



HAL
open science

Au-delà de la performance des modèles : la prédiction de liens peut-elle enrichir des graphes lexico-sémantiques du français ?

Hee-Soo Choi, Priyansh Trivedi, Mathieu Constant, Karën Fort, Bruno Guillaume

► To cite this version:

Hee-Soo Choi, Priyansh Trivedi, Mathieu Constant, Karën Fort, Bruno Guillaume. Au-delà de la performance des modèles : la prédiction de liens peut-elle enrichir des graphes lexico-sémantiques du français ?. Actes de JEP-TALN-RECITAL 2024. 31ème Conférence sur le Traitement Automatique des Langues Naturelles, volume 1 : articles longs et prises de position, Jul 2024, Toulouse, France. pp.36-49. hal-04623008

HAL Id: hal-04623008

<https://inria.hal.science/hal-04623008v1>

Submitted on 9 Jul 2024

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

Au-delà de la performance des modèles : la prédiction de liens peut-elle enrichir des graphes lexico-sémantiques du français ?

Hee-Soo Choi^{1,2} Priyansh Trivedi⁴

Mathieu Constant¹ Karën Fort^{2,3} Bruno Guillaume²

(1) ATILF, CNRS, Université de Lorraine, 54000 Nancy, France

(2) LORIA, Université de Lorraine, 54506 Vandoeuvre-lès-Nancy, France

(3) Sorbonne Université, 75006 Paris, France, (4) Chercheur indépendant

hee-soo.choi@loria.fr, mail@priyansh.page,

mathieu.constant@atilf.fr, karen.fort@loria.fr, bruno.guillaume@loria.fr

RÉSUMÉ

Cet article présente une étude sur l'utilisation de modèles de prédiction de liens pour l'enrichissement de graphes lexico-sémantiques du français. Celle-ci porte sur deux graphes, `RezoJDM16k` et `RL-fr` et sept modèles de prédiction de liens. Nous avons étudié les prédictions du modèle le plus performant, afin d'extraire de potentiels nouveaux triplets en utilisant un score de confiance que nous avons évalué avec des annotations manuelles. Nos résultats mettent en évidence des avantages différents pour le graphe dense `RezoJDM16k` par rapport à `RL-fr`, plus clairsemé. Si l'ajout de nouveaux triplets à `RezoJDM16k` offre des avantages limités, `RL-fr` peut bénéficier substantiellement de notre approche.

ABSTRACT

Beyond Model Performance : Can Link Prediction Enrich French Lexical Graphs ?

This paper presents a resource-centric study of link prediction approaches over French lexical-semantic graphs. Our study incorporates two graphs, `RezoJDM16k` and `RL-fr` and seven link prediction models. We studied the predictions of the best-performing model to extract potential new triples based on a confidence score that we evaluated with manual annotations. Our findings highlight different benefits for the dense graph `RezoJDM16k` compared to the sparser graph `RL-fr`. While the addition of new triples to `RezoJDM16k` offers limited advantages, `RL-fr` can benefit substantially from our approach.

MOTS-CLÉS : graphes lexicaux, ressources du français, prédiction de liens, complétion de graphes.

KEYWORDS: lexical graphs, French resources, link prediction, graph completion.

1 Introduction

Les ressources lexicales ont longtemps été essentielles pour le développement des systèmes de Traitement Automatique des Langues (TAL). Malgré les récentes avancées de l'apprentissage non supervisé et des modèles de langue, ces ressources et plus spécifiquement les graphes de connaissances ou lexico-sémantiques restent des sources de connaissances linguistiques permettant l'amélioration des systèmes dans la résolution de tâches de TAL (Hwang *et al.*, 2021; d'Amato *et al.*, 2023). En outre, ces objets décrivant la langue sont également importants dans des domaines tels que la linguistique

ou l’enseignement des langues. Dans cet article, nous nous intéressons à l’utilisation de modèles de prédiction de liens pour enrichir des graphes lexico-sémantiques du français. Plus précisément, nous cherchons ici à améliorer la complétude du Réseau Lexical du français, $RL-fr$ (Lux-Pogodalla & Polguère, 2011), un graphe lexico-sémantique relativement peu dense créé manuellement par des lexicographes. Pour mieux comparer l’efficacité de notre approche, nous avons également appliqué nos expériences sur $RezoJDM16k$ (Mirzapour *et al.*, 2022), un graphe lexico-sémantique très dense. Les principales contributions de l’article sont les suivantes : i) nous avons évalué les performances de sept modèles de prédictions de liens sur deux graphes lexico-sémantiques du français, dont un modèle présentant des résultats état-de-l’art, ii) nous avons ajouté un score de confiance sur les prédictions générées par le modèle $CompGCN-ConVE$ pour extraire de potentiels nouveaux triplets et iii) nous avons analysé qualitativement ces prédictions à partir d’annotations manuelles. Notre expérience produit des résultats encourageants et ouvre une voie nouvelle vers l’enrichissement semi-automatique des ressources lexico-sémantiques.

2 Tâche de prédiction de liens

Les graphes de connaissances sont généralement incomplets en raison de l’impossibilité de décrire le monde ou la langue de manière exhaustive. La tâche de prédiction de liens consiste alors à prédire des triplets manquants dans un graphe. Il existe deux variantes principales de cette tâche : la prédiction transductive et la prédiction inductive. Dans la prédiction transductive, l’entraînement et l’inférence se font sur le même graphe. Inversement, dans la prédiction inductive, l’inférence peut avoir lieu sur un graphe différent et l’échantillon de test peut inclure des nœuds inconnus (Galkin *et al.*, 2022). Dans cet article, nous nous concentrons uniquement sur la prédiction transductive. Les graphes de connaissances peuvent être décrits comme un ensemble de triplets, désignés par (h, r, t) pour *head* (tête), *relation* et *tail* (queue). Étant donné des triplets incomplets tels que $(h, r, ?)$ ou $(?, r, t)$, le modèle doit prédire l’entité manquante. Pour ce faire, les modèles neuronaux sont entraînés à donner un meilleur score aux triplets positifs qu’aux triplets négatifs créés *negative sampling* (Bordes *et al.*, 2013), technique consistant à corrompre les triplets positifs en remplaçant l’entité de tête ou de queue par une autre entité choisie au hasard. Les scores sont calculés à l’aide d’une fonction score, qui dépend du type de modèle. Dans cette section, nous donnons un aperçu général de certains de ces types. Nous renvoyons les lecteurs intéressés à l’étude complète de Chen *et al.* (2020) pour une compréhension approfondie des approches neuronales de prédiction de liens.

Modèles translationnels Ces modèles utilisent la distance entre les plongements de nœuds et de relations comme fonction de score. $TransE$ (Bordes *et al.*, 2013) utilise la distance euclidienne, les modèles suivants comme $TransH$ (Wang *et al.*, 2014), $TransR$ (Lin *et al.*, 2015) et $TransD$ (Ji *et al.*, 2015) offrant diverses extensions. D’autres, comme $RotateE$ (Sun *et al.*, 2019), utilisent des espaces vectoriels complexes pour représenter les entités et définissent les relations comme des rotations entre elles. Cela permet de modéliser des modèles de relations plus complexes comme la symétrie/asymétrie, les inversions et les compositions.

Modèles *semantic-matching* Ces modèles utilisent une fonction de score dérivée de la similarité sémantique pour découvrir les connexions sémantiques potentielles entre les entités et les relations. Parmi les exemples notables, nous pouvons citer $RESCAL$ (Nickel *et al.*, 2011), qui capture les interactions par paire entre les entités, $DistMult$ (Yang *et al.*, 2015) qui réduit la charge de calcul mais est limité aux relations symétriques et $Complex$ (Trouillon *et al.*, 2016), qui introduit des plongements basés sur un espace vectoriel complexe pour une capacité de modélisation plus large.

Architectures neuronales profondes Les approches présentées ci-dessus se limitent à l'utilisation d'opérations mathématiques simples, telles que le produit scalaire ou les multiplications de matrices sur les plongements d'entités et de relations. Par conséquent, leur capacité de modélisation ne peut être augmentée qu'en modifiant les dimensions des plongements. L'application des réseaux neuronaux profonds à ces graphes peut être une alternative pour pallier ces limitations. Non triviale, elle a été sous-explorée jusqu'aux deux développements suivants. Premièrement, [Dettmers et al. \(2018\)](#) ont proposé `ConvE` qui applique des couches de convolution sur l'espace d'intégration latent pour modéliser les interactions entité-relation et utilise une couche dense pour calculer le score. Parallèlement, les réseaux de convolution de graphes (GCN), proposés dans [Kipf & Welling \(2017\)](#), ont permis de propager les informations de différents nœuds à travers les chemins du graphe, ce qui a conduit à des représentations d'entités et de relations tenant compte du voisinage. Des modèles comme `R-GCN` ([Schlichtkrull et al., 2018](#)) et `CompGCN` ([Vashishth et al., 2020](#)) proposent d'autres modifications pour gérer les graphes multi-relationnels. Toutefois, les GCN eux-mêmes ne résolvent pas la tâche de prédiction des liens, mais fournissent des moyens plus riches d'encoder le graphe.

Métriques Inspirées de la recherche d'information (RI), les métriques traditionnelles pour la tâche de prédiction de liens sont basées sur le classement des scores des prédictions correctes parmi toutes les prédictions générées :

- **Mean Rank (MR)** : étant donné un ensemble de triplets classés par leur score de prédiction, le MR calcule le rang moyen des triplets corrects. Un MR faible signifie une meilleure performance.
- **Mean Reciprocal Rank (MRR)** : MRR est la moyenne des inverses des rangs des triplets corrects. Un MRR élevé indique une meilleure performance.
- **Hits@k** : cette mesure calcule la proportion de triplets corrects apparaissant dans le top k de la liste classée des triplets prédits. Une valeur Hits@k plus élevée indique une précision de prédiction supérieure pour les triplets les plus importants.

3 État de l'art

Si la tâche de complétion des graphes lexico-sémantiques reste sous-explorée, il existe une vaste littérature sur l'enrichissement des graphes lexicaux. De manière générale, ces approches visent à accroître la couverture de la ressource, c'est-à-dire à ajouter de nouveaux nœuds au graphe, en s'appuyant sur des ressources externes. L'une des premières approches se base sur les co-occurrences statistiques pour amorcer les graphes lexico-sémantiques existants ([Biemann et al., 2004](#)). Ces approches, ainsi que d'autres antérieures ([Riloff & Shepherd, 1997](#)), nécessitaient toutefois une intervention manuelle importante en raison de la faible qualité des prédictions. L'une des avancées majeures dans ce domaine a été la création de la ressource multilingue `BabelNet` ([Navigli & Ponzetto, 2012](#)), qui a étendu `WordNet` ([Miller, 1995](#)) en croisant les sens des mots avec les articles de `Wikipedia`. Des modèles de traductions automatiques ainsi que d'autres approches ([Oliver & Climent, 2012](#); [Lam et al., 2014](#)) ont également été utilisées à des degrés divers. Les progrès de la levée d'ambiguïté lexicale et l'existence de corpus ou de dictionnaires parallèles ont également permis d'enrichir les `Wordnets` d'autres langues à partir de l'anglais ([Taghizadeh & Faili, 2016](#); [Arcan et al., 2016](#)). En outre, les approches visant à optimiser la complétude de ces graphes adoptent rarement une perspective centrée sur les ressources. En effet, bien qu'il existe de nombreux travaux utilisant des modèles transductifs de prédiction de liens sur ces graphes, très peu d'entre eux ont conduit à des ajouts concrets à ladite ressource, à l'exception des travaux de [Fellbaum \(1998\)](#) sur l'ajout de liens entre les nœuds de `WordNet` par extraction de motifs (*pattern mining*) à partir de corpus.

Il est également important de souligner que ce domaine a un biais important en faveur des ressources de l'anglais, comme le montre l'utilisation de jeux de données de référence tels que WN18RR (Dettmers *et al.*, 2018) et FB15K-237 (Toutanova & Chen, 2015). Si la plupart des techniques sont en théorie transférables à n'importe quel graphe, très peu d'efforts empiriques ont été faits. Pour les ressources du français, Mirzapour *et al.* (2022) ont étudié l'efficacité des modèles de prédiction de liens sur un graphe lexico-sémantique, mais se concentrent davantage sur la création du jeu de données et l'évaluation des modèles plutôt que sur l'enrichissement de la ressource.

4 Graphes lexico-sémantiques du français

RezoJDM et RezoJDM16k RezoJDM (Lafourcade & Joubert, 2008; Lafourcade & Le Brun, 2020) est un réseau lexico-sémantique du français développé *via* des jeux ayant un but, des approches contributives et des mécanismes d'inférence. La plateforme JeuxDeMots¹ propose des jeux permettant de développer le réseau en ajoutant de nouvelles entrées et de vérifier des informations du réseau. Le jeu principal invite les joueurs à saisir des termes dans un délai imparti selon un terme et un type de relation donnés. L'utilisation des jeux ayant un but a ainsi permis de créer un graphe dirigé très dense, puisqu'il comprend actuellement plus de 537 millions de relations et six millions de nœuds². Dans RezoJDM, les nœuds représentent principalement des termes (type `n_term`) mais contiennent aussi des informations telles que des étiquettes de partie du discours (type `n_pos`) ou des formes fléchies (`n_form`). Les relations sont divisées en trois catégories : lexicales (synonymie, antonymie...), ontologiques (hyponymie, méronymie...) et prédicatives (agent, conséquences...). Les nœuds et les relations ont la particularité de présenter des poids en fonction de la dynamique du jeu. Pour les relations, un poids positif code une relation vraie et un poids négatif code une relation fautive, attribut relativement rare dans les graphes de connaissances. La polysémie d'un terme est exprimée en distinguant un nœud générique de ses nœuds de raffinement. Cependant, comme le jeu demande au joueur d'entrer autant de termes que possible dans un temps limité, les joueurs ont tendance à ne pas affiner leurs réponses, ce qui conduit à une faible densité des nœuds de raffinement. Par exemple, le nœud générique `accord` a un degré de 10 549, tandis que les degrés des nœuds de raffinement `accord>pacte` et `accord>acceptation` sont respectivement de 194 et 123.

Mirzapour *et al.* (2022) ont créé le sous-graphe RezoJDM16k en appliquant divers filtres aux nœuds et relations de RezoJDM. Seuls les nœuds de type `n_term` et de poids supérieur à 50 ont été retenus. Le même filtre de poids a été appliqué aux relations et certains types ont également été supprimés. En outre, les types de relations apparaissant moins de 100 fois et les nœuds ayant un degré inférieur à 45 ont été exclus, afin d'améliorer l'efficacité des modèles. Le graphe final est composé de 15 746 nœuds et de 832 093 relations.

RL-fr Le Réseau Lexical du Français RL-fr est un réseau lexico-sémantique du français créé par des lexicographes, où les nœuds correspondent à des unités lexicales et les arêtes à des relations lexico-sémantiques ou combinatoires (Lux-Pogodalla & Polguère, 2011). Dans cette section, nous présentons des informations concernant la version 2.1 de RL-fr, utilisée pour nos expériences³.

Les unités lexicales dans RL-fr sont les entités fondamentales pour la description lexicographique et peuvent être soit un lexème, soit un idiomme. Les lexèmes sont des unités lexicales monolexémiques

1. <https://www.jeuxdemots.org>

2. En octobre 2023.

3. <https://www.ortolang.fr/market/lexicons/lexical-system-fr/v2.1>

et correspondent aux sens des mots. Par conséquent, un mot polysémique, appelé vocable, est représenté comme une collection d'unités lexicales interconnectées par une relation de copolysémie. La version 2.1 du RL-fr contient 29 220 unités lexicales et 18 625 vocables. Voici les lexèmes du vocable *jambe*, à titre d'exemple :

- *Jambe* I.1 : Marc attend patiemment, les **jambes** croisées.
- *Jambe* I.2a : Le cheval s'est blessé à la **jambe**.
- *Jambe* I.2b : Il y a de la **jambe** de porc au menu.
- *Jambe* II : La **jambe** droite du pantalon est déchirée.
- *Jambe* III : Une des **jambes** de suspension doit être changée.

Contrairement aux dictionnaires classiques, où les sens d'un mot sont généralement seulement listés, le RL-fr représente les types de relation entre les sens avec la notion de copolysémie, décrite comme la relation entre les différents sens d'un mot, opposée à la polysémie qui est la propriété des mots d'exprimer plusieurs sens (Polguère, 2018). Il existe donc plusieurs relations de copolysémie, telle que la métonymie ou la métaphore (cf. Annexe A).

La ressource est basée sur la théorie Sens-Texte qui permet d'encoder les relations paradigmatiques et syntagmatiques avec les fonctions lexicales (Mel'čuk, 1996). Quelques exemples de fonctions lexicales paradigmatiques sont donnés ci-dessous :

- Synonymie (Syn) : *vélo* → *bicyclette*
- Antonymie (Anti) : *accord* → *désaccord*
- Hyperonymie (Gener) : *amour* → *sentiment*

Les relations syntagmatiques comprennent, entre autres, les collocations et les verbes supports :

- Intensifieur (Magn) : *boire* → *comme un trou*
- Verbe support (Oper) : *danger* → *courir*

Les fonctions lexicales peuvent être utilisées pour représenter des relations sémantiques simples ou complexes. Dans la version 2.1 du RL-fr, il existe 686 fonctions lexicales différentes. Pour nos expériences, nous avons utilisé les familles de relations pour réduire le nombre de types de relations, ce qui correspond à 95 types de fonctions lexicales et 11 types de relations de copolysémie. Au total, il existe 62 641 relations encodées par des fonctions lexicales et 9 413 relations de copolysémie.

Deux graphes différents mais complémentaires Malgré un format de réseau commun, RezoJDM16k et RL-fr présentent des différences significatives en termes de création, de portée et de représentation de la polysémie. Le tableau 1 illustre le contraste topologique entre ces deux graphes : RezoJDM16k est très dense, avec un degré moyen de nœuds de 105,7 et un nombre d'arêtes 10 fois supérieur à celui de RL-fr. Par ailleurs, RL-fr présente davantage de nœuds, en raison de la représentation des sens des mots dans des nœuds distincts. La figure 1 montre un aperçu du réseau RL-fr autour du vocable *accord*, représenté par deux nœuds *accord 1* et *accord 2*⁴.

La méthode de création des ressources a un impact sur la structure du graphe, l'enrichissement manuel étant un processus chronophage. Cependant, elle garantit un contenu contrôlé et de qualité qui est vérifié par des experts du domaine. Concernant RezoJDM16k, la qualité de la ressource n'est pas nécessairement inférieure, car elle est fournie par des bénévoles motivés par leur intérêt pour la langue (Lafourcade & Le Brun, 2020), mais elle présente un certain bruit malgré les vérifications semi-automatiques.

4. https://spiderlex.atilf.fr/fr/q/*accord***

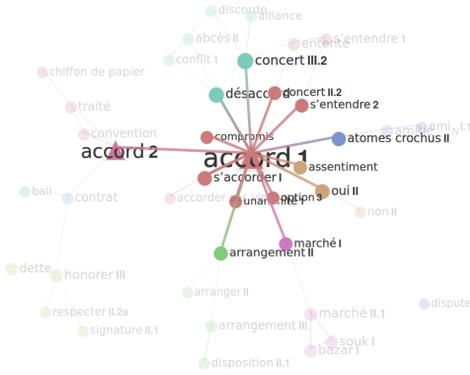


FIGURE 1 – Vocabulaire `accord` représenté dans le `RL-fr`.

	RezoJDM16k	RL-fr
# nœuds	15 746	29 220
# arêtes	832 093	72 054
# types d'arête	150	106
Degré moyen d'un nœud	105,7	5,2
Degré minimum	2	1
Degré maximum	10 403	187

TABLE 1 – Statistiques sur `RezoJDM16k` et `RL-fr`.

5 Expériences

Pré-traitement des jeux de données La prédiction transductive des liens implique que tous les nœuds des échantillons de validation et de test soient présents dans l'échantillon d'entraînement. Nous avons décidé de diviser aléatoirement les triplets entre les échantillons d'entraînement, de validation et de test (80 %, 10 %, 10 %) et de vérifier si tous les nœuds de la validation et du test sont présents dans l'entraînement. Si ce n'est pas le cas, nous supprimons le triplet de l'ensemble. Pour un graphe très dense comme `RezoJDM16k`, aucun triplet n'a été perdu même après dix divisions aléatoires différentes. En revanche, un graphe peu dense tel que `RL-fr` implique des pertes d'arêtes et de nœuds (- 2 152 nœuds, - 1 037 arêtes). Nous obtenons un graphe final de 27 068 nœuds et 71 017 arêtes, que nous appelons `RLF27k`. Le tableau 2 présente `RezoJDM16k` et `RLF27k` et leurs échantillons d'entraînement, de validation et de test respectifs.

	RezoJDM16k	RLF27k
# nœuds	15 746	27 068
# arêtes	832 093	71 017
# triplets entraînement	665 674	57 643
# triplets validation	83 209	6 674
# triplets test	83 210	6 700

TABLE 2 – Statistiques sur les échantillons d'entraînement, validation et test de `RezoJDM16k` et `RLF27k` après une division en 80 %, 10 %, 10 %.

Performances des modèles Nous avons évalué les performances de modèles de prédiction de liens sur les deux ressources en utilisant les six mêmes modèles que ceux décrits dans [Mirzapour et al. \(2022\)](#). Nous avons reproduit leurs expériences sur `RezoJDM16k` et nous les avons menées sur `RLF27k`. De plus, nous avons utilisé un modèle `ConvE` avec l'encodeur `CompGCN` ([Vashishth et al., 2020](#)) (appelé `CompGCN-ConvE`), afin d'explorer l'efficacité d'un modèle `GNN` (*Graph Neural Networks*) sur des données en français. Les résultats pour `RezoJDM16k` sont présentés dans le tableau 3. Nous constatons que `CompGCN-ConvE` surpasse tous les autres modèles dans presque

Modèle	MRR ↑	MR ↓	Hits@10 ↑	Hits@3 ↑	Hits@1 ↑
TransE (Bordes <i>et al.</i> , 2013)	0,180	200,78	0,437	0,242	0,040
TransH (Wang <i>et al.</i> , 2014)	0,217	173,28	0,503	0,293	0,064
TransD (Ji <i>et al.</i> , 2015)	0,216	168,18	0,500	0,290	0,065
DistMult (Yang <i>et al.</i> , 2015)	0,219	194,16	0,446	0,252	0,109
ComplEx (Trouillon <i>et al.</i> , 2016)	0,256	190,79	0,539	0,309	0,119
RotatE (Sun <i>et al.</i> , 2019)	0,312	177,04	0,587	0,409	0,155
CompGCN-ConvE (Vashishth <i>et al.</i>, 2020)	0,461	171,26	0,659	0,514	0,357

TABLE 3 – Résultats des modèles de prédiction de liens sur RezoJDM16k.

Modèle	MRR ↑	MR ↓	Hits@10 ↑	Hits@3 ↑	Hits@1 ↑
TransE (Bordes <i>et al.</i> , 2013)	0,278	2594,24	0,624	0,497	0,033
TransH (Wang <i>et al.</i> , 2014)	0,250	2957,59	0,581	0,465	0,011
TransD (Ji <i>et al.</i> , 2015)	0,255	2752,03	0,587	0,472	0,016
DistMult (Yang <i>et al.</i> , 2015)	0,373	2748,25	0,613	0,502	0,216
ComplEx (Trouillon <i>et al.</i> , 2016)	0,413	3447,98	0,593	0,524	0,284
RotatE (Sun <i>et al.</i> , 2019)	0,399	3650,92	0,490	0,454	0,336
CompGCN-ConvE (Vashishth <i>et al.</i>, 2020)	0,515	2808,68	0,627	0,559	0,450

TABLE 4 – Résultats des modèles de prédiction de liens sur RLF27k.

toutes les métriques. Il atteint notamment un Hits@1 de 0,357, qui est plus de deux fois supérieur à celui de RotatE. Les scores Hits@3, Hits@10 et MRR sont également les plus élevés et le MR reste proche du meilleur score obtenu par TransD. Le tableau 4 présente les résultats pour RLF27k. Une fois de plus, CompGCN-ConvE se distingue avec un MRR de 0,515 et les meilleurs scores Hits@k. Cependant, il convient de noter que le MR élevé indique une disparité dans le classement des triplets corrects : 60 % d’entre eux se classent parmi les dix premiers, tandis que les 40 % restants sont nettement moins bien classés.

Score de confiance Nous avons effectué une analyse approfondie des prédictions du meilleur modèle, CompGCN-ConvE. Au-delà d’évaluer la performance du modèle dans la prédiction des triplets de test, nous avons observé toutes les prédictions du modèle pour une entité et une relation données en classant les prédictions selon leur score. La figure 2 montre les 20 meilleures prédictions pour la tête bonnet I et la relation synonymie de RLF27k. Le modèle prédit logiquement un triplet présent dans l’entraînement avec un score élevé (0,893), puis prédit deux triplets de test avec des scores d’environ 0,08. Nous supposons que les triplets restants, qui n’existent pas dans le graphe original, pourraient être des nouveaux potentiels triplets. Néanmoins, la fonction de score ne permet pas d’évaluer significativement la pertinence de ces triplets, dans la mesure où tous les scores sont très faibles (environ 0,01).

Nous avons donc décidé d’utiliser un meilleur algorithme d’inférence, à savoir la technique de Monte-Carlo (MC) Dropout (Gal & Ghahramani, 2016). Le dropout fonctionne en désactivant de manière aléatoire des neurones d’un réseau de neurones. Généralement, lors d’une inférence standard, nous désactivons le dropout pour obtenir des prédictions déterministes et moins bruitées. Pour l’inférence basée sur MC Dropout, nous générons plusieurs prédictions pour la même entrée, en échantillonnant un masque dropout différent à chaque fois. On obtient ainsi une distribution prédictive pour un modèle et des entrées donnés, qui produit un ensemble plus riche d’informations sur les prédictions du modèle, comme la possibilité de calculer des scores de confiance pour n’importe quelle prédiction (cf.

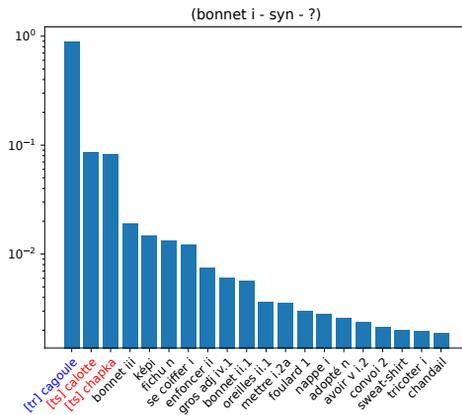


FIGURE 2 – Prédications pour la tête bonnet I et la relation synonymie. Les triplets qui existent déjà dans RLF27k sont en rouge ([tr] pour *training* (entraînement) et [ts] pour test). L’axe des ordonnées est mis à l’échelle logarithme pour une meilleure lisibilité.

Annexe B). Considérons par exemple le triplet incomplet (bonnet I, syn, ?). Étant donnée la distribution prédictive originale du modèle, nous appliquons le MC Dropout pour générer $n = 100$ nouvelles distributions pour le triplet. Ensuite, nous définissons un critère d’inclusion pour transformer les prédictions à valeur réelle en décisions binaires, ici nous vérifions si l’entité prédite figure parmi les 10 premières prédictions notées. Nous calculons enfin le score de confiance comme le rapport entre le nombre de fois où l’entité prédite apparaît dans le top-10 et le nombre de distributions prédictives n . Cela nous permet d’établir des scores de confiance pour les prédictions du modèle et de faire des affirmations telles que « Selon ce modèle, l’entité queue kepi est dans le top-10 des prédictions pour l’entité tête bonnet I et la relation synonymie avec 75 % de confiance ».

6 Analyse qualitative

Nous cherchons à déterminer si le score de confiance permet d’identifier des triplets pertinents susceptibles d’être ajoutés aux graphes. Pour cela, nous avons généré toutes les combinaisons possibles de nœuds et de relations pour chaque jeu de données et nous avons supprimé les triplets déjà présents dans les graphes. Au total, 533 551 triplets ont été générés pour RLF27k et 1 720 454 pour RezoJDM16k. Nous nous sommes concentrés sur les triplets dont les entités ne sont pas reliées par un chemin orienté. Pour RLF27k, nous avons obtenu 95 766 triplets finaux. Pour RezoJDM16k, en raison de la forte densité du graphe, le plus court chemin entre deux entités est au maximum de longueur 4. Pour appliquer une méthodologie similaire à celle utilisée pour RLF27k, nous avons conservé les triplets avec des chemins de longueurs 3 et 4, ce qui donne un total de 154 168 triplets.

Pour évaluer le score de confiance, 240 triplets par jeu de données ont été annotés par quatre annotateurs, chacun ayant annoté 120 triplets ce qui permet d’obtenir deux annotations pour chaque triplet et calculer un accord inter-annotateurs (AIA). Les scores de confiance sont représentés de manière homogène dans les échantillons de chaque annotateur, regroupés par intervalle de 0,1. La tâche d’annotation consiste à déterminer si un lien sémantique ou syntaxique existe entre deux entités.

Les annotateurs ne disposent d’aucune information en dehors du triplet et trois étiquettes d’annotation sont possibles : (1) il existe un lien entre les entités, (-1) il n’y a pas de lien, (0) le lien est ambigu ou discutable. L’AIA est calculé avec un kappa de Cohen et s’avère bien plus élevé pour le RLF27k avec l’accord le plus fort à 0,84 contre 0,61 pour RezoJDM16k (cf. Annexe C). L’accord minimum pour RezoJDM16k est particulièrement faible, à 0,1, contrairement à 0,49 pour le RLF27k. La différence de l’AIA entre RezoJDM16k et RLF27k peut être attribuée à la distribution inégale des étiquettes d’annotation, et notamment à la surreprésentation de l’étiquette -1 dans RezoJDM16k. Sur les 240 triplets et sur les 183 où les annotateurs sont en accord, 85 % (156 triplets) ont été annotés comme -1 et 15 % (27) comme 1. Cette proportion s’explique par la forte densité du graphe qui implique que des nœuds sémantiquement distants sont connectés par un chemin relativement court de 4.

La figure 3 présente la comparaison entre les annotations manuelles et les scores de confiance des triplets des échantillons de RezoJDM16k et RLF27k. Pour RezoJDM16k, les rares triplets annotés comme exacts ont tendance à présenter des scores de confiance élevés. Toutefois, en raison de la prévalence des annotations -1, il est difficile d’établir une corrélation solide. D’autre part, dans l’échantillon RLF27k, sur les 200 triplets où les deux annotateurs étaient d’accord, 56,5 % (113) ont été annotés comme -1, 39 % (78) comme 1 et 4,5 % (9) comme 0. Nous pouvons noter que les étiquettes d’annotation dans RLF27k présentent une distribution plus équilibrée par rapport à RezoJDM16k et qu’on observe une plus forte corrélation entre les étiquettes d’annotation et les scores de confiance sur les triplets de RLF27k. Les triplets annotés comme -1 ont un score de confiance faible, tandis que ceux annotés comme 1 ont un score de confiance plus élevé.

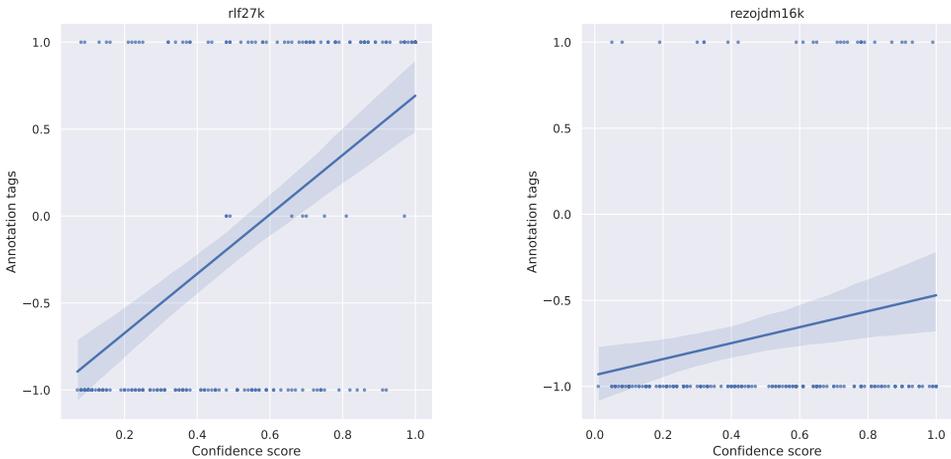


FIGURE 3 – Corrélations entre les valeurs d’annotation et les scores de confiance dans RLF27k et de RezoJDM16k. Les triplets considérés sont ceux pour lesquels les annotateurs sont d’accord.

La figure 5 présente le ratio des triplets corrects, i.e. annotées par les annotateurs comme ayant un lien entre ses entités, en fonction du seuil de confiance. Nous remarquons qu’un seuil de confiance élevé conduit à une plus grande proportion de triplets corrects. Par exemple, poser un seuil de confiance de 0,95 permet de faire en sorte que tous les triplets prédits soient corrects. Sur le graphe RLF27k complet, il existe 95 766 triplets dont les entités ne sont pas connectées par un chemin orienté et 398 d’entre eux ont un score de confiance supérieur à 0,95, qui peuvent donc être considérés comme de nouveaux triplets potentiels.

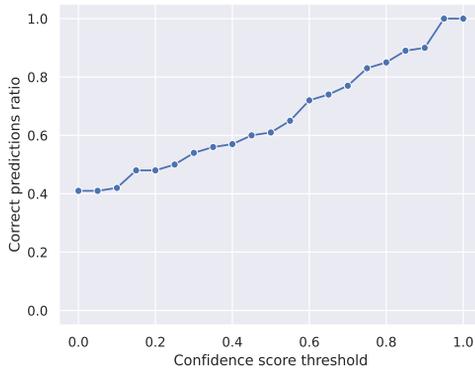


FIGURE 4 – Ratio des prédictions correctes (annotées 1) du RLF27k selon le seuil de confiance.

7 Discussion

Dans cette étude, nous avons cherché à identifier de nouveaux triplets potentiels pour deux graphes lexico-sémantiques, RezoJDM16k et RLF27k, en utilisant des modèles de prédiction de liens. Sur les sept modèles, le modèle CompGCN-Conve a obtenu les meilleures performances sur les deux graphes, surpassant les modèles état-de-l’art pour RezoJDM16k. Nous avons également étudié l’utilisation de l’algorithme MC Dropout pour générer des prédictions basées sur un score de confiance que nous avons évalué avec des annotations manuelles.

L’analyse qualitative que nous avons menée montre que les triplets de RLF27k ayant un score de confiance élevé sont de potentiels candidats à l’intégration dans la ressource, sous réserve d’une validation par des experts. Pour RezoJDM16k, du fait de sa forte densité, l’ajout de nouveaux triplets n’apporte pas d’avantages significatifs dans la mesure où même les entités ayant une faible proximité sémantique sont connectées par des chemins courts. Cependant, notre approche peut s’avérer utile pour identifier des erreurs ou affiner des relations génériques. En effet, RezoJDM16k présente une forte proportion de la relation générique *associated*, constituant 31 % des arêtes, qui pourrait être affinée grâce aux prédictions du modèle.

Bien que les résultats de notre approche soient prometteurs, nous maintenons que la vérification manuelle est une étape importante car la représentation de la polysémie dans des nœuds distincts affecte directement les prédictions. Étant donné que le modèle ne s’appuie que sur la structure du graphe et les nœuds voisins pour appréhender la sémantique, il existe des limites inhérentes à la prédiction de l’entité précise parmi les différents sens d’un vocable. Dans des recherches ultérieures, au-delà des méthodologies d’évaluation intrinsèque, nous avons pour perspective d’évaluer de manière extrinsèque en utilisant des graphes lexico-sémantiques augmentés ou corrigés dans des tâches de TAL, telle que la levée d’ambiguïté lexicale. En outre, d’un point de vue ressource, nous souhaitons explorer les avantages mutuels et l’amélioration entre les deux réseaux lexico-sémantiques français. Le code associé à nos expériences et les graphiques qui en résultent sont disponibles librement sur le lien suivant : <https://github.com/hschoi4/fr-link-prediction>.

Remerciements

Nous remercions les relecteurs pour leurs commentaires avisés qui ont permis l'amélioration de cet article. Nous remercions également Vincent Tourneur (LORIA, INRIA) pour son aide dans les manipulations techniques. Les expériences présentées dans cet article ont été réalisées sur le banc de tests Grid'5000, soutenu par un groupe inter-scientifique hébergé par l'Inria et comprenant le CNRS, le RENATER et plusieurs universités ainsi que d'autres organisations (<https://www.grid5000.fr>).

Références

- ARCAN M., MCCRAE J. P. & BUITELAAR P. (2016). Expanding wordnets to new languages with multilingual sense disambiguation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics : Technical Papers*, p. 97–108, Osaka, Japon : The COLING 2016 Organizing Committee.
- BIEMANN C., SHIN S.-I. & CHOI K.-S. (2004). Semiautomatic extension of CoreNet using a bootstrapping mechanism on corpus-based co-occurrences. In *COLING 2004 : Proceedings of the 20th International Conference on Computational Linguistics*, p. 1227–1232, Genève, Suisse : COLING.
- BORDES A., USUNIER N., GARCIA-DURÁN A., WESTON J. & YAKHNENKO O. (2013). Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13*, p. 2787–2795, Red Hook, États-Unis : Curran Associates Inc.
- CHEN Z., WANG Y., ZHAO B., CHENG J., ZHAO X. & DUAN Z. (2020). Knowledge graph completion : A review. *IEEE Access*, **8**, 192435–192456. DOI : [10.1109/ACCESS.2020.3030076](https://doi.org/10.1109/ACCESS.2020.3030076).
- D'AMATO C., MAHON L., MONNIN P. & STAMOU G. (2023). Machine Learning and Knowledge Graphs : Existing Gaps and Future Research Challenges. *Transactions on Graph Data and Knowledge*, **1**(1), 1–35. DOI : [10.4230/TGDK.1.1.8](https://doi.org/10.4230/TGDK.1.1.8), HAL : [hal-04353543](https://hal.archives-ouvertes.fr/hal-04353543).
- DETTMERS T., MINERVINI P., STENETORP P. & RIEDEL S. (2018). Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'18/IAAI'18/EAAI'18* : AAAI Press.
- FELLBAUM C. (1998). *WordNet : An Electronic Lexical Database*. The MIT Press. DOI : [10.7551/mitpress/7287.001.0001](https://doi.org/10.7551/mitpress/7287.001.0001).
- GAL Y. & GHAHRAMANI Z. (2016). Dropout as a bayesian approximation : Representing model uncertainty in deep learning. In M. F. BALCAN & K. Q. WEINBERGER, Édts., *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 de *Proceedings of Machine Learning Research*, p. 1050–1059, New York, États-Unis : PMLR.
- GALKIN M., BERRENDORF M. & HOYT C. T. (2022). An open challenge for inductive link prediction on knowledge graphs.
- HWANG J. D., BHAGAVATULA C., LE BRAS R., DA J., SAKAGUCHI K., BOSSELUT A. & CHOI Y. (2021). Comet-atomic 2020 : On symbolic and neural commonsense knowledge graphs. In *AAAI*.

- Ji G., HE S., XU L., LIU K. & ZHAO J. (2015). Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1 : Long Papers)*, p. 687–696, Pékin, Chine : Association for Computational Linguistics. DOI : [10.3115/v1/P15-1067](https://doi.org/10.3115/v1/P15-1067).
- KIPF T. N. & WELLING M. (2017). Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Conference Track Proceedings*, Toulon, France : OpenReview.net.
- LAFOURCADE M. & JOUBERT A. (2008). JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes. In *JADT'08 : Journées internationales d'Analyse statistiques des Données Textuelles*, p. 657–666, France. HAL : [lirmm-00358848](https://hal.archives-ouvertes.fr/lirmm-00358848).
- LAFOURCADE M. & LE BRUN N. (2020). Jeuxdemots : Un réseau lexico-sémantique pour le français, issu de jeux et d'inférences. *Revue Lexique*, **27**, 47–86.
- LAM K. N., AL TAROUTI F. & KALITA J. (2014). Automatically constructing Wordnet synsets. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2 : Short Papers)*, p. 106–111, Baltimore, Maryland : Association for Computational Linguistics. DOI : [10.3115/v1/P14-2018](https://doi.org/10.3115/v1/P14-2018).
- LIN Y., LIU Z., SUN M., LIU Y. & ZHU X. (2015). Learning entity and relation embeddings for knowledge graph completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, **29**(1). DOI : [10.1609/aaai.v29i1.9491](https://doi.org/10.1609/aaai.v29i1.9491).
- LUX-POGODALLA V. & POLGUÈRE A. (2011). Construction of a French Lexical Network : Methodological Issues. In *First International Workshop on Lexical Resources, WoLeR 2011*, p. 54–61, Ljubljana, Slovénie. HAL : [hal-00686467](https://hal.archives-ouvertes.fr/hal-00686467).
- MEL'ČUK I. (1996). Lexical functions in lexicography and natural language processing. *Lexical Functions : A Tool for the Description of Lexical Relations in the Lexicon*, p. 37–102.
- MILLER G. A. (1995). WordNet : A lexical database for English. *Communications of the ACM*, **38**(11), 39–41.
- MIRZAPOUR M., RAGHEB W., SAEEDIZADE M. J., COUSOT K., JACQUENET H., CARBON L. & LAFOURCADE M. (2022). Introducing RezoJDM16k : a French KnowledgeGraph DataSet for link prediction. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, p. 5163–5169, Marseille, France : European Language Resources Association.
- NAVIGLI R. & PONZETTO S. P. (2012). Babelnet : The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, **193**, 217–250. DOI : <https://doi.org/10.1016/j.artint.2012.07.001>.
- NICKEL M., TRESP V. & KRIEGEL H.-P. (2011). A three-way model for collective learning on multi-relational data. In L. GETOOR & T. SCHEFFER, Éd., *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, ICML '11, New York, États-Unis : ACM.
- OLIVER A. & CLIMENT S. (2012). Parallel corpora for wordnet construction : Machine translation vs. automatic sense tagging. In A. GELBUKH, Éd., *Computational Linguistics and Intelligent Text Processing*, p. 110–121, Berlin, Heidelberg : Springer Berlin Heidelberg.
- POLGUÈRE A. (2018). A Lexicographic Approach to the Study of Copolysemy Relations *. *Russian Journal of Linguistics = Vestnik Rossijskogo universiteta družby narodov. Seria Lingvistika*, **22**(4), 788 – 820. DOI : [10.22363/2312-9182-2018-22-4-788-820](https://doi.org/10.22363/2312-9182-2018-22-4-788-820), HAL : [halshs-02089585](https://hal.archives-ouvertes.fr/halshs-02089585).
- RILOFF E. & SHEPHERD J. (1997). A corpus-based approach for building semantic lexicons. In *Second Conference on Empirical Methods in Natural Language Processing*.

SCHLICHTKRULL M., KIPF T. N., BLOEM P., VAN DEN BERG R., TITOV I. & WELLING M. (2018). Modeling relational data with graph convolutional networks. In A. GANGEMI, R. NAVIGLI, M.-E. VIDAL, P. HITZLER, R. TRONCY, L. HOLLINK, A. TORDAI & M. ALAM, Éd., *The Semantic Web*, p. 593–607, Cham : Springer International Publishing.

SUN Z., DENG Z., NIE J. & TANG J. (2019). Rotate : Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations, ICLR 2019*, La Nouvelle-Orléans, États-Unis : OpenReview.net.

TAGHIZADEH N. & FAILI H. (2016). Automatic wordnet development for low-resource languages using cross-lingual wsd. *J. Artif. Int. Res.*, **56**(1), 61–87.

TOUTANOVA K. & CHEN D. (2015). Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, p. 57–66, Pékin, Chine : Association for Computational Linguistics. DOI : [10.18653/v1/W15-4007](https://doi.org/10.18653/v1/W15-4007).

TROUILLON T., WELBL J., RIEDEL S., GAUSSIER E. & BOUCHARD G. (2016). Complex embeddings for simple link prediction. In M. F. BALCAN & K. Q. WEINBERGER, Éd., *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 de *Proceedings of Machine Learning Research*, p. 2071–2080, New York, États-Unis : PMLR.

VASHISHTH S., SANYAL S., NITIN V. & TALUKDAR P. P. (2020). Composition-based multi-relational graph convolutional networks. In *8th International Conference on Learning Representations, ICLR 2020*, Addis-Abeba, Éthiopie : OpenReview.net.

WANG Z., ZHANG J., FENG J. & CHEN Z. (2014). Knowledge graph embedding by translating on hyperplanes. *Proceedings of the AAAI Conference on Artificial Intelligence*, **28**. DOI : [10.1609/aaai.v28i1.8870](https://doi.org/10.1609/aaai.v28i1.8870).

YANG B., YIH W., HE X., GAO J. & DENG L. (2015). Embedding entities and relations for learning and inference in knowledge bases. In Y. BENGIO & Y. LECUN, Éd., *3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings*, San Diego, États-Unis.

Annexes

A Exemple de relations de copolysémie

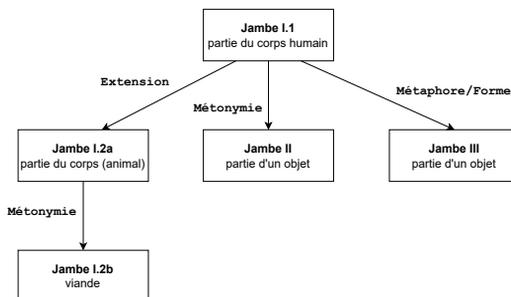


FIGURE 5 – Relation de copolysémie du vocable jambe.

B Monte Carlo Dropout

Notre objectif est d'obtenir une approximation de la distribution prédictive pour un point donné à partir de notre modèle paramétré. La distribution prédictive $p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{X}, \mathbf{Y})$ pour un nouveau point de données \mathbf{x}^* à partir d'un jeu de données (\mathbf{X}, \mathbf{Y}) et un modèle paramétré par θ est :

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{X}, \mathbf{Y}) = \int p(\mathbf{y}^*|\mathbf{x}^*, \theta)p(\theta|\mathbf{X}, \mathbf{Y}) d\theta \quad (1)$$

Ici, le premier terme est la probabilité d'observer la sortie y^* compte tenu de l'entrée x^* et des paramètres du modèle θ , et le second terme est la distribution *a posteriori* des paramètres compte tenu des données. Le calcul de cette dernière est difficile à réaliser pour les réseaux neuronaux profonds en raison de leur espace paramétrique de grande dimension. Monte-Carlo (MC) Dropout (Gal & Ghahramani, 2016) nous permet d'approximer cette intégrale. En effectuant T passages stochastiques à travers le réseau avec le *dropout* activé, nous obtenons T prédictions pour chaque entrée de test \mathbf{x}^* . La distribution empirique de ces prédictions se rapproche de la distribution prédictive $p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{X}, \mathbf{Y})$. Mathématiquement, cela peut s'exprimer comme suit :

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{X}, \mathbf{Y}) \approx \frac{1}{T} \sum_{t=1}^T p(\mathbf{y}^*|\mathbf{x}^*, \theta_t) \quad (2)$$

C Accords inter-annotateurs

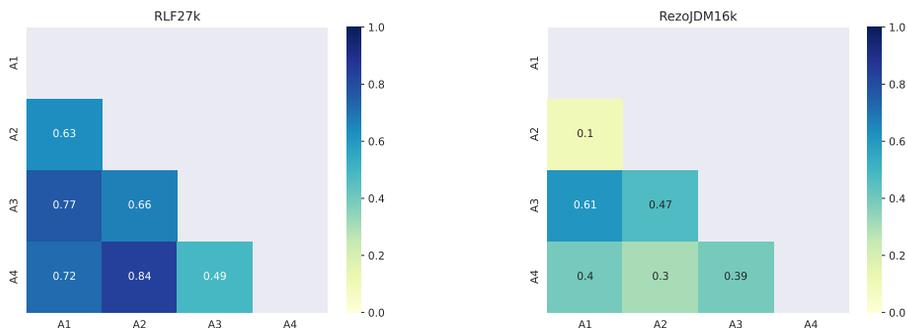


FIGURE 6 – Kappa de Cohen sur RLF27k et RezoJDM16k.