Regroupement de structures de dérivations lexicales par raisonnement analogique

Sandrine Ollinger
CNRS, ATILF, UMR 7118 Nancy, F-54063, France
Université de Lorraine, ATILF, UMR 7118 Nancy, F-54063, France
sandrine.ollinger@atilf.fr

Résumé. Cet article propose une méthode de regroupement de structures de dérivations lexicales par raisonnement analogique. Nous présentons les caractéristiques générales d'un graphe lexical issu du Réseau Lexical du Français, dont nous exploitons par la suite les composantes faiblement connexes. Ces composantes sont regroupées en trois étapes : par isomorphisme, par similarité de relations, puis par similarité d'attributs. Les résultats du dernier regroupement sont analysés en détail.

Abstract. This paper presents a method for merging structures of lexical derivations by analogical reasoning. Following the presentation of general features of a lexical graph from the French Lexical Network, we focus on the weak connected components of this graph. This components are grouped together in three steps : by isomorphism, by relational similarity and finally by attributional similarity. The results of the last merging are analyzed in detail.

Mots-clés : graphe lexical, composantes connexes, analogie, raisonnement analogique, dérivation lexicale.

Keywords: lexical graph, analogy, connected components, analogical reasoning, lexical derivation.

1 Introduction

Cet article rend compte d'une expérimentation réalisée dans le cadre du projet RELIEF (REssource Lexicale Informatisée d'Envergure sur le Français), dont le but principal est le développement d'une modélisation informatisée à large couverture du lexique français : le Réseau Lexical du Français, ou RL-fr. Il s'inscrit dans une volonté d'assister les lexicographes dans la suite de leur travail en proposant des méthodes permettant le développement d'outils d'aide à la rédaction lexicographique [1] et l'enrichissement automatique de la ressource, sous couvert de validation manuelle.

Nous émettons l'hypothèse que le travail lexicographique s'effectue pour une part importante par raisonnement analogique et que le lexique d'une langue regorge de sous-parties analogues [2]. Une telle hypothèse sous-entend que le lexique s'organise en sous-groupes d'unités entretenant des relations privilégiées et que la structuration de ces relations se répète à l'intérieur du lexique.

Nous commencerons ici par présenter la structuration du RL-fr en graphe lexical. Nous ferons ensuite une brève présentation du raisonnement analogique, avant de nous concentrer sur l'application d'un tel raisonnement aux sous-groupes d'unités facilement isolables que sont les composantes faiblement connexes du graphe. Nous discuterons ensuite de la pertinence de l'exploitation des structures détectées pour l'enrichissement du RL-fr.

2 Le Réseau Lexical du Français

Le RL-fr s'apparente à la famille des réseaux lexicaux du type des WordNet (Fellbaum, 1998), de BabelNet (Navigli & Ponzetto, 2010) et de JeuxDeMots (Lafourcade & Joubert, 2010). Il s'en distingue par une visée de description lexico-

1. Ces outils pourront être développés sous la forme de fonctionnalités de l'éditeur lexicographique MvsDicet (Gader *et al.*, 2012).
2. Cette hypothèse est à rapprocher des considérations de (Grimes, 1990) sur les fonctions lexicales inverses et les régularités à motifs dans le lexique, ainsi que des travaux sur l'inférence de relations lexicales dans le réseau lexical JeuxDeMots (Zarrouk *et al.*, 2014).

graphique dans la lignée des dictionnaires virtuels (Atkins, 1996; Spohr, 2012) et des travaux en Lexicologie Explicative et Combinatoire (Mel'čuk *et al.*, 1995)[3]. Sa réalisation, manuelle, a débuté en 2011, au laboratoire ATILF. Sa structuration suit le modèle de système lexical introduit par (Polguère, 2009). Il s'agit d'un graphe orienté encapsulé dans une base de données contenant des entités et des relations de natures variées. Nous choisissons ici de nous concentrer sur le graphe lexical extrait de cette base. Après une brève présentation de ses différents éléments, nous montrerons en quoi il se rapproche d'un graphe «petit monde» et pourquoi une telle particularité nous intéresse.

2.1 Éléments du réseau

Le RL-fr fournit une description détaillée des unités lexicales du français. Conformément au cadre de la Lexicologie Explicative et Combinatoire, une unité lexicale s'entend ici comme une *lexie*, ayant un sens, une forme phonique/graphique et un ensemble de traits de combinatoire (Mel'čuk *et al.*, 1995, p.16). L'approche classique d'une entrée de dictionnaire regroupant différents sens d'une même unité est ici abandonnée au profit d'une approche consacrant une entrée indépendante à chaque sens. Le regroupement des lexies partageant une forme phonique/graphique et liés sémantiquement reste toute fois accessible par le biais de la notion de *vocable*. Un vocable est dit *monosémique* s'il ne correspond qu'à une seule lexie, *polysémique* dans le cas inverse. Les lexies partageant une forme phonique/graphique, mais aucun lien sémantique sont traitées comme des *homonymes* et réparties dans des vocables distincts.

Le graphe lexical issu de la base de données du RL-fr, désormais G_{RLfr}, correspond à l'ensemble fini des unités lexicales du RL-fr, muni de l'ensemble des relations directes entre ces unités.

Il contient trois types de sommets :

- des unités lexicales monolexématiques, ou *lexèmes* : VACHE **I.1** [*Dans le pré, des vaches broutent de l'herbe.*] ;
- des unités polylexématiques, ou *phrasèmes* :
  - des *locutions* : ⌜ PLANCHER DES VACHES ⌝ ;
  - des expressions phraséologiques non lexicalisées, telles que les *clichés linguistiques* : *Comment ça va ?* .

Cette granularité le distingue notamment des graphes lexicaux issus de dictionnaires papier exploités par (Gaume, 2004; Gaillard *et al.*, 2011a), dont les sommets sont des formes phoniques/graphiques. De plus, chaque sommet du G_{RLfr} est associé à une description lexicographique formelle. Celle des lexèmes et des locutions contient un ensemble de caractéristiques grammaticales, une étiquette sémantique (paraphrase minimale), une forme propositionnelle (structure prédicative), une combinatoire lexicale et des exemples d'emplois. Celle des phrasèmes contient, en plus, l'ensemble des lexies qu'ils incluent formellement. La description des expressions phraséologiques non lexicalisées est simplifiée. Elle ne comporte ni combinatoire lexicale, ni étiquette sémantique, ni forme propositionnelle.

Les relations entre ces unités, qui correspondent aux arcs de G_{RLfr}, sont également de trois types :

- des liens de *fonctions lexicales* (Mel'čuk *et al.*, 1995, p.125-152), désormais FL, qui rendent compte de la combinatoire lexicale des unités ;
- des liens de *copolysémie*, désormais CP, qui rendent compte des liens sémantiques entre les unités d'un vocable polysémique ;
- des liens d'*inclusion formelle*, qui rendent compte des différents lexèmes présents dans la forme d'un phrasème.

Les liens de FL sont les plus nombreux. Ils se répartissent en deux grandes classes : ceux mettant en jeu des FL servant à encoder des relations paradigmatiques – comme la synonymie – et ceux mettant en jeu des FL servant à encoder des relations syntagmatiques – comme la cooccurrence entre un nom et ses verbes supports. Selon (Mel'čuk *et al.*, 1995, p.126), «une **fonction lexicale** [=FL] est une fonction au sens mathématique». Elle se note traditionnellement :

F(*lexie 1*) = {*lexie 2*, *lexie 3*, ...}

L'ensemble de lexies {*lexie 2*, *lexie 3*, ...} est alors appelé *valeur d'application* de la FL **F** à son *argument*, la *lexie 1*.

Les liens de FL, pour leur part, mettent en relation les lexies deux à deux. Ainsi, si, comme en (1), la valeur d'application de la FL **Magn** (FL syntagmatique encodant la relation entre une lexie et ses cooccurrents d'intensification) contient une seule lexie, il existe un seul lien. En revanche si, comme en (2), elle comprend plus d'une lexie, il existe autant de liens que de lexies contenues dans cet ensemble.

(1) **Magn**(*coma*) = {*profond*$_{Adj}$ **II**}

3. Une étude comparative du RL-fr et du Wordnet de Princeton est disponible dans les actes de la conférence GWC 2014 (Gader *et al.*, 2014).

(2) **Magn**(*aboyer* I) = {*furieusement* I ; *férocement*}

De plus, chaque classe de FL est subdivisée en familles, correspondant à des types de relations. Ainsi, la famille **Syn** regroupe l'ensemble des FL relevant de la synonymie. Elle comporte notamment les FL suivantes :

– synonymie exacte : **Syn**(*pull*) = {*pull-over*} ;
– synonymie plus riche : $\mathbf{Syn}_{\supset}$(*fixer*) = {*clouer* I} ;
– synonymie plus riche relative au sexe : $\mathbf{Syn}_{\supset}^{\mathbf{sex}}$(*sénateur*) = {*sénatrice*} ;
– synonymie à intersection de sens : $\mathbf{Syn}_{\cap}$(*pull*) = {*sweat, sweat-shirt*}.

Nous parlerons des liens *sortants* d'une lexie pour désigner l'ensemble des liens correspondant à des applications de FL dont elle est l'argument. À l'inverse, nous parlerons de liens *entrants* pour désigner l'ensemble des liens correspondant à des applications de FL dont elle est un élément de la valeur.

2.2 Analyse topologique formelle

Nous émettons l'hypothèse que G_{RLfr} s'organise en sous-groupes d'unités entretenant des relations privilégiées. Une telle structure se rapproche de celle des graphes de données réelles observés dans de nombreux domaines (Watts & Strogatz, 1998; Newman, 2003; Gaume, 2004). Afin de caractériser G_{RLfr} et de déterminer si sa structure est celle d'un tel graphe, dit graphe petit monde, une analyse topologique, appelée *pedigree de graphe*, a été réalisée [4], visible dans le tableau 1.

sommets	21 992	coefficient d'agrégation	0,1327
arcs	42 626	Distribution des degrés entrants	
degré sortant moyen	1,9383	a	-2,3977
boucles	36	r^2	0,9397
arcs multiples	577	Plus grande composante connexe	
arcs symétriques	19 906	sommets	15 302
sommets isolés	3 226	arcs	38 274
composantes connexes	4 311	L	13,0402

TABLE 1: Pedigree du RL-fr

2.2.1 Caractéristiques formelles

La partie gauche du tableau 1 présente une première partie du pedigree de G_{RLfr}. Il s'agit d'un multigraphe orienté, qui comporte 21 992 sommets et 42 626 arcs. Chacun de ses sommets est, en moyenne, la source de près de deux relations. En tant que multigraphe [5], il comporte des boucles et des arcs multiples [6]. Les boucles, peu nombreuses, correspondent à des phénomènes lexicaux particuliers, tels que celui observable pour la lexie POIDS I, qui désigne à la fois une caractéristique physique et le deuxième actant de celle-ci. Les arcs multiples sont un peu plus nombreux. Ils correspondent, également, à des phénomènes lexicaux particuliers, tels que celui qui est observable entre les lexies ABOYER I et JAPPER, où il existe à la fois une relation de quasi-synonymie et une relation d'atténuation.

Les arcs symétriques sont beaucoup plus nombreux [7]. Il s'agit majoritairement d'arcs correspondant à des FL de la famille des synonymes et de dérivations syntaxiques. C'est le cas, par exemple, pour les lexies DANSER et DANSE I, pour lesquelles les relations de nominalisation et de verbalisation suivantes sont encodées [8] : $\mathbf{S}_0$(*danser*) = {*danse* I} et $\mathbf{V}_0$(*danse* I) = {*danser*}.

La proportion de sommets isolés (14,69%) doit être considérée dans le temps. Le tableau 2 montre comment elle a diminué au cours des six derniers mois, tandis que le nombre global de sommets a augmenté. Il montre également que la

4. Nous avons utilisé à cette fin le script *pedigree.py*, développé par Emmanuel Navarro (Gaillard *et al.*, 2011b).
5. À la suite de (Tabourier, 2010), nous ignorons ici la distinction entre multigraphes et pseudographes.
6. Attention, le nombre d'arcs multiple fourni par *pedigree.py* est calculé de façon séquentielle. L'ensemble des arcs du graphe est parcouru et c'est uniquement lorsqu'un arc correspond à un couple de sommets déjà reliés que le compteur d'arcs multiples est incrémenté.
7. Nous appelons arcs symétriques les arcs $a \rightarrow b$ pour lesquels il existe un arc $b \rightarrow a$.
8. Notez que les relations entre les lexies sont considérées en synchronie et qu'il n'est pas question, ici, d'encoder une dérivation morphologique orientée. Les lexies entrant en relation de dérivation syntaxique ne sont d'ailleurs pas nécessairement morphologiquement liées.

connectivité du RL-fr croît plus rapidement que sa nomenclature.

	28/10/13	10/03/14	Évolution
sommets isolés	3 540	3 226	-9%
sommets	20 793	21 992	+6%
arcs	34 922	42 626	+22%
composantes connexes	4 832	4 311	-11%

TABLE 2: Évolution du RL-fr.

La décomposition du graphe en composantes connexes réalisée ici consiste à le partitionner en sous-ensembles maximaux de sommets tous reliés entre eux, sans prendre en considération l'orientation des arcs. Il s'agit d'une décomposition en composantes faiblement connexes. Chaque sommet isolé est considéré comme une composante. Comme le montre le tableau 2, le nombre de composantes tend à diminuer et nous estimons que le RL-fr deviendra entièrement connexe avant d'avoir atteint sa maturité. Cependant, ces composantes constituent des sous-groupes de lexies facilement isolables et nous pensons que leur observation constitue une première étape intéressante dans la recherche et l'analyse de sous-ensembles de connexions lexicales privilégiées et récurrentes.

2.2.2 Graphe petit monde ?

Les graphes petits mondes se distinguent par la concomitance des quatre caractéristiques suivantes :

1. une faible densité, c.-à-d. un petit nombre d'arcs relativement au nombre de sommets ;
2. un coefficient d'agrégation élevé, c.-à-d. une forte probabilité que deux sommets voisins d'un même sommet soient eux-mêmes voisins ;
3. une distribution des degrés sortants et entrants (distribution des probabilités du nombre d'arcs associés à un sommet) qui suit une loi de puissance ;
4. une faible moyenne des plus courts chemins entre deux sommets quelconques du graphe.

Pour déterminer la densité d'un graphe, il faut s'intéresser aux nombres d'arcs (m) et de sommets (n) qui le constituent. Si G_{RLfr} était un graphe simple (sans boucle ni arcs multiples), son nombre maximal d'arcs vaudrait $n \times (n-1)$, soit environ 484 x 10^6. Nous pouvons donc affirmer que sa densité est faible. De plus, selon (Gaume, 2004), le nombre d'arcs observés dans les graphes petits mondes est généralement inférieur à $nlog(n)$. Pour un graphe de 21 992 sommets, il ne doit donc pas excéder 95 495 arcs, soit plus du double du nombre présent dans G_{RLfr} (42 626).

Pour déterminer si G_{RLfr} possède les trois autres caractéristiques des graphes petits mondes, nous nous intéressons à la seconde partie de son pedigree, à droite du tableau 1.

Le coefficient d'agrégation doit être considéré par rapport à celui d'un graphe aléatoire classique[9] de même densité (Newman, 2003), soit 0,00018. Nous pouvons donc affirmer que G_{RLfr} présente un coefficient d'agrégation élevé.

La distribution des degrés et la moyenne des plus courts chemins permettent de se faire une idée sur l'organisation des agrégats au sein du graphe. La distribution des degrés entrants est ici fortement corrélée (0,9397) à une loi de puissance de coefficient -2,3977. Cela signifie que la probabilité pour un sommet quelconque d'avoir beaucoup de voisins est faible et que celle d'en avoir peu est forte. Dans le cas de G_{RLfr}, les lexies fortement connectées sont des lexies carrefour, telles que FAIRE **II.1** [*Il fait du ping-pong.*], jouant un rôle central dans l'organisation du lexique et les lexies très faiblement connectées sont des lexies rares, telles que MONOGRAMME [*Cette assiette est signée PAK, monogramme de Pieter Adriaensz Kocks.*].

(Bollobás & Riordan, 2004) ont montré que la longueur moyenne des plus courts chemins (L) des graphes petits mondes n'excède pas $logn/loglogn$. Une telle valeur signifie qu'il est possible de passer rapidement d'un sommet du graphe à n'importe quel autre. G_{RLfr} n'étant pas connexe, une telle mesure est problématique (Newman, 2003). Une alternative possible consiste à effectuer cette mesure sur la plus grande partie connexe du graphe. Dans le cas de G_{RLfr}, cette composante comporte 15 302 sommets, son L ne devrait donc pas excéder 6,8091. Il est pourtant de près du double (13,0402). Cependant, si nous considérons cette valeur dans le temps, à l'aide du tableau 3, nous constatons que le nombre de sommets de la plus grande composante connexe du RL-fr a augmenté de 17% au cours des six derniers mois, alors que la longueur moyenne de ses plus courts chemins a diminué de 20%.

9. Pour un graphe aléatoire classique, la valeur de C est estimée à $2m/n^2$.

	28/10/13	10/03/14	Evolution
n de la plus grande partie connexe	13 082	15 302	+17%
L de la plus grande partie connexe	16,3112	13,0402	-20%

TABLE 3: Évolution de la L_lcc du RL-fr.

En conclusion, nous pouvons dire que G_{RLfr} est proche d'un graphe petit monde. La moyenne des plus courts chemins de sa plus grande composante connexe est beaucoup plus grande qu'attendu, mais diminue au fur et à mesure du développement de la ressource. L'hypothèse d'un lexique s'organisant en sous-groupes d'unités entretenant des relations privilégiées est, pour sa part, d'ores et déjà confirmée. La jeunesse du RL-fr ne nous permet pas de déterminer avec exactitude si la modélisation du lexique qu'il propose aboutira à un graphe petit monde. Mais elle nous permet d'exploiter son absence de connectivité pour tester l'hypothèse de structures locales analogues et mettre en place un protocole de détection automatique de celles-ci.

3 À la recherche de configurations de dérivations lexicales

Comme nous l'avons vu en 2.2, G_{RLfr} s'organise en agrégats lexicaux [10]. Nous émettons l'hypothèse que la nature de ces agrégats varie en fonction de leur taille. Ainsi, les agrégats de grandes tailles correspondraient à des champs sémantiques, tandis que les agrégats plus denses et plus petits correspondraient à des connexions lexicales particulières.

3.1 Configurations de dérivations lexicales

Le travail présenté ici porte sur la seconde catégorie d'agrégats. Nous pensons que des connexions lexicales particulières se répètent à l'intérieur du graphe et qu'il est possible d'en élaborer des modèles. De tels modèles, que nous nommons *configurations de dérivations lexicales*, pourraient être exploités pour enrichir le RL-fr et intégrer de nouvelles fonctionnalités à l'éditeur lexicographique utilisé pour son développement. Ils seraient constitués d'un ensemble de relations orientées entre lexies, ou plus exactement entre profils de lexies détaillant les caractéristiques nécessaires au déclenchement d'une configuration. Deux axes d'enrichissement automatique seraient alors envisageables : la génération de liens entre lexies correspondant aux profils et l'enrichissement des descriptions incomplètes de lexies d'ores et déjà interconnectées.

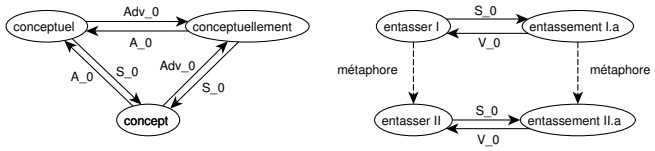


FIGURE 1: Exemples de connexions lexicales.

La figure 1 présente deux exemples d'agrégats lexicaux. Les relations en œuvre dans le premier sont la nominalisation [$\mathbf{S}_0$], l'adverbialisation [$\mathbf{Adv}_0$] et l'adjectivation [$\mathbf{A}_0$]. Une telle structure de relations ne concerne pas que les lexies CONCEPT, CONCEPTUEL et CONCEPTUELLEMENT. Nous pouvons facilement prédire qu'il s'agit d'une configuration de dérivations lexicales, qu'il serait intéressant de pouvoir propager dans le réseau.

Le second exemple illustre un autre type de configuration. Elle met en œuvre, d'une part, une symétrie de dérivations syntaxiques entre deux couples de lexies (nominalisation [$\mathbf{S}_0$] et verbalisation [$\mathbf{V}_0$]) et, d'autre part, une symétrie de dérivation métaphorique entre ces mêmes lexies réorganisées en couples différents. Ici aussi, nous pouvons facilement prédire qu'il s'agit d'une structure récurrente dans le RL-fr, pour laquelle il serait intéressant de définir des profils de lexies.

Nous pensons que de nombreuses autres configurations de dérivations lexicales existent, mettant en jeu des ensembles de

10. Ces agrégats sont à rapprocher des notions de communautés (Borgatti *et al.*, 1990; Navarro *et al.*, 2010) et de motifs locaux (Milo *et al.*, 2004; Wernicke, 2006) que nous ne détaillerons pas ici.

plus de deux lexies. Dans un premier temps, nous avons choisi de chercher à identifier celles en jeu dans les composantes connexes de G_{RLfr}. Ces composantes présentent l'avantage d'être facilement accessibles. Nous leur consacrons donc la présente expérience, qui pose les bases d'une procédure d'identification de configurations de dérivations lexicales par regroupement de microstructures analogues.

3.2 Regroupement de composantes connexes analogues

Comme nous l'avons vu en 2.2, G_{RLfr} est partitionnable en 4 311 composantes connexes, désormais CC. Afin d'identifier parmi elles des configurations de dérivations lexicales, nous avons choisi de procéder par regroupements successifs, visant l'automatisation d'un raisonnement analogique.

3.2.1 Raisonnement analogique

À la suite de (Gentner, 1983; Medin *et al.*, 1990), nous considérons le raisonnement analogique comme un appariement structurel. Les lexies s'apparentent alors à des objets disposant d'un certain nombre d'attributs, éléments de leur description lexicographique, et entretenant des relations, représentées par les arcs de G_{RLfr}. Dans une telle approche, une analogie s'établit entre une CC source et une CC cible. La « bonne qualité » d'une analogie implique que les relations présentes dans la CC source soient mises en correspondance avec les relations de la CC cible. La projection des attributs est, elle, de moindre importance.

Nous empruntons à (Turney, 2006) les notions de similarités de relations et d'attributs, ainsi que de mesures de celles-ci. Rapportée aux données que nous exploitons, la similarité de relations entre deux CC, CC_1 et CC_2, dépend du degré de correspondance entre les relations qu'elles mettent en jeu. La mesure de cette similarité est une fonction qui associe les deux CC à un nombre réel, $sim_r(CC_1, CC_2) \in \Re$. La similarité d'attributs, pour sa part, s'établit entre deux lexies L_1, L_2 et dépend du degré de correspondance entre leurs descriptions lexicographiques. La mesure de cette similarité est une fonction qui associe les deux lexies à un nombre réel, $sim_a(L_1, L_2) \in \Re$.

Tout comme le fait (Lepage, 2003), nous avons choisi de restreindre l'ensemble des valeurs possibles de sim_r et sim_a en les ramenant à des nombres réels compris entre 0 et 1 ; 0 équivalent à l'absence de similarité, 1 à une similarité complète.

À partir de ces considérations, l'identification de CC analogues, susceptibles de faire émerger des configurations de dérivations lexicales, s'est déroulé en trois étapes de regroupement : par isomorphisme (3.2.2), par similarité de relations (3.2.3), puis par similarité d'attributs (3.2.4).

3.2.2 Regroupement par isomorphisme

La première étape a consisté à regrouper les CC par structures mathématiques [11]. Chaque CC a alors été considérée comme un graphe indépendant et comparée aux autres CC en vue d'établir des ensembles isomorphes [12].

Deux graphes sont isomorphes s'ils comportent le même nombre de sommets, le même nombre d'arcs et que leurs arcs se répartissent entre les sommets de manière identique. Ainsi, dans la figure 2, seuls les deux premiers graphes le sont.

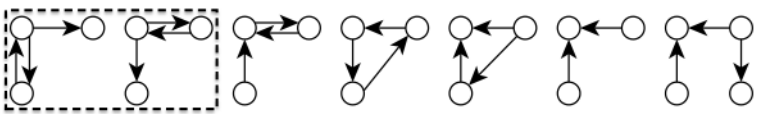


FIGURE 2: Exemple d'isomorphisme de graphes.

Lors de ce traitement, 3 226 lexies isolées et 517 CC ne contenant que deux sommets ont été exclus. 140 CC ne partageant leur structure avec aucune autre ont également été écartées. 428 CC ont été conservées, réparties en 36 groupes.

11. Nous avons choisi de laisser de côté la question de l'existence de sous-structures analogues à l'intérieur d'une ou plusieurs CC et de nous concentrer sur les CC directement manipulables.

12. Nous avons utilisé à cette fin la librairie python *igraph* et sa fonctionnalité *isomorphic*. G_{RLfr} et ses CC étant orientés, cette fonctionnalité a eu recours à l'algorithme VF2 (Cordella *et al.*, 2001).

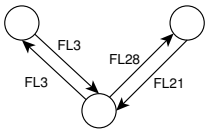
3.2.3 Regroupement par similarités de relations

La deuxième étape a consisté à subdiviser les groupes de CC isomorphes obtenues précédemment en fonction des relations qu'elles mettaient en œuvre.

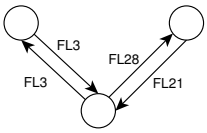
À cette fin, nous avons étiqueté l'ensemble des arcs de la manière suivante :

– les liens de FL sont étiquetés à l'aide du préfixe **FL**, suivi de l'identifiant unique de la FL dans la base de données du RL-fr : un lien de nominalisation $\mathbf{S}_0$ devient **FL21** ;
– les liens de CP sont étiquetés à l'aide du préfixe **CP**, suivi de l'identifiant unique du type de co-polysémie dans la base de données du RL-fr : un lien de métaphore devient **CP1** ;
– les liens d'inclusion formelle sont étiquetés **PH**.

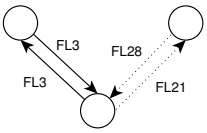
Nous avons ensuite comparé les ensembles d'étiquettes d'arcs des CC. Si les ensembles étaient identiques, nous avons estimé être dans une situation de similarité de relations complète, équivalent à un $sim_r = 1$. Ceci n'est cependant pas nécessairement exact, car l'orientation des relations n'est pas prise en compte dans ce traitement. La figure 3 montre ainsi trois CC regroupées par similarité de relations, alors que seulement deux d'entre-elles le sont réellement.



(a) CC2718



(b) CC2058



(c) CC1496

FIGURE 3: Exemple de regroupement par similarité de relations.

Les relations en jeu dans ces CC sont la synonymie exacte (**FL3**), la nominalisation (**FL21**) et l'adjectivation (**FL28**). Nous pouvons donc en conclure que les CC 2718 et 2058 sont chacune composées de deux noms en relation de synonymie exacte et d'un adjectif en lien de dérivation syntaxique avec l'un d'entre eux, tandis que la CC 1496 est composée de deux adjectifs et d'un nom en lien de dérivation syntaxique avec l'un d'entre eux. Ces conclusions sont confirmées par l'observation des lexies effectivement concernées : TOMBEAU, TOMBE et TOMBAL pour la CC 2718 ; PEUPLE**1**, ETHNIE et ETHNIQUE pour la CC 2058 ; ISRAÉLITE$_{Adj}$, JUÏF$_{Adj}$ **1** et JUDAÏSME pour la CC 1496.

À l'issue de cette étape, 192 CC ont été conservées, réparties en 47 groupes. Les CC isolées ont été exclues.

3.2.4 Regroupement par similarité d'attributs

La dernière étape de notre automatisation du raisonnement analogique a consisté à subdiviser les groupes de CC en similarité de relations complète en fonction des lexies qu'elles mettaient en jeu.

Comme nous l'avons souligné, l'orientation des relations n'a pas été prise en compte dans le précédent regroupement. Nous pensons qu'une ultime étape, de regroupement par similarité d'attributs des lexies, permet de remédier à ce manquement. Ainsi, en comparant l'ensemble des lexies de la CC 2718 à celles de la CC 2058, trois couples de lexies en situation de similarité d'attributs complète seraient obtenus : $sim_a(\text{TOMBEAU, PEUPLE1}) = 1$, $sim_a(\text{TOMBE, ETHNIE}) = 1$, $sim_a(\text{TOMBAL, ETHNIQUE}) = 1$. En revanche, en comparant l'ensemble des lexies de la CC 2718 à celle de la CC 1496, seuls deux couples de lexies le seraient : $sim_a(\text{TOMBE, JUDAISME}) = 1$, $sim_a(\text{TOMBAL, JUIF}_{Adj}\,1) = 1$. Les CC 2718 et 2058 seraient alors regroupées, en tant que CC analogues, relevant d'une même configuration de dérivations lexicales, tandis que la CC 1496 se retrouverait isolée.

L'ensemble des éléments de description lexicographique disponible pour chaque lexie n'est pas pertinent pour effectuer de tels regroupements. Aussi, nous avons choisi de nous concentrer sur deux types d'éléments de description à notre disposition : les caractéristiques grammaticales et la combinatoire lexicale.

Les caractéristiques grammaticales encodées dans le RL-fr sont variées. Il s'agit de caractéristiques fondamentales, de marques d'usage langagier, stylistique et rhétorique, de caractéristiques formelles, de positions syntaxiques et d'informations de linéarisation. Chacune de ces informations ne semble pas pertinente pour déterminer si les connexions lexicales entre lexies sont analogues. Par exemple, la différence entre caractéristiques fondamentales de genre pour deux lexies

nominales n'est significative que dans des cas particuliers, comme la dérivation entre un nom de fonction masculin et son équivalent féminin. Nous avons donc souhaité concentrer notre attention sur les parties du discours. La granularité de ces dernières (50 parties du discours de surface et 9 parties du discours profondes [13]) risquait cependant de nous amener à considérer comme différentes des CC que nous aurions souhaité conserver regroupées. Nous avons donc eu recours à un artifice élaboré en collaboration avec les lexicographes : un ensemble de 13 méta-parties du discours, auxquelles a été rapportée chacune des 59 parties du discours existantes.

La combinatoire lexicale, quant à elle, nous a semblé être un élément essentiel pour déterminer la présence d'analogie entre CC. En effet, elle fournit des informations sur le rôle que joue chaque lexie dans l'organisation générale du lexique. Une lexie verbale comme FAIRE **I**, par exemple, joue un rôle carrefour. Elle n'est associée qu'à un seul lien de FL sortant – encodant une relation de synonymie – mais a 109 liens de FL entrants – dont 82 rendant compte de son utilisation en tant que verbe support et 18 en tant que verbe de réalisation. Elle se distingue en cela d'une lexie verbale plus classique comme KIDNAPPER **I**, qui est associée à 14 liens de FL sortants – encodant des relations de synonymie, de nominalisation et de dérivation sémantique nominale actancielle – et à quatre liens de FL entrants – encodant des relations de synonymie, de verbalisation et de causation. La comparaison de ces deux lexies conduit à penser que trois propriétés méritent d'être considérées comme pertinentes : la nature des liens de FL sortants, la nature des liens de FL entrants et le rapport arithmétique entre le nombre de liens entrants et le nombre de liens sortants. Cependant, la granularité des FL (673 FL distinctes en jeu dans notre graphe) risquait, ici aussi, de nous amener à différencier des CC qui ne devraient pas l'être. Nous avons donc choisi d'établir la comparaison de nature des liens de FL au niveau des familles de FL. De plus, les familles permettant de rendre compte de relations de synonymie (**Syn**), d'antonymie (**Anti**) et de contrastivité (**Contr**) ont été exclues de l'ensemble des points de comparaison. En effet, nous estimons que les liens relevant de ces familles amèneraient à des distinctions inappropriées. Ainsi, deux CC en similarité de relations complète mettant en jeu l'une la lexie FRÉQUEMMENT et l'autre la lexie EXTRÊMEMENT seraient considérées comme différentes, car la lexie FRÉQUEMMENT entretient une relation de synonymie avec SOUVENT, tandis que la lexie EXTRÊMEMENT ne compte aucun synonyme.

Nous avons finalement associé à chaque lexie un ensemble d'attributs constitué de la manière suivante :

- un attribut rendant compte de sa méta-partie du discours ;
- autant d'attributs FLout que de familles de FL en jeu dans l'ensemble de ses liens de FL sortants ;
- autant d'attributs FLin que de familles de FL en jeu dans l'ensemble de ses liens de FL entrants ;
- un attribut rendant compte du rapport arithmétique entre le nombre de liens entrants et le nombre de liens sortants, valant **out+** en cas de supériorité numérique des liens sortants, **in+** en cas de supériorité numérique des liens entrants ou **in=out** en cas d'égalité numérique.

Nous avons ensuite mesuré la similarité d'attributs des lexies en jeu dans chacun des 47 groupes de CC en situation de similarité de relations complète et les avons subdivisés de la manière suivante :

Soit un groupe composé des deux CC CC_1 et CC_2, comportant chacune trois lexies. Chacun des ensembles d'attributs des lexies de CC_1 a été comparé à chacun des ensembles d'attributs des lexies de CC_2. 9 mesures de sim_a ont donc été réalisées, pour lesquelles $sim_a(Lexie_1, Lexie_2) = \frac{nbr\ d'attributs\ communs}{nbr\ d'attributs\ Lexie_1 + nbr\ d'attributs\ Lexie_2}$. Si exactement trois $sim_a = 1$ ont été trouvées, les CC ont été considérées comme analogues et maintenues dans un seul groupe. Si plus de trois $sim_a = 1$ ont été trouvées, le groupe a été maintenu, mais la question de l'analogie des CC est restée en suspens. Enfin, si moins de trois $sim_a = 1$ ont été trouvées, les CC ont été considérées comme non analogues et chacune s'est retrouvée isolée.

Pour les groupes de plus de deux CC le regroupement effectué peut offrir plusieurs possibilités. En effet, une même CC peut partager un nombre différent de $sima = 1$ avec chacune des CC de son groupe. Nous avons alors décidé de regrouper les CC par nombre maximal de $sim_a = 1$. Une fois les CC partageant le plus de $sim_a = 1$ regroupées, le nombre de $sim_a = 1$ partagées par les CC restantes est considéré, etc.

4 Analyse des résultats

À l'issue de la dernière étape, 92 CC ont été réparties en 24 groupes de CC analogues et 80 CC ont été réparties en 20 groupes sans que la question de leur analogie soit tranchée. Les 20 CC restantes ont été isolées. Nous avons observé en dé-

13. Les parties du discours profondes se distinguent des parties du discours de surface. Ainsi, un nom commun (partie du discours de surface) comme la lexie BŒUF **IV** [*Ryan Gosling lui fait un effet bœuf.*] a un emploi appositif, il a donc la valence passive d'un adjectif (partie du discours profonde). Pour une introduction détaillée de ces notions, nous vous invitons à consulter (Mel'čuk, 2006).

tail les CC ainsi regroupées et isolées. L'objectif de cette analyse était à la fois de vérifier la pertinence des regroupements et de se faire une idée sur l'exploitation possible de ces résultats.

4.1 Groupes de composantes analogues

Chacun des 24 groupes constitués automatiquement contient bien des CC analogues. Elles mettent toutes en jeu des ensembles de trois lexies, reliées par des liens de FL. Cinq groupes ont la particularité de rassembler des CC comportant deux lexies de même méta-partie du discours. Cependant, la comparaison des attributs des lexies de ces CC deux à deux aboutit à seulement trois $sim_a = 1$, correspondant aux lexies occupant une place identique dans la structure des CC. Ainsi, les CC de la figure 4 présentent une structure de dérivations lexicales par verbalisation (**FL23**), nominalisation (**FL21**) et dérivation sémantique nominale du premier actant (**FL31**). Elles mettent chacune en œuvre un verbe et deux noms. Dans ce cas précis, les trois similarités d'attributs complètes comptabilisées concernent les couples (SUCCÉDER **I**, FOUILLER), (SUCCESSION **I**, FOUILLE) et (SUCCESSEUR, FOUILLEUR).

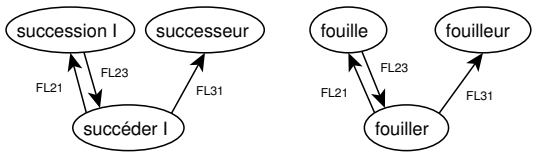


FIGURE 4: Exemple de CC analogues comportant deux lexies de même méta-pdd.

Quelles que soient les CC regroupées, les lexies qu'elles contiennent sont peu décrites dans le RL-fr. Il ne nous semble donc pas pertinent de les exploiter pour définir des profils de lexies susceptibles de déclencher des relations particulières. En revanche, certains groupes de CC présentent des structures imbricables. Cette particularité nous amène à nous interroger sur la granularité des configurations de dérivations lexicales. Faut-il chercher à établir les modèles les plus denses possible et exploiter les imbrications de CC analogues pour enrichir automatiquement les CC comportant le moins de liens de FL ? Le tableau 4 présente les imbrications observées [14] et les suggestions d'ajout de liens qui en découlent.

Parmi les huit groupes de CC comportant des lexies ayant pour méta-partie du discours adjectif, nom et verbe, deux imbrications apparaissent. Parmi les neuf groupes de CC comportant des lexies ayant pour méta-partie du discours adjectif, nom et adverbe, les imbrications sont plus nombreuses. Elles se divisent en deux chaînes distinctes.

Nous avons consulté l'équipe de lexicographes pour savoir si les CC les moins denses étaient toujours valides une fois enrichies. Un seul des cas que nous leur avons présentés a été rejeté. Il s'agit du résultat de la première chaîne d'imbrications concernant des groupes lexies « adjectif, nom et adverbe ». L'ajout d'une relation de dérivation sémantique adjectivale du premier actant (**FL104**) est considéré comme une erreur. Cette observation nous met en garde contre la propagation automatique de mauvais liens.

4.2 Groupes de composantes à l'analogie incertaine

L'observation des 20 groupes de CC dont la question de l'analogie était restée en suspens nous amènent à constater qu'il s'agit de groupes de CC analogues. Deux d'entre eux concernent des relations d'insertions formelles. Ils sortent donc du cadre de la dérivation lexicale qui nous intéresse. Il est cependant intéressant de constater que chacun de ces groupes correspond à une structure syntaxique de locution nominale particulière : **N + de + N** (LEVÉE DE BOUCLIERS) pour l'un, **Adj + N** (TIERS ÉTAT) pour l'autre.

L'ensemble des autres groupes comporte des CC qui mettent en œuvre des relations de synonymie ou d'antonymie. Deux d'entre eux ont même la particularité de rassembler des CC ne comportant que des liens de synonymie. L'ensemble des lexies qui les composent sont en similarité d'attributs complète, $nbr\,de\,sim_a = 1 : (nbr\,lexies)^2$. Ces groupes ne sont pas exploitables pour l'enrichissement automatique du RL-fr. Les CC d'un seul groupe comportent des liens d'antonymie.

14. Les FL en jeu dans ces groupes de CC sont : la dérivation sémantique adjectivale du premier et du deuxième actant (**FL104** et **FL111**), la dérivation sémantique adjectivale potentielle du deuxième actant (**FL107**), la nominalisation simple et prédicative (**FL21** et **FL366**), la verbalisation (**FL23**), l'adjectivisation (**FL28**) et l'adverbalisation (**FL28**).

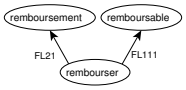
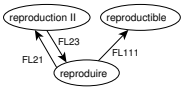
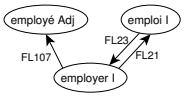
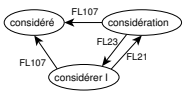
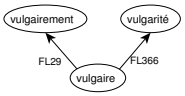
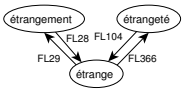
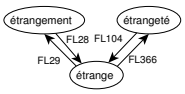
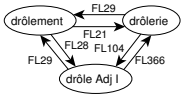
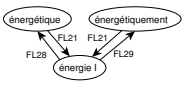
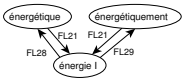
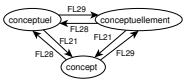
méta-pdd	liens communs	groupe 1	groupe 2	ajouts
adjectif, nom, verbe	FL111, FL21			FL23
	FL107, FL21, FL23			FL107
adjectif, nom, adverbe	FL29, FL366			FL28
	FL28, FL29, FL366			FL104
	FL104, FL28, FL29, FL366			FL21, FL29
	FL28, FL29			2 x FL21
	2 x FL21, FL28, FL29			FL28, FL29

TABLE 4: Imbrication de groupes de composantes analogues.

Il s'agit de CC de quatre lexies, entre lesquelles nous observons huit situations de similarité d'attributs complète. Dans 14 des groupes restants, la dérivation entre un nom masculin et son équivalent féminin [15] est présente dans l'ensemble des CC. Quel que soit le nombre de lexies en jeu dans ces CC, nous observons que le nombre de similarités d'attributs complètes vaut toujours deux de plus, $nbr\,de\,sim_a = 1 : nbr\,lexies + 2$. Seulement deux cas d'imbrication de structures sont constatés dans ces groupes. Le dernier groupe rassemble des CC comportant, entre autres, des relations de synonymie exacte. Ici aussi, le nombre de similarités d'attributs complètes vaut deux de plus que le nombre de lexies interconnectées.

À l'issue de cette analyse, nous pensons que le critère selon lequel deux CC regroupées par similarité de relations sont analogues si leurs lexies sont en situation de similarité d'attributs strictement deux à deux doit être affiné en fonction des relations mises en œuvre.

4.3 Composantes isolées

À l'issue de la dernière étape du traitement, 20 CC se sont retrouvées isolées, alors qu'elles étaient en situation de similarité de relations complète avec au moins une autre CC. Ces CC ne comportent que des liens de FL. En 3.2.3, nous avons émis l'hypothèse que l'absence de prise en compte de l'orientation des relations pouvait être à l'origine de mauvais regroupements. L'analyse des CC isolées nous permet de vérifier cette hypothèse et d'observer d'autres cas de figure.

Afin d'effectuer cette analyse, nous nous sommes intéressée aux couples constitués d'une CC isolée et de chacune des autres CC avec lesquelles elle était précédemment regroupée. Le tableau 5 montre le résultat de cette analyse. L'ensemble des cas de figure rencontrés peut être subdivisé à partir de deux critères : les méta-parties du discours des lexies en jeu dans les CC et la répartition des liens entre ces lexies.

15. Cette dérivation sémantique est encodée à l'aide des FL **Syn**$_{\subset}^{\textbf{sex}}$ et **Syn**$_{\supset}^{\textbf{sex}}$. Sur cette question précise, nous vous invitons à consulter (Delaite & Polguère, 2013).

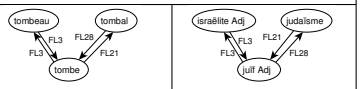
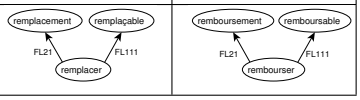
méta-pdd	liens	CC isolées	couples	$sim_a = 1$	exemple
$\neq$	$\neq$	1	2	2/3	
$\neq$	$=$	14	31	0/3	
$=$	$\neq$	4	19	1/3	
?	$=$	1	2	2/3	

TABLE 5: Répartition des groupes de composantes non analogues.

La première catégorie ainsi obtenue correspond à l'exemple illustré par la figure 3 de la section 3.2.3, elle ne permet d'envisager aucun enrichissement automatique. La deuxième n'en permet pas davantage. Elle comporte toutefois un cas intéressant de CC contenant une lexie mal catégorisée [16]. Cette observation a été transmise aux lexicographes et la description de la lexie corrigée. La troisième catégorie concerne, dans deux cas sur trois, des groupes de dix CC, dont une seulement n'est pas analogue aux autres. Cette catégorie nous semble exploitable pour générer automatiquement des liens manquants. Ainsi, dans le cas utilisé comme exemple, un lien de FL23 pourrait être ajouté à chacune des CC. Il correspondrait à la verbalisation du nom dans le cas de la CC non analogue aux neuf autres de son groupe – $\mathbf{V}_0$(*simplification*) = {*simplifier*} – et à la verbalisation des adjectifs dans les neuf autres cas – $\mathbf{V}_0$(*reproductible*) = {*reproduire*}. La dernière catégorie permet la détection d'une anomalie dans le réseau. En effet, elle est due à l'absence de partie du discours dans la description de la lexie REMPLACEMENT.

5 Conclusion

Les résultats de notre expérimentation confirment l'hypothèse d'un lexique s'organisant en sous-groupes d'unités correspondant à des structures de relations récurrentes, identifiables par automatisation du raisonnement analogique.

Les lexies présentes dans les composantes faiblement connexes exploitées ici sont cependant trop peu décrites pour permettre d'établir des configurations de dérivations lexicales comportant à la fois un ensemble de relations et des profils de lexies. De plus, nous n'avons pu observer aucune configuration mettant en jeu des relations de co-polysémie. Pour remédier à ces manquements, nous envisageons de nous désintéresser des composantes connexes au profit des sous-structures moins aisément isolables que sont les motifs locaux (Milo *et al.*, 2004; Wernicke, 2006). De surcroît, une méthode reste à développer en sortie de notre procédure, pour abstraire des configurations de dérivations lexicales des regroupements de structures analogues effectués.

Parallèlement à cela, une évaluation des configurations de dérivations lexicales par les lexicographes doit être mise en place. Elle devra permettre de déterminer des critères de bonne granularité des configurations et d'évaluer les risques de propagation d'erreurs liés à leur exploitation.

Références

ATKINS S. B. T. (1996). Bilingual dictionaries : Past, present and future. In M. GELLERSTAM, J. JÄRBORG, S.-G. MALMGREN, K. NORÉN, L. ROGSTRÖM & C. R. PAPMEHL, Eds., *Proceedings of the 7th EURALEX International Congress*, p. 515–546, Göteborg, Sweden : Novum Grafiska AB.

BOLLOBÁS B. & RIORDAN O. (2004). The diameter of a scale-free random graph. *Combinatorica*, **24**(1), 5–34.

BORGATTI S. P., EVERETT M. G. & SHIREY P. R. (1990). LS sets, lambda sets and other cohesive subsets. *Social Networks*, **12**(4), 337–357.

16. Il s'agit de la lexie verbale DONNER **I.3**, pour laquelle les caractéristiques grammaticales «nom commun» et «masc» étaient encodées.

CORDELLA L. P., FOGGIA P., SANSONE C. & VENTO M. (2001). An improved algorithm for matching large graphs. In *In : 3rd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition, Cuen*, p. 149–159.

DELAITE C. & POLGUÈRE A. (2013). Sex-Based Nominal Pairs in the French Lexical Network : It's Not What You Think. In *Proceedings of the 6th International Conference on Meaning-Text Theory (MTT'13)*, p. 29–40, Prague, Tchèque, République.

FELLBAUM C. (1998). *WordNet : an electronic lexical database*. Language, Speech and Communication. MIT Press.

GADER N., LUX-POGODALLA V., POLGUÈRE A. *et al.* (2012). Hand-crafting a lexical network with a knowledge-based graph editor. In *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex˜ III)*, p. 109–125.

GADER N., OLLINGER S. & POLGUÈRE A. (2014). One Lexicon, Two Structures : So What Gives ? In *Proceedings of the Seventh Global Wordnet Conference (GWC2014)*, p. 163–171, Tartu, Estonie : Global WordNet Association.

GAILLARD B., GAUME B. & NAVARRO E. (2011a). Invariants and variability of synonymy networks : Self mediated agreement by confluence. In *Proceedings of TextGraphs-6 : Graph-based Methods for Natural Language Processing*, TextGraphs-6, p. 15–23, Stroudsburg, PA, USA : Association for Computational Linguistics.

GAILLARD B., GAUME B. & NAVARRO E. (2011b). Invariants and variability of synonymy networks : Self mediated agreement by confluence. In *Proc. of TextGraphs-6 : Graph-based Methods for NLP*, p. 15–23, Portland : ACL.

GAUME B. (2004). Balades aléatoires dans les petits mondes lexicaux. *Information interaction intelligence*, **4**(2), 39–96.

GENTNER D. (1983). Structure-mapping : A theoretical framework for analogy. *Cognitive Science*, **7**(2), 155–170.

GRIMES J. E. (1990). Inverse lexical functions. In *Meaning-text theory : Linguistics, lexicography, and implications*, p. 350–364. University of Ottawa Press, James Steele edition.

LAFOURCADE M. & JOUBERT A. (2010). Computing trees of named word usages from a crowdsourced lexical network. *Investigationes Linguisticae*, **21**, 39–56.

LEPAGE Y. (2003). *De l'analogie rendant compte de la commutation en linguistique*. PhD thesis, Université Joseph-Fourier - Grenoble I.

MEDIN D. L., GOLDSTONE R. L. & GENTNER D. (1990). Similarity involving attributes and relations : Judgments of similarity and difference are not inverses. *Psychological Science*, **1**(1), 64–69.

MEL'ČUK I. (2006). Parties du discours et locutions. *Bulletin de la Société de linguistique de Paris*, **101**(1), 29–65.

MEL'ČUK I., CLAS A. & POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve : Duculot.

MILO R., ITZKOVITZ S., KASHTAN N., LEVITT R., SHEN-ORR S., AYZENSHTAT I., SHEFFER M. & ALON U. (2004). Superfamilies of evolved and designed networks. *Science*, **303**(5663), 1538–1542.

NAVARRO E., CAZABET R., CAZABET R. & CAZABET R. (2010). Détection de communautés, étude comparative sur graphes réels.

NAVIGLI R. & PONZETTO S. P. (2010). Babelnet : Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, p. 216–225, Stroudsburg, PA, USA : Association for Computational Linguistics.

NEWMAN M. E. J. (2003). The structure and function of complex networks. *SIAM REVIEW*, **45**, 167–256.

POLGUÈRE A. (2009). Lexical systems : graph models of natural language lexicons. *Language Resources and Evaluation*, **43**(1), 41–55.

SPOHR D. (2012). *Towards a Multifunctional Lexical Resource*, volume 141 of *Lexicographica. Series Maior*. De Gruyter.

TABOURIER L. (2010). *Méthode de comparaison des topologies de graphes complexes : applications aux réseaux sociaux*. PhD thesis, Paris 6.

TURNEY P. D. (2006). Similarity of semantic relations. *Computational Linguistics*, **32**(3), 379–416.

WATTS D. J. & STROGATZ S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, **393**(6684), 440–442.

WERNICKE S. (2006). Efficient detection of network motifs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, **3**(4), 347–359.

ZARROUK M., LAFOURCADE M. & JOUBERT A. (2014). About inferences in a crowdsourced lexical-semantic network. In *proc of 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg, Sweden.