HAL open science

Inférences déductives et réconciliation dans un réseau lexico-sémantique

Manel Zarrouk, Mathieu Lafourcade, Alain Joubert

To cite this version:

Manel Zarrouk, Mathieu Lafourcade, Alain Joubert. Inférences déductives et réconciliation dans un réseau lexico-sémantique. TALN: Traitement Automatique des Langues Naturelles, Jun 2013, Sables d'Olonne, France. lirmm-00816229

HAL Id: lirmm-00816229

https://hal-lirmm.ccsd.cnrs.fr/lirmm-00816229v1

Submitted on 11 Jun 2013

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.

Inférences déductives et réconciliation dans un réseau lexico-sémantique

Manel Zarrouk Mathieu Lafourcade Alain Joubert
LIRMM, 161, rue ADA 34095 Montpellier Cedex 5
manel.zarrouk@lirmm.fr,mathieu.lafourcade@lirmm.fr,Alain.Joubert@lirmm.fr

RÉSUMÉ

La construction et la validation des réseaux lexico-sémantiques est un enjeu majeur en TAL. Indépendamment des stratégies de construction utilisées, inférer automatiquement de nouvelles relations à partir de celles déjà existantes est une approche possible pour améliorer la couverture et la qualité globale de la ressource. Dans ce contexte, le moteur d'inférences a pour but de formuler de nouvelles conclusions (c'est-à-dire des relations entre les termes) à partir de prémisses (des relations préexistantes). L'approche que nous proposons est basée sur une méthode de triangulation impliquant la transitivité sémantique avec un mécanisme de blocage pour éviter de proposer des relations douteuses. Les relations inférées sont proposées aux contributeurs pour être validées. Dans le cas d'invalidation, une stratégie de réconciliation est engagée pour identifier la cause de l'inférence erronée : une exception, une erreur dans les prémisses, ou une confusion d'usage causée par la polysémie.

ABSTRACT

Inductive and deductive inferences in a Crowdsourced Lexical-Semantic Network

In Computational Linguistics, validated lexical-semantic networks are crucial resources. Regardless the construction strategies used, automatically inferring new relations from already existing ones may improve coverage and global quality of the resource. In this context, an inference engine aims at producing new conclusions (i.e. potential relations) from premises (pre-existing relations). The approach we propose is based on a triangulation method involving the semantic transitivity with a blocking mechanism to avoid proposing dubious relations. Inferred relations are then proposed to contributors to be validated or rejected. In cas of invalidation, a reconciliation strategy is implemented to identify the cause of the erroneous inference : an exception, an error in the premises, or a confusion caused by polysemy.

MOTS-CLÉS : inférence de relations, réconciliation, enrichissement, réseau lexical, peuplonomie.

KEYWORDS: relation inferences, reconcialiation, enrichment, lexical network, crowdsourcing.

1 Introduction

Développer un réseau lexico-sémantique pour le TAL est l'un des enjeux majeurs du domaine. La plupart des ressources existantes ont été construites à la main, comme dans le cas de WordNet (Miller *et al.*, 1990). Bien entendu, quelques outils sont généralement utilisés pour la vérification de la cohérence, mais cependant la tâche reste coûteuse en temps et en prix. Des

approches entièrement automatisées sont généralement limitées à la co-occurrence des termes car l'extraction des relations sémantiques précises entre termes à partir d'un texte reste difficile. De nouvelles approches impliquant l'externalisation ouverte (*crowdsourcing*) émergent dans le TAL spécialement avec l'avènement de Amazon Mechanical Turk ou plus largement avec Wikipédia et le Wiktionnaire pour ne citer que les plus connus. Wordnet ((Miller *et al.*, 1990) et (Fellbaum et Miller, 1998)) est un réseau lexical basé sur des synsets qui peuvent être globalement considérés comme des concepts. (Vossen, 1998) avec EuroWordnet, une version multi-langues de Wordnet et (Sagot et Fier, 2008) avec WOLF, une version française de Wordnet, ont utilisé des croisements automatiques de Wordnet avec d'autres ressources lexicales suivi d'une vérification manuelle partielle. (Navigli et Ponzetto, 2012) a construit BabelNet, un grand réseau lexical multilingue à partir de l'encyclopédie Wikipédia mais en se basant sur les co-occurences entre termes. Dans le domaine de l'intelligence artificielle, Cyc (Lenat, 1995) est un exemple de base de connaissances très redondante ayant demandé un effort manuel particulièrement important. Hownet (Dong et Dong, 2006) est un autre exemple d'une grande base de connaissances bilingue (anglais et chinois) contenant des relations sémantiques entre les formes de mots, les concepts et les attributs. Dans Hownet, il existe d'avantage de types de relations différents que dans Wordnet bien que les deux projets aient démarré pendant les années 80 et aient été manuellement construits par des linguistes et des psychologues.

Un réseau lexico-sémantique très lexicalisé peut contenir des concepts, mais aussi des termes, des formes de mots ainsi que des expressions composées comme points d'entrée vers des sens ou des usages. L'idée elle-même de *sens du mot* issue de la tradition lexicographique, peut être discutable dans le cas de ressources pour le TAL et l'analyse sémantique, et on préfèrera généralement considérer les mots dans leurs usages. Par *usage des mots*, on entend que le raffinement d'un mot donné est clairement identifiable par les locuteurs mais qu'il peut ne pas être séparé totalement des autres usages de la même entrée. Un usage de mot met l'accent sur le contexte d'utilisation utilisé par les locuteurs. Un terme polysémique peut avoir beaucoup d'usages substantiellement différents des définitions classiquement trouvées dans un dictionnaire. Un usage donné peut aussi avoir plusieurs raffinements. Par exemple, *frégate* peut être un oiseau ou un bateau. Une frégate>bateau peut être distinguée comme un bateau moderne ou un navire à voiles ancien. Dans le contexte d'une approche collaborative, une telle ressource lexicale peut être considérée comme étant constamment en cours de construction. Pour un terme polysémique, certains raffinements peuvent manquer et une règle générale est de n'avoir aucune certitude autour de l'état d'une entrée. Il est presque impossible (sauf par inspection manuelle) de savoir si l'ensemble des raffinements d'une entrée est exhaustif, voire même si cette question a vraiment du sens dans le contexte d'une ressource dynamique.

La construction d'un réseau lexical collaboratif (ou de n'importe quelle ressource similaire) peut être catégorisée selon deux stratégies. Premièrement, comme un système contributif du type Wikipédia où des volontaires complètent les entrées (cas du Wiktionnaire). Dans un second cas, les contributions sont faites indirectement par l'entremise de jeux, connus sous le nom de GWAP (Game With A Purpose) (von Ahn et Dabbish, 2008). Dans ce cas, les joueurs n'ont pas spécialement besoin d'être conscients qu'ils sont en train de participer à la construction d'une ressource lexicale. En aucun cas il ne faudrait croire que le réseau construit serait dépourvu d'erreurs, erreurs qui sont corrigées au fur et à mesure de leur découverte. L'expérience montre que les joueurs/contributeurs complètent le réseau sur ce qui leur parait intéressant. Ce faisant, un grand nombre de relations *triviales* ne sont pas

présentes bien qu'elles demeurent pourtant nécessaires à un réseau de qualité devant être utilisé dans diverses applications du TAL dont notamment l'analyse sémantique. Par exemple, les joueurs n'indiquent que rarement pour un type d'oiseau particulier que celui-ci peut voler tant cela parait une généralité évidente. Seuls les faits notables et peu facilement déductibles sont renseignés par les contributeurs. Les exceptions sont également renseignées par les contributeurs, et prennent la forme d'une relation ayant un poids négatif (*voler* $\xrightarrow{agent:-100}$ *autruche*).

Afin de consolider le réseau lexical issu du projet JeuxDeMots, nous utilisons une approche par inférence qui permet de déduire de nouvelles relations à partir de celles existantes. L'approche est uniquement endogène en ce qu'elle ne s'appuie sur aucune ressource externe. Les relations inférées sont soumises aux contributeurs pour vote et par la suite soumises à une validation ou invalidation par un expert. Une grande majorité des inférences se révèle correcte. Toutefois, une part non négligeable se révèle fausse et il convient de déterminer pourquoi. Ce processus d'explication constitue la réconciliation entre le moteur d'inférences et le validateur, mené à l'aide d'un dialogue lui permettant d'expliciter en quoi la relation considérée est incorrecte. Les causes possibles sont de trois ordres : erreur dans une des prémisses, exception, ou confusion liée à la polysémie.

Dans cet article, nous présentons tout d'abord les principes de construction du réseau lexical par externalisation ouverte et GWAP (*games with a purpose* ou *human-based computation game*) et nous les illustrons grâce au projet JeuxDeMots. Ensuite, nous détaillons un moteur d'élicitation composé d'un moteur d'inférences et d'un moteur de réconciliation. Une expérimentation sur les performances du système est ensuite rapportée.

2 Réseau lexical et externalisation ouverte

Il existe beaucoup de méthodes pour construire un réseau lexical en tenant compte des facteurs principaux tels que la qualité des données, le coût et le temps de développement. En marge des stratégies manuelles et automatiques, les approches contributives connaissent une popularité croissante en ce qu'elles se révèlent à la fois peu coûteuses et efficaces en qualité. Plus précisément, l'intérêt donné aux GWAP (Thaler *et al.*, 2011) comme méthode d'acquisition de tels réseaux augmente considérablement. Le réseau de JeuxDeMots (JDM) est un réseau lexical construit à partir d'un ensemble de jeux en ligne. Dans ces jeux, les joueurs sont invités à faire des associations entre termes qui se traduisent en relations lexicales et sémantiques entre les nœuds d'un graphe. Les informations dans le réseau JDM sont récoltées via un accord non négocié entre les joueurs qui sont les acteurs d'une externalisation ouverte [1].

2.1 JeuxDeMots : un GWAP pour la construction d'un réseau lexical

Lancé en septembre 2007, JeuxDeMots [2] (Lafourcade, 2007) est un GWAP [3] associant les joueurs par paires, et visant à construire un grand réseau lexico-sémantique. Le réseau lexical contruit est composé de termes (noeuds) et de relations typées (arcs). Il contient des termes potentiellement raffinés de manière analogue aux synsets de WordNet. Il y a plus de 50 types de relations et chaque occurrence de relation est pondérée indiquant une force d'association. Une occurrence de relation peut éventuellement avoir un poids négatif indiquant dans ce cas que la relation est fausse bien que pertinente (ex : une autruche ne vole pas).

1. http ://fr.wikipedia.org/wiki/Crowdsourcing
2. http ://www.jeuxdemots.org/
3. http ://en.wikipedia.org/wiki/Human-based_computation_game

Quand un joueur A commence une partie, une consigne concernant le type de la relation lexicale (synonymie, antonymie, domaine, etc.) est affichée, ainsi qu'un terme cible T choisi dans le réseau lexical. Ce joueur A a un temps limité pour saisir des termes qui, selon lui, correspondent au terme T et à la relation lexicale qu'indique la consigne. Le nombre maximum des termes que le joueur peut saisir lors d'une partie est limité, ce qui l'incite à réfléchir soigneusement à ses choix. Le même terme T, avec les mêmes instructions, est donné ultérieurement à un autre joueur B, pour qui le processus est identique. Pour rendre le jeu plus amusant, les deux joueurs obtiennent des points pour les mots qu'ils ont donné en commun. Le calcul du score est expliqué dans (Joubert et Lafourcade, 2008) et est conçu pour augmenter à la fois la précision et le rappel lors de la construction du réseau. Les réponses présentées par les deux joueurs sont affichées, celles en commun sont mises en évidence ainsi que leur score.

Pour un terme cible T, les réponses communes entre les deux joueurs sont insérées dans la base de données. Celles données seulement par un seul des deux joueurs ne le sont pas ce qui réduit considérablement le bruit et les chances de dégradation du réseau. Le réseau sémantique est donc construit par des termes connectés par des relations typées et pondérées validées par paires de joueurs. Ces relations sont étiquetées selon les instructions données aux joueurs et sont pondérées selon le nombre de paires de joueurs ayant choisi ces relations. Initialement et avant la mise du jeu en ligne, la base de données a été remplie avec des termes, cependant si une paire de joueurs suggère un terme non-existant, le nœud correspondant à ce dernier est ajouté à la base de données.

Afin de préserver la qualité et la consistance du réseau lexical, il a été décidé que le processus de la validation implique des joueurs anonymes jouant ensemble. Une relation est considérée valide si et seulement si elle est proposée par au moins une paire de joueurs. Ce processus de validation est similaire à celui utilisé par (von Ahn et Dabbish, 2008) pour l'indexation des images et par (Lieberman *et al.*, 2007) pour la collecte de la connaissance du "sens commun" et (Siorpaes et Hepp, 2008) pour l'extraction de connaissance. A notre connaissance, cette technique n'a jamais été utilisée lors de la construction des réseaux sémantiques. Dans le TAL, d'autres jeux accessibles via le web existent, comme Open Mind Word Expert (Mihalcea et Chklovski, 2003) qui vise a créer un grand corpus étiqueté sémantiquement avec l'aide des utilisateurs du Web ou comme SemKey (Marchetti *et al.*, 2007) qui utilise WordNet et Wikipédia pour désambiguïser les formes lexicales se référant à des concepts, ainsi identifiant les mots-clés sémantiques. Plus de 1300000 parties ont été jouées depuis le lancement pour environ 25000 heures de temps cumulé de jeu.

2.2 Diko : un outil contributif pour le réseau JDM

Diko est un outil basé sur le web permettant d'afficher les informations contenues dans le réseau lexical JDM ainsi qu'un outil contributif. La nécessité de ne pas dépendre seulement de jeux pour construire le réseau lexical vient du fait qu'une part non négligeable des types de relations de JDM soit sont difficiles à saisir pour un joueur non expert ou soit sont peu contributifs (il n'existe pas beaucoup de termes qui peuvent lui être associés). En outre, le besoin d'un outil contributif vient historiquement des joueurs eux-mêmes qui ont voulu devenir des contributeurs du réseau JDM.

Le principe du processus de la contribution est qu'une proposition faite par un joueur sera soumise aux votes des autres joueurs. Une fois un certain nombre de votes donnés, un expert validateur est averti et finit par inclure (ou éventuellement exclure) la relation proposée dans le réseau. L'expert peut rejeter totalement une proposition de relation ou l'inclure dans le

réseau avec un poids négatif s'il trouve cela pertinent. Un système de points et de classement encourage les contributeurs non seulement à proposer de nouvelles relations mais également à voter pour (ou contre) celles des autres. Les contributeurs les plus prolifiques dépassent les 150000 relations proposées ou votées (depuis septembre 2010). Les contributions peuvent aussi être proposées par un processus automatique pour être vérifiées et votées ultérieurement par des utilisateurs. Ce que nous proposons dans cet article se range sous cette catégorie de scénario.

2.3 Quelques caractéristiques du réseau JDM

En janvier 2013, le réseau de JDM contient environ 250000 termes et plus de 1500000 relations (dont environ 15000 qui sont négatives). Plus de 4500 termes ont quelques raffinements (usages variés) pour un total d'environ 15000 usages. Bien que JDM ait des relations ontologiques, il ne constitue pas une ontologie proprement dite avec des concepts ou des termes soigneusement hiérarchiques. Un terme donné peut avoir une collection substantielle d'hyperonymes qui couvre une large partie de la chaîne ontologique jusqu'aux concepts supérieurs. Par exemple : hyperonyme(chat) = {félin, mammifère, être vivant, animal de compagnie, vertébré, ...}. Dans la liste d'hyperonymes précédente, nous avons omis les poids pour simplifier, mais en toute généralité, les termes les plus *lourds* sont ceux considérés par les utilisateurs comme les plus pertinents.

3 Inférence et réconciliation dans un moteur d'élicitation

Les travaux sur l'inférence dans les réseaux lexicaux sont curieusement assez peu répandus, en particulier si on les compare aux approches concernant diverses formes de déduction logique dans les textes (pour la détermination du référent dans les groupes circonstanciels, ou la résolution d'anaphores, par exemple). Dans (Blanco et Moldovan, 2011), des relations sémantiques entre concepts présents dans des textes sont inférées. Les termes du texte sont considérés de facto comme des concepts en ignorant de façon quelque peu artificielle tout problème lié à la polysémie lexicale. Dans (Harabagiu et Moldovan, 1998) du raisonnement par inférence est effectué sur le WordNet étendu où non seulement les synsets mais également les termes des définitions associés sont des concepts délexicalisés.

Pour augmenter le nombre de relations dans le réseau lexical JDM, nous avons conçu un système d'élicitation [4] ayant deux principales composantes : (1) un moteur d'inférences et (2) un réconciliateur. Le moteur d'inférences propose des relations, tout comme un contributeur, qui vont être évaluées par la suite par un autre contributeur humain. Dans le cas d'invalidation d'une relation inférée, le réconciliateur est appelé pour essayer d'évaluer pourquoi la relation a été trouvée fausse. L'élicitation est ici le processus de formalisation de connaissances implicites de l'utilisateur en relations explicites dans le réseau lexical.

3.1 Moteur d'inférences

Les idées principales sous-tendant le moteur d'inférences sont les suivantes. Inférer pour le moteur c'est dériver des conclusions logiques sous la forme de relations entre les termes à partir de prémisses (les relations existantes dans le réseau). Les inférences candidates peuvent

4. L'élicitation en Gestion des Connaissances est l'action d'aider un expert à formaliser ses connaissances pour permettre de les sauvegarder et/ou les partager. Celui ou celle qui élicite va donc inviter l'expert à rendre ses connaissances tacites en connaissances aussi explicites que possible.

être logiquement bloquées en se basant sur la présence ou l'absence de quelques autres relations. Les inférences candidates peuvent également être filtrées et rejetées de prime abord en se basant sur le niveau de force de l'évaluation. Les conclusions faites par le moteur sont supposées être correctes, mais une fois proposées à un validateur humain peuvent s'avérer tout aussi bien incorrectes, correctes avec un degré de précision (polysémie), correctes dans certains cas (exception) ou correctes mais pas pertinentes. Dans cet article, le type d'inférence qui nous intéresse est basé sur la transitivité de la relation ontologique *is-a* (hyperonyme). Si un terme A est un type de B et B a une relation R avec le terme C, alors on peut espérer que A entretienne la même relation avec C. Le schéma de l'inférence est le suivant :

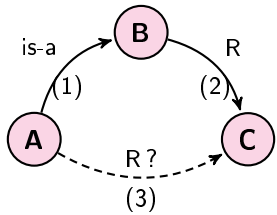

FIGURE 1 – Schéma d'inférence triangulaire simple appliqué à la transitivité de l'hyperonymie (la relation *is-a*). Les relations (1) et (2) sont les prémisses et la relation (3) est la conclusion logique proposée dans le réseau lexical en attendant d'être validée.

Plus formellement on peut écrire : $\exists\, A \xrightarrow{is-a} B \quad \wedge \quad \exists\, B \xrightarrow{R} C \quad \Rightarrow \quad A \xrightarrow{R} C$

Par exemple, $chat \xrightarrow{is-a} félin \quad \wedge \quad félin \xrightarrow{has-part} griffe \quad \Rightarrow \quad chat \xrightarrow{has-part} griffe$

3.1.1 Traitement global

Le moteur d'inférences est appliqué sur les termes ayant au minimum un hyperonyme (sinon le schéma ne peut pas être appliqué). Considérons un terme T avec un assortiment d'hyperonymes pondérés. A partir de chaque hyperonyme, le moteur d'inférences déduit un ensemble d'inférences. Généralement ces inférences ne sont pas disjointes et le poids d'une inférence proposée dans plusieurs ensembles est la moyenne géométrique incrémentale de chaque occurrence (c'est-à-dire que la présence d'un poids négatif suffit à rendre la moyenne invalide).

Par exemple, nous avons l'ensemble pondéré d'hyperonymes suivants pour *chat* : *félin - animal - être vivant - mammifère - animal de compagnie - félidé - animal domestique - vertébré*. L'inférence $chat \xrightarrow{has-part} squelette$ peut provenir de plusieurs hyperonymes mais fort probablement de *vertébré*. L'inférence $chat \xrightarrow{location} maison$ ne peut provenir que de l'hyperonyme *animal de compagnie*.

3.1.2 Filtrage logique

Bien sûr, le schéma ci-dessus est très naïf, spécialement si nous tenons compte de la ressource que nous traitons. En effet, B est possiblement un terme polysémique et des méthodes pour bloquer les inférences certainement fausses peuvent et doivent être conçues. Si le terme B relié à la première et la deuxième relation a deux sens distincts, alors probablement l'inférence est fausse. La condition pour proposer l'inférence peut être formalisée comme suit (la relation *raff-of* correspondant à celle de raffinement/usage pour un terme) :

$$
\begin{array}{c}
\exists\, A \xrightarrow{is-a} B \quad \wedge \quad \exists\, B \xrightarrow{R} C \\
\bigwedge \quad \exists\, B_i \xrightarrow{raff-of} B \quad \wedge \quad \exists\, B_j \xrightarrow{raff-of} B \\
\bigwedge \quad (\quad \nexists\, A \xrightarrow{is-a} B_i \quad \vee \quad \nexists\, B_j \xrightarrow{R} C \quad) \\
\Rightarrow \quad A \xrightarrow{R} C
\end{array}
$$

En d'autres termes, l'inférence est bloquée si le terme central B posséde deux raffinements distincts B_i et B_j tels que $A \xrightarrow{is-a} B_i$ et $B_j \xrightarrow{R} C$ (schéma présenté à la figure 2).

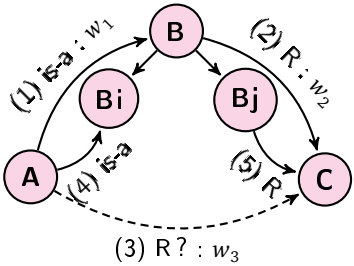

FIGURE 2 – Schéma d'inférence déductive triangulaire avec un bloquage logique se basant sur la polysémie du terme B du milieu. Les termes B_i et B_j sont des raffinements/usages de B.

Par ailleurs, si l'une des prémisses est annotée comme "correcte mais pas pertinente", l'inférence est bloquée, évitant ainsi de propager des inférences certes vraies mais ne présentant *a priori* que peu d'intérêt en pratique.

3.1.3 Filtrage statistique

Il est possible d'évaluer un niveau de confiance (ou d'intensité) pour les inférences produites, de façon à ce que les inférences douteuses puissent être filtrées et rejetées. Le poids P d'une relation inférée est la moyenne géométrique des poids des prémisses (relations (1) et (2) dans figure 1). Si ce poids est trop faible, l'inférence est rejetée. Si la deuxième relation a une valeur négative, le poids n'est pas calculable et la proposition est rejetée. Puisque la moyenne géométrique est moins tolérante aux petites valeurs que la moyenne arithmétique, les inférences qui ne sont pas basées sur deux relations de poids raisonnables ne vont probablement pas passer ce type de filtrage.

$$
\begin{array}{c}
\mathrm{P}(A \xrightarrow{R} C) = (\mathrm{P}(A \xrightarrow{is-a} B) * \mathrm{P}(B \xrightarrow{R} C))^{1/2} \\
\Rightarrow \quad w3 = (w1 * w2)^{1/2}
\end{array}
$$

3.2 Moteur de réconciliation

Le raisonneur propose des inférences déduites et ces relations inférées sont présentées au validateur pour décider si elles sont "plutôt vraies" , "plutôt fausses", "possibles" ou " vraies mais pas pertinentes". Dans le cas d'invalidation, le réconciliateur essaie de diagnostiquer les raisons : erreur (une des relations déjà existantes sur lesquelles le moteur s'est basé pour inférer est fausse), exception (un cas rare), polysémie (l'inférence est faite en se basant sur un terme de milieu polysémique) ou un "cas général" qui va être rencontré principalement lors des premières validations. Cela est réalisé par un dialogue avec l'utilisateur dont le but ici est de découvrir les raisons de l'invalidation et essayer de réconcilier le réseau avec des informations issues de l'utilisateur en rapport avec une relation inférée invalidée. Ce

dialogue doit être aussi minimal que possible (le moins possible de questions/le plus possible d'informations tirées). Pour savoir dans quel ordre procéder, le réconciliateur détermine si les poids des relations (1) et (2) sont relativement forts ou faibles et ce en comparant chaque poids au seuil de confiance correspondant à son terme (Figure 3). Par exemple, le terme A écolier ; B humain ; C visage ; relation 1 : $écolier \xrightarrow{is-a} humain$

La figure 3 est la courbe de poids/distributions pour toutes les relations sortantes ayant comme terme source A (qui est dans notre cas *écolier*) avec un pic séparant en deux parts égales la surface sous la courbe. Le seuil de confiance est l'intersection entre la courbe de distribution et ce pic (valant ici $\simeq 70$).

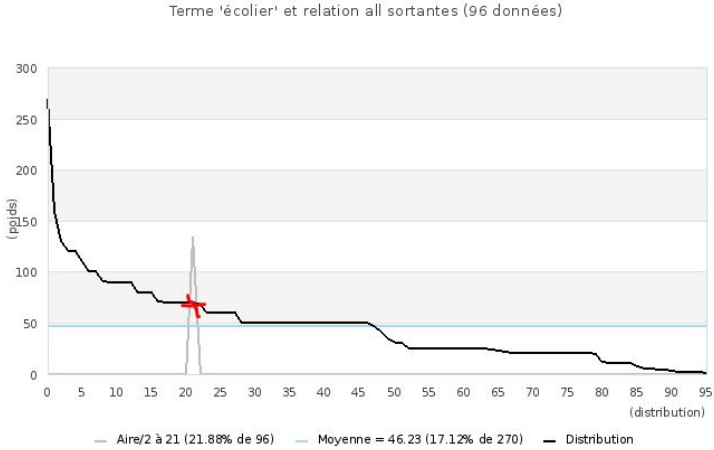

FIGURE 3 – Courbe poids/distribution des relations sortantes du terme A

- Si P($A \xrightarrow{is-a} B$) >= seuil-confiance(A) $\Rightarrow$ $A \xrightarrow{is-a} B$ est une relation vraisemblable.
- Si P($A \xrightarrow{is-a} B$) < seuil-confiance(A) $\Rightarrow$ $A \xrightarrow{is-a} B$ est une relation douteuse.

3.2.1 Exception → Polysémie → Erreur

Dans le cas où les deux relations de base (1) et (2) sont des relations vraisemblables, le réconciliateur essaie en premier lieu, en entamant un dialogue avec les validateurs/utilisateurs, de vérifier si la relation R inférée est une exception. Si ce n'est pas le cas, il vérifie si le terme B (terme central) est polysémique. Finalement si aucun des deux cas précédents ne s'avère vrai, il demande si c'est un cas d'erreur. Ce cas d'erreur est vérifié en dernier lors de cette démarche car les niveaux de confiance des deux relations les rendent plutôt vraisemblables. Soit l'exemple suivant : A :autruche ; B :oiseau ; C :voler ; R :carac

(1) : $autruche \xrightarrow{is-a} oiseau$ (2) : $oiseau \xrightarrow{carac} voler$ $\Rightarrow$ (3) : $autruche \xrightarrow{carac} voler$

Dans cet exemple, il est plutôt vrai que (1) "l'autruche est un oiseau" et que (2) "un oiseau peut voler" d'où les deux relations ont probablement des niveaux de confiance dépassant le seuil. Or la relation inférée (3)" l'autruche peut voler " est fausse et constitue une exception.

3.2.2 Erreur → Exception → Polysémie

Dans le cas où l'une des relations (1) et (2) est douteuse, le réconciliateur suspecte que c'est un cas d'erreur et que cette relation avec le faible niveau de confiance est l'origine de l'invalidation de la relation inférée. Alors il demande à l'utilisateur de la confirmer ou de l'invalider. Dans le cas de l'invalidation de l'une des relations, un cas d'erreur se présente. Sinon, on procède à la vérification des deux autres cas : cas d'exception ou de polysémie. Soit

l'exemple suivant : A :enfant ; B :humain ; C :aile ; R :has-part

(1) : $enfant \xrightarrow{is-a} humain$ (2) : $humain \xrightarrow{has-part} aile$ $\Rightarrow$ (3) : $enfant \xrightarrow{has-part} aile$

Evidemment, la relation $enfant \xrightarrow{has-part} aile$ est fausse et la relation $humain \xrightarrow{has-part} aile$ en est la cause.

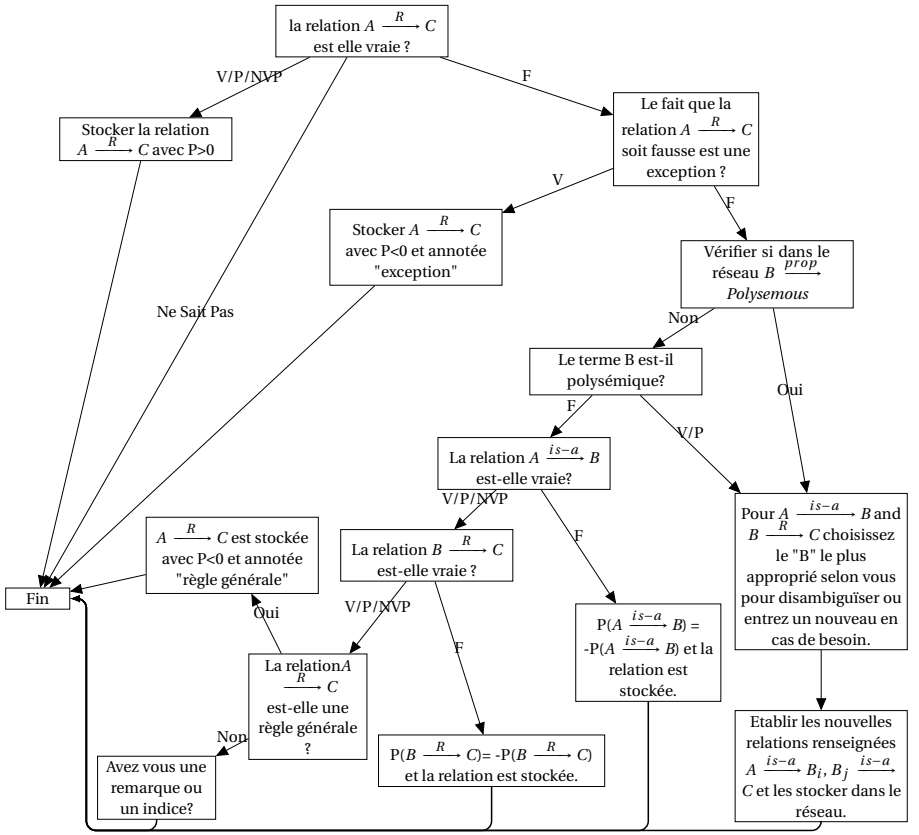

FIGURE 4 – Schéma de la procédure de validation/réconciliation. Légende : plutôt **V**rai, **P**ossible, **V**rai **N**on **P**ertinent, plutôt **F**aux.

Cas d'erreur dans les prémisses

Supposons que la relation (1) (figure 1) a un faible poids. Alors le réconciliateur demande au validateur si la relation (1) est vraie. Si la réponse est négative, l'opposé du poids actuel de la relation (1) lui est attribué (soit P = -1*(P)) et la réconciliation se termine. Si la réponse est positive, le réconciliateur demande si la relation (2) est vraie et il procède comme précédemment en cas de réponse négative. Sinon enfin, il vérifie les autres cas (exception, polysémie). (Figure 4)

Cas d'exception

Dans le cas de deux relations vraisemblables, le réconciliateur demande au validateur si la relation inférée $A \xrightarrow{R} C$ constituerait une exception. Si c'est le cas, la relation est stockée dans le réseau avec un poids négatif mais annotée avec une méta-information qui indique que c'est une exception. (Figure 4)

Cas de polysémie

Dans ce cas, le terme B est soit marqué dans le réseau comme polysémique ou indiqué comme tel par le validateur. Il s'agit alors de lister dans le dialogue les raffinements B_1, B_2, ...,B_n en les ordonnant selon une fonction de similarité et ainsi de permettre au validateur de choisir le plus approprié selon lui pour les deux relations $A \xrightarrow{is-a} B_i$ et $B_j \xrightarrow{R} C$. Il est possible à l'utilisateur de préciser un nouveau raffinement en cas d'insatisfaction vis-à-vis de ceux présentés. Après cette procédure, le réseau sera réconcilié par deux nouvelles relations $A \xrightarrow{is-a} B_i$ et $B_j \xrightarrow{R} C$ qui pourront être utilisées ultérieurement par le moteur d'inférences. (Figure 4)

4 Expérimentation

Nous avons mené une expérience consistant à produire en masse et en une seule fois le nombre maximum d'inférences possible par application du moteur sur le réseau lexical JDM. L'objectif est d'évaluer les productions du moteur ainsi que les mécanismes de blocage logique et de filtrage. A partir de cet ensemble d'inférences proposées, nous en avons sélectionné aléatoirement 400 pour chacun des types de relations et nous les avons soumises au processus de validition/réconciliation. Cette expérience a été menée dans un but d'évaluation car en pratique le moteur d'élicitation fonctionne conjointement avec les jeux et l'approche contributive de façon incrémentale. Dans cette expérience, une proposition a été validée soit manuellement par un contributeur de confiance (un validateur), soit par des joueurs/contributeurs si au moins 4 votes ont été effectués avec au minimum 75% de concordance. Les mêmes seuils ont été appliqués pour la réconciliation.

4.1 Productions du moteur d'inférences

Nous avons appliqué le moteur d'inférences sur environ 23000 termes aléatoirement choisis parmi ceux ayant au minimum un hyperonyme. Environ 1500000 inférences ont été produites et 77000 autres ont été bloquées. Le seuil de filtrage a été fixé à un poids de 25 (seules les inférences avec un poids égal ou supérieur à 25 ont été prises en considération). Dans la figure 5 la distribution a tendance à suivre une loi de puissance, ce qui n'est pas surprenant car la distribution dans le réseau lexical des relations est elle-même gouvernée par une loi de puissance.

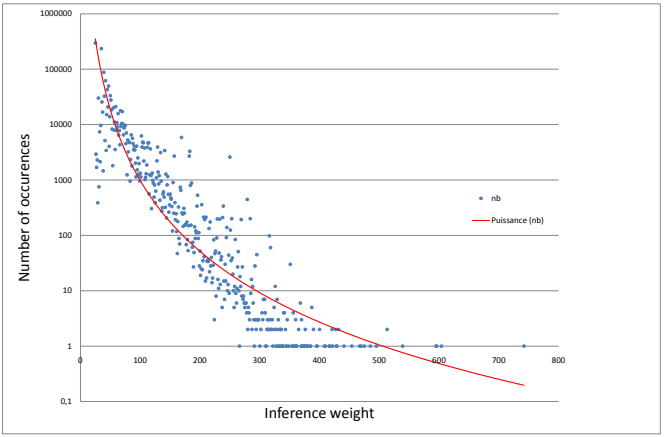

FIGURE 5 – Distribution des inférences proposées selon leur poids. La distribution semble une loi de puissance contravariante au poids. En pratique, les relations candidates de très fort poids sont validées les premières.

Les tables 1 et 2 présentent le nombre de relations proposées par le moteur d'inférences.
Les différents types de relations pour la seconde prémisse (la relation générique R dans le schéma d'inférence triangulaire) sont productifs à des degrés divers. Bien entendu, cette variation est en partie due au nombre de relations déjà existantes pour chaque type dans le réseau.
La productivité d'un type de relation est le ratio entre le nombre d'inférences proposées et le nombre d'occurences de ce type de relations dans le réseau.

Type de relation	**nb proposés**	**nb existants**	**productivité**
est-un	91 037	91 799	99,16%
parties	372 688	21 886	1702,86%
holonyme	108 191	13 124	824,37%
lieu	271 717	26 346	1031,34%
carac	203 095	24 180	839,92%
agent-1	198 359	6 820	2908,48%
instr-1	24 957	4 797	520,26%
patient-1	14 658	3 930	372,97%
lieu-1	145 159	8 835	1642,99%
lieu-action	50 035	4 559	1097,49%
matière	4 313	3 097	139,26%

TABLE 1 – Productivité des types de relations selon le moteur d'inférences.

L'inférence transitive pour *is-a* est la moins productive ce qui peut sembler surprenant en première analyse. En fait, la relation *is-a* est d'ors et déjà fortement renseignée dans le réseau lexical JDM, et ce faisant, relativement peu de nouvelles inférences peuvent être proposées. Les données sont quelques peu inversées pour les autres types de relation qui ne sont pas suffisamment représentés mais qui sont pourtant potentiellement valides.

Le rôle sémantique d'agent (la relation *agent-1*) est de loin le type le plus productif, avec 30 fois plus de propositions que ce qui existe actuellement dans le réseau JDM.

La productivité d'une relation est covariante avec deux facteurs : (1) la taille de la population pour ce type de relation dans le réseau et, (2) le nombre d'hyponymes dont disposent les termes renseignés pour cette relation.

En termes de filtrages, nous constatons une grande disparité entre les types de relations, qui provient essentiellement de la force de transitivité du type de relation, et pour les termes concernés du taux de polysémie et du poids des relations.

Par exemple, *is-a* est fortement transitive et se retrouve peu bloquée par rapport à *parties* ou *holo*. Les rôles sémantiques (*agent-1*, *instr-1*, etc.), peu productifs par ailleurs car toutes proportions gardées peu renseignées, semblent peu sensibles aux deux filtrages.

Type de relation	nb proposées	%	nb bloquées	%	nb filtrées	%
est-un (is-a/hperonyme de x)	91 037	6,13	4 034	5,23	53 586	26,32
parties (constitutives de x)	372 688	25,11	31 421	40,76	100 297	49,26
holo (tout de x)	108 191	7,28	17 944	23,27	26 818	13,17
lieu (typique pour x)	271 717	18,30	11 502	14,92	14 174	6,96
carac(téristiques de x)	203 095	13,68	2 647	3,43	6 576	3,23
agent-1 (que peut faire x ?)	198 359	13,36	9052	11,74	1122	0,55
instr-1 (que peut-on faire avec x ?)	24 957	1,68	127	0,16	391	0,19
patient-1 (que peut-on faire à x ?)	14 658	0,98	7	0,01	13	0,00
lieu-1 (que peut-on trouver sur/dans x ?)	145 159	9,78	129	0,17	206	0,10
lieu-action (que peut-on faire sur/dans x ?)	50 035	3,379	91	0,12	132	0,06
matière (de quelle matière/substance est fait x ?)	4 313	0,29	135	0,17	262	0,12
Total	**1 484 209**	100	**77 089**	100	**203 577**	100

TABLE 2 – Statut des inférences proposées par type de relation. *Bloqué* fait référence au filtrage logique et *filtré* au filtrage statistique.

4.2 Quelques données sur la réconciliation

Le dialogue avec les joueurs permet de déterminer le type d'erreur (dans les prémisses, exceptions ou à cause de la polysémie). La table 3 présente une évaluation du statut des inférences proposées par le moteur d'inférences. Les inférences sont valides pour environ 80-90% d'entres elles avec aux alentours de 10% d'inférences valides mais non pertinentes (comme par exemple, $chien \xrightarrow{has-parts} proton$). Nous observons que les erreurs dans les prémisses sont relativement peu nombreuses, et quoiqu'il en soit ces erreurs peuvent être aisément corrigées. Bien sûr, tous les types d'erreur ne sont pas détectables par ce processus. De façon plus intéressante, la réconciliation permet dans 5% des cas d'identifier les termes polysémiques et de sélectionner ou proposer des raffinements. Globalement, les inférences fausses négatives (celles votées fausses mais valides) et les inférences fausses positives (celles votées vraies mais invalides) sont évaluées à moins de $0,5\%$ du total. Mener le dialogue à son terme n'étant pas une obligation, les utilisateurs ne sont pas encluns à donner des réponses au hasard en cas de difficulté. Nous avons également mené une expérience *in vivo* et donc moins artificielle où les inférences sont produites à la volée sur les termes joués ou contribués par les joueurs, afin de leur présenter des relations pour lesquelles ils peuvent se prononcer. L'expérience a démarré en décembre 2012 et se poursuit encore au moment de l'écriture de cet article (soit une durée de 2 mois). Environ 10000 propositions ont été validées et 250 invalidées et réconcilées par les contributeurs. Les propositions ayant fait l'objet de votes contradictoires de la part des contributeurs ne portent que sur 8 termes (la décision finale à leur rejet a été donnée par l'expert). Le nombre de votes total correspondant aux relations inférées a été supérieur à 32000. Ces données semblent confirmer la viabilité de la démarche consistant à solliciter les contributeurs pour valider ou réconcilier les propositions faites automatiquemment.

	% valides		% d'erreur		
Types de relation	**pertinent**	**non pertinent**	**prémisses**	**exception**	**polysémie**
est-un	76%	13%	2%	0%	9%
parties	65%	8%	4%	13%	10%
holonyme	57%	16%	2%	20%	5%
lieu	78%	12%	1%	4%	5%
carac	82%	4%	2%	8%	4%
agent-1	81%	11%	1%	4%	3%
instr-1	62%	21%	1%	10%	6%
patient-1	47%	32%	3%	7%	11%
lieu-1	72%	12%	2%	10%	6%
lieu-action	67%	25%	1%	4%	3%
matière	60%	3%	7%	18%	12%

TABLE 3 – Résultats de la validation/réconciliation selon le type de relation inférée.

5 Conclusion

Dans cet article, nous avons présenté quelques enjeux concernant la construction des réseaux lexico-sémantiques à l'aide de jeux et de contributions. Un tel réseau est fortement lexicalisé et les usages des termes sont découverts incrémentalement au fur et à mesure de sa construction. Des erreurs sont évidemment présentes dans ce type de ressources puisqu'elles peuvent provenir des parties jouées sur des relations difficiles, mais elles sont généralement découvertes par les contributeurs, seulement cependant pour les termes qui les intéressent. Pour être capable d'augmenter la qualité et la couverture du réseau lexical, nous avons proposé un système d'élicitation basé sur des inférences de relations et des réconciliations en cas d'invalidation. Les inférences ici sont construites sur la base d'une triangulation simple basée sur la transitivité de l'hyperonymie associée à des mécanismes de blocage logique et de filtrage statistique. La réconciliation est appliquée dans le cas où la relation inférée est prouvée fausse et ce dans le but d'en identifier la cause. Globalement, nous pouvons conclure que les relations inférées sont correctes et pertinentes dans 78% des cas et correctes mais non pertinentes dans 10% des cas. En général, les inférences fausses suite à une faute dans les prémisses représentent 2% des cas, les exceptions autour de 5% des cas et les confusions à cause d'une polysémie environ 5%. La philosophie de notre approche est de ne jamais valider automatiquement des propositions même si bon nombre d'entre-elles semblent certaines, car les exceptions ne sont pas prédictibles. Toutefois, la validation en corpus pour une inclusion dans le réseau lexical semble une approche possible à l'automatisation. De même, la validation temporaire d'inférences à niveau de confiance élevé serait tout à fait pertinente lors d'une analyse automatique de textes afin d'aider à la mener à terme. En plus d'être un outil d'augmentation du nombre de relations dans un réseau lexical, le système d'élicitation est un détecteur efficace d'erreurs et de polysémie. Les mesures prises durant la phase de réconciliation empêchent une inférence fausse d'être réinférée en boucle. Une telle approche doit être développée plus avant avec d'autres types de schémas d'inférences et possiblement avec une évaluation de la distribution des classes sémantiques des termes sur lesquels les inférences sont élaborées. En effet, certaines classes comme les objets concrets ou les êtres animés peuvent être considérablement plus productives sur certains types de relations que, par exemple, les noms d'objets abstraits ou les termes d'événements/processus. Quoi qu'il en soit de telles variations dans la productivité d'inférences méritent certainement d'être explorées.

Références

BLANCO, E. et MOLDOVAN, D. (2011). A model for composing semantic relations. *Ninth International Conference on Computational Semantics (IWCS'11), Oxford, United Kingdom*, pages 45–54.

CHAMBERLAIN, J., POESIO, M. et KRUSCHWITZ, U. (2008). Phrase detectives : A web-based collaborative annotation game. *In Proceedings of the International Conference on Semantic Systems (I-Semantics'08)*.

DONG, Z. et DONG, Q. (2006). *HowNet and the Computation of Meaning.* WorldScientific, London.

FELLBAUM, C. et MILLER, G. (1998). *(eds) WordNet.* The MIT Press.

FENG, D., BESANA, S. et ZAJAC, R. (2009). Acquiring high quality non-expert knowledge from on-demand workforce. *In Proceedings of the 2009 Workshop on The People's Web Meets NLP : Collaboratively Constructed Semantic Resources, People's Web '09, Morristown, NJ, USA. Association for Computational Linguistics.*, pages 51–56.

HARABAGIU, S. et MOLDOVAN, D. (1998). Knowledge processing on an extended wordnet. *WordNet : An Electronic Lexical Database, MIT Press.*, pages 381–405.

JOUBERT, A. et LAFOURCADE, M. (2008). Jeuxdemots : un prototype ludique pour l'émergence de relations entre termes. *In proc of JADT'2008, Ecole normale supérieure Lettres et sciences humaines , Lyon, France, 12-14 mars 2008*, page 8 p.

LAFOURCADE, M. (2007). Making people play for lexical acquisition. *In Proc. SNLP 2007, 7th Symposium on Natural Language Processing. Pattaya, Thailande, 13-15 December 2007*, page 8 p.

LAFOURCADE, M. et JOUBERT, A. (2012). Increasing long tail in weighted lexical networks. *In proc of Cognitive Aspects of the Lexicon (CogAlex-III), COLING, Mumbai, India, December 2012*, page 16 p.

LAFOURCADE, M., JOUBERT, A., SCHWAB, D. et ZOCK, M. (2011). Evaluation et consolidation d'un réseau lexical grâce à un assistant ludique pour le mot sur le bout de la langue. *In proc of TALN'11, Montpellier, France, 27 juin-1er juillet 2011*, pages 295–306.

LENAT, D. (1995). Cyc : A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38.

LIEBERMAN, H., SMITH, D. A. et TEETERS, A. (2007). Common consensus : a web-based game for collecting commonsense goals. *In Proc. of IUI, Hawaii.*, page 12 p.

MARCHETTI, A., TESCONI, M., RONZANO, F., MOSELLA, M. et MINUTOLI, S. (2007). Semkey : A semantic collaborative tagging system. *in Procs of WWW2007, Banff, Canada*, page 9 p.

MIHALCEA, R. et CHKLOVSKI, T. (2003). Open mindword expert : Creating large annotated data collections with web users help. *In Proceedings of the EACL 2003, Workshop on Linguistically Annotated Corpora (LINC 2003)*, page 10 p.

MILLER, G., BECKWITH, R., FELLBAUM, C., GROSS, D. et MILLER, K. (1990). Introduction to wordnet : an on-line lexical database. *International Journal of Lexicography*, 3(4):235–244.

NAVIGLI, R. et PONZETTO, S. (2012). Babelnet : Building a very large multilingual semantic network. *In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 11-16 July 2010*, pages 216–225.

SAGOT, B. et FIER, D. (2008). Construction d'un wordnet libre du français à partir de ressources multilingues. *TALN 2008, Avignon, France, 2008.*, page 12.

SIORPAES, K. et HEPP, M. (2008). Games with a purpose for the semantic web. *In IEEE Intelligent Systems*, 23(3):50–60.

THALER, S., SIORPAES, K., SIMPERL, E. et HOFER, C. (2011). A survey on games for knowledge acquisition. *STI Technical Report.*, page 19.

von AHN, L. et DABBISH, L. (2008). Designing games with a purpose. *Communications of the ACM*, 51(8):58–67.

VOSSEN, P. (1998). Eurowordnet : a multilingual database with lexical semantic networks. *Kluwer Academic Publishers, Norwell, MA, USA*, page 200.

ZESCH, T. et GUREVYCH, I. (2009). Wisdom of crowds versus wisdom of linguists measuring the semantic relatedness of words. *Natural Language Engineering, Cambridge University Press.*, pages 25–59.