Construction de l'arbre des usages nommés d'un terme dans un réseau lexical évolutif

Mathieu Lafourcade, Alain Joubert

LIRMM – Univ. Montpellier 2 - CNRS

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
161, rue Ada – 34392 Montpellier Cédex 5 – France

Résumé

Grâce à la participation d'un grand nombre de personnes via des jeux accessibles sur le web, nous avons construit un réseau lexical évolutif de grande taille. A partir de cette ressource, nous avons abordé la question de la détermination des sens d'usage d'un terme, puis nous avons introduit la notion de similarité entre ces différents usages. Nous avons pu ainsi obtenir pour un terme son arbre des usages : la racine regroupe tous les usages du terme et une descente dans l'arbre correspond à un raffinement de ces usages. Le nommage des différents nœuds de l'arbre des usages d'un terme est effectué lors d'une descente en largeur, depuis la racine jusqu'aux feuilles : la racine est étiquetée par le terme lui-même et chaque nœud de l'arbre est étiqueté par un terme issu de la clique ou quasi-clique que ce nœud représente. Nous montrons sur un exemple que tous les nœuds de l'arbre, souvent des feuilles, ne sont pas nécessairement étiquetables, du moins sans ambiguïté. Ces nœuds représentent généralement des usages très proches qui pourraient éventuellement à terme être regroupés en raison du caractère évolutif de notre réseau lexical.

Abstract

Thanks to the participation of a large number of persons via web-based games, we built a large-sized evolutionary lexical network. From this resource, we approached the question of the determination of the word usages of a term, and then we introduced the notion of similarity between these various word usages. So, we were able to build for a term its word usage tree: the root groups together all the word usages of the term and a search in the tree corresponds to a refinement of these word usages. The labelling of the various nodes of the word usage tree of a term is made during a width-first search, from the root to leaves: the root is labelled by the term itself and each node of the tree is labelled by a term stemming from the clique or quasi-clique this node represents. We show on an example that it is possible that some nodes of the tree, often leaves, cannot be labelled without ambiguity. These nodes represent generally very close word usages which could be grouped together later because of the evolutionary character of our lexical network.

Keywords : Natural Language Processing, evolutionary lexical network, tree of word usages for a term, labelling nodes of word usages tree

1. Introduction

Dans cet article, nous rappelons tout d'abord brièvement les principes de deux jeux visant à construire une base de relations entre termes. Le premier de ces deux jeux (JeuxDeMots [1]) permet

[1] JeuxDeMots et PtiClic sont accessibles à l'adresse http://jeuxdemots.org et http://pticlic.org. Il existe également une version anglaise, ainsi qu'une version thaï et une version japonaise (toutes deux en cours de développement), à l'adresse http://www.lirmm.fr/jeuxdemots/world-of-jeuxdemots.php.

la construction d'un réseau lexical : c'est un jeu ouvert grâce auquel les utilisateurs fournissent des associations entre termes, en proposant des termes destination à partir d'un terme origine et d'une consigne correspondant à une relation typée. Le deuxième jeu (PtiClic [1]) permet aux utilisateurs de consolider les associations acquises grâce à JeuxDeMots en leur proposant de valider, ou non, des associations potentielles entre termes : c'est un jeu fermé, les utilisateurs ne proposant pas de terme. L'un des éléments-clés de ces deux jeux est qu'une relation ne peut être validée que si elle est proposée par au moins deux utilisateurs. Dans une deuxième partie, utilisant le réseau ainsi obtenu, nous abordons la problématique de la détermination de la polysémie d'usage, en analysant les relations de chaque terme avec ses voisins immédiats. La similarité entre les différents usages d'un même terme est alors abordée; elle nous permet, dans une troisième partie, de construire l'arbre des usages d'un terme constituant un raffinement dans la différentiation de ses usages. Le nommage des différents nœuds de cet arbre est effectué lors d'une descente en largeur. Nous obtenons donc, pour chaque terme, l'ensemble de ses différents usages, structuré en arbre dont les nœuds sont nommés. L'intérêt d'une telle structuration est qu'elle devrait permettre de réaliser une désambiguïsation ; en particulier, l'un des objectifs est de pouvoir relier des termes entre eux à l'aide de relations, dans le réseau lexical obtenu, non pas seulement sur les termes eux-mêmes (avec ambiguïtés pour les termes polysémiques), mais également entre leurs usages (en levant donc ces ambiguïtés).

2. Constitution du réseau lexical

2.1. Principe du logiciel initial : JeuxDeMots

Le principe de base du logiciel, le déroulement d'une partie, ainsi que la construction progressive du réseau lexical, ont déjà été décrit par (Lafourcade and Joubert, 2008) et (Lafourcade et al., 2009). Une partie se déroule entre deux joueurs, en asynchrone. Pour un même terme [2] cible T et une même consigne (synonymes, domaines, association libre…), les deux joueurs proposent des termes correspondant, selon eux, à cette consigne appliquée à ce terme T. Ces propositions sont limitées en nombre, ce qui a pour effet d'augmenter leur pertinence, mais également dans le temps pour favoriser leur caractère spontané. Nous mémorisons alors les réponses communes à ces deux joueurs [3]. Les validations sont donc faites par concordance des propositions entre paires de joueurs. Ce processus de validation rappelle celui utilisé par (von Ahn and Dabbish, 2004) pour l'indexation d'images ou plus récemment par (Lieberman et al., 2007) pour la collecte de « connaissances de bon sens ». A notre connaissance, il n'avait jamais été mis en œuvre dans le domaine des réseaux lexicaux.

Le logiciel développé permet la construction d'un réseau lexical reliant les termes par des relations typées et pondérées [4]. Ces relations sont typées par la consigne imposée aux joueurs ; elles sont pondérées en fonction du nombre de paires de joueurs qui les ont proposées, comme indiqué dans (Lafourcade and Joubert, 2008). La structure du réseau lexical que nous cherchons ainsi à obtenir s'appuie sur les notions de nœuds et de relations entre nœuds, selon un modèle

[2] Un terme peut être constitué de plusieurs mots (par exemple : *sapin de Noël*).

[3] La limitation dans le temps de la saisie des propositions des joueurs peut favoriser les fautes d'orthographe, mais, comme nous ne mémorisons que les réponses communes aux deux joueurs d'une même partie, l'expérience montre que ce risque est très limité et que seules subsistent les fautes d'orthographes qui de toutes façons auraient été faites par les joueurs (par exemple : *chat* pour *chas*, en parlant d'*aiguille*).

[4] Une relation peut donc être considérée comme un quadruplet : terme origine, terme destination, type et poids de la relation. Entre deux mêmes termes, plusieurs relations de types différents peuvent exister.

initialement présenté par (Collins and Quillian, 1969) et plus récemment explicité par (Polguère, 2006). Chaque nœud du réseau est constitué d'une unité lexicale (terme, raffinement d'un terme ou segment textuel) et les relations entre nœuds traduisent des fonctions lexicales, telles que présentées par (Mel'čuk et al., 1995). Plusieurs exemples de relations acquises ont été donnés dans (Lafourcade and Joubert, 2008).

2.2. Consolidation du réseau lexical : logiciel PtiClic

De manière analogue à JeuxDeMots (JDM), une partie de PtiClic se déroule en asynchrone entre deux joueurs. Un premier joueur se voit proposer un terme cible T, origine de relations, ainsi qu'un nuage de mots provenant de l'ensemble des termes reliés à T dans le réseau lexical produit par JDM. Plusieurs consignes correspondant à des types de relations sont également affichées. Le joueur associe, par cliquer-glisser, des mots du nuage aux consignes auxquelles il pense qu'ils correspondent. Ce même terme T, ainsi que le même nuage de mots et les mêmes consignes, sont également proposées à un deuxième joueur. Selon un principe analogue à celui mis en place pour JDM, seules leurs propositions communes aux deux joueurs sont prises en compte, renforçant ainsi les relations du réseau lexical.

Contrairement à JDM, PtiClic est un jeu fermé où les utilisateurs ne peuvent pas proposer de nouveaux termes, mais sont contraints de choisir parmi ceux affichés. Ce choix de conception doit permettre de réduire le bruit dû aux termes mal orthographiés ou aux confusions de sens. PtiClic réalise donc une consolidation des relations produites par JDM et permet de densifier le réseau lexical. Notons également que PtiClic permet de créer de nouvelles relations entre termes précédemment reliés par au moins une relation d'un autre type (même si ce n'est pas l'objectif premier de ce logiciel).

Afin de réduire le silence correspondant aux termes non proposés par les utilisateurs de JDM, (Zampa and Lafourcade, 2009) ont suggéré de générer le nuage de mots à l'aide de LSA, en utilisant un corpus externe de grand volume [5]. Cette solution permet d'augmenter le réseau lexical par l'ajout de nouvelles relations, en proposant aux joueurs de nouveaux termes destinations non encore reliés à T dans notre réseau.

3. Similarité « fine » entre usages d'un même terme

3.1. Détermination des usages

La majorité des termes courants est polysémique. Lorsqu'un terme T est polysémique, les termes qui lui sont directement reliés forment plusieurs groupes distincts, chacun de ces groupes constituant un sens d'usage de T. La notion de sens d'usage (appelée plus communément usage) est beaucoup plus fine que celle de sens qui, comme l'a montré (Véronis, 2001), est relativement pauvre lorsqu'on se réfère aux dictionnaires traditionnels ou à des ressources comme WordNet. L'usage est donc en TALN une notion plus importante que le sens. Pour ne citer qu'un exemple extrait de Fig. 1, *sapin-vert-arbre* et *sapin-arbre-forêt* constituent deux usages distincts du terme *sapin*, alors que, selon les dictionnaires traditionnels, il s'agit manifestement du même sens de ce terme. Les usages correspondent donc non seulement aux différents sens, mais également aux champs lexicaux « forts » donnés par les joueurs (cette notion sera reprise en section 5, lors de la discussion des résultats obtenus pour le terme *sapin*). Nous faisons l'hypothèse que

[5] L'expérimentation réalisée utilise un corpus comportant une année du journal « Le Monde ».

les sens d'usage d'un terme correspondent dans le réseau aux différentes cliques auxquelles appartient ce terme.

Comment déterminer une clique ? C'est un ensemble de termes « fortement » reliés entre eux constituant un sous-graphe induit complet (ou clique) dans le réseau lexical. Les liens qui sont considérés ici sont des relations de type « *idée associée* » ou synonymie, ainsi que des relations ontologiques (hyperonymie, hyponymie, …). Nous faisons actuellement abstraction des autres types de relation en raison de leur caractère non symétrique. Deux termes appartiennent à une même clique s'il existe au moins une relation de l'un vers l'autre et au moins une relation de l'autre vers l'un.

Evaluer la pertinence d'un usage consiste à obtenir une mesure de son importance à la fois en fréquence d'utilisation mais aussi en couverture lexicale. On émettra l'hypothèse que pour un terme donné et en dehors de tout contexte spécifique, l'usage le plus pertinent est celui auquel on pense en premier en général. Ainsi donc, lors d'une analyse sémantique de texte, les usages peuvent être pondérés par défaut en fonction de leur pertinence a priori. Compte tenu du principe de la pondération des relations dans notre réseau lexical, le poids d'un usage est corrélé aux poids des relations entre les termes de la clique qui caractérise cet usage. Ainsi, pour une clique C de m termes et comprenant le terme T, la pertinence (Rel) de l'usage correspondant sera égale à :

$$\mathrm{Rel}(C) = \ln(m) * P(C) / [m*(m-1)] \quad \text{où} \quad P(C) = \sum_{i,j} \mathrm{poids}(T_i, T_j)$$

où *poids* (T_i, T_j) est la somme des poids des relations entre les termes T_i et T_j. La pertinence est la moyenne des poids des relations existant entre les termes de la clique, valeur qui en exprime la cohérence, que multiplie le logarithme du nombre de termes impliqués dans cette clique.

Il est à noter que la pertinence est une valeur relative entre cliques d'un même terme, et non une valeur absolue permettant la comparaison entre cliques de termes différents, car les poids des relations à partir d'un terme cible dépendent du nombre de fois où ce terme a été joué.

Fig. 1 présente les différents usages obtenus pour le terme *sapin*, ainsi que la pertinence de chacun de ces usages.

	P	*nl*	*moy*	*REL*
0: 'sapin' 'épicéa' 'ginkgo' 'conifère' 'cèdre' 'mélèze' 'résineux'	1400	42	33	65
1: 'sapin' 'épicéa' 'épinette' 'conifère'	480	12	40	55
2: 'sapin' 'montagne'	110	2	55	38
3: 'sapin' 'vert' 'arbre'	920	6	153	168
4: 'sapin' 'conifère' 'arbre'	1015	6	169	186
5: 'sapin' 'arbre' 'forêt'	1520	6	253	278
6: 'sapin' 'aiguille'	110	2	55	38
7: 'sapin' 'cercueil'	75	2	38	26
8: 'sapin' 'guirlande' 'Noël'	605	6	101	111
9: 'sapin' 'boule' 'boules'	280	6	47	51
10: 'sapin' 'boule' 'Noël'	590	6	98	108
11: 'sapin' 'Noël' 'sapin de Noël' 'sapin de noël'	720	12	60	83
12: 'sapin' 'Noël' 'fête'	770	6	128	141

Figure 1 : Les 13 usages du terme sapin actuellement décelés

3.2. Notion de similarité « fine »

La similarité entre deux objets peut être définie selon (Tversky, 1977) comme étant fonction de leurs caractéristiques communes par rapport à l'ensemble de leurs caractéristiques. En TALN,

on trouve plusieurs définitions de la similarité, par exemple (Manning and Schütze, 1999), ou plus récemment (Fairon and Ho, 2004). Dans notre cas, elle correspond au rapport du poids des relations reliant deux cliques sur le poids total des relations sur l'ensemble des termes de ces deux cliques. En considérant chaque clique comme un ensemble de termes et de relations pondérées les reliant, la similarité entre deux cliques C_1 et C_2 du même terme t s'écrira :

$$Sim(C_1,C_2) = [P(C_1 \cap C_2) + P(C_1\text{-}\{t\}, C_2\text{-}\{t\}) + P(C_2\text{-}\{t\}, C_1\text{-}\{t\}) - 2\bullet P(C_1\text{-}\{t\} \cap C_2\text{-}\{t\})] / P(C_1 \cup C_2)$$

ou plus simplement :

$$Sim(C_1,C_2) = 1 - [\, P(C_1\text{-}C_2+\{t\}) + P(C_2\text{-}C_1+\{t\}) \,] / P(C_1 \cup C_2)$$

où P(E) désigne la somme des poids des relations entre les termes composant l'ensemble E

et P(E,F) désigne la somme des poids des relations (orientées) unissant un terme de E à un terme de F [6], y compris donc les termes appartenant à E∩F.

Dans la deuxième expression de la similarité, nous exprimons le fait qu'elle peut être calculée comme étant égale à 1 moins le poids relatif des relations différentiant les cliques C_1 et C_2.

Nous parlons ici de similarité fine, car nous tenons compte des relations entre les termes de $C_1 \cap C_2$, mais également des relations éventuelles entre les termes de C_1 et les termes de C_2, même si ces termes n'appartiennent pas à l'intersection de ces deux cliques. En reprenant l'exemple de Fig. 1, à savoir les usages du terme *sapin*, un calcul de la similarité qui ne tiendrait compte que de l'intersection des cliques conduirait à une similarité nulle entre les cliques *2* : 'sapin''montagne' et *3* : 'sapin''vert''arbre' tout comme entre les cliques *2* : 'sapin' 'montagne' et *8* : 'sapin' 'guirlande' 'Noël', alors que dans notre réseau il existe une relation entre les termes *montagne* et *arbre*, alors qu'il n'en existe pas entre les termes *montagne* et *guirlande* ou *montagne* et *Noël* ; la similarité entre les cliques *2* et *3* n'est pas nulle, contrairement à celle entre les cliques *2* et *8*.

4. Arbre des usages

4.1. Construction de l'arbre des usages

Notre objectif est d'obtenir une représentation des différents usages d'un terme T sous forme d'un arbre, la racine regroupant tous les sens de T, les branches correspondant à ses différents usages. De manière générale, la plupart des termes possèdent plusieurs cliques non disjointes. Dans ce cas, plus on s'éloigne de la racine de l'arbre, plus on rencontre des distinctions fines d'usages. En réalité, nous construisons l'arbre des usages d'un terme T selon une méthode « bottom-up », depuis l'ensemble de ses cliques, c'est-à-dire depuis ses feuilles en remontant jusqu'à sa racine qui regroupe tous les sens de T. Pour cela, nous fusionnons ses cliques, deux à deux, en commençant par celles dont le coefficient de similarité est le plus élevé : nous constituons ainsi des quasi-cliques représentant des regroupements d'usages, proches lors des premières fusions, de moins en moins proches lors des fusions successives. L'algorithme de fusion s'arrête lorsque tous les coefficients de similarité sont nuls ; nous faisons l'hypothèse que les quasi-cliques obtenues alors pourraient correspondre aux différents sens de T, tels qu'on pourrait les trouver dans un dictionnaire. Fig. 2 présente, pour le terme *sapin*, les quasi-cliques

[6] Remarquons que pour le calcul de P(E,F) nous considérons les relations des termes de E « vers » les termes de F et donc $P(E,F) \neq P(F,E)$.

obtenues grâce à cet algorithme de fusion. Pour ce terme, nous avons calculé son arbre des usages ; celui-ci est présenté à Fig. 3.

Compte tenu de sa construction par fusions successives deux à deux de cliques, l'arbre des usages d'un terme est un arbre binaire, hormis à sa racine où le nombre de fils est égal au nombre de cliques fusionnées dont les coefficients de similarité sont nuls.

L'arbre des usages d'un terme est une structure exprimant les raffinements de ses différents sens. Il constitue donc un arbre de décision, structure de données qui pourra être exploitable en désambiguïsation.

13 : 'sapin' 'vert' 'arbre' 'forêt'	3 et 5 / 0.8333
14 : 'sapin' 'boule' 'Noël' 'fête'	10 et 12 / 0.7777
15 : 'sapin' 'vert' 'conifère' 'arbre' 'forêt'	4 et 13 / 0.7241
16 : 'sapin' 'guirlande' 'boule' 'Noël' 'fête'	8 et 14 / 0.7164
17 : 'sapin' 'guirlande' 'boule' 'Noël' 'fête' 'sapin de Noël' 'sapin de noël'	11 et 16 / 0.6067
18 : 'sapin' 'épicéa' 'ginkgo' 'épinette' 'conifère' 'cèdre' 'mélèze' 'résineux'	0 et 1 / 0.5789
19 : 'sapin' 'épicéa' 'ginkgo' 'vert' 'épinette' 'conifère' 'cèdre' 'arbre' 'mélèze' 'résineux' 'forêt'	15 et 18 / 0.2942
20 : 'sapin' 'guirlande' 'boule' 'Noël' 'fête' 'boules' 'sapin de Noël' 'sapin de noël'	9 et 17 / 0.1882
21 : 'sapin' 'montagne' 'aiguille'	2 et 6 / 0.12
22 : 'sapin' 'épicéa' 'montagne' 'ginkgo' 'vert' 'épinette' 'conifère' 'cèdre' 'arbre' 'aiguille' 'mélèze' 'résineux' 'forêt'	19 et 21 / 0.0224

Figure 2 : Quasi-cliques obtenues pour le terme sapin. Pour chacune d'entre elles, nous indiquons la liste des termes qui la composent, ainsi que les deux cliques ou quasi-cliques dont elle est la fusion avec leur similarité. Les numéros des cliques correspondent à ceux de la figure 1

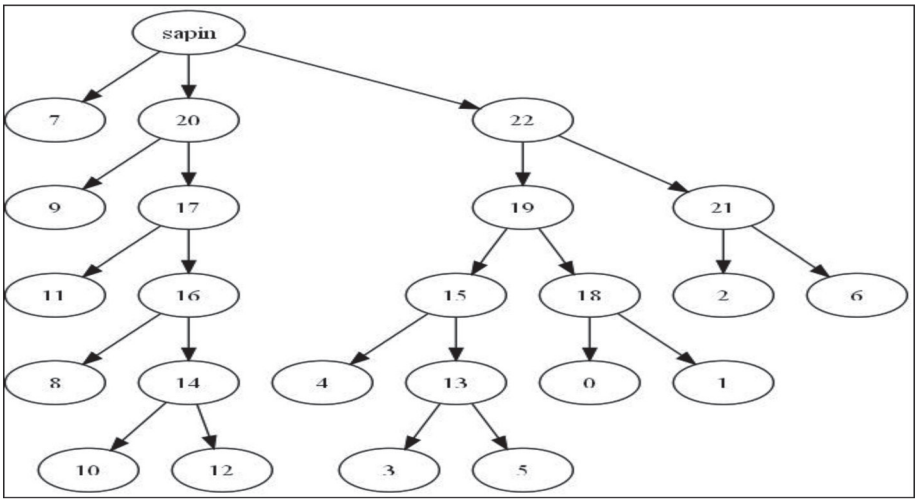


Figure 3 : Arbre des usages du terme sapin. Les valeurs numériques correspondent aux numéros des différentes cliques ou quasi-cliques des figg.1 et 2

4.2. Nommage de l'arbre des usages

Le nommage des différents nœuds de l'arbre des usages d'un terme est effectué lors d'une descente en largeur, donc selon une méthode « top-down », depuis la racine jusqu'aux feuilles. La racine est étiquetée par le terme lui-même. Chaque nœud de l'arbre est étiqueté par un terme issu de la clique ou quasi-clique que ce nœud représente ; le terme retenu est celui dont le poids

de la (des) relation(s) avec le terme racine est le plus élevé, après élimination de tous les termes étiquetant les nœuds de l'arbre situés à des profondeurs inférieures à celle du nœud concerné. Si tous les termes possibles sont reliés au terme racine avec des relations dont les sommes des poids sont identiques, nous retenons le terme dont le poids dans le réseau (fonction du nombre de fois où il a été proposé par les joueurs) est le plus élevé. Il peut donc arriver qu'un nœud ne soit pas étiquetable, si tous les termes qui le définissent ont déjà été utilisés dans le nommage des nœuds précédemment traités. Dans ce cas, ce nœud, mais également son frère ainsi que tous leurs successeurs, ne seront pas étiquetés. Fig. 4 montre le résultat obtenu pour le terme *sapin*. Sur cet exemple, nous constatons qu'effectivement tous les nœuds de l'arbre, souvent des feuilles, ne sont pas nécessairement étiquetables, du moins sans ambiguïté. Ces nœuds représentent généralement des usages très proches qui pourraient éventuellement à terme être regroupés en raison du caractère évolutif de notre réseau lexical. Dans ces cas là, nous préférons limiter l'arbre des usages nommés aux nœuds étiquetables, considérant comme éventuellement « instables » les nœuds qui ne le sont pas.

En fait, un nœud correspond à un usage (ou un regroupement d'usages) du terme racine, usage qui peut être nommé par la liste des étiquettes du chemin depuis la racine jusqu'au nœud considéré.

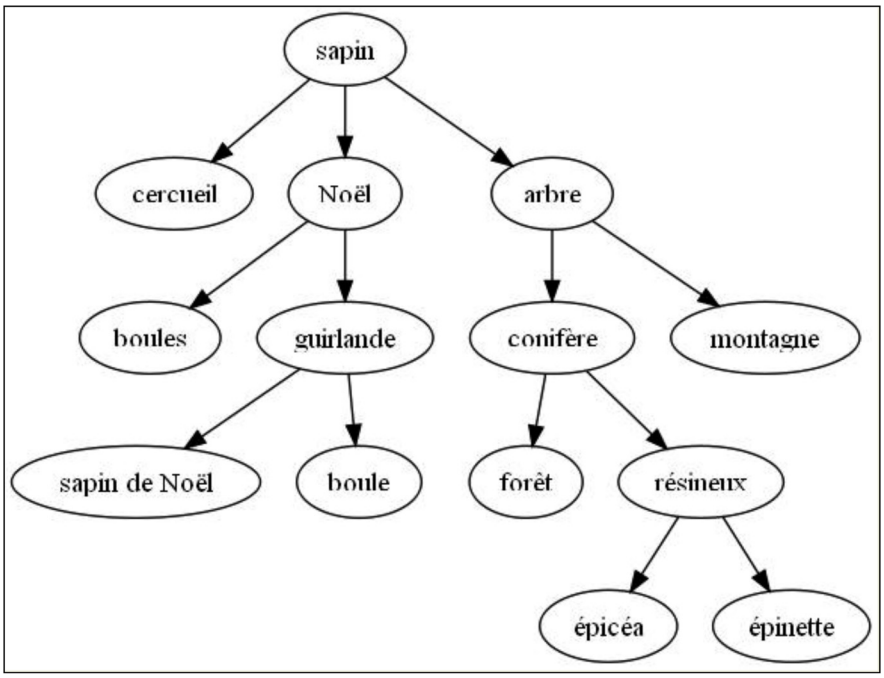


Figure 4 : Arbre des usages nommés du terme sapin. Nous avons élagué cet arbre en supprimant les nœuds non étiquetables, d'où une simplification de la structure de l'arbre par rapport à celui de Fig.3

5. Résultats

Au vu des premiers résultats obtenus, ces arbres des usages nommés paraissent conformes dans leurs structures générales à ceux que construirait un locuteur. En particulier, les branches principales, directement issues de la racine, correspondent dans la majorité des cas aux sens du terme racine tels qu'on pourrait les rencontrer dans un dictionnaire (voir par exemple les arbres obtenus pour les termes *frégate* et *blaireau* donnés en annexe). Ces branches principales se subdivisent en sous-branches qui sont autant de raffinements dans les usages.

On peut cependant remarquer que dans certains cas les branches principales issues de la racine ne correspondent pas uniquement à des sens tels qu'on les rencontre dans des dictionnaires, mais également à des champs lexicaux « forts » exprimés par les joueurs. Par exemple, l'arbre des usages du terme *sapin* (Fig. 4) possède une branche principale dont le premier terme est étiqueté *Noël*, alors que ce sens n'existe pas en tant que tel ni dans le dictionnaire en ligne Wiktionnaire ni dans les dictionnaires traditionnels (Larousse ou Robert).

Dans les structures détaillées de ces arbres des usages, on constate des éléments différents de ce qu'aurait écrit un individu humain : par exemples, les termes *boule* et *boules* distincts dans l'arbre de *sapin* et situés à des profondeurs différentes (Fig. 4), ou le terme *voilier* situé sous le terme *navire de guerre* dans l'arbre de *frégate* (voir en annexe). Ces quelques différences (peut-on réellement parler d'anomalies ?) sont-elles dues à notre méthode de construction des arbres des usages nommés ? ou au fait que notre réseau ne soit pas encore suffisamment « complet » ? Bien qu'il soit encore trop tôt pour répondre, il semblerait que plus on descende dans l'arbre des usages nommés, plus la « non complétude » de notre réseau soit prépondérante, et donc, à terme, nombre de ces anomalies devraient disparaître.

6. Conclusion

La plupart des termes courants sont polysémiques. Nous avons vu qu'un terme polysémique possède au moins autant d'usages que le nombre des feuilles de son arbre des usages nommés. Chaque terme est relié à d'autres par des relations typées. Mais ces relations ne concernent généralement pas tous les usages de ce terme. En considérant l'un des exemples donnés en annexe, les relations liées au terme *frégate* en tant que *navire* (par exemple, *domaine : Marine*) ne concernent pas les usages de *frégate* en tant qu'*oiseau* (pour lesquels la relation *domaine : Zoologie* existe). Il faudrait donc lier la relation *domaine : Marine*, non pas au terme *frégate* lui-même, mais uniquement aux usages correspondant à *frégate* en tant que *navire*, c'est-à-dire à l'un des nœuds de l'arbre des usages, et non pas à sa racine. De plus, dans cet exemple, ce n'est pas non plus à l'une des feuilles de l'arbre des usages qu'il faut lier la relation *domaine : Marine*, mais au nœud interne le plus proche de la racine regroupant tous les usages correspondant à cette notion de *navire*. Est-il possible de déterminer automatiquement ce nœud ? Comment procéder ? C'est l'état actuel de nos recherches. A suivre…

Références

von Ahn L. and Dabbish L. (2004). Labelling Images with a Computer Game. In *ACM Conference on Human Factors in Computing Systems* (CHI), pp. 319-326.

Collins A. and Quillian M.R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behaviour*, 8 (2) : 240-248.

Fairon C. and Ho N.D. (2004). Quantité d'information échangée : une nouvelle mesure de la similarité des mots. In *JADT2004*, Louvain-la-Neuve (Belgique).

Lafourcade M. and Joubert A. (2008). JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes. In *JADT2008*, Lyon, pp.657-666.

Lafourcade M., Joubert A. and Riou S. (2009). Sens et usages d'un terme dans un réseau lexical évolutif. In *Conférence sur le Traitement Automatique des Langues Naturelles (TALN'09)*, Senlis.

Lieberman H., Smith D.A. and Teeters A. (2007). Common Consensus: a web-based game for collecting commonsense goals. In *International Conference on Intelligent User Interfaces (IUI'07)*, Hawaii, USA.

Manning C.D. and Schütze H. (1999). *Foundations of Statistical Natural Language Processing*. Canbridge (MA) : The MIT Press.

Mel'čuk I.A., Clas A. and Polguère A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain : Editions Duculot AUPELF-UREF

Polguère A. (2006). Structural properties of Lexical Systems: Monolingual and Multilingual Perspectives. In *Proceedings of the Workshop on Multilingual Language Resources and Interoperability* (COLING/ACL 2006), Sydney, pp. 50-59.

Tversky A. (1977). Features of similarity. *Psychological Review*, 84 : 327-352.

Véronis J. (2001). Sense tagging: does it make sense?. In *Corpus linguistics' 2001 Conference*, Lancaster, U.K.

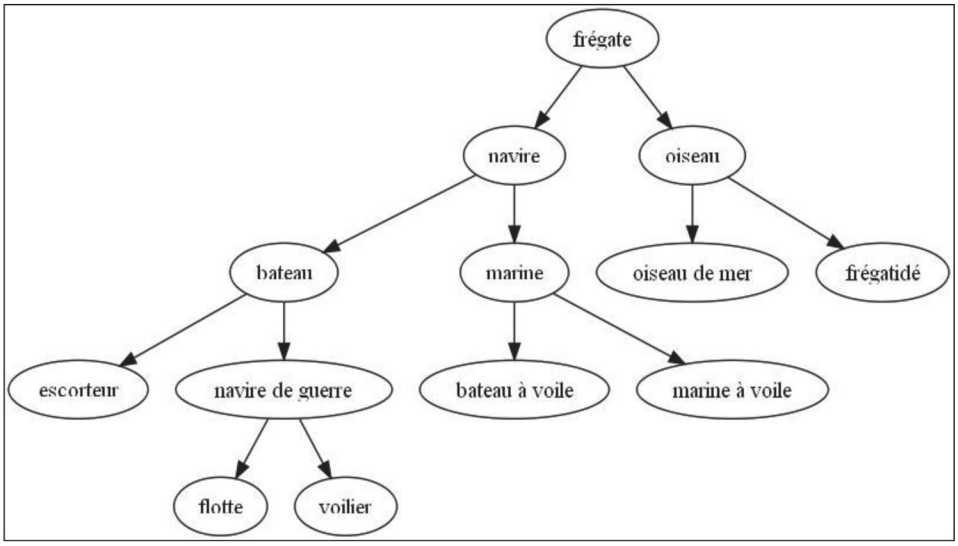
Zampa V. and Lafourcade M. (2009) Evaluations comparées de deux méthodes d'acquisitions lexicale et ontologique : JeuxDeMots vs Latent Semantic Analysis. In *XVIèmes rencontres de Rochebrune : ontologie et dynamique des systèmes complexes, perspectives interdisciplinaires.*

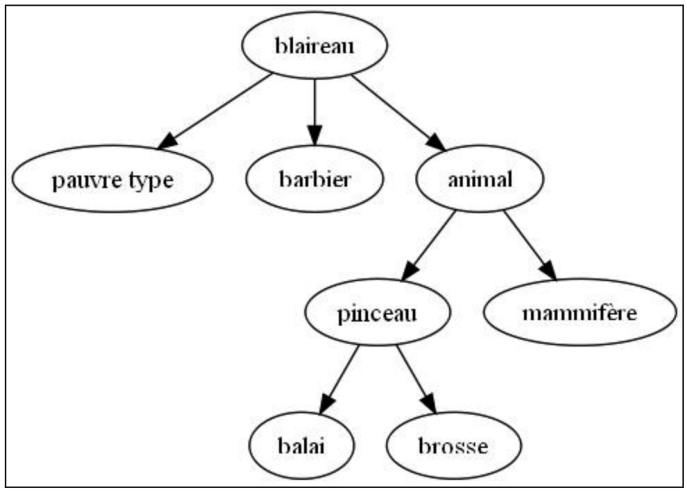
Annexes

Cliques et arbres des usages nommés pour les termes *frégate* (A) et *blaireau* (B)

A: Clique et arbre des usages nommés du terme frégate

		P	*nl*	*moy*	*REL*
0:	'frégate' 'destroyer' 'navire' 'bateau' 'navire de guerre'	5000	20	250	402
1:	'frégate' 'escorteur' 'navire'	1145	6	191	210
2:	'frégate' 'navire' 'voilier' 'bateau'	6830	12	569	789
3:	'frégate' 'navire' 'flotte' 'bateau' 'marine'	5705	20	285	459
4:	'frégate' 'bateau' 'navire de guerre' 'bateau à voile' 'marine à voile'	1765	20	88	142
5:	'frégate' 'bateau' 'marine' 'marine à voile'	1720	12	143	199
6:	'frégate' 'oiseau' 'oiseau de mer'	550	6	92	101
7:	'frégate' 'frégatidé'	50	2	25	17



B - Clique et arbre des usages nommés du terme blaireau

		P	*nl*	*moy*	*REL*
0:	'blaireau' 'carcajou'	40	2	20	14
1:	'blaireau' 'animal' 'mammifère'	625	6	104	114
2:	'blaireau' 'frottoir' 'goupillon' 'balai' 'époussette' 'pinceau' 'veinette' 'hérisson' 'saie' 'écouvillon' 'brosse'	2580	110	23	56
3:	'blaireau' 'pauvre type'	50	2	25	17
4:	'blaireau' 'faisceau' 'pinceau' 'touffe' 'brosse'	690	20	35	56
5:	'blaireau' 'barbier' 'mousse'	195	6	33	36)