les chercheurs à tester plusieurs types de *prompts* et d’en proposer de nouveaux. Nous présentons donc notre approche avec des exemples de *prompts* pour la détection des métaphores :

1. Un *prompt few-shot* fournit quelques exemples avant de soumettre au modèle une tâche similaire (Brown *et al.*, 2020). L’objectif est d’améliorer la performance de la détection des métaphores en permettant au modèle d’apprendre un schéma à partir des exemples fournis.

Voici deux phrases contenant des métaphores, dans la première le terme "bourgeonnante" est utilisé métaphoriquement et dans la seconde il s'agit du mot "envahissement".

Phrase 1 : "A l'examen clinique nous retrouvons une lésion tumorale volumineuse, indolore, exophytique, bourgeonnante, irrégulière du pénis latéro-postérieure supérieure"

Phrase 2 : "La compression urétérale pelvienne était secondaire à un envahissement ganglionnaire iliaque massif visible sur le scanner abdomino-pelvien."

Analyse la phrase suivante et identifie les métaphores : <phrase>

2. Un *prompt basé sur le rôle* oriente le modèle en lui attribuant un rôle spécifique : linguiste, chercheur, expert dans le domaine médical (Kong *et al.*, 2024). L’idée consiste à adapter la réponse du modèle en lui imposant une expertise.

Agis comme un linguiste ayant de solides connaissances dans le domaine médical et détecte (toutes) les (mots utilisés métaphoriquement) / métaphores dans ce texte : <phrase>.

3. Un *prompt chain of thought* permet de décortiquer une tâche complexe en plusieurs étapes de raisonnement (Wei *et al.*, 2022b) et de fournir ainsi une réponse plus précise.

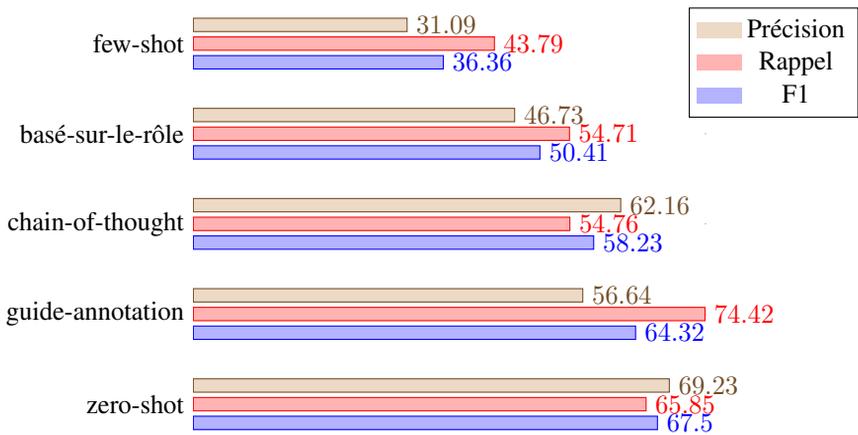


FIGURE 2 – Résultats de la détection automatique des métaphores

Explique-moi étape par étape pourquoi la phrase suivante contient une ou plusieurs métaphores : <phrase>

- Un *prompt* basé sur le guide d'annotation, que nous proposons, fournit un guide détaillé issu de l'annotation manuelle (comme dans la section 3.2). L'objectif consiste à imposer au modèle une méthode de recherche des métaphores en expliquant les critères que nous avons établis et en présentant des exemples. La longueur du *prompt* ne présente pas de difficultés au modèle.

Voici le guide d'annotation pour la détection des métaphores dans les textes médicaux. Je vais par la suite te donner des phrases, j'aimerais que tu t'appuies sur ce guide pour détecter les métaphores dans les phrases.

- Un *prompt* zéro-shot ne fournit aucun exemple (Wei et al., 2022a). Le modèle doit donc comprendre la tâche par lui-même et générer une réponse à partir de son apprentissage général, sans contexte particulier.

Analyse la phrase suivante et identifie les mots utilisés métaphoriquement : <phrase>.

Concernant la tâche de typage, nous avons fourni au modèle la typologie élaborée lors de l'annotation manuelle (section 3.2), en lui précisant les différentes catégories possibles. Pour chaque occurrence métaphorique, une phrase contenant l'expression lui était donnée, avec pour consigne d'assigner l'une des catégories de la typologie à cette expression.

Dans la phrase suivante, le mot 'envahissement' est utilisé métaphoriquement, dis-moi à quel type de métaphore il appartient : <phrase>

Pour les deux tâches, l'évaluation repose sur la comparaison entre la catégorie proposée par le modèle et celle établie manuellement lors de l'annotation. L'évaluation est mesurée en termes de précision, de rappel et de F-mesure au niveau des occurrences (version micro de ces mesures).

4 Résultats

L'évaluation est effectuée avec dix cas cliniques. La figure 2 présente les résultats de détection des métaphores. Nous pouvons voir que le *prompt zero-shot* a la meilleure précision (69,23) avec un rappel équilibré, ce qui fournit la meilleure F-mesure (67,50). De manière intéressante, le *prompt basé sur le guide d'annotation* montre le meilleur rappel (74,42), ce qui est sans doute dû à sa sensibilisation à la reconnaissance des métaphores grâce au guide d'annotation assez complet. Cependant, il obtient une précision plus faible (56,64), ce qui donne la F-mesure de 64,32. Le *prompt chain-of-thought* arrive en troisième position, avec une F-mesure de 58,23. Il manque certaines métaphores et souvent génère des explications trop longues. Finalement, le *prompt basé sur le rôle* et le *prompt few-shot* obtiennent les résultats les plus faibles. Le *prompt few-shot* par exemple, avec une F-mesure de 36,36, manque trop de métaphores tout en fournissant un grand nombre de fausses détections. Comparé aux résultats du *prompt zero-shot*, nous nous attendions que les exemples aideraient la détection, alors que nous voyons que ce n'est pas le cas. Il est possible que le nombre d'exemples fournis soit insuffisant ou bien leur présentation inappropriée pour le modèle. Il est également possible que les quelques exemples fournis biaisent le modèle en le rendant sensible à certains types de métaphores seulement.

La manière dont les *prompts* sont formulés est importante car chaque mot peut impacter les réponses. Par exemple, avec le *prompt few-shot*, lorsque certaines phrases ne contiennent pas de métaphores, le modèle montre parfois tendance à les sur-annoter et en identifier quand même. Les extractions du *prompt few-shot* sont sans doute guidées par les exemples donnés, ce qui peut biaiser les résultats et donc impacter les performances. Par exemple, si les exemples fournis mettent en avant certaines métaphores, le modèle cherche des métaphores similaires et passe à côté d'autres métaphores. Le modèle fonctionne mieux sur des phrases isolées plutôt que sur des paragraphes entiers. Par exemple, lorsqu'un texte plus long est analysé, le modèle peut *oublier* certaines métaphores et devient alors moins précis. De plus, avec certains prompts, une métaphore peut être correctement détectée dans une phrase, mais ignorée dans d'autres, ce qui montre le caractère contextuel de l'analyse des phrases en emplois métaphoriques. Notons aussi qu'au fil de l'exécution des *prompts*, le modèle s'améliore et détecte certaines métaphores avec plus de précision.

En ce qui concerne le typage des métaphores, le modèle génératif propose un typage correct pour 76 métaphores (45,51 %) et un typage incorrect pour 91 métaphores (54,49 %). Nous avons observé que le modèle catégorise correctement les métaphores appartenant aux types militaire, voyage et architecture. Les autres catégories, étant moins fréquentes et peut-être moins spécifiques, présentent plus de difficultés. Ce constat peut s'expliquer par le déséquilibre des catégories dans le corpus annoté, où certaines métaphores comme les personnifications sont largement majoritaires. Ce déséquilibre semble avoir un impact sur les performances du modèle, qui parvient plus facilement à typer les catégories fréquentes et lexicalement distinctives. À l'inverse, les métaphores appartenant à des catégories moins représentées, comme celles liées au domaine vestimentaire ou financier, sont souvent mal classées. Même si notre évaluation actuelle ne distingue pas les performances par type, nos observations suggèrent que la répartition déséquilibrée des catégories influence les résultats.

Il est difficile de comparer nos résultats avec l'état de l'art car nous utilisons un corpus différent : un corpus médical de cas cliniques en français, alors que dans l'état de l'art les chercheurs utilisent des corpus de langue générale en anglais. Si nous faisons l'abstraction des corpus, les résultats que nous obtenons restent comparables avec les travaux existants mais n'établissent pas un nouvel état de l'art dans l'identification des emplois métaphoriques. Actuellement, nous ne pouvons pas non plus définir si les emplois métaphoriques que nous avons dans notre corpus sont *difficiles* (Li *et al.*, 2024).

5 Conclusion et travaux futurs

Dans ce travail, nous nous sommes intéressés aux métaphores employées dans les textes du domaine médical. Nous utilisons une typologie inspirée des travaux existants (Salager-Meyer, 1990) et adaptée à notre corpus. Nous obtenons ainsi une typologie composée de plusieurs catégories de métaphores (voyage, politique, architecture, nature, guerre, finance, vestimentaire, mécanique), personnification et autres (cas ambigus et non catégorisés). Nous annotons manuellement un corpus de 30 cas cliniques, où nous relevons 430 emplois métaphoriques. Nous effectuons ensuite la détection automatique des métaphores avec un giga-modèle génératif ChatGPT 4o. Plusieurs types des *prompts* sont testés. Le *prompt basé sur guide d'annotation*, que nous proposons dans notre travail, est le plus performant en termes de rappel. Il détecte un plus grand nombre de métaphores, mais il effectue une sur-annotation et fournit donc un certain nombre de faux positifs. Le *prompt zero-shot* a une meilleure précision, avec moins d'erreurs de détection, mais un taux de détection plus faible. Il montre la meilleure F-mesure (67,50). Le *prompt few-shot*, en revanche, s'est avéré inefficace : il génère un grand nombre de fausses détections et manque les vraies métaphores. En ce qui concerne le typage des métaphores, les résultats montrent une concordance partielle entre l'annotation manuelle et automatique, avec 76 métaphores (45,51 %) typées correctement. Ce constat peut s'expliquer par le déséquilibre des catégories dans le corpus annoté, où certaines métaphores comme les personnifications sont largement majoritaires. Ce déséquilibre semble favoriser la reconnaissance des types fréquents (militaire, voyage, architecture), tandis que les catégories plus rares sont souvent mal identifiées.

Le déséquilibre entre les différentes catégories de métaphores présentes dans le corpus est à prendre en compte pour interpréter les performances du modèle. Comme dans d'autres tâches de classification, une forte inégalité entre les classes favorise la détection des types majoritaires, au détriment des catégories moins représentées. Une piste d'amélioration consisterait à calculer séparément les métriques (précision, rappel, F-mesure) pour chaque type de métaphores.

Enfin, l'utilisation d'un seul modèle génératif pose plusieurs limites méthodologiques. Certaines de ses caractéristiques techniques, comme le nombre de paramètres, ne sont pas rendues publiques, ce qui complique la reproductibilité des résultats et limite les possibilités de comparaison avec d'autres approches ou de réutilisation dans des contextes similaires. L'usage d'un modèle open-source reste une perspective de notre travail. Cela facilitera l'interprétation des performances et renforcera nos conclusions.

Nous avons plusieurs perspectives à ce travail. Tout d'abord, il est nécessaire d'augmenter le corpus annoté et de consolider la typologie des métaphores. Cela permettra également de tester d'autres approches et d'adapter les modèles de langue à cette tâche aux données en français. La typage des métaphores, en particulier de la personnification, peut aussi impliquer les informations syntaxiques. La détection et le typage, effectué indépendamment dans notre travail, peuvent être effectués simultanément ou enchaînés dans un même *prompt*.

Par ailleurs, nous voulons aussi étudier dans quelle mesure les métaphores aident la compréhension des informations médicales par les patients. Une autre piste concerne l'étude des réseaux de métaphores : si le domaine médical utilise des métaphores provenant d'autres domaines, ces autres domaines emploient aussi des métaphores, y compris celles provenant du domaine médical.

Remerciements

Ce travail a été partiellement financé par le projet *CLEAR* (*Communication, Literacy, Education, Accessibility, Readability*) financé par l'ANR sous la référence ANR-17-CE19-0016-01. Nous remercions aussi les relecteurs anonymes pour leurs commentaires qui ont permis d'améliorer la qualité de la version finale de ce papier et d'ouvrir de nouvelles pistes pour les travaux futurs.

Références

- ALEJO R., PIQUER-PÍRIZ A., CASTELLANO-RISCO I., MARTÍN-GILETE M., FIELDEN-BURNS L., MACARTHUR F., NACEY S., PHILIPS G., KRENNMAYR T., COELHO M., LITTLEMORE J. & ÄDEL A. (2021). Metclic corpus v1.
- BADRYZLOVA Y., SHEKHTMAN N., ISAEVA Y. & KERIMOV R. (2013). Annotating a Russian corpus of conceptual metaphor : a bottom-up approach. In E. SHUTOVA, B. BEIGMAN KLEBANOV, J. TETREAUULT & Z. KOZAREVA, Éd.s., *Proceedings of the First Workshop on Metaphor in NLP*, p. 77–86, Atlanta, Georgia : Association for Computational Linguistics.
- BIRKE J. & SARKAR A. (2006). A clustering approach for the nearly unsupervised recognition of nonliteral language. In *Proceedings of EACL-06*, p. 329–336, Trento, Italia.
- BROWN T., MANN B., RYDER N., SUBBIAH M., KAPLAN J. D., DHARIWAL P., NEELAKANTAN A., SHYAM P., SASTRY G., ASKELL A., AGARWAL S., HERBERT-VOSS A., KRUEGER G., HENIGHAN T., CHILD R., RAMESH A., ZIEGLER D., WU J., WINTER C., HESSE C., CHEN M., SIGLER E., LITWIN M., GRAY S., CHESS B., CLARK J., BERNER C., MCCANDLISH S., RADFORD A., SUTSKEVER I. & AMODEI D. (2020). Language models are few-shot learners. In H. LAROCHELLE, M. RANZATO, R. HADSELL, M. BALCAN & H. LIN, Éd.s., *Advances in Neural Information Processing Systems*, volume 33, p. 1877–1901 : Curran Associates, Inc.
- CHARTERIS-BLACK J. (2004). *Corpus Approaches to Critical Metaphor Analysis*. Palgrave Macmillan. DOI : [10.1057/9780230000612](https://doi.org/10.1057/9780230000612).
- COHEN J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**(1), 37–46.
- DANKERS V., MALHOTRA K., KUDVA G., MEDENTSIY V. & SHUTOVA E. (2020). Being neighbourly : Neural metaphor identification in discourse. In ACL, Éd., *Proc of the Second Workshop on Figurative Language Processing*, p. 227–234.
- DIVASSON L. & LEÓN I. (2006). Metaphors in english, french, and spanish medical written discourse. *Encyclopedia of Language & Linguistics*, p. 58–63. DOI : [10.1016/B0-08-044854-2/02358-0](https://doi.org/10.1016/B0-08-044854-2/02358-0).
- DO DINH E.-L. & GUREVYCH I. (2016). Token-level metaphor detection using neural networks. In B. BEIGMAN KLEBANOV, E. SHUTOVA & P. LICHTENSTEIN, Éd.s., *Proceedings of the Fourth Workshop on Metaphor in NLP*, p. 28–33, San Diego, California : Association for Computational Linguistics. DOI : [10.18653/v1/W16-1104](https://doi.org/10.18653/v1/W16-1104).
- DUNN J. (2013). What metaphor identification systems can tell us about metaphor-in-language. In E. SHUTOVA, B. BEIGMAN KLEBANOV, J. TETREAUULT & Z. KOZAREVA, Éd.s., *Proceedings of the First Workshop on Metaphor in NLP*, p. 1–10, Atlanta, Georgia : Association for Computational Linguistics.

GEDIGIAN M., BRYANT J., NARAYANAN S. & CIRIC B. (2006). Catching metaphors. In *Proc of the 3rd Workshop on Scalable Natural Language Understanding*, p. 41–48, New York.

GHOSH D., BEIGMAN KLEBANOV B., MURESAN S., FELDMAN A., PORIA S. & CHAKRABARTY T., Éd.s. (2022). *Proc of the 3rd Workshop on Figurative Language Processing (FLP)*, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

GINGRAS HARVEY M. (2014). *Les modes de conceptualisation métaphorique en neuroanatomie : étude de la description anatomique du cervelet et de la moelle spinale*. Mémoire de master, Université de Montréal, Montréal, Canada.

GRABAR N., GROUIN C., HAMON T. & CLAVEAU V. (2019). Corpus annoté de cas cliniques en français (annotated corpus with clinical cases in French). In E. MORIN, S. ROSSET & P. ZWEIGENBAUM, Éd.s., *Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN) PPIA 2019. Volume I : Articles longs*, p. 71–84, Toulouse, France : ATALA.

GROUP P. (2007). Mip : A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, **22**(1), 1–39. DOI : [10.1080/10926480709336752](https://doi.org/10.1080/10926480709336752).

HEINTZ I., GABBARD R., SRIVASTAVA M., BARNER D., BLACK D., FRIEDMAN M. & WEISCHDEL R. (2013). Automatic extraction of linguistic metaphors with LDA topic modeling. In *Proc of the First Workshop on Metaphor in NLP*, p. 58–66, Atlanta, GA.

JAMET D. (2008). *La perception d'Internet via ses métaphores*, In D. JAMET, Éd., *Métaphore et perception. Approches linguistiques, littéraires et philosophiques*, p. 39–56. L'Harmattan.

KONG A., ZHAO S., CHEN H., LI Q., QIN Y., SUN R., ZHOU X., WANG E. & DONG X. (2024). Better zero-shot reasoning with role-play prompting. In ACL, Éd., *NAACL-HLT*, p. 4099–4113.

KRENNMAYR T. & STEEN G. (2017). *VU Amsterdam Metaphor Corpus*, In N. IDE & J. PUSTEJOVSKY, Éd.s., *Handbook of Linguistic Annotation*, p. 1053–1071. Springer.

LAKOFF G. & JOHNSON M. (2003). *Metaphors We Live By*. Chicago : University of Chicago Press. Réédition de l'édition originale de 1980.

LAKOFF G. & JOHNSON M. (1980). *Metaphors we live by*. Chicago, USA : The University of Chicago Press.

LEONG C. W. B., BEIGMAN KLEBANOV B., HAMILL C., STEMLE E., UBALÉ R. & CHEN X. (2020). A report on the 2020 VUA and TOEFL metaphor detection shared task. In B. B. KLEBANOV, E. SHUTOVA, P. LICHTENSTEIN, S. MURESAN, C. WEE, A. FELDMAN & D. GHOSH, Éd.s., *Proc of the Second Workshop on Figurative Language Processing*, p. 18–29, Online : Association for Computational Linguistics. DOI : [10.18653/v1/2020.figlang-1.3](https://doi.org/10.18653/v1/2020.figlang-1.3).

LEONG C. W. B., BEIGMAN KLEBANOV B. & SHUTOVA E. (2018). A report on the 2018 VUA metaphor detection shared task. In B. BEIGMAN KLEBANOV, E. SHUTOVA, P. LICHTENSTEIN, S. MURESAN & C. WEE, Éd.s., *Proceedings of the Workshop on Figurative Language Processing*, p. 56–66, New Orleans, Louisiana : Association for Computational Linguistics. DOI : [10.18653/v1/W18-0907](https://doi.org/10.18653/v1/W18-0907).

LI Y., GUERIN F. & LIN C. (2024). *Finding Challenging Metaphors that Confuse Pretrained Language Models*. Rapport interne, ARXIV.

LI Y., WANG S., LIN C., GUERIN F. & BARRAULT L. (2023). Framebert : Conceptual metaphor detection with frame embedding learning. In ACL, Éd., *Proc of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, p. 1558–1563, Dubrovnik, Croatia.

LIMAIEM F., BENJEMAA N., BELKACEM O., TALBI G. & BOURAOUI S. (2021). Métaphores : outil conceptuel et didactique en anatomie pathologique. *Annales de Pathologie*, **41**(5), 470–5.

MAO R., LIN C. & GUERIN F. (2019). End-to-end sequential metaphor identification inspired by linguistic theories. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, p. 3888–3898.

MARQUER B. (2019). *La maladie comme métaphore : littérature et médecine de l'esprit dans la seconde moitié du XIXe siècle*, In *Passerelles. Entre sciences et littératures, Classiques Garnier*, p. 215–246.

MOHLER M., BRACEWELL D., TOMLINSON M. & HINOTE D. (2013). Semantic signatures for example-based linguistic metaphor detection. In *ACL, Éd., Proc of the First Workshop on Metaphor in NLP*, p. 27–35, Atlanta, GA.

PTIČEK M. & DOBŠA J. (2023). Methods of annotating and identifying metaphors in the field of natural language processing. *Future Internet*, **15**(6). DOI : [10.3390/fi15060201](https://doi.org/10.3390/fi15060201).

RICOEUR P. (1975). *La Métaphore vive*. Paris : Éditions du Seuil.

ROLLO A. (2015). Les métaphores conceptuelles dans la science médicale : outil cognitif et communicatif. Les avatars de la métaphore. *Publifarum*, **23**, 1–19.

SALAGER-MEYER F. (1990). Metaphors in medical english prose : A comparative study with french and spanish. *English for Specific Purposes*, **9**(2), 145–159. DOI : [https://doi.org/10.1016/0889-4906\(90\)90004-V](https://doi.org/10.1016/0889-4906(90)90004-V).

SANCHEZ-BAYONA E. & AGERRI R. (2022). Leveraging a new Spanish corpus for multilingual and cross-lingual metaphor detection. In A. FOKKENS & V. SRIKUMAR, Éds., *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL)*, p. 228–240, Abu Dhabi, United Arab Emirates (Hybrid) : Association for Computational Linguistics. DOI : [10.18653/v1/2022.conll-1.16](https://doi.org/10.18653/v1/2022.conll-1.16).

SEMINO E., DEMJÉN Z. & DEMMEN J. (2016). An Integrated Approach to Metaphor and Framing in Cognition, Discourse, and Practice, with an Application to Metaphors for Cancer. *Applied Linguistics*, **39**(5), 625–645. DOI : [10.1093/applin/amw028](https://doi.org/10.1093/applin/amw028).

SHUTOVA E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics*, **41**(1), 579–623. DOI : [10.1162/COLI_a_00233](https://doi.org/10.1162/COLI_a_00233).

SHUTOVA E. (2017). *Annotation of Linguistic and Conceptual Metaphor*, In N. IDE & J. PUSTEJOVSKY, Éds., *Handbook of Linguistic Annotation*, p. 1073–1100. Springer.

SHUTOVA E., SUN L. & KORHONEN A. (2010). Metaphor identification using verb and noun clustering. In *ACL, Éd., Proc of Conference on Computational Linguistics 2010*, p. 1002–1010, Beijing.

SKODA F. (1988). *Médecine ancienne et métaphore : le vocabulaire de l'anatomie et de la pathologie en grec ancien*. Paris : Peeters-Selaf.

STEEN G., DORST L., HERRMANN J., KAAL A., KRENNMAYR T. & PASMA T. (2010a). *A method for linguistic metaphor identification : From MIP to MIPVU*. John Benjamins. DOI : [10.1075/celcr.14](https://doi.org/10.1075/celcr.14).

STEEN G. J., DORST A. G., HERRMANN J. B., KAAL A. A. & KRENNMAYR T. (2010b). Metaphor in usage. *Cognitive Linguistics*, **21**(4), 765–796. DOI : [doi:10.1515/cogl.2010.024](https://doi.org/10.1515/cogl.2010.024).

STENETORP P., PYYSALO S., TOPIĆ G., OHTA T., ANANIADOU S. & TSUJII J. (2012). brat : a web-based tool for nlp-assisted text annotation. In *European Chapter of the Association for Computational Linguistics*, p. 102–107, Avignon, France : Association for Computational Linguistics.

VANDAELE S., BOUDREAU S., LUBIN L. & MARSHMAN E. (2006). La conceptualisation métaphorique en biomédecine : indices de conceptualisation et réseaux lexicaux. *GLOTTOPOL*, **8**, 73–94.

VEALE T. & HAO Y. (2008). A fluid knowledge representation for understanding and generating creative metaphors. In *Proc of the 22nd International Conference on Computational Linguistics (Coling 2008)*, p. 945–952.

VEYNE P. & TAMBA I. (1979). Metaphora et comparaison selon aristote. *Revue des Études Grecques*, **92**(436), 77–98. DOI : [10.3406/reg.1979.4218](https://doi.org/10.3406/reg.1979.4218).

WANG Z. M., PENG Z., QUE H., LIU J., ZHOU W., WU Y., GUO H., GAN R., NI Z., YANG J., ZHANG M., ZHANG Z., OUYANG W., XU K., HUANG S. W., FU J. & PENG J. (2023). *RoleLLM : Benchmarking, Eliciting, and Enhancing Role-Playing Abilities of Large Language Models*. Rapport interne, ARXIV. arXiv preprint :2310.00746v3.

WEI J., BOSMA M., ZHAO V. Y., GUU K., YU A. W., LESTER B., DU N., DAI A. M. & LE Q. V. (2022a). Finetuned language models are zero-shot learners. *arXiv preprint arXiv :2109.01652*.

WEI J., WANG X., SCHUURMANS D., BOSMA M., ICHTER B., XIA F., CHI E., LE Q. & ZHOU D. (2022b). Chain-of-thought prompting elicits reasoning in large language models. In *NeurIPS*, p. 1–14.

WHITE J., FU Q., HAYS S., SANDBORN M., OLEA C., GILBERT H., ELNASHAR A., SPENCER-SMITH J. & SCHMIDT D. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. Rapport interne, ARXIV. DOI : [10.48550/arXiv.2302.11382](https://doi.org/10.48550/arXiv.2302.11382).

ZHANG S. & LIU Y. (2022). Metaphor detection via linguistics enhanced Siamese network. In N. CALZOLARI, C. R. HUANG, H. KIM, J. PUSTEJOVSKY, L. WANNER, K.-S. CHOI, P.-M. RYU, H.-H. CHEN, L. DONATELLI, H. JI, S. KUROHASHI, P. PAGGIO, N. XUE, S. KIM, Y. HAHM, Z. HE, T. K. LEE, E. SANTUS, F. BOND & S.-H. NA, Édts., *Proceedings of the 29th International Conference on Computational Linguistics*, p. 4149–4159, Gyeongju, Republic of Korea : International Committee on Computational Linguistics.