



HAL
open science

Transitivité partielle de la synonymie : application aux dictionnaires de synonymes

Jean-Luc Manguin

► **To cite this version:**

Jean-Luc Manguin. Transitivité partielle de la synonymie : application aux dictionnaires de synonymes. CORELA - COgnition, REprésentation, LAngage, 2004, 2 (2), [15 p.]. 10.4000/corela.611 . hal-00012325

HAL Id: hal-00012325

<https://hal.science/hal-00012325>

Submitted on 20 Oct 2005

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.

Transitivité partielle de la synonymie : application aux dictionnaires de synonymes

Jean-Luc MANGUIN
CRISCO – CNRS et Université de Caen

Résumé

A partir de la formalisation de la relation de synonymie par un graphe dont les sommets sont les unités lexicales, et les arêtes les liaisons synonymiques, nous proposons une méthode d'enrichissement et de contrôle du dictionnaire représenté par ce graphe. Notre procédé est basé sur les indices de similitude entre sommets du graphe ; suivant certains critères que nous définirons, il permet de relier directement à des sommets certains de leurs synonymes d'ordre 2. Nous testerons ensuite la procédure sur un dictionnaire réel (celui de René Bailly) dans sa forme électronique afin d'examiner la pertinence des résultats obtenus. Enfin nous appliquerons la méthode à un dictionnaire plus vaste et plus consensuel, résultant d'une fusion de dictionnaires existants, avant de conclure.

Abstract

As a starting point, we assume that our model of the relation of synonymy is an undirected graph, in which the nodes are lexical units, and the edges are synonymic links ; then we propose an automatic method for enriching or checking the dictionary figured by this graph. Our method is based on similarity indexes between the nodes of the graph ; according to some criteria that will be defined, our process can add direct links from a node to some of its 2-order synonyms. We will test this process on a real dictionary (René Bailly's) in its electronic form, and we will examine the quality of the results. Finally we will apply the process to a larger dictionary (made by merging some real dictionaries) before we conclude.

Mots-clefs

Synonymie, dictionnaire

0. Introduction

Dans le domaine du traitement automatique des langues, de nombreux travaux appliqués prennent en compte l'aspect sémantique au moyen de ressources lexicales importantes (dictionnaires, ontologies, etc.). L'essor de ces travaux accroît la demande de telles ressources, et en stimule la production (parfois automatique) ainsi que les recherches sur la structure même de ces ressources lexicales. D'une manière générale, ce type de matériel est une structure algébrique appliquée à un ensemble assez large d'unités lexicales ; en particulier, une des structures les plus utilisées consiste en un graphe dont les sommets sont ces unités, et où les arcs ou arêtes représentent les relations sémantiques¹. L'introduction de ce modèle semble revenir à l'école scandinave, que ce soit chez Brodda et Karlgren (1969) ou chez Kahlmann (1975), dans le domaine des graphes de synonymie ; plus près de nous, on trouve entre autres chez Véronis (2002) des relations syntagmatiques, chez Bourigault (2002) des relations basées sur des critères distributionnels, chez Gaume (2002) une utilisation conjointe de l'axe paradigmatique et syntagmatique, et enfin, de Ploux et Victorri (1998) à Jacquet (2003) ou Venant (2004) une exploitation de plus en plus subtile des relations paradigmatiques d'un dictionnaire des synonymes construit pour la circonstance.

¹ On parle d'arc lorsque la relation est orientée, et d'arête si elle ne l'est pas (cf. Berge)

1. La relation de synonymie et ses propriétés

Dans sa thèse, André Kahlmann (op. cit.) aborde ce problème en discutant notamment de la symétrie de la relation². De notre côté, nous reprendrons les propriétés de la relation de synonymie dans l'ordre classique de l'algèbre, à savoir réflexivité, symétrie et transitivité.

En préambule à cette analyse, il est intéressant de rappeler la définition de la synonymie selon Kahlmann (op. cit., p. 13) :

« *Un synonyme est un mot qui peut en remplacer un autre dans au moins un contexte sans changer la portée du message* »,

qui est à rapprocher de celle de S. Ploux et B. Victorri (1998) qui exige plus qu'un seul contexte :

« *Deux unités lexicales sont en relation de synonymie si toute occurrence de l'une peut être remplacée par une occurrence de l'autre dans un certain nombre d'environnements sans modifier notablement le sens de l'énoncé dans lequel elle se trouve.* ».

A partir de cela, nous pouvons affirmer sans risque que la relation de synonymie est réflexive, puisque toute unité lexicale peut se substituer à elle-même ; ce point, qui semble anodin, est, comme nous le verrons par la suite, important pour les calculs que nous effectuerons à partir des relations présentes dans le graphe.

La question de la symétrie est un peu plus délicate ; dans un premier temps, Kahlmann (op. cit.) considère que le lexicographe a placé dans son ouvrage des relations orientées, et c'est ce qu'il fait figurer dans son modèle, en justifiant cela par la présence de renvois. Cependant, par une étude du « graphe inversé » (c'est-à-dire obtenu en inversant les arcs d'origine), il remarque certaines anomalies, engendrées entre autres par ce système de renvois, et par l'orientation du graphe. Il en arrive ainsi à considérer que la symétrisation améliore sensiblement la qualité du dictionnaire, sans pour autant engendrer de relations aberrantes. De même, lors de la construction du dictionnaire des synonymes du CRISCO, la structure de graphe non-orienté s'est imposée pour plusieurs raisons (voir Ploux, 1997), et a notamment permis de donner naissance à des travaux sur la structure même du graphe (Venant, op. cit.). En résumé, la symétrie de la relation synonymique est globalement plus satisfaisante, tout en restant compatible avec la définition que nous avons citée.

La troisième propriété classique est la transitivité, qui consisterait à dire « si A est synonyme de B, et B est synonyme de C, alors A est synonyme de C » . Il est facile de trouver des exemples, mais encore plus facile de trouver des contre-exemples, en raison de la polysémie (voire de la quasi-homonymie) des unités du lexique : *fenêtre* est synonyme de *baie*, *baie* est synonyme de *golfe*, mais *golfe* et *fenêtre* ne sont jamais substituables. Cette question des synonymes d'ordre 2 a aussi été étudiée par Ploux (1997), dans une perspective d'obtention de termes génériques, ainsi que par Brodda et Karlgren pour une résolution théorique et appliquée de la « synonymie croisée », c'est-à-dire la recherche des synonymes les plus appropriés pour remplacer deux termes conceptuellement voisins. Nous allons maintenant nous pencher sur ce problème, cela en raison des « lacunes » présentes dans les dictionnaires de synonymes ; en cherchant un moyen de les combler, nous allons définir une « fermeture partielle » du graphe de synonymie qui aboutit à une amélioration du dictionnaire existant ; cette fermeture partielle est très proche du concept de fermeture transitive de la théorie des graphes ; nous montrerons, en outre, que cette méthode s'avère être un moyen de contrôle des relations déjà présentes dans la ressource lexicale étudiée.

² Cette question fait l'objet du chapitre 4 de sa thèse.

2. Exemple de lacune dans les dictionnaires

Comme nous venons de l'annoncer, nous ne parlons ici que des lacunes dans le corps d'un article de dictionnaire ; il est en effet toujours possible de trouver des entrées « absentes » d'un dictionnaire à l'autre. Dans le cas qui nous préoccupe, qui est celui des dictionnaires de synonymes, les lacunes se traduisent par l'absence flagrante d'un synonyme dans la liste des substituts possibles d'un mot-vedette. Dans le dictionnaire des synonymes du CRISCO, que nous avons construit en fusionnant les versions symétrisées de sept dictionnaires³, un exemple nous est fourni avec la vedette *curieux*, dans son sens adjectival ; en effet, cinq dictionnaires⁴ ne mentionnent pas la synonymie avec *insolite*, qui est pourtant attestée par ailleurs dans le Trésor de la Langue Française, dont voici un extrait de l'article « curieux » :

II. Au sens passif. [Concerne des pers. ou des choses] Qui suscite un intérêt particulier.

A. Vieilli. Digne d'intérêt en raison de sa valeur. Synon. *intéressant*. *Maigre collection qui, je pense, n'est pas curieuse pour un savant* (FLAUB., *Champs et grèves*, 1848, p. 203).

B. Qui suscite l'intérêt par sa nouveauté ou son étrangeté. Synon. *insolite*, *singulier*, *bizarre*. *Un travail curieux et certainement non dépourvu d'intérêt scientifique* (RUYER, *Esq. philos. struct.*, 1930, p. 333). *Naturellement il avait couru d'abord aux animaux les plus curieux de la ménagerie, tapirs et tatous, surréalistes et néothomistes* (BLOCH, *Dest. du S.*, 1931, p. 169).

Les 54 synonymes de *curieux*, obtenus en cumulant les cinq dictionnaires sont :

amateur, amusant, anxieux, attachant, attentif, avide, badaud, beau, bizarre, bizarroïde, chercheur, collectionneur, dilettante, drolatique, drôle, déconcertant, désireux, extraordinaire, extravagant, flâneur, fouille-au-pot, fouilleur, fouinard, fouineur, fureteur, incompréhensible, inconcevable, incroyable, indiscret, inouï, inquisiteur, intéressant, intéressé, inusité, investigateur, juge, original, paradoxal, piquant, pittoresque, plaisant, rare, regardant, sensationnel, singulier, soigneux, soucieux, spectateur, surprenant, truculent, unique, voyeur, étonnant, étrange.

Par ailleurs, *insolite* possède 26 synonymes selon les mêmes sources :

abracadabrant, anormal, baroque, bizarre, bizarroïde, drôle, déroutant, excentrique, extraordinaire, extravagant, inaccoutumé, incohérent, inhabituel, inusité, irrégulier, nouveau, nouvel, original, pittoresque, rare, saugrenu, sensationnel, singulier, surprenant, étonnant, étrange.

Parmi ceux-ci, 14 sont aussi synonymes de *curieux* :

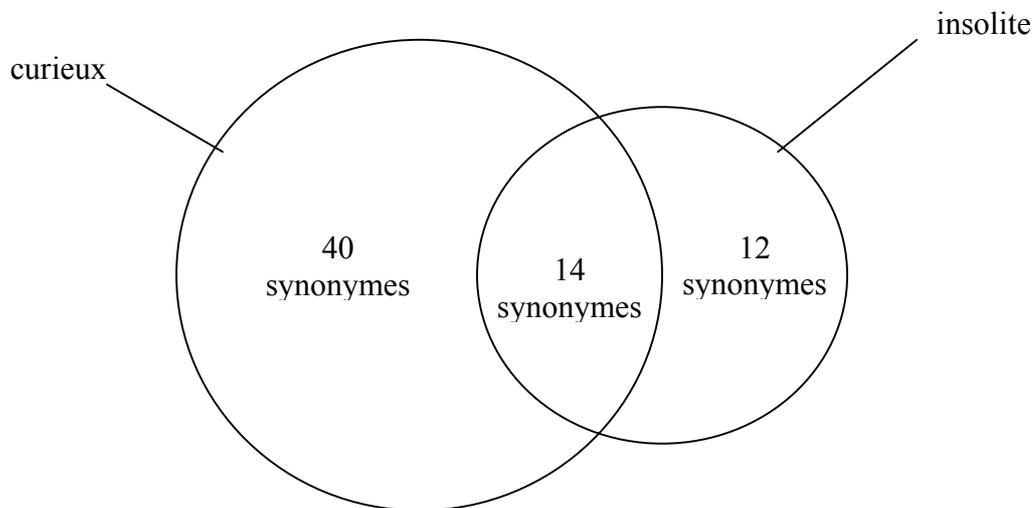
bizarre, bizarroïde, drôle, extraordinaire, extravagant, inusité, original, pittoresque, rare, sensationnel, singulier, surprenant, étonnant, étrange.

Autrement dit, plus de la moitié des synonymes d'*insolite* sont des synonymes directs de *curieux* ; on voit ainsi que l'adjectif *insolite*, bien qu'étant synonyme d'ordre 2 de *curieux*, est lié à ce dernier d'une manière assez forte, et l'on peut pour ce cas reprendre un diagramme de synonymie parmi ceux proposés au milieu du 19^{ème} siècle par P.B. Lafaye (dans sa

³ Il s'agit des dictionnaires de Guizot, Lafaye, Bailly, Bénac, Du Chazaud, et des liens synonymiques relevés dans le Grand Larousse de la Langue Française, et dans le Grand Robert de la Langue Française. Pour la construction du dictionnaire du CRISCO, voir Ploux et Victorri (1998).

⁴ Lafaye et Guizot n'ont pas retenu *curieux* parmi la liste des vedettes.

monumentale introduction⁵ à son dictionnaire déjà cité), que nous avons complété avec nos données :



Si un dictionnaire de synonymes présente des lacunes, il est probable qu'elles seront révélées par la même propriété que celle qui lie *insolite* à *curieux*, c'est-à-dire que les deux mots qui devraient être des synonymes directs sont synonymes d'ordre 2, et que la moitié (ou plus) des synonymes de l'un sont synonymes directs de l'autre.

3. Les indices de similitude

3.1 Principe de calcul

Pour aller plus loin, nous allons quantifier la relation qui lie un mot à un de ses synonymes d'ordre 2 (la quantification présentée peut aussi s'appliquer aux synonymes directs). Pour cela, nous ferons appel à une mesure bien connue : les indices de Jaccard, aussi appelés indices de communauté. Si l'on considère les deux ensembles qui correspondent aux synonymes respectifs des deux mots concernés, cet indice peut se définir comme le quotient du cardinal de l'intersection des deux ensembles, par le cardinal de l'union de ces deux mêmes ensembles. Si l'on nomme :

A et B les deux mots étudiés

SynA (resp. SynB) l'ensemble des synonymes de A (resp. B)

la formule s'écrit⁶ :

$$S = \frac{\text{card} (\text{SynA} \cap \text{SynB})}{\text{card} (\text{SynA} \cup \text{SynB})}$$

Pour *curieux* et *insolite*, S vaut donc $14/66=0,212$.

Cependant, comme nous l'avons dit plus haut, ce qu'il y a de remarquable dans le cas de ces deux mots, c'est que plus de la moitié des synonymes de l'un sont synonymes de l'autre ;

⁵ Pages XXXIX à XLI.

⁶ On en déduit facilement les valeurs extrêmes de S : il vaut 0 si A et B n'ont rien en commun, et 1 si leurs ensembles de synonymes coïncident parfaitement.

nous sommes donc amenés à définir des indices S_A et S_B **relatifs**⁷ à chacun des deux sommets, selon les formules ci-dessous :

$$S_A = \frac{\text{card}(\text{Syn}A \cap \text{Syn}B)}{\text{card}(\text{Syn}A)}$$

et

$$S_B = \frac{\text{card}(\text{Syn}A \cap \text{Syn}B)}{\text{card}(\text{Syn}B)}$$

Et si nous faisons correspondre *curieux* à A et *insolite* à B, alors S_A vaut $14/54=0,259$, mais S_B vaut $14/26=0,538$.

3.2 Justification de la réflexivité

Le moment est maintenant venu de revenir sur la réflexivité de la relation de synonymie ; en effet, considérons le cas théorique suivant, où A et B ne possèdent qu'un seul synonyme :

A ————— B

Intuitivement, il va de soi que A et B sont parfaitement identiques ; toutefois, si nous calculons l'indice de Jaccard entre A et B, il est nul puisque A et B n'ont pas de synonyme commun : A a pour seul synonyme B, et inversement. Ce résultat absurde se résout si nous postulons que la relation de synonymie est réflexive : A a alors pour synonymes A et B, et il en est de même pour B ; l'indice de communauté vaut alors 1, ce qui est conforme à la logique du graphe.

Par conséquent, les valeurs précédemment trouvées avec *curieux* et *insolite* sont à rectifier, et deviennent $14/68$ pour S, $14/55$ pour S_A et $14/27$ pour S_B .

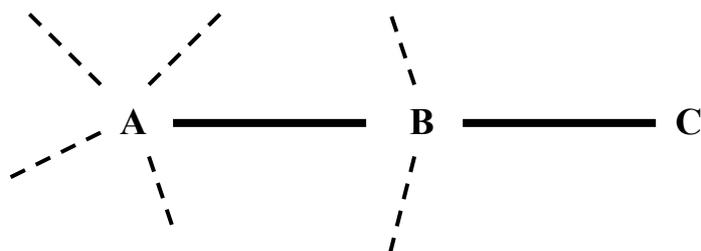
Cela dit, le cas de ces deux mots n'est pas unique dans notre dictionnaire, et des cas semblables se retrouvent dans les dictionnaires sources pris isolément. Aussi, nous nous proposons d'étudier ce phénomène de manière exhaustive dans un seul dictionnaire (symétrisé), et ensuite d'examiner en détail les liaisons que notre méthode a détectées comme « manquantes » ; mais auparavant, nous nous attarderons sur cette notion, pour déterminer quelles sont les conditions structurelles locales qui déterminent la possibilité d'ajouter une liaison.

4. Méthode proposée

4.1 Principe de détection automatique de liaisons manquantes

Comme nous l'avons vu dans l'exemple de *curieux* et *insolite*, le phénomène troublant est la présence d'un synonyme d'ordre 2 qui possède un indice de similitude relatif supérieur à 0,5 avec le mot-vedette. Cette valeur constitue un seuil en dessous duquel on passe à la transitivité directe ; en effet, considérons le schéma suivant :

⁷ Il est à noter que Bourigault (2002) définit lui aussi des indices de Jaccard relatifs selon une méthode quasi similaire.



Dans cette situation, C est un synonyme d'ordre 2 de A qui a pour seul synonyme direct B (et lui-même) ; son indice de similitude relatif avec A est donc $S_C = 1/2$, puisqu'il ne partage qu'un seul synonyme avec A, et qu'il ne possède que deux synonymes.

On se rend compte aisément que ce cas représente la situation minimale (1 seul synonyme), et que nous devons éviter de considérer qu'il manque ici une liaison entre A et C, sinon cela conduirait à des aberrations, comme celle de relier *golfe* à *fenêtre* dans l'exemple décrit plus haut. Cela implique qu'un mot qui pourra être relié à un autre devra posséder **au moins 2 synonymes** (en plus de lui-même), et que ceux-ci seront communs aux deux unités à relier.

Dans le schéma qui nous sert d'exemple, nous avons supposé que A est riche en synonymes, ce qui pose un problème si nous fixons le « seuil de liaison » à 0,5 ; il faudrait en effet que A partage plus de la moitié de ses synonymes avec C, ce qui est relativement rare dans le cas d'unités qui ont des synonymes nombreux. Dans l'exemple de *curieux* et *insolite*, nous avons considéré la liaison acceptable alors que seuls 25 % des synonymes de A (*curieux*) étaient partagés par C (*insolite*). Par conséquent, nous pouvons fixer la règle suivante : **deux sommets situés à une distance 2 dans le graphe de synonymie pourront être reliés si l'un au moins de leurs deux indices de similitude relatifs est strictement supérieur à 0,5**. Par la suite, nous verrons dans l'application pratique, que cette règle peut être assouplie en tenant compte du nombre de synonymes communs.

Nous allons maintenant détailler la procédure applicable à un dictionnaire complet.

4.2 Description de la procédure complète

La procédure que nous avons mise en œuvre utilise des programmes écrits en langage Perl, et traite un dictionnaire sous la forme d'un fichier texte ; ce fichier comprend une ligne par entrée, suivant la structure ci-dessous :

vedette : *syno1*, *syno2*, *syno3*, ..., *synoN*

où *syno1*, etc. sont les synonymes directs de *vedette*.

Les étapes successives sont les suivantes :

- Repérage, pour chaque entrée, des synonymes d'ordre 2 qui satisfont au(x) critère(s).
- Mise en fichier des liaisons à créer.
- Ajout des nouvelles liaisons au dictionnaire.

Le nouveau dictionnaire ainsi créé va subir à son tour le traitement décrit, et ainsi de suite, jusqu'à ce que le nombre de liaisons à ajouter soit nul⁸. Par ailleurs, on garde une trace du dictionnaire dans ses états successifs, ainsi que des fichiers qui contiennent les liaisons créées à chaque étape.

⁸ Du point de vue mathématique, on réalise ainsi une fermeture transitive *partielle* du graphe de synonymie.

5. Application pratique

5.1 Traitement du dictionnaire de René Bailly (1946)

Le dictionnaire des synonymes de René Bailly est un dictionnaire de type « distinctif », c'est-à-dire qu'il contient un nombre moyen d'entrées, et qu'il donne un nombre assez faible de synonymes, voire parfois un simple renvoi vers une autre entrée, mais que chaque article contient des commentaires sur les emplois des différents synonymes. L'Institut National de la Langue Française nous en a donné vers 1996 une version électronique résultant d'une saisie manuelle des relations synonymiques sans les commentaires ; après symétrisation des relations, nous arrivons à un dictionnaire qui contient 12698 entrées, et 28334 relations synonymiques, soit une moyenne relativement faible de 2,23 synonymes par entrée.

L'application réitérée de notre principe de fermeture partielle du graphe de synonymie avec le critère mentionné ci-dessus s'arrête à la 5^{ème} étape, laquelle n'ajoute aucune liaison ; le tableau ci-dessous indique le nombre de liaisons ajoutées à chaque étape :

Étape	Nombre de liaisons ajoutées
1	387
2	63
3	24
4	4
total	478

Il ne faut pas oublier que les liaisons sont symétriques, donc une liaison ajoutée représente 2 relations de synonymie ; nous avons donc ajouté au total 956 relations, faisant passer le nombre de relations dans tout le dictionnaire à 29290, soit une progression de 3,4 %, ce qui est relativement faible.

A titre de comparaison, le simple fait de tolérer un seuil « supérieur OU égal à 0,5 » au lieu de « strictement supérieur à 0,5 » pour la valeur de l'indice de similitude relatif provoquerait l'ajout de 47130 relations dès la première étape ; nous discuterons par la suite de l'importance de ce seuil en donnant quelques exemples de liaisons ajoutées lors de l'abaissement de ce seuil.

5.2 Pertinence des liaisons ajoutées

Il faut remarquer la qualité des relations synonymiques ajoutées par notre méthode ; en voici quelques-unes, parmi les 387 de la première étape :

abaisser, avilir
abrégé, concis
acariâtre, renfrogné
accepter, subir
accouchement, délivrance
accusateur, mouche⁹
accuser, signaler
acerbe, acrimonieux
affreux, horrible
aider, servir
ajouter, rallonger
ambigu, suspect
angoisse, affolement

⁹ La *mouche* prend ici son sens argotique de *mouchard*, ou d'*espion*.

antédiluvien, archaïque
apparition, fantasma
armée, légion
armée, régiment
assommer, estourbir
assommer, tanner
attrayant, captivant

A noter que les relations sont tout aussi pertinentes aux étapes suivantes ; voici celles qui sont rajoutées à la quatrième et dernière phase :

ballot, gourde
baluchon, gourde
détourner, s'emparer
heurter, taper

Il faut toutefois signaler l'apparition de quelques relations aberrantes, comme *femme, homme* qui est relativement difficile à étayer ; nous y reviendrons dans un paragraphe ultérieur.

5.3 Intérêt du seuil

Comme nous l'avons dit précédemment, le seuil de 0,5 correspond à celui de la « transitivité directe », autrement dit à l'apparition d'une liaison entre les sommets A et C s'il existe UN SEUL SOMMET B tel que A et B d'une part, et B et C d'autre part, sont reliés.

Dans le cas du dictionnaire de Bailly, voici un exemple des relations qui seraient ajoutées pour le mot *abrégé* :

abrégé, concis
abrégé, cursif
abrégé, laconique
abrégé, lapidaire
abrégé, mathématique
abrégé, montant
abrégé, succinct

On retrouve en première ligne celle qui est ajoutée quand le seuil n'admet pas la valeur limite de 0,5 ; pour les autres, nous pouvons les discuter à la lumière des synonymes proposés par le dictionnaire :

abrégé : *analyse, compendium, condensé, court, digest, extrait, notice, précis, raccourci, résumé, sommaire, somme, épitomé*
concis : *court, précis*
cursif : *court*
laconique : *court*
lapidaire : *court*
succinct : *court*
mathématique : *précis*
montant : *somme*

Nous voyons en effet que si la liaison avec *abrégé* demeure acceptable pour *cursif, laconique, lapidaire* ou *succinct*, à cause de leur relation avec *court*, en revanche elle devient fautive pour

mathématique car elle se base sur une variation de catégorie (un *abrégé* (Nom) est un *précis* (Nom), et quelque chose de *mathématique* (Adjectif) est souvent *précis* (Adjectif)), et enfin erronée pour *montant*, à cause de la quasi-homonymie de *somme*. On s'aperçoit ainsi du risque qu'il y a à accepter la transitivité directe lors de l'ajout de nouvelles relations, surtout si l'on a ôté du graphe les données de catégorisation.

5.4 Cas des relations inexactes

Si nous revenons aux résultats de la procédure normale, nous remarquons malgré tout la présence de liaisons discutables (voire carrément erronées) parmi les ajouts proposés ; par exemple, nous signalions le cas de *femme*, *homme* et celui de *filles*, *garçon*. Il est en effet étonnant de trouver des résultats aussi aberrants au milieu d'autres qui sont très satisfaisants ; l'examen du fichier symétrisé montre que *femme* et *homme* ont deux synonymes en commun, à savoir *époux* et *personne*. Une investigation plus poussée révèle que c'est le fichier d'origine qui est la source de ces erreurs, autrement dit que notre procédé met à jour des fautes d'interprétation commises lors de la saisie manuelle. Le fichier source indique en effet :

époux : *mari*, *femme*, *conjoint*, *seigneur et maître*, *homme*, *compagne*, *bourgeoise*, *ménagère*, *moitié*, *légitime*.

Un regard sur l'article d'origine du dictionnaire explique les erreurs commises à la réalisation typographique du fichier :

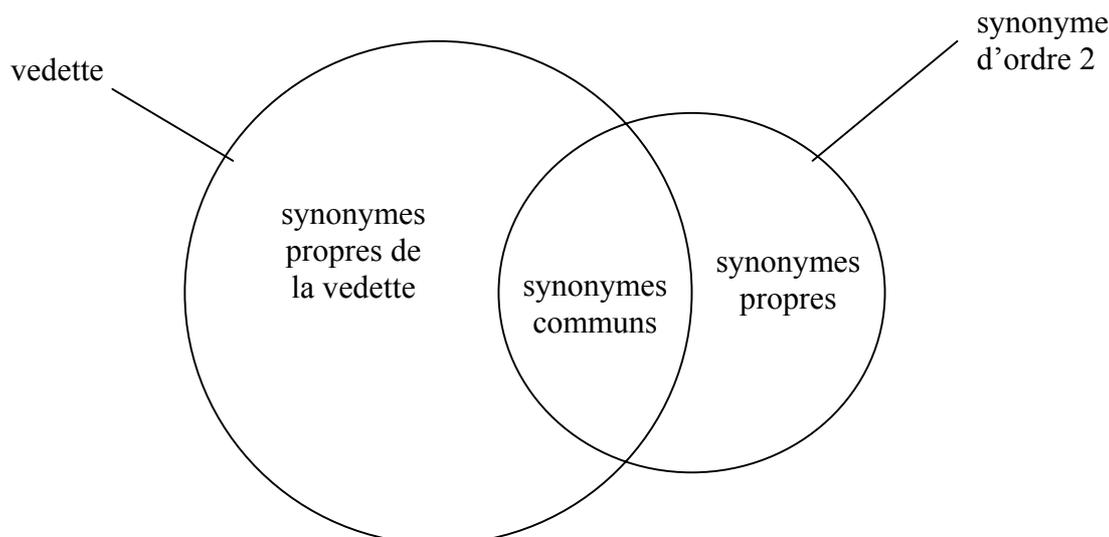
époux désigne l'homme conjoint par mariage, dans le style relevé, et a rapport aux liens d'affection, de caractère. (Dans le lang. ord., ce mot présente une idée d'emphase qui rend ridicule celui qui l'emploie.) **Mari** (ou FEMME, au fém.) est du style familier et concerne le lien légal ou l'union physique. **Conjoint** est un terme de droit qui s'applique à chacun des époux considéré par rapport à l'autre. **Seigneur et maître** se dit par plaisanterie d'un mari par rapport à sa femme. **Homme**, syn. de *mari*, est populaire. - **Compagne**, syn. *d'épouse*, est du style recherché. **Bourgeoise** est le nom que le mari donne parfois à sa femme, considérée comme la maîtresse de maison, dans les milieux populaires. **Ménagère**, syn. de *bourgeoise*, s'emploie plutôt par plaisanterie. **Moitié**, nom donné à la femme par rapport au mari, est très fam. **Légitime**, syn. de *mari* et surtout de *femme*, est populaire.

On se rend compte que ce type de dictionnaire suppose que le lecteur connaisse le féminin d'*époux*, et soit capable en outre d'attribuer les synonymes féminins de l'article à *épouse* ! Dans le cas d'une saisie rapide, cela conduit aux erreurs rencontrées.

Le cas de la relation entre *filles* et *garçon* est semblable et provient d'erreurs de saisie dans l'article *filles*. On s'aperçoit ici que, loin de dévaloriser notre méthode, ces faits ne font que lui donner un intérêt supplémentaire : celui de constituer un outil de contrôle ; ainsi, une relation proposée qui éveille les soupçons pourra conduire à une vérification fructueuse.

5.5 La question du nombre de synonymes

Si nous généralisons le diagramme de Venn tracé avec l'exemple de *curieux* et *insolite*, nous obtenons la figure ci-dessous, à partir de laquelle nous pouvons faire certaines remarques :



Comme nous l'avons dit, s'il n'y a qu'un seul synonyme commun, et que le synonyme d'ordre 2 ne possède pas de synonymes autres que lui-même, nous n'ajoutons pas de liaison, car cela équivaldrait à admettre la transitivité directe (ou totale). C'est pourquoi nous avons choisi la valeur 0,5 comme seuil du rapport entre « synonymes communs » et « synonymes du synonyme d'ordre 2 ». Cependant, cette clause apparaît finalement assez restrictive : on voit mal pourquoi l'on devrait refuser une liaison entre un synonyme d'ordre 2 qui posséderait 20 synonymes dont 10 seraient communs avec la vedette. Le but de notre méthode étant de récupérer des liaisons « manquantes », il semble plus juste de tenir compte du nombre de synonymes impliqués dans le calcul du rapport. C'est pourquoi nous proposons d'abaisser la valeur seuil du nombre de synonymes communs N_c en calculant, non pas le rapport N_c/N (si nous appelons N le nombre de synonymes du synonyme d'ordre 2) mais le rapport suivant :

$$S = \frac{N_c + 2 \log\left(\frac{N_c}{2}\right)}{N}$$

L'effet est le suivant : si N est supérieur à 6, la valeur seuil de N_c est diminuée de 1 ; si N est supérieur à 12, elle est diminuée de 2, etc. On peut ainsi favoriser la connexion d'unités plus riches en synonymes, et accroître encore l'efficacité de la récupération des liaisons. L'application du nouveau procédé au dictionnaire de Bailly se déroule comme indiqué dans le tableau suivant :

Etape	Nombre de liaisons ajoutées
1	399
2	67
3	33
4	6
5	7
6	3
7	4
8	1
total	520

Nous avons donc récupéré 42 liaisons supplémentaires, soit une augmentation de 9 % ; parmi ces nouvelles liaisons, certaines (marquées d'un astérisque dans le tableau ci-après) sont contestables, et sont dues, dans le cas de celles qui impliquent *pincer*, au système de « renvois » pratiqué dans certains dictionnaires et en particulier dans celui que nous avons soumis à notre traitement. Ce système de renvois, s'il répond à des impératifs d'économie, est pour le moins très discutable au point de vue linguistique, comme l'a montré Kahlmann (op. cit. chapitre 5).

<i>accepter, permettre</i>	<i>compagnie, corporation</i>	<i>idée, imagination</i>
<i>allonger, augmenter</i>	<i>conception, imagination</i>	<i>idée, savoir</i>
<i>*barboter, pincer</i>	<i>conserver, retenir</i>	<i>malice, vilenie</i>
<i>barboter, prendre</i>	<i>corporation, fédération</i>	<i>ordonnance, règle</i>
<i>battre, heurter</i>	<i>délié, mince</i>	<i>*pincer, s'approprier</i>
<i>cellule, emprisonnement</i>	<i>*dérober, pincer</i>	<i>pincer, s'emparer</i>
<i>*chauffer, pincer</i>	<i>dérober, prendre</i>	<i>*pincer, voler</i>
<i>chauffer, piquer</i>	<i>*détourner, pincer</i>	<i>piquer, prendre</i>
<i>*chiper, pincer</i>	<i>détourner, piquer</i>	<i>piquer, s'approprier</i>
<i>chiper, prendre</i>	<i>détourner, prendre</i>	<i>piquer, s'emparer</i>
<i>choper, pincer</i>	<i>enfantement, mise bas</i>	<i>prendre, voler</i>
<i>choper, piquer</i>	<i>*faucher, pincer</i>	<i>s'amouracher, s'engouer</i>
<i>cogner, tambouriner</i>	<i>faucher, prendre</i>	<i>*tambouriner, tamponner</i>
<i>*colleter, faucher</i>	<i>heurter, tambouriner</i>	<i>tambouriner, toquer</i>

Néanmoins, pour les trois quarts des relations nouvelles, la qualité est comparable à celle obtenue précédemment, et se retrouve jusqu'à la fin du processus, puisque la dernière relation ajoutée est « *prendre : voler* ».

5.6 Essai avec un autre dictionnaire

Nous avons prolongé l'application de notre méthode au dictionnaire des synonymes présent sur notre site, qui résulte de la fusion et de l'harmonisation de sept dictionnaires sources¹⁰ ; ce dictionnaire possède 49074 entrées reliées par 398988 relations, ce qui lui confère un degré moyen de 8,13 , valeur nettement supérieure à celle de 2,23 du dictionnaire de Bailly, qui s'explique par le fait que la fusion fait disparaître les effets pervers du système des renvois. En effet, les renvois ne sont pas les mêmes d'un dictionnaire à l'autre, et l'on remarque ainsi que le nombre d'entrées qui ne possèdent qu'un seul synonyme devient plus faible : pour le dictionnaire de Bailly, la proportion est de 63,7 %, tandis que pour le dictionnaire qui résulte de la simple fusion des sept dictionnaires sources, elle n'est que de 40,8 % (pour un degré moyen égal à 6,93) , et elle tombe à 29,7 % dans le cas du dictionnaire « harmonisé »¹¹. Pour ce dictionnaire « harmonisé », notre protocole propose 37384 nouvelles relations, soit une augmentation de 18,7 % du nombre de connexions existantes ; on voit par là que le degré moyen du graphe de synonymie de ce dictionnaire, autrement dit sa densité plus élevée, permet un enrichissement beaucoup plus conséquent que pour le dictionnaire de Bailly¹².

¹⁰ Ces dictionnaires sont mentionnés à la note 2.

¹¹ Cette « harmonisation » est décrite dans Ploux, 1997.

¹² Sans nous écarter de notre propos, nous pouvons néanmoins signaler que nous avons relevé sur nos sept dictionnaires sources une relation directe entre la densité du graphe de synonymie et le nombre de relations ajoutées à la première passe de notre procédé.

A titre d'illustration, voici les liaisons proposées pour les vedettes de plus de 100 synonymes¹³ :

<u>arrêter, brider</u>	<u>dire, enseigner</u>	<i>extraordinaire, titanesque</i>
<i>arrêter, coincer</i>	<i>dire, faire connaître</i>	<i>grossier, indécent</i>
<u>arrêter, obstruer</u>	<i>dire, proclamer</i>	<i>marquer, déceler</i>
<i>association, amalgame</i>	<u>division, dissension</u>	<i>partir, s'évader</i>
<i>attacher, coincer</i>	<i>doux, accommodant</i>	<u>porter, déclencher</u>
<u>augmenter, surfaire</u>	<i>désordre, méli-mélo</i>	<i>porter, faire naître</i>
<i>bas, ignominieux</i>	<u>détruire, détériorer</u>	<i>sot, aberrant</i>
<i>bas, immonde</i>	<u>exciter, inspirer</u>	<u>soutenir, affermir</u>
<u>bas, ordurier</u>	<i>exciter, réchauffer</i>	<i>tenir, atteler</i>
<i>chagrin, désespérance</i>	<i>extraordinaire, effarant</i>	<u>tomber, expirer</u>
<i>dire, attester</i>	<i>extraordinaire, monumental</i>	

Il est clair à la lecture de ce tableau que les liaisons sont très pertinentes, encore plus que dans le cas du dictionnaire unique ; certaines peuvent toutefois appeler un commentaire ou un exemple de substituabilité, comme « *arrêter* : *obstruer* », valable dans un exemple tel que :

Cet immeuble arrête/obstrue la vue.

ou bien « *porter* : *déclencher, faire naître* », dans :

La révolution a porté/déclenché/fait naître la terreur dans tout le pays.

6. Conclusion et perspectives

Nous avons donc présenté une méthode de transformation d'un graphe de synonymie qui permet de rajouter automatiquement une partie des liaisons auxquelles les lexicographes n'ont pas pensé lors de la rédaction de leurs ouvrages ; il faut cependant préciser que ce procédé ne permet pas de raccorder des parties du graphe global qui ne sont pas connectées entre elles, autrement dit des « composantes connexes » indépendantes, et dont l'indépendance serait le reflet d'un oubli. Là encore, la fusion de dictionnaires ou l'examen humain deviennent impératifs pour trouver une solution, quand celle-ci existe.

Pour terminer, nous dirons que cette méthode, dont l'application aux graphes de synonymie s'avère intéressante, nous paraît pouvoir être étendue à d'autres graphes, pas forcément pour ajouter des liaisons, mais pour détecter dans ceux-ci des arcs ou arêtes « en gestation ».

Bibliographie

BERGE, Claude (1958) : *Théorie des graphes et ses applications*. Paris, Dunod.

BOURIGAULT, Didier (2002) : « UPERY : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus », *Actes de la conférence TALN 2002*, pp. 75-84. Nancy.

¹³ Les relations soulignées sont celles qui résultent de la condition utilisant le logarithme du nombre de synonymes communs.

- BRODDA, Benny et Hans KARLGRÉN (1969) : « Synonyms and synonyms of synonyms », *SMIL*, 5, pp. 3-17. Stockholm.
- GAUME, Bruno, Karine DUVIGNAU, Olivier GASQUET et Marie-Dominique GINESTE (2002) : « Forms of meaning, meaning of forms », *Journal of Experimental and Theoretical Artificial Intelligence (JETAI)*, V. 14 N. 1, pp. 61-74. Binghampton, Taylor & Francis.
- JACQUET, Guillaume (2003), « Polysémie verbale et construction syntaxique : étude sur le verbe *jouer* », *Actes de la conférence TALN 2003*, pp. 469-479, Batz-sur-mer.
- KAHLMANN, André (1975), *Traitement automatique d'un dictionnaire de synonymes*, Stockholm, Université de Stockholm.
- PLOUX, Sabine (1997). « Modélisation et traitement informatique de la synonymie ». *Linguisticae Investigationes*, XXI (1), Amsterdam, John Benjamins.
- PLOUX, Sabine & Bernard VICTORRI (1998), « Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes », *TAL*, Vol 39/1, pp. 161-182.
- VENANT, Fabienne (2004), « Géométrer le sens », *Actes de la conférence TALN 2004*, Fès.
- VERONIS, Jean (2003), « Cartographie lexicale pour la recherche d'information », *Actes de la conférence TALN 2003*, pp. 266-275, Batz-sur-mer.

Dictionnaires

- BAILLY, René (1946) : *Dictionnaire des synonymes*. Paris, Larousse.
- BÉNAC, Henri (1956) : *Dictionnaire des synonymes*. Paris, Hachette.
- BERTAUD DU CHAZAUD, Henri (1971) : *Nouveau dictionnaire des synonymes*. Paris, Robert.
- GUIZOT, François (1864) : *Dictionnaire Universel des synonymes de la Langue Française*. Paris, Didier (7ème édition).
- LAFAYE, Pierre-Benjamin (1858) : *Dictionnaire des synonymes de la Langue Française*. Paris, Hachette.
- Grand Larousse de la Langue Française* (1971) : Paris, Larousse.
- Le Grand Robert, dictionnaire de la langue française* (1985) : sous la dir. d'A. REY. Paris, Robert.
- Trésor de la Langue Française informatisé* (2001) : CNRS, ATILF (Analyse et traitement informatique de la langue française), UMR CNRS-Université Nancy2, <http://atilf.atilf.fr/>.