

# Consolidation de réseaux lexico-sémantiques par des inférences déductives et inductives

Manel Zarrouk<sup>1</sup>, Mathieu Lafourcade<sup>1</sup>, Alain Joubert<sup>1</sup>

<sup>1</sup>Lirmm – prénom.nom@lirmm.fr

## Abstract

In Computational Linguistics, building and validating lexical-semantic networks as well as adding some reasoning skills are crucial issues. We devise here an inference engine which produces new "potential" relations from already existing ones in the JeuxDeMots network. This network is constructed with a GWAP (game with a purpose) thanks to thousands of players. It contains terms and over 2 million relation occurrences. Terms may be refined in several senses (bank may be a bank>financial institution or a bank>river) but as the network is indefinitely under construction (in the context of a Never Ending Learning approach) some senses may be missing at a given time. The approach we proposed here is founded through two kinds of inference schemes: deduction (topdown from generic to specific terms) and induction (bottom-up from specific to generic terms). A blocking mechanism, whose purpose is to avoid proposing highly dubious new relations, is based on logical and statistical constraints. Automatically inferred relations are then proposed to human contributors to be validated. In case of invalidation, a reconciliation dialog is undertaken to identify the cause of the wrong inference: an exception, an error in the premises or a previously undetected confusion due to polysemy on the central term common to both premises.

## Résumé

La construction d'un réseau lexico-sémantique, ainsi que sa validation, sont des tâches primordiales en Traitement Automatique des Langues. Il est également possible de densifier un tel réseau de manière endogène, grâce à des processus de raisonnement sur les données qu'il contient. Dans cet article, nous présentons un moteur d'inférences dont le but est de produire de nouvelles relations « potentielles » à partir de relations préexistantes dans le réseau JeuxDeMots. Ce réseau, en perpétuelle évolution, a été construit par plusieurs milliers de joueurs grâce à une série de jeux en ligne (*Games With A Purpose* ou GWAP) dont le principal est JeuxDeMots. Il relie plus de 250 000 termes par environ 3 millions d'occurrences de relations typées et pondérées. Les termes polysémiques peuvent être raffinés en plusieurs sens, mais comme le réseau est toujours en construction (dans le contexte d'un apprentissage sans fin), certains sens peuvent être absents à un instant donné. L'approche que nous proposons ici repose sur une méthode de triangulation à travers deux schémas généraux d'inférences : la déduction (dans une approche descendante, du générique vers le spécifique) et l'induction (dans une approche ascendante, du spécifique vers le générique). Un mécanisme de blocage, dont le but est d'éviter au système de proposer des relations qui pourraient être douteuses, repose sur un filtrage logique et statistique. Les relations ainsi automatiquement inférées sont proposées pour validation/invalidation à un contributeur humain. En cas d'invalidation par ce dernier, un dialogue de réconciliation est proposé pour déterminer la cause de cette inférence erronée : il s'agit soit d'une exception, soit d'une erreur dans l'une des prémisses, soit d'une confusion provenant d'un cas de polysémie du terme central commun aux deux prémisses non encore renseigné dans le réseau.

**Mots-clés :** inférence, induction, déduction, réseau lexico-sémantique, réconciliation, élicitation.

## 1. Introduction

La constitution de ressources est l'une des tâches primordiales en TALN. La plupart des réseaux lexico-sémantiques existants ont été construits manuellement ; c'est le cas par exemple du plus célèbre d'entre eux, le réseau *WordNet*. Des outils ont été conçus afin d'en vérifier la validité et la complétude, mais leur mise en œuvre nécessite généralement beaucoup de temps. Les approches automatiques sont le plus souvent limitées aux

cooccurrences de termes, et la détermination précise de relations sémantiques entre termes reste délicate. De nouvelles approches, fondées sur la peuplologie, apparaissent en TALN, avec en particulier le développement de *Amazon Mechanical Turk* (Adda et Mariani, 2010) ou dans un contexte plus étendu l'émergence de *Wikipédia* et *Wiktionary* pour ne citer que les plus connus. *WordNet* ((Miller et al., 1990) et (Fellbaum, 1988)) est un réseau lexical basé sur la notion de *synsets* qui peuvent être en première approximation assimilés à des concepts. *EuroWordnet* (Vossen, 1998), une version multilingue de *WordNet*, et *WOLF* (Sagot, 2008), une version française de *WordNet*, ont été construits automatiquement par croisement de *WordNet* avec d'autres ressources lexicales. (Navigli, 2012) a construit automatiquement *BabelNet*, un réseau lexical multilingue de grande taille, à partir de cooccurrences de termes dans *Wikipédia*.

Un réseau lexico-sémantique fortement lexicalisé est constitué de termes (des concepts, mais également des mots ou des expressions multi-termes), chacun possédant une signification. L'idée elle-même de sens d'un terme peut être sujette à débat dans le contexte des ressources pour l'analyse sémantique, c'est pourquoi nous lui préférons la notion d'usage d'un terme. Par usage, nous voulons exprimer un raffinement d'un terme, clairement identifié par un locuteur. Un terme polysémique est raffiné en plusieurs usages qui peuvent différer notablement des sens tels qu'on les trouverait dans une ressource dictionnaire. Un usage donné peut lui-même être raffiné en usages plus précis, et l'ensemble des usages d'un terme donné peut être représenté sous la forme d'un arbre de décision (Lafourcade et Joubert, 2010). Dans le principe d'une construction collaborative, une telle ressource doit être considérée comme en perpétuelle évolution, et une règle générale est de n'avoir aucune certitude quant à l'état de complétude d'une entrée.

Afin de construire de manière collaborative un réseau lexical (ou n'importe quelle ressource similaire), deux grandes stratégies peuvent être mises en œuvre. La première résulte en un système contributif tel que *Wikipédia* où les utilisateurs effectuent leurs contributions de manière directe, ajoutant de nouvelles entrées ou complétant des entrées existantes, selon leur propre décision. Dans la seconde stratégie, les contributions sont faites indirectement au travers de jeux, plus connus sous le nom de *Games With A Purpose* (ou GWAP (vonAhn, 2008) et (Law et al., 2009)). Dans ce dernier cas, les joueurs ne sont généralement pas conscients du fait qu'ils contribuent à la construction d'une ressource. Quelle que soit la stratégie mise en œuvre, la ressource construite n'est pas exempte d'erreurs qui doivent être corrigées au fur et à mesure de leurs découvertes. Un grand nombre de relations, pourtant pertinentes, ne sont pas présentes dans le réseau, bien que nécessaires pour une ressource de qualité suffisante pour être utilisée dans diverses applications en TALN, en particulier en analyse sémantique. Par exemple, les contributeurs n'indiquent que très rarement que tel type d'oiseau peut voler, car c'est une généralité manifeste. Seuls les faits notables qui ne sont pas aisément déductibles sont indiqués par les contributeurs. En particulier, les exceptions connues sont souvent indiquées et prennent la forme d'une relation avec un poids négatif (par exemple, la relation « voler a\_pour\_agent autruche » possède un poids égal à -100, exprimant ainsi clairement le fait qu'une autruche ne peut pas voler).

Dans le but de densifier le réseau JeuxDeMots (JDM), nous adoptons ici une stratégie qui repose sur un moteur d'inférences simple qui propose de nouvelles relations à partir de celles préexistantes. Cette approche est strictement endogène par le fait qu'elle ne repose sur aucune ressource externe. Les relations ainsi inférées sont soumises soit à des contributeurs qui donnent leurs avis par un mécanisme de votes, soit à des experts linguistes pour une validation / invalidation directe. Une grande proportion des relations inférées ont été jugées

correctes. Malgré tout, une quantité non négligeable de ces relations ont été jugées fausses et il est important d'en comprendre la (ou les) raison(s). Ce processus d'explication peut être considéré comme une réconciliation entre le moteur d'inférences et le validateur qui est guidé au travers d'un dialogue lui permettant d'exprimer pourquoi il a jugé fausse la relation. Une relation inférée peut être incorrecte pour l'une des trois raisons suivantes : l'une des prémisses peut être erronée, la relation inférée peut correspondre à une exception ou il peut s'agir d'une confusion due à la polysémie.

Dans cet article, tout d'abord, nous rappelons brièvement les principes sur lesquels repose la construction du réseau lexical JDM par peuplonomie grâce à des jeux en ligne. Ensuite, nous présentons un moteur d'élicitation basé sur un moteur d'inférences utilisant les principes de déduction et d'induction, ainsi qu'un moteur de réconciliation. Nous terminons par une expérimentation permettant d'évaluer les performances du système.

## 2. Réseaux lexicaux et peuplonomie

Il existe de multiples méthodes pour construire un réseau lexical, elles ne présentent pas toutes les mêmes caractéristiques de qualité des données, coût et délai de réalisation. Parallèlement aux stratégies automatiques ou manuelles, les approches contributives sont de plus en plus populaires, par le fait qu'elles sont à la fois peu coûteuses à mettre en œuvre et qu'elles conduisent à des résultats de qualité. Plus spécifiquement, on constate une tendance croissante à l'utilisation de jeux en lignes ((Thaler et al., 2011) et (Crawford et al., 2007)).

Le réseau lexical JDM, en évolution permanente, a été construit grâce à un ensemble de jeux en ligne collaboratifs. Les informations dans le réseau JDM sont collectées par agréments non négociés entre les joueurs; il est à noter que les systèmes contributifs les plus courants reposent sur des agréments négociés.

### 2.1. *JeuxDeMots: un GWAP pour construire un réseau lexico-sémantique*

JeuxDeMots<sup>1</sup> est un jeu en ligne à deux joueurs dont le but est la construction d'un réseau lexico-sémantique de grande taille (Lafourcade, 2007). Le réseau est constitué de termes (les nœuds) reliés entre eux par des relations typées et pondérées (les arcs entre ces nœuds). Il contient des termes, avec des raffinements possibles, analogues aux *synsets* que l'on rencontre dans *WordNet* (Miller et al., 1990).

Les relations typées et pondérées reliant les termes sont validées par paires de joueurs. Ces relations sont étiquetées par les types de relation proposés aux joueurs (synonymie, hyperonymie, idée associée...) et elles sont pondérées en fonction du nombre de paires de joueurs qui les ont proposées. D'autres systèmes basés sur le web existent, tels *Open Mind Word Expert* (Mihalcea et al., 2003) dont le but est de créer un corpus tagué sémantiquement avec l'aide de contributeurs internautes, ou *SemKey* (Marchetti et al., 2007) qui utilise les ressources de *WordNet* et *Wikipédia* pour désambiguïser les formes lexicales.

---

<sup>1</sup> JeuxDeMots est accessible à l'adresse <http://jeuxdemots.org>

## 2.2. *Diko: un outil contributif*

Diko<sup>2</sup> est initialement un outil permettant d'afficher les informations contenues dans le réseau JDM. Il peut également être utilisé en tant qu'outil collaboratif. La nécessité de ne pas reposer uniquement sur le jeu JDM pour construire le réseau provient du fait qu'un nombre non négligeable de types de relation sont soit délicats à proposer aux joueurs (car relativement complexes à appréhender), soit très peu productifs (car peu de termes peuvent leur être associés).

Le principe du processus de contribution est qu'une proposition faite par un joueur va entraîner un processus de vote par d'autres joueurs, avant d'être validée ou invalidée par un expert lexicographe. Ce que nous proposons dans cet article correspond à ce type de scénario par contributions/validations.

## 3. Elicitation par inférence et réconciliation

Nous avons développé un système afin d'accroître le nombre de relations dans le réseau lexical JDM. Ce système possède deux principaux composants : (a) un moteur d'inférences et (b) un système réconciliateur. Le moteur d'inférences propose des relations, de la même manière que ne le fait un contributeur: ces relations sont présentées pour validation ou invalidation à des contributeurs humains ou à un expert valideur. En cas d'invalidation d'une relation inférée, le système réconciliateur est invoqué afin d'essayer de déterminer la raison du rejet de cette relation. L'élicitation doit ici être comprise comme étant le processus permettant de transformer une connaissance implicite d'un utilisateur en relation explicite dans le réseau lexical.

### 3.1. Réalisation des inférences

Les idées principales aboutissant à l'élaboration des inférences sont les suivantes :

- en ce qui concerne le moteur, inférer consiste à émettre de nouvelles prémisses (des relations entre termes) à partir de prémisses préexistantes ;
- les inférences potentielles peuvent être bloquées logiquement sur la base de la présence ou absence de certaines autres relations ;
- les inférences potentielles peuvent être filtrées si elles sont jugées trop faibles, c'est-à-dire d'un poids trop peu élevé.

#### 3.1.1. Schéma déductif

Dans cet article, le premier type d'inférence que nous allons présenter est la déduction. Il s'agit d'un schéma descendant, de type top-down, reposant sur la transitivité de la relation ontologique *is-a* (ou hyperonymie). Si un premier terme A est une sorte de B et que B possède une relation R avec C, alors nous pouvons supposer que A puisse avoir la même relation R avec C. Plus formellement, ce schéma peut être écrit :

$$\exists A \xrightarrow{is-a} B \quad \wedge \quad \exists B \xrightarrow{R} C \quad \Rightarrow \quad A \xrightarrow{R} C$$

---

<sup>2</sup> Diko est accessible à l'adresse <http://www.jeuxdemots.org/diko.php>

Processus global - Considérons un terme T possédant un ensemble pondéré d'hyperonymes. A partir de chacun de ces hyperonymes, le moteur d'inférences déduit un ensemble d'inférences (des relations potentielles). Ces ensembles d'inférences ne sont généralement pas disjoints, et le poids d'une inférence proposée dans plusieurs ensembles est la moyenne géométrique de chaque occurrence.

Filtrage logique - Bien évidemment, le schéma ci-dessus est quelque peu naïf, compte tenu de la ressource sur laquelle il est mis en œuvre. En effet, il est tout à fait possible que le terme central B soit polysémique, et nous devons trouver un moyen de bloquer des inférences qui sont fort probablement erronées. Supposons que le terme B possède deux sens distincts qui interviennent respectivement dans la première et dans la deuxième relation, tel que représenté en figure 1, alors l'inférence est très probablement erronée.

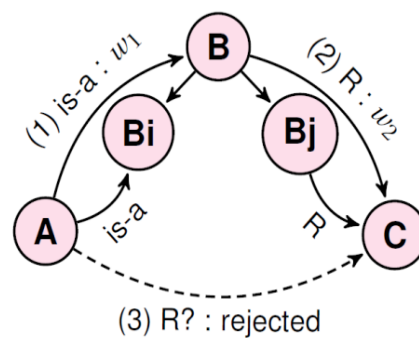


Figure 1. Schéma d'inférence triangulaire avec blocage logique basé sur la polysémie de B.

Dans ce cas, une relation R à inférer doit remplir les contraintes énoncées ci-dessous :

$$\begin{aligned} & A \xrightarrow{is-a} B \quad \wedge \quad B \xrightarrow{R} C \\ \wedge \quad & (\exists B_i \xrightarrow{meaning-of} B \quad \wedge \quad \exists B_j \xrightarrow{meaning-of} B) \\ \wedge \quad & (\nexists A \xrightarrow{is-a} B_i \quad \vee \quad \nexists B_j \xrightarrow{R} C) \\ \Rightarrow & A \xrightarrow{R} C \end{aligned}$$

De plus, si l'une des prémisses est étiquetée vraie mais non pertinente, alors l'inférence est bloquée.

Filtrage statistique - Il est possible d'évaluer, sur une échelle ouverte, un niveau de confiance pour chaque inférence produite ; de cette manière, les inférences douteuses peuvent être filtrées. Le poids w (pour *weight*) d'une relation inférée est la moyenne géométrique des poids des prémisses (les relations (1) et (2) de la figure 1). Si la seconde prémisses a un poids négatif, le poids de la relation potentiellement inférée ne peut pas être calculé : cette inférence est alors rejetée. Comme la moyenne arithmétique est moins sensible aux faibles valeurs que la moyenne géométrique, les inférences qui ne sont pas obtenues à partir de deux relations estimées vraies sont rejetées.

$$\begin{aligned} w(A \xrightarrow{R} C) &= ( w(A \xrightarrow{is-a} B) * w(B \xrightarrow{R} C) )^{1/2} \\ \Rightarrow \quad w3 &= (w1 * w2)^{1/2} \end{aligned}$$

### 3.1.1. Schéma inductif

Tout comme pour le processus d'inférence déductive, celui de l'inférence inductive exploite la transitivité de la relation is-a. Si un terme A est une sorte de B et que A soit source d'une relation R vers C, alors il est probable que B soit source d'une relation de même type vers C. De manière plus formelle, nous pouvons écrire :

$$\exists A \xrightarrow{is-a} B \quad \wedge \quad \exists A \xrightarrow{R} C \quad \Rightarrow \quad B \xrightarrow{R} C$$

Ce schéma correspond à une inférence de généralisation. Le processus global est similaire à celui appliqué pour la déduction et, de manière analogue, nous pouvons appliquer des filtrages logiques et statistiques. Il est possible que le terme commun aux deux prémisses (appelé le terme central, A dans le cas de l'induction) soit polysémique. Si ce terme A possède deux sens distincts qui interviennent respectivement dans la première et dans la deuxième relation, tel que représenté en figure 2, alors l'inférence est très probablement erronée.

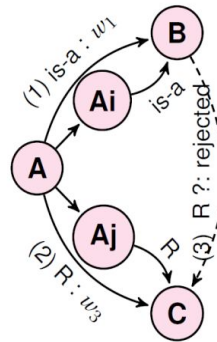


Figure 2. Les relations (1) et (2) sont les prémisses, la relation (3), obtenue par induction, sera proposée pour validation. Le terme central A peut être polysémique, avec possibilité d'induire une relation potentiellement erronée.

Le filtrage logique peut être formalisé comme suit :

$$\begin{aligned} & A \xrightarrow{is-a} B \quad \wedge \quad A \xrightarrow{R} C \\ \wedge & \quad (\exists A_i \xrightarrow{\text{meaning-of}} A \quad \wedge \quad \exists A_j \xrightarrow{\text{meaning-of}} A) \\ \wedge & \quad (\exists A_i \xrightarrow{is-a} B \quad \vee \quad \exists A_j \xrightarrow{R} C) \\ \Rightarrow & \quad B \xrightarrow{R} C \end{aligned}$$

Le filtrage statistique : comme dans le cas du schéma déductif, le niveau de confiance de la relation induite peut être évalué. Conformément à l'estimation du poids calculé pour le schéma déductif, le poids de la relation induite est ici :

$$\begin{aligned} w(B \xrightarrow{R} C) &= (w(A \xrightarrow{R} C))^2 / w(A \xrightarrow{is-a} B) \\ \Rightarrow w_2 &= \frac{(w_3)^2}{w_1} \end{aligned}$$

### 3.2. Réconciliation

Les relations inférées, que ce soit par déduction ou induction, sont ultérieurement présentées au validateur qui décide de leur statut : plutôt vraie, plutôt vraie mais non pertinente, possible ou majoritairement fausse. En cas d'invalidation, un processus de réconciliation est exécuté dans le but de tenter d'en diagnostiquer la raison : une erreur dans l'une des prémisses (les relations préexistantes peuvent être erronées), un cas d'exception ou une confusion due à la polysémie (l'inférence a été réalisée sur un terme central polysémique). Un dialogue avec l'utilisateur est alors amorcé. Ce dernier est libre de choisir d'y répondre entièrement, partiellement ou pas du tout. Afin de déterminer dans quel ordre procéder, le système réconciliateur détermine si les poids des prémisses sont plutôt élevés ou faibles. Le degré de confiance d'une relation est évalué en comparant le poids de cette relation à un seuil de confiance qui correspond au niveau du début de la longue traîne dans la distribution de cette relation. Pour l'ensemble des relations sortantes d'un terme, la longue traîne débute au point où la somme des poids des relations de la traîne est égale à la somme des poids des relations qui n'appartiennent pas à cette traîne (Lafourcade et Joubert, 2012).

Si  $w(A \text{ is-a } B) \geq (\text{conf} - \text{seuil}(A))$  alors la relation est digne de confiance ;

Si  $w(A \text{ is-a } B) < (\text{conf} - \text{seuil}(A))$  alors la relation est douteuse.

Dans le cas où les deux relations (1) et (2) sont dignes de confiance, le système réconciliateur tente, par un dialogue avec le validateur, de vérifier si la relation inférée est une exception. Si ce n'est pas le cas, il vérifie si le terme central est polysémique. Si ce n'est pas le cas, il vérifie finalement s'il s'agit d'un cas d'erreur. Nous vérifions le cas d'erreur en dernier car le niveau de confiance des relations (1) et (2) était élevé.

Dans le cas où l'une des relations (1) et (2) est jugée douteuse, le système réconciliateur suspecte qu'il s'agit d'un cas d'erreur. Ainsi, il est d'abord demandé au validateur de confirmer ou infirmer cette hypothèse. Si le validateur réfute l'une des prémisses, il s'agit effectivement d'un cas d'erreur. Dans le cas contraire, le dialogue se poursuit en demandant au validateur s'il s'agit d'une exception, puis éventuellement d'une polysémie.

#### 3.2.1. Erreurs dans les prémisses

Supposons par exemple que le poids de la relation (1) soit relativement faible. Le système réconciliateur pose au validateur la question de la validité de cette relation (1).

- Si elle est fausse, un poids négatif est attribué à (1) et la réconciliation est terminée. Cette relation ne sera plus utilisée par la suite en tant que prémisses dans le calcul de nouvelles inférences ;
- Si elle est vraie, il est demandé au validateur si la relation (2) est également vraie et il est procédé comme ci-dessus si la réponse est négative ;
- Si les prémisses sont vraies, le dialogue se poursuit pour vérifier les autres cas possibles (exception ou polysémie).

#### 3.2.2. Erreurs par exception

Dans le cas d'une inférence par déduction, si le validateur indique que la relation inférée est une exception relativement au terme B, cette relation est mémorisée dans le réseau lexical avec un poids négatif et avec une méta-information indiquant qu'il s'agit d'une exception.

Dans le cas d'une inférence par induction, si le validateur indique que la relation A-R-C (qui est utilisée en tant que prémisse) est une exception relativement au terme B, en plus de mémoriser la relation inférée fausse (B-R-C) avec un poids négatif dans le réseau, la relation (A-R-C) est étiquetée avec une méta-information indiquant qu'il s'agit d'une exception. Dans le cas d'une inférence par induction, une exception est une prémisse vraie qui conduit à une relation induite fausse.

Dans les deux cas, déduction et induction, l'étiquette exception concerne toujours la relation (A-R-C) : c'est le spécifique qui est une exception par rapport au général, général qui est pointé par l'étiquette. Par exemple, nous avons: ((autruche \*agent-1 voler) exception oiseau), la notation \*agent-1 indique que la relation agent-1 est fausse, et c'est bien la relation fausse qui est annoté avec une relation exception. Dès qu'une relation est étiquetée en tant qu'exception, elle ne participera plus comme prémisse pour l'inférence de généralisation (modèle ascendant ou *bottom-up*), par contre elle peut toujours être utilisée pour l'inférence de spécialisation (modèle descendant ou *top-down*).

### 3.2.3. Erreurs dues à la polysémie

Si le terme central (B dans le cas de la déduction et A pour l'induction) présentant une polysémie est mentionné comme tel dans le réseau, ses différents raffinements terme<sub>1</sub>, terme<sub>2</sub>, terme<sub>n</sub>... sont soumis au validateur qui peut alors choisir celui qui lui semble approprié. Si aucun des raffinements proposés ne satisfait le validateur, il peut lui-même en proposer un nouveau (induisant ainsi la création d'un nouveau raffinement). A l'issue de cette procédure, deux nouvelles relations (A is-a B<sub>i</sub> et B<sub>j</sub> R C dans le cas de la déduction, ou A<sub>i</sub> is-a B et A<sub>j</sub> R C dans le cas de l'induction) seront incluses dans le réseau avec un poids positif et le moteur d'inférences pourra les utiliser par la suite en tant que prémisses.

## 4. Expérimentation

Nous avons réalisé une expérimentation avec une passe unique du moteur sur le réseau lexical JDM. Le but de cette expérimentation est de mesurer la production du moteur d'inférences comportant le blocage et le filtrage décrits précédemment. A partir de l'ensemble des relations inférées (à la fois par déduction et induction), nous avons sélectionné aléatoirement un ensemble de 400 propositions pour chaque type de relation et nous avons exécuté le processus de validation/réconciliation. Cette expérimentation a été conduite uniquement dans un but d'évaluation ; actuellement, le système fonctionne itérativement grâce aux contributeurs.

### 4.1. Résultats du moteur d'inférences

Nous avons appliqué le moteur d'inférences sur approximativement 23 000 termes, sélectionnés au hasard, qui possèdent au moins un hyperonyme ou un hyponyme ; nous avons ainsi produit par déduction 1 484 209 inférences (77 089 inférences supplémentaires ont été bloquées). Le niveau de seuil pour le filtrage correspondait à un poids de 25. Cette valeur est pertinente, en effet, quand une relation proposée par un contributeur humain est validée par un expert, son poids est arbitrairement fixé à 25. Par le mécanisme d'induction, le moteur d'inférences a produit 353 371 relations potentielles. La table 1 montre le nombre de relations proposées par le moteur d'inférences en utilisant le schéma déductif. Les différents types de relation pour la seconde prémisse conduisent à des résultats variés. Bien évidemment, ceci est en grande partie du au nombre de relations existantes et à la distribution de leurs types dans le réseau.



Relation type	Proposed	Blocked	Filtered
is-a	91k (6,1)	4 k (5.2)	53 k (26,3)
has-parts	372k (25.1)	31 k (40.7)	100 k (49.3)
holonym	108k (7.2)	17 k (23.3)	26 k (13.2)
place	271k (18.3)	11 k (15)	14 k (7)
charac	203k (13.7)	2 k (3.4)	6 k (3.2)
agent-1	198k (13.3)	9 k (11.7)	1122 (0.5)
instr-1	24k (1.7)	127 (0.2)	391 (0.2)
patient-1	14k (1)	7 (0.01)	13 (0)
place-1	145k (9.8)	129 (0.2)	206 (0.1)
place >action	50k (3.4)	91 (0.1)	132 (0.06)
obj >mater	4k (0.3)	135 (0.2)	262 (0.1)
<b>Total</b>	<b>1 484k</b>	<b>77 k</b>	<b>203 k</b>

Table 1. Nombres et pourcentages des inférences (proposées, bloquées, filtrées) obtenues par déduction.

La relation transitive is-a est la moins productive, ce qui pourrait sembler étonnant au premier abord. En réalité, ce type de relation est déjà bien renseigné dans le réseau, et en conséquence relativement peu de nouvelles relations peuvent être inférées. Les résultats sont opposés pour certains autres types de relation qui ne sont que peu renseignés dans le réseau, mais sont cependant potentiellement valides. Le rôle sémantique agent (type de relation agent-1) est de loin le plus productif, avec 30 fois plus de propositions que le nombre de relations de ce type actuellement existantes dans le réseau.

#### 4.2. Résultats de la réconciliation

La table 2 présente une évaluation sur le statut des inférences proposées par le moteur d'inférence utilisant le schéma déductif. Globalement, les inférences sont valides dans 80 à 90% des cas, avec approximativement 10% d'inférences valides mais non pertinentes (comme par exemple chien has-parts proton). Nous pouvons remarquer que le nombre d'erreurs dans les prémisses est particulièrement faible, ces erreurs pouvant être aisément corrigées. Bien évidemment, toutes les erreurs possibles ne peuvent pas être détectées par ce processus. La réconciliation permet dans environ 5% des cas d'identifier des termes polysémiques. Globalement, les faux négatifs (inférences votées fausses, mais vraies en réalité) et les faux positifs (inférences votées vraies, mais fausses en réalité) sont évaluées à moins de 0,5%. En ce qui concerne le processus inductif (table 3), la relation is-a n'est pas pertinente, un réseau lexical ne peut pas être réduit à une ontologie et des héritages multiples sont possibles. Les résultats semblent meilleurs d'environ 5% par rapport à ceux obtenus en utilisant le processus déductif : globalement, les inférences sont valides dans 85 à 95% des cas. Le nombre d'erreurs est très faible. La différence principale avec la déduction réside dans le taux d'erreurs dues à la polysémie qui est bien inférieur avec le processus inductif.

Deduction	% valid		% error		
	rlvt	$\neg$ rlvnt	prem	excep	pol
is-a	76%	13%	2%	0%	9%
has-parts	65%	8%	4%	13%	10%
holonym	57%	16%	2%	20%	5%
typical place	78%	12%	1%	4%	5%
charac	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%
typical place-1	72%	12%	2%	10%	6%
place >action	67%	25%	1%	4%	3%
object >mater	60%	3%	7%	18%	12%

Table 2. Résultats pour la validation/réconciliation en fonction des types de relation pour le schéma déductif. Les relations valides peuvent être pertinentes (rlvt) ou non (rlvt), et les erreurs peuvent provenir des prémisses, des exceptions ou de la polysémie.

Induction	% valid		% error		
	rlvt	$\neg$ rlvnt	prem	excep	pol
has-parts	78%	10%	3%	2%	7%
holonym	68%	17%	2%	8%	5%
typical loc	81%	13%	1%	2%	3%
carac	87%	6%	2%	2%	3%
agent-1	84%	12%	1%	2%	1%
instr-1	68%	24%	1%	4%	3%
patient-1	57%	36%	3%	2%	2%
typical loc-1	75%	16%	2%	5%	2%
lieu-action	67%	28%	1%	3%	1%
object mater	75%	10%	7%	5%	3%

Table 3. Résultats pour la validation/réconciliation en fonction des types de relation pour le schéma inductif. La relation is-a est inappropriée pour l'induction.

## 5. Conclusion

Nous avons présenté les principes, ainsi que quelques résultats concernant l'inférence de nouvelles relations à partir de relations préexistantes dans un réseau lexico-sémantique dans lequel les sens d'usage sont découverts incrémentalement au fur et à mesure de sa construction. Des erreurs sont naturellement présentes, dans la mesure où elles peuvent résulter de parties jouées sur des types de relation difficiles, mais elles sont habituellement détectées et corrigées par les contributeurs jouant sur les termes qui les intéressent. Dans le but d'améliorer la qualité du réseau, nous proposons un moteur d'élicitation basé sur un système d'inférences et réconciliations. Les inférences sont proposées en utilisant deux

schémas différents (déduction et induction), avec un blocage logique et un filtrage statistique pour éliminer des inférences a priori erronées. Si une relation inférée est estimée fautive par un validateur, un dialogue de réconciliation lui est proposé dans le but d'en identifier la raison. De manière globale, nous pouvons conclure que les relations inférées par déduction sont pertinentes dans environ 78% des cas et correctes mais non pertinentes dans 10% des cas. La totalité des inférences déductives erronées représente 12% des cas, dont 2% avec au moins une erreur dans les prémisses, 5% de cas d'exception et 5% d'erreurs de confusion due à la polysémie. Le processus inductif est naturellement moins productif, mais globalement plus sûr. En plus d'être un outil permettant d'augmenter le nombre de relations dans le réseau lexical, le moteur d'élicitation est à la fois un détecteur d'erreurs et un identificateur de polysémie. Les inférences jugées fautes ou de caractère d'exception pendant la phase de réconciliation ne pourra pas servir de prémisses par la suite. Une telle approche devrait être appliquée à d'autres types de schémas d'inférence, tels que l'abduction, avec une étude de l'évaluation de la distribution des classes sémantiques sur lesquelles les inférences sont réalisées. En effet, certaines classes, comme par exemple les objets concrets ou les êtres vivants, devraient être plus productives pour certains types de relation que les noms abstraits ou les événements. Quoi qu'il en soit, de telles différences dans la production d'inférences entre les classes sont dignes d'intérêt et devraient être étudiées plus en détail.

## Références

- Adda G. et Mariani J. (2010). Language resources and amazon mechanical turk: legal, ethical and other issues. In LISLR2010, "Legal Issues for Sharing Language Resources workshop", LREC2010, Malta, 17 May.
- von Ahn L. et Dabbish L. (2008). Designing games with a purpose. *Communications of the ACM*, number 8, volume 51. pp. 58-67.
- Chamberlain J., Fort K., Kruschwitz U., Lafourcade M. et Poesio M. (2013). Using Games to Create Language Resources: Successes and Limitations of the Approach. Gurevych, Iryna; Kim, Jungi (Eds.), Springer, ISBN 978-3-642-35084-9, 2013, 42 p.
- Fellbaum C. ed (1988). *WordNet : An Electronic Lexical Database*. TheMIT Press.
- Law E., Luis von Ahn L. et Mitchell T. (2009). Search war: a game for improving web search. *KDD Workshop on Human Computation 2009*. 31p.
- Law E., Luis von Ahn L., Dannenberg R. B. et Crawford M. (2007). TAGATune: A Game for Music and Sound Annotation. *ISMIR 2007*. pp. 361-364.
- Lafourcade M. et Joubert A. (2012). Long Tail in Weighted Lexical Networks. In *proc of Cognitive Aspects of the Lexicon (CogAlex-III)*, COLING, Mumbai, India, December 2012. 16 p.
- Lafourcade M. et Joubert A. (2010). Construction de l'arbre des usages nommés d'un terme dans un réseau lexical évolutif. In *proc of JADT'2010*, Sapienza, University of Rome, Italie, 9-11 juin 2010.
- Mihalcea R. et Chklovski T. (2003). Open Mind-Word Expert: Creating large annotated data collections with web users help. In *Proceedings of the EACL 2003, Workshop on Linguistically Annotated Corpora (LINC)*. 10 p.
- Miller G.A., Beckwith R., Fellbaum C., Gross D. et Miller K.J. (1990). Introduction to WordNet: an online lexical database. *International Journal of Lexicography*. Volume 3, pp. 235-244.
- Miller G.A. (1995). *WordNet : A Lexical Database for English*. *Communications of the ACM* Vol. 38, No. 11, pp. 39-41.

- Navigli R. et Ponzetto S. (2010). 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. pp: 216-225.
- Navigli R. et Ponzetto S. (2012). BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network, Artificial Intelligence. 193: pp. 217-150.
- Sagot B. et Fier D. (2010). Construction d'un wordnet libre du français à partir de ressources multilingues. In Proceedings of TALN 2008, Avignon, France, 2008.12 p.
- Thaler S., Siorpaes K., Simperl E. et Hofer C. (2011). A Survey on Games for Knowledge Acquisition. STI Technical Report, May 2011.19 p.
- Vossen P. (1998). EuroWordNet: a multilingual database with lexical semantic networks.. Kluwer Academic Publishers. Norwell, MA, USA. 200 p.