HAL
open science

La polysémie régulière dans WordNet

Lucie Barque, François-Régis Chaumartin

To cite this version:

Lucie Barque, François-Régis Chaumartin. La polysémie régulière dans WordNet. TALN 2008 - 15eme conférence sur le Traitement Automatique des Langues Naturelles, Jun 2008, Avignon, France. pp. 101-108. hal-00611243

HAL Id: hal-00611243

https://hal.science/hal-00611243v1

Submitted on 26 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La polysémie régulière dans WordNet

Lucie Barque (1), François-Régis Chaumartin (2,3)

(1) Lattice – Université Paris 7, UFRL, Case 7003, 75251 Paris cedex 5
lucie.barque@linguist.jussieu.fr

(2) Proxem – 7 impasse Dumur, 92110 Clichy
frc@proxem.com

(3) Alpage – Université Paris 7, UFRL, Case 7003, 75251 Paris cedex 5
fchaumartin@linguist.jussieu.fr

Résumé Cette étude propose une analyse et une modélisation des relations de polysémie dans le lexique électronique anglais WordNet. Elle exploite pour cela la hiérarchie des concepts (représentés par des synsets), et la définition associée à chacun de ces concepts. Le résultat est constitué d'un ensemble de règles qui nous ont permis d'identifier d'une façon largement automatisée, avec une précision voisine de 91%, plus de 2100 paires de synsets liés par une relation de polysémie régulière. Notre méthode permet aussi une désambiguïsation lexicale partielle des mots de la définition associée à ces synsets.

Abstract This paper presents an analysis and modeling of polysemy in the WordNet English lexical database. It exploits the concepts hierarchy (constituted by synsets), and the gloss defining each of these concepts. The result consists of rules set which enabled us to identify in a largely automated way, with a precision close to 91%, more than 2100 synsets pairs, connected by a regular polysemy relation. Our method also allows a partial word sense disambiguation of the definition associated with these synsets.

Mots-clés : polysémie régulière, métaphore, métonymie, WordNet, désambiguïsation lexicale

Keywords: regular polysemy, metaphor, metonymy, WordNet, word sense disambiguation

1 Introduction

WordNet a déjà été exploitée en vue d'une caractérisation de la polysémie (Peters, 2006 ; Veale, 2006). Dans cette lignée, notre étude propose une démarche pour créer des patrons de polysémie régulière, puis pour détecter automatiquement leurs occurrences dans ce lexique. WordNet est construit sous la forme d'une hiérarchie de *synsets* (ensemble de lexies synonymes entre elles) ; ces concepts sont liés entre eux par différents types de relations sémantiques (hypéronymie, méronymie, antonymie, etc.). Chaque synset est également associé à une définition lexicographique (Fellbaum, 1998), comme l'illustre l'exemple suivant (où nous notons lexie#i la i^{ème} lexie d'un vocable dans WordNet.):

- **{**treachery#2, **betrayal**#1, treason#3, perfidy#2} = *an act of deliberate* ***betrayal***

Comme on peut le remarquer dans cet exemple, un des éléments du **synset** se retrouve dans sa définition (betrayal#1, *betrayal*). En fait il ne s'agit pas, à proprement parler, du même élément. Dans la suite de notre article, nous distinguerons systématiquement les notions de **lexie** (unité lexicale, association d'un signifiant et d'un signifié) et de **vocable** (unité polysémique regroupant différentes lexies de même signifiant). Nous dirons donc que la lexie betrayal#1 contient dans sa définition une autre lexie du vocable BETRAYAL, mais on ignore laquelle avant d'avoir désambiguïsé les éléments de la définition.

Notre article se constitue de trois sections. La première dresse un rapide état des lieux des recherches dédiées à la description de la polysémie régulière. La deuxième section présente les objectifs de cette étude et les moyens mis en œuvre pour y parvenir ; nous y proposerons notamment une méthode de construction des patrons de polysémie « assistée par ordinateur ». Nous présenterons enfin les résultats obtenus, sous forme d'une classification de ces patrons, et d'une mesure de leur régularité dans la version la plus récente de WordNet[1].

2 Etat de l'art

2.1 Description de la polysémie régulière

L'intérêt d'expliciter la polysémie régulière lors du développement d'un lexique a souvent été mis en évidence, notamment dans le cadre du TAL. Qu'elle soit représentée sous forme de règles lexicales (Ostler et Atkins, 1991; Copestake et Briscoe, 1995) ou de mécanismes transformationnels agissant lors de la composition de mots en syntagmes (Pustejovsky, 1995), la description de la polysémie régulière présente au moins deux avantages.

D'un point de vue théorique tout d'abord, il s'agit d'offrir une représentation de l'un des aspects de la formation du lexique, la polysémie régulière constituant une source importante de créativité lexicale. Par exemple, en disposant d'une règle de polysémie régulière entre une unité de type `animal` et une unité de type `individu ayant une fonction`, (alternance qui s'applique par exemple aux vocables GORILLE, LIEVRE, TAUPE, etc.) le lexique dispose potentiellement d'entrées de type `individu ayant une fonction` dérivées à partir d'entrée existantes de type `animal`. L'unité de type `individu ayant une fonction`

[1] WordNet 3.0 compte 117 659 synsets et 206 941 lexies.

pour le vocable MULE (individu chargé de transporter de la drogue) n'est, par exemple, pas (encore) référencée dans les dictionnaires. Autrement dit, les règles de polysémie constituent l'un des moyens de rendre compte de l'aspect dynamique du lexique.

Le second avantage, d'ordre pratique, concerne la valorisation du lexique à partir duquel s'effectue l'étude de la polysémie régulière, en l'occurrence WordNet. Les règles lexicales de polysémie régulière permettent en effet de systématiser l'encodage des données en fournissant au lexicographe un canevas définitionnel. Par exemple, le schéma de définition indiqué en gras dans l'exemple ci-dessous pourra servir à la définition d'autres paires de lexies de type `pièce de vaisselle ~ quantité (de qqch.)` liées par une métonymie régulière.

- **L2 = quantité de X contenue dans L1**
- assiette#2 de X = quantité de X contenue dans une assiette#1
- bol#2 de X = quantité de X contenue dans un bol#1

2.2 Définition de la polysémie régulière

Selon Apresjan (1974), une polysémie est régulière s'il existe au moins deux vocables (A et B) ayant chacun deux lexies (A#1~A#2 et B#1~B#2) liées par la même relation sémantique. Les lexies A#1 et B#1 ne doivent pas être synonymes, pas plus que les lexies A#2 et B#2. Illustrons cette notion de polysémie régulière avec des données extraites de WordNet :

- {cerise#1, **cherry**#4, cherry red#1} = *the red color of* ***cherries***
- {**chestnut**#4} = *the brown color of* ***chestnuts***

Les vocables CHERRY et CHESTNUT présentent la même alternance sémantique entre un `fruit` et une `couleur` (respectivement le rouge des cerises et le marron des châtaignes). On peut donc d'ores et déjà dire que ce lien est régulier et chercher d'autres occurrences dans WordNet, afin de déterminer son degré de régularité (Barque, 2008). Bien entendu, ce dernier dépendra du degré de spécificité de la caractérisation sémantique du lien. Par exemple, le lien entre un `fruit` et une `couleur` sera associé à moins d'occurrences que le lien entre une `entité` et une `couleur`, le type `entité` étant plus général que le type `fruit`.

Une chose est de déterminer le degré de régularité d'un lien de polysémie, une autre de déterminer à quelle catégorie il appartient. Nous distinguons, pour la suite de notre étude, trois grandes catégories de liens de polysémie régulière (Fass, 1988) :

1. La **spécialisation** : une lexie L2 est une spécialisation d'une lexie L1 si son sens est plus spécifique que celui de L1. Ci-dessous, on peut voir que la lexie pressure#7 dénote un cas particulier de ce à quoi renvoie « pressure » dans la définition.
   - {**pressure**#7} = *the* ***pressure*** *exerted by the atmosphere*
2. La **métaphore** : deux lexies L1 et L2 sont liées par métaphore si le référent de L1 et celui de L2 sont en relation d'*analogie*, autrement dit s'ils se ressemblent sur au moins un de leurs aspects. Par exemple, le rire dénoté par la lexie cackle#3 ressemble, du point de vue du son, au gloussement de la poule (*"hen's cackle*").
   - {**cackle**#3} = *a loud laugh suggestive of a hen's* ***cackle***

3. La **métonymie** : deux lexies L1 et L2 sont liées par métonymie si le référent de L1 et celui de L2 sont en relation de *contiguïté*, autrement dit si les deux référents « se touchent », de façon plus ou moins concrète. Par exemple, le lien entre les deux sens de CHESNUT, déjà évoqué plus haut, relève de la métonymie puisque la couleur dénotée par la lexie chesnut#4 est celle du fruit dénoté par *"chesnuts*" dans la définition.
   - {**chesnut**#4} = *the brown color of* ***chestnuts***

3 Objectif et méthode

L'objectif principal de notre étude est de décrire les liens de polysémie réguliers de WordNet et de mesurer leur régularité en détectant automatiquement leurs occurrences dans la base lexicale. L'un des enrichissements possibles du lexique sera ainsi la création de nouvelles relations sémantiques, en l'occurrence des relations de métaphore et de métonymie[2].

3.1 Description des règles

Notre objectif étant de rendre compte de la polysémie **régulière** représentée dans WordNet, nous avons pris le parti de nous intéresser, dans un premier temps, aux seuls synsets dont la définition contient une lexie (L1) appartenant au même vocable que l'une des lexies du synset défini (L2) ; ce parti pris repose sur l'idée communément admise qu'un lien de sens entre deux lexies de même forme est d'autant plus évident que l'une est définie au moyen de l'autre : {…, **L2**, …} = ... ***L1*** ... En voici deux exemples :

- {**driver**#3} = *a golfer who hits the golf ball with a* ***driver***
- {**falsify**#4} = ***falsify*** *knowingly*

Dans le premier exemple, la lexie **driver**#3 est définie au moyen d'une autre lexie du vocable DRIVER. On rappellera qu'à ce stade, cette dernière n'est pas identifiée, les éléments utilisés dans les définitions de WordNet n'étant pas désambiguïsés[3]. Sur le plan informatique, nous avons procédé à l'étiquetage morphosyntaxique des définitions de tous les synsets[4] pour les filtrer et retenir un premier ensemble de 1984 synsets où L1 et L2 appartiennent à la même partie du discours[5].

---

[2] Dans WordNet, les relations de spécialisation sont déjà présentes pour les noms et verbes. Par exemple, pressure#7 a explicitement pour hyperonyme pressure#1.

[3] Notons l'existence du projet eXtended WordNet, mené en 2003 à l'Université de Dallas. Ce projet enrichit WordNet 2.0 avec une représentation logique et une analyse syntaxique de la définition de chaque *synset*. Après l'analyse syntaxique, les mots d'une définition font l'objet d'une désambiguïsation lexicale, avec une qualité *gold*, *silver* ou *bronze* (*gold* indiquant une validation manuelle). Or, pour un nombre total de 1 127 695 mots utilisés dans les définitions, seuls 14 657 mots sont de qualité *gold* (1,3%). Ce ratio nous semblant faible, nous n'avons pas utilisé cette ressource dans le cadre de notre étude.

[4] Nous avons utilisé pour cela la plateforme de traitement linguistique Antelope (www.proxem.com).

[5] En toute rigueur, nous avons imposé des contraintes supplémentaires. Nous avons éliminé les synsets où L1 désigne en fait L2. Cela correspond aux cas où la définition contient *"equal*" (comme dans {kopek, kopeck, copeck} = *100 kopecks equal 1 ruble in Russia*), *"trademark*" ou *"trade name*" (L1 et L2 représentent alors un nom commercial, par exemple {sildenafil, sildenafil citrate, Viagra} = *virility drug, trade name Viagra*), et "*capital of*" (comme dans {Bern, Berne, capital of Switzerland} = *the capital of Switzerland*).

La méthode adoptée pour attribuer à une occurrence L1~L2 une catégorie de liens de polysémie repose sur différents critères formels appliqués aux définitions de WordNet (Martin, 1972 ; Fass, 1988). Nous avons regardé, tout d'abord, si l'inclusion de L1 figurait dans la première partie de la définition de L2 (*i.e.* en tant que genre prochain) ou bien dans sa seconde partie (*i.e.* en tant que différence spécifique), comme illustré ci-dessous :

- {**behave**#3} = ***behave*** *well or properly*
- {**sweep**#6} = *clean by* ***sweeping***

Dans le premier exemple, L1 apparaît dans la première partie de la définition de L2 : le troisième sens du vocable TO BEHAVE est défini au moyen d'un autre sens du même vocable, qui en constitue le genre prochain (« *se comporter* » signifie, dans un de ses sens, « *se comporter d'une certaine manière* », en l'occurrence d'une manière appropriée). Dans le second exemple, L1 apparaît dans la seconde partie de la définition de L2 : le sixième sens du verbe TO SWEEP ne veut pas dire « *balayer d'une certaine manière* » mais « *faire quelque chose en balayant* », en l'occurrence nettoyer en balayant.

Outre la place de l'inclusion de L1 dans la définition de L2, nous retenons des sous-chaînes récurrentes dans les définitions. Voici trois exemples d'inclusion dans la seconde partie de la définition, distingués selon certains éléments pertinents de leur définition :

- {**mint**#5} = *a candy that is flavored with a* ***mint*** *oil* → "that is flavored with L1"
- {**bluefish**#2} = *fatty bluish flesh of* ***bluefish*** → "flesh of L1"
- {**fin**#5} = *a stabilizer on a ship that resembles the* ***fin*** *of a fish* → "that resembles L1"

On peut ainsi, en mêlant ces deux critères (la place de l'inclusion et les éléments définitionnels qui entourent cette inclusion), attribuer de manière automatique une catégorie de lien de polysémie à une occurrence donnée. De manière informelle, disons que si l'inclusion a lieu dans la première partie de la définition, il s'agit soit d'une spécialisation, soit d'une métaphore. Les deux exemples ci-dessous montrent en effet deux cas d'inclusion de L1 dans la première partie de la définition ; mais le premier exemple relève de la spécialisation, tandis que le second relève de la métaphore.

- {**arrange**#5} = ***arrange*** *attractively*
- {**grow**#9} = ***grow*** *emotionally*

Si l'inclusion a lieu dans la seconde partie de la définition, il s'agit soit d'une métonymie, soit d'une métaphore, selon le type d'élément qui introduit l'inclusion. Parmi les trois exemples déjà présentés plus haut, les deux premiers ({**mint**#5} et {**bluefish**#2}) sont des cas de métonymie, le troisième ({**fin**#5}) est un cas de métaphore. Ici, l'ambiguïté sur la catégorie d'appartenance d'un lien donné peut être levée grâce aux canevas définitionnels. Par exemple, si l'inclusion est précédée de la séquence *"that resembles*" comme c'est le cas pour le vocable FIN (*nageoire*) ci-dessus, on sait qu'il s'agit d'un cas de métaphore et non de métonymie.

3.2 Recherche d'occurrences

Notre méthode de détection des liens de polysémie régulière s'applique, dans un premier temps, à l'ensemble des 1984 synsets dont la définition inclut un synset de même forme. Nous

avons créé manuellement une soixantaine de patrons, en analysant des définitions de ces synsets qui correspondaient manifestement à un cas de polysémie régulière.

L'application de ces patrons a permis d'obtenir un premier classement de 1427 synsets[6]. Nous allons détailler ce processus en montrant notamment comment il permet de désambiguïser L1. Nous verrons ensuite comment cette méthode peut se généraliser aux autres synsets ne présentant pas la particularité d'inclure un synset de même forme dans leur définition.

3.2.1 Définition de patrons de polysémie régulière

Voici quelques lignes de code[7] définissant un patron appelé `colorOf` (« couleur de ») :

```
patterns.Add(new Pattern("colorOf")
  .AddType("color", "fruit")
  .AddType("color", "gem")
  .AddType("color", "metal")
  .AddMatchingRule("color of *"));
```

La première ligne de code définit le patron de polysémie. Les trois lignes suivantes indiquent que les paires de lexies susceptibles d'instancier ce patron sont de type `color` pour L2 et `fruit`, `gem` ou `metal` pour L1. Enfin, la dernière ligne indique que la définition de L2 doit, pour être déclarée occurrence du patron, contenir la chaîne de caractères *"color of"* suivie de L1 (indicateur `*` à droite)[8].

Considérons un autre des quelques soixante patrons de polysémie produits par notre étude :

```
patterns.Add(new Pattern("causedBy")
  .AddMatchingRule("resulting from *")
  .AddMatchingRule("caused by *"));
```

Ce patron, contrairement au précédent, n'impose pas de contrainte sur le type des lexies susceptibles de l'instancier mais nécessite que L2 inclue dans sa définition une des deux séquences indiquées ("*resulting from*" ou "*caused by*"). On voit ici que la seule utilisation de la hiérarchie des concepts conduirait à exclure un certain nombre de patrons de polysémie régulière, et donc à diminuer le rappel de notre méthode.

3.2.2 Application des patrons

Les exemples ci-après sont reconnus comme des occurrences du patron `colorOf` :

- {**emerald**#3} = *the green color of an* ***emerald***
- {tan#2, **topaz**#3} = *a light brown, the color of* ***topaz***
- {**copper**#4} = *a reddish-brown color resembling the color of polished* ***copper***

[6] Les autres 557 synsets correspondent soit à des situations n'étant pas un cas de polysémie régulière, soit à des situations où la définition d'un patron de polysémie ne permettrait de couvrir qu'un faible nombre de cas.

[7] Cet exemple est codé en C#, langage proche de Java.

[8] Plusieurs paraphrases peuvent être précisées ; par exemple, la détection des métaphores testera plusieurs cas de figure : *"suggestive of *", "similar to *", "corresponds to *", "that suggests *"*, "*imitating **"…

Grâce aux informations de typage associées au patron, la lexie L1 peut être ensuite désambiguïsée. Pour ce faire, le système énumère tous les sens possible de L1 (autre que L2), et s'arrête quand le couple (L2, L1) est compatible avec l'un des couples de types définis dans le patron (la hiérarchie d'hyperonymie des noms ou des verbes est explorée si besoin).

On obtient alors une définition dans laquelle le sens de L1 est désambiguïsé :

- {**emerald**#3} = *the green color of an* ***emerald****#1* [gem]
- {tan#2, **topaz**#3} = *a light brown, the color of* ***topaz****#2* [gem]
- {**copper**#4} = *a reddish-brown color resembling the color of polished* ***copper****#1* [metal]

3.2.3 Généralisation par élargissement du champ d'application des patrons

Les patrons sont ensuite appliqués à l'ensemble des lexies ayant plusieurs sens, sans imposer aux synsets la contrainte d'inclure un synset de même forme dans leur définition. Cette étape permet d'identifier 367 synsets supplémentaires. Par exemple, *"gold"* (« or ») a cinq sens : les pièces d'or, la couleur, le métal, une bonne santé, quelque chose de précieux. La paire de lexies (gold#2, gold#3) est de type (`color, metal`) ; elle est donc compatible avec la règle `colorOf`, et peut donc s'appliquer même si la définition de {gold#2} (*"a deep yellow color*") ne contient pas directement le mot "*gold*". Le même traitement est appliqué pour {coral#1} (« corail »). On obtient alors aussi :

- {amber#1, **gold**#2} = *a deep yellow color* (lien implicite vers gold#3$_{[metal]}$)
- {**coral**#1} = *a variable color averaging a deep pink* (lien implicite vers coral#2$_{[gem]}$)

Cette généralisation de l'application des patrons doit toutefois se faire en prenant des précautions. On constate expérimentalement que cette généralisation donne de bons résultats sur certains patrons, mais pas sur tous. En effet, quand les patrons sont contraints par des types trop généraux (entité, artefact, abstraction…), le fait de ne plus imposer L1 dans la définition de L2 va se traduire par une multiplication de couples de synsets qui ne sont pas liés par une relation de polysémie régulière. Pour minimiser ce risque, un patron peut tester si les deux synsets portent effectivement sur le même sujet, ou des sujets voisins. Cette vérification est implémentée par une classique mesure de similarité entre les deux définitions, qui peut être astreinte à respecter un seuil minimal. Nous avons utilisé une mesure vectorielle de recouvrement des mots entre définitions, avec une pondération de type TF-IDF[9]. Par exemple, parmi les occurrences de la relation entre un mouvement et le son associé, nous obtenons pour « *(bruit de) pas* » une paire de synsets dont les définitions comportent deux mots en commun (donnant une similarité égale à 48,5%) :

- {footstep#1} = *the sound of a* ***step*** *of someone* ***walking***
- {footstep#2} = *the act of taking a* ***step*** *in* ***walking***

Imposer une telle contrainte, avec un seuil minimal, favorise la précision au détriment du rappel. Par exemple, pour la métaphore entre animal et personne, notre système identifie

[9] TF-IDF (*term frequency-inverse document frequency*) est une méthode de pondération souvent utilisée dans la fouille de textes. Cette mesure statistique nous permet ici d'évaluer l'importance d'un mot au sein d'une définition. Le poids augmente proportionnellement en fonction du nombre d'occurrences du mot dans la définition. Il varie également en fonction de la fréquence du mot dans l'ensemble des définitions.

« *tigre* » (et il existe bien une métaphore entre les deux lexies, basée sur la férocité de l'animal) ; néanmoins, il ne le retient pas car les deux définitions ne partagent aucun mot :

- {tiger#1} = *a fierce or audacious person*
- {tiger#2} = *large feline of forests in Asia having a tawny coat with black stripes*

4 Classification (non exhaustive) des relations de polysémie

Nous proposons ici une classification des relations de métonymie et de métaphore[10] sur la base des patrons identifiés pendant l'étude. Dans cette classification, nous indiquons entre parenthèses deux nombres (occurrences correctes / nombre total d'occurrences détectées), suivis d'exemples significatifs choisis pour illustrer le caractère régulier de la relation.

4.1 Classification des relations de métonymie

L2 *représente* L1
└─ Carte à jouer *représente* Figure ou Nombre entier (5/6 ; queen#7, king#9 ; ten#2, nine#3)

L2 *est causé par* L1
├─ Dépense *causée par* Action (27/27 ; admission#3, anchorage#2)
└─ Maladie *causée par* Organisme (13/17 ; ergot#1, herpes#1)

L2 *est produit par* L1
├─ Son *produit par* Instrument, Mouvement ou Appareil (15/15 ; drum#2, whistle#1 ; snap#5 ; bell#3)
├─ Œuvre *écrite par* Personne
│ ├─ Livre *écrit par* Auteur (pas d'exemple dans WordNet ; ce pourrait être "Shakespeare")
│ │ └─ Livre *écrit par* Prophète (15/15 ; Job#12, Jeremiah#2)
│ └─ Musique *écrite par* Compositeur (9/9 ; Mozart#2, Wagner#3)
└─ L2 *produit par* Plante ou Arbre
   ├─ Fruit *produit par* Arbre (orange#1, citrus#1)
   └─ Fleur *produite par* Plante (50/51 ; chrysanthemum#1, cottonweed#1)

L2 *produit* L1
└─ Entreprise *produit* Media (2/2 ; newspaper#2, magazine#3)

L2 *est dérivé de* L1
├─ L2 *est dérivé d'*Animal
│ ├─ Chair *d'*Animal, Poisson, Volaille ou Crustacé (303/303 ; rabbit#3 ; trout#1 ; pheasant#2 ; lobster#1)
│ ├─ Fourrure *d'*Animal (17/17 ; fox#3, chinchilla#1)
│ └─ Laine *d'*Animal (2/2 ; alpaca#1, vicuna#1)
└─ L2 *est dérivé de* Plante, Feuille, Arbre…
   ├─ Boisson *dérivée de* Feuille (3/3 ; tea#1, mate#9)
   ├─ Fibre *dérivée de* Plante (13/13 ; cotton#1, flax#1)
   ├─ Bois *dérivé d'*Arbre (70/70 ; bamboo#1, balsa#1)
   └─ Vin *dérivé de* Vigne (2/2 ; Tokay#1, Verdicchio#2)

L2 *a pour sujet* L1
├─ Discipline *a pour* Sujet (56/64 ; literature#2, physics#1)
├─ L2 *est responsable de* L1 (4/6)
│ ├─ Ministère *est responsable de* Sujet (Education#6, Energy#7)
│ └─ Division *est responsable de* Sujet (personnel#2, security#6)
└─ Livre *a pour sujet* Personne (6/6 ; Jonah#3, Joshua#2)

L2 *accompagne* L1
└─ Musique *accompagne* Danse (32/32 ; polka#1, mazurka#1)

[10] Les résultats contiennent également une proportion significative d'occurrences de liens de spécialisation (approximativement 12%). Toutefois, il nous semble que cette catégorie de lien de polysémie se prête mal à une classification, dans la mesure où il est difficile d'identifier un typage régulier pour L1 et L2.

L2 *est inclus dans* L1
- Substance *contenue dans* Medicament (17/17 ; arnica#2, menthol#1)
- Personne *membre de* Groupe (37/39 ; samurai#1, ninja#1)
- Personne *occupant une* Construction (6/6 ; building#4, floor#7)
- Quantité *contenue dans* Conteneur (39/39 ; teaspoon#1, bag#5)
  - Nourriture *contenue dans* Plat (5/5 ; plate#8, casserole#1)
- Rivière *passant dans* Région (6/6 ; Alabama#3, Delaware#1)
- Pays *situé dans* Île (22/22 ; Ireland#1, Malta#1)

L2 *est caractérisé par* L1
- Balle *caractérisé par* Jeu (9/10 ; paintball#1, volleyball#2)
- Boisson *provenant de* Région (4/4 ; Chablis#2, Bordeaux#2)
- Couleur *caractéristique de* L1 (7/7)
  - Couleur *caractéristique de* Gemme (topaz#3, emerald#3)
  - Couleur *caractéristique de* Métal (gold#2, copper#4)
  - Couleur *caractéristique de* Fruit (cherry#4, chestnut#4)
- Nourriture *au goût de* L1 (13/25)
  - Nourriture *au goût d'*Herbe (mint#5, ratafia#2)
- Vêtement *caractérisé par* Partie du corps (12/14 ; back#7, shoulder#4)
- Personne *caractérisée par* L1
  - Sportif *caractérisé par* Position (31/31 ; center#13, wingback#1)
  - Chanteur *caractérisé par* Voix (11/11 ; contralto#1, soprano#1)

Langue *parlée par* Personne (199/223 ; Korean#2, Portuguese#1)

4.2 Classification des relations de métaphore

L2 *est analogue à* L1
- Communication humaine *est analogue à* Communication animale (3/4 ; bark#1, cackle#1)
- Partie du corps animal *correspond à* Partie du corps humain (3/3 ; leg#2, throat#4)
- Individu *ressemble, dans son comportement, à* Animal (36/54 ; piranha#1, popinjay#1 (*perroquet*))
- Objet *ressemble, par sa forme, à* Objet naturel (38/38 ; moon#2, snake#5)
  - Artefact *ressemble à* Partie du corps (5/5 ; nose#2, throat#3)

5 Evaluation des résultats

A notre connaissance, il n'existe pas de *gold standard* pour ce type d'expérience. A titre indicatif, nous avons évalué manuellement les 2351 relations proposées par notre système. Nous estimons que 2140 d'entre elles sont correctes, ce qui donne une précision de 91,03%.

Nous n'avons pas identifié de méthode permettant une évaluation automatique précise du rappel. Toutefois, nous avons calculé manuellement le rappel pour deux des patrons de polysémie présentés ci-dessus : la métaphore *Individu ressemble à Animal*, ainsi que la métonymie *Bois dérivé d'Arbre*. Nous avons identifié manuellement 142 occurrences du lien de métaphore dans WordNet (rappel de 36/142=25,3%) et 79 occurrences du lien de métonymie (rappel de 70/79=88,6%). Comme on le voit, le rappel dépend aussi de la nature de la relation, qui peut être plus ou moins régulière.

6 Conclusion et perspective

Nous avons présenté ici une méthode permettant de créer automatiquement dans WordNet, avec une bonne précision, deux nouvelles catégories de relations sémantiques : des métaphores et des métonymies. (Une ressource contenant ces nouvelles relations est

disponible en ligne[11].) Cette méthode peut aussi être mise à contribution dans une tâche de désambiguïsation lexicale, pour inférer des sens de mots n'existant pas un lexique.[12]

Ce travail a été effectué sur l'anglais ; il pourrait aussi être décliné pour des WordNet en d'autres langues (une fois, bien sûr, les patrons adaptés à la langue décrite) pour (*i*) aider à valider l'homogénéité des définitions produites et (*ii*) comparer les polysémies régulières partagées entre différentes langues.

Références

APRESJAN J. (1974). Regular Polysemy. *Linguistics* 142, 5-32.

BARQUE L. (2008). *Description et formalisation de la polysémie régulière du français*. Université Paris 7, Thèse de doctorat.

COPESTAKE A., BRISCOE T. (1995). Semi-productive polysemy and Sense Extension. *Journal of Semantics* 1, 15-67.

FASS D. (1988). Metonymy and Metaphor : What's the difference ?. *Coling-88*, 177-181.

FELLBAUM C. (1998). *WordNet, An Electronic Lexical Database*. Cambridge : MIT Press.

FELLBAUM C. (2000). Autotroponymy. In Ravin Y., LEACOCK C. (EDS.): *Polysemy*, pp. 52-67. Cambridge: Cambridge University Press.

KILGARRIFF A., GAZDAR G. (1995). Polysemous relations. In PALMER F.R. (ED.). *Grammar and Meaning: Essays in Honour of Sir John Lyons*, pp. 1-25. Cambridge University Press.

MARTIN R. (1972). Esquisse d'une analyse formelle de la polysémie. *Travaux de linguistique et de littérature* 10, 125-136.

OSTLER N., ATKINS B. (1991). Predictable Meaning Shift : Some Linguistic Properties of Lexical Implication Rules. In PUSTEJOVSKY J., BERGLER S. (ED.). *Lexical Semantics and Knowledge Representation : First SIGLEX Workshop Proceedings.* Berlin : Springer-Verlag.

PETERS W. (2006). In Search for More Knowledge: Regular Polysemy and Knowledge Acquisition. *Proceedings of GWC2006.*

PUSTEJOVSKY J. (1995). *The Generative Lexicon*. Cambridge : MIT Press.

VEALE T. (2006) A typology of Lexical Analogy in WordNet. *Proceedings of GWC2006.*

[11] http://www.chaumartin.fr/download/wnpolysemy.zip

[12] Par exemple, si {Bordeaux#1} (*a port city in southwestern France*) et {Bordeaux#2, Bordeaux wine#1} existent tous deux dans WordNet, « Bourgogne » n'y est présent qu'en tant que région. L'exploitation des informations présentes dans les patrons définis pour l'étude pourrait permettre de créer dynamiquement des nouveaux sens de mots, quand le contexte s'y prête ("*they convince you to drink Bourgogne*").