HAL
open science

Représentation de graphes par ACP granulaire

Bruno Gaume, Louis Ferré

To cite this version:

Bruno Gaume, Louis Ferré. Représentation de graphes par ACP granulaire. journées d'Extraction et de Gestion des Connaissances (EGC), Jan 2004, Clermont Ferrand, France. hal-01321959

HAL Id: hal-01321959

https://hal.science/hal-01321959v1

Submitted on 26 May 2016

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.

Représentation de graphes par ACP granulaire

Bruno Gaume*, Louis Ferré**

*IRIT-UPS, 118 route de Narbonne, F-31062 Toulouse Cedex 4, France
gaume@irit.fr,
http://dilan.irit.fr
**GRIMM-EA2254, Université Toulouse II, 5 Allées Antonio Machado, 31058 Toulouse Cedex
loferre@univ-tlse2.fr
http://www.univ-tlse2.fr/grimm/smash/ferre/index.html

Résumé. L'extraction d'information de grands graphes repose le plus souvent sur leur représentation dans des espaces de dimension réduite et on utilise généralement des méthodes factorielles appliquées à des mesures de dissimilarités calculées à partir des matrices associée du graphe ou l'analyse spectrale de leur Laplacien discret. Efficaces pour dégager les structures globales, ces représentations sont parfois peu exploitables dès lors que l'on s'intéresse à une perspective du graphe à partir de certains sommets privilégiés. Or l'information recherchée a souvent un caractère "local ". Pour représenter le graphe du point de vue d'un ou de plusieurs sommets sélectionnés, nous proposons une méthode d'Analyse en Composantes Principales "Granulaire " consistant à appliquer une A.C.P. "filtrée " à un tableau de proximités. La visualisation d'un graphe de dictionnaire dont la mesure de proximité est obtenue à partir d' un algorithme original illustre notre propos.

1 Introduction

De par l'étude des grands réseaux, la visualisation de grands graphes est un problème crucial qui se rencontre dans des domaines aussi variés que la linguistique, l'électronique, la biologie, la "webologie" (étude et représentation du Web), les neurosciences etc... Selon la problématique, on cherche à révéler tel ou tel aspect privilégié du graphe au travers de sa représentation. Souvent, on vise à le représenter globalement et à identifier les composantes connexes ou les composantes "les plus" connexes.

Pour atteindre cet objectif, il est possible de représenter un graphe par les coordonnées des sommets dans un espace propre de son Laplacien (voir, e.g., [Chung, 1997], [Kuntz *et al.*, 2001], [Mohar, 1991]) : il est notamment connu que la multiplicité de la valeur propre nulle est égale au nombre de composantes connexes du graphe. Citons également, l'analyse de contiguïté [Lebart, 2001] qui conduit à une représentation du graphe par Analyse des Correspondances de la matrice associée (voir également [Benzecri *et al.*, 1973]) ce qui revient encore à projeter sur des sous-espaces propres du Laplacien. Une deuxième solution est fournie par la construction d'une mesure de similarité ou de dissimilarité entre sommets du graphe à partir de sa matrice associée (en utilisant un indice de type Dice, Jacquard,...). On peut ainsi appliquer à cette

matrice une méthode de type principal co-ordinates analysis (diagonalisation de la matrice de Torgerson) ([Jouve *et al.* 1998], [Jouve *et al.*, 2001] et [Ferré et Jouve, 2002]) ou une méthode de type Multidimensional Scaling pour positionner les sommets du graphe dans des espaces de dimension réduite. Ces méthodes peuvent s'appliquer à des graphes orientés ou non, pondérés ou non.

Cependant, si elles peuvent s'avérer pertinentes lorsque l'objectif est notamment le partionnement du graphe (par exemple, en appliquant une méthode de classification aux coordonnées des sommets projetés), ces méthodes n'ont cependant pas été conçues à la base pour représenter le graphe localement du point de vue d'un ou de plusieurs sommets.

C'est pourtant ce à quoi nous aspirons, voir [Gaume, 2003a], dans le cadre de l'étude de graphes lexicographiques[1] constitués à partir de dictionnaires de langues comme présenté en Section 2. En effet, obtenir une représentation globale par l'une ou l'autre des méthodes est tout à fait concevable pour un tel graphe, cependant, cela n'offre que peu d'intérêt car la représentation de plusieurs dizaines de milliers de sommets dans un espace de dimension deux ou trois est difficilement lisible. De plus, notre objectif premier n'est pas celui-là. Il s'agit d'obtenir une représentation du graphe à partir d'un mot particulier, de représenter une région du graphe, le voisinage d'un sommet particulier, dont la forme géométrique est une bonne traduction de la forme sémantique[2] du sommet en question. Par exemple en Section 4, nous produirons des représentations tri-dimensionnelles d'un voisinage de verbes mettant en évidence leurs structures conceptuelles, notamment leurs structures polysémiques.

Pour obtenir, une telle représentation nous procédons en deux étapes. La première consiste à construire une mesure de proximité sur les sommets du graphe, appelée Proxémie et notée $Prox$. On trouvera dans [Gaume, 2003b] une présentation détaillée de $Prox$ et dans [Gaume, 2003a] une utilisation de celle-ci à la désambiguïsation des entrée homonymiques. Alors qu'habituellement les similarités (ou dissimilarités) entre paires de sommets déduites de la matrice associée ne prennent en compte que les voisins immédiats, la Proxémie est construite en prenant en compte l'ensemble du graphe dans sa globalité[3]. Nous présentons le graphe et la mesure $Prox$ dans la Section 2. Nous appliquons ensuite au tableau des Proxémies une analyse en Composantes Principales filtrée qui permet de sélectionner les individus en fonction de leurs Proxémies au mot étudié. Cette approche est présentée en Section 3. Enfin, nous présenterons en Section 4 la mise en oeuvre de la méthode sur le graphe du dictionnaire. Nous nous limiterons à ce point de vue bien que la portée de notre propos s'étende à d'autres graphes et d'autre situations comme celles mentionnées au début cette introduction.

---

1. Travaux financés en partie par le projet " DiLan, Les Dictionnaires de Langue, des graphes aux propriétés topologico-sémantiques " Programme CNRS interdisciplinaire: Société de l'information http://dilan.irit.fr/

2. " dans un état de langue, tout repose sur des rapports" Saussure

3. Seul le graphe tout entier a un sens, ou plutôt le rapport des sommets entre eux. Ce n'est pas le sens des parties qui donne le sens du tout, mais le sens du tout qui donne le sens à ses parties. Ce n'est pas le sens des sommets qui donne le sens du graphe, mais le sens du graphe qui donne du sens à ses sommets. Ce n'est pas le sens des mots qui donne le sens au dictionnaire, mais le sens du dictionnaire qui donne du sens aux mots qui sont ses entrées. Notre approche est donc de type Gestaltiste, voir [Gaume *et al.*, 2002].

2 Le graphe et sa mesure de proximité

Nous considérons un graphe de dictionnaire $\Gamma = (V,E)$ où V est l'ensemble de ses n sommets et E est celui de ses arcs.

L'ensemble V est celui des n entrées du dictionnaire étudié et E est défini de la façon suivante [4] :

$$\forall (x,y) \in V^2 \ (xy) \in E \Leftrightarrow \text{y est dans la définition de x.}$$

Par exemple, la Figure1 présente la partie du graphe correspondant aux voisins immédiats du verbe "écorcer" dans le Robert, alors qu'une partie du sous-graphe des voisins d'ordre 1, 2 et 3 d'écorcer restreint aux sommets de type "Verbe" est représentée en Figure2.

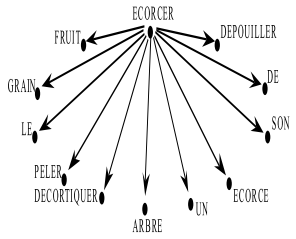

FIG. 1 – *Extrait du graphe autour de Ecorcer.*

Soit A la matrice associee Γ, (voir e.g. [Berge, 1958]), on pose $D = diag(A\mathbf{1}_n)$ où $n = card(V)$ et $\mathbf{1}_n$ est le vecteur de $\Re^n$ dont toutes les coordonnées sont égales à 1. Considérons la matrice

$$G = D^{-1}A.$$

Cette matrice est stochastique et peut être interprétée comme la matrice de transition d'une chaîne de Markov homogène, correspondant à une marche aléatoire sur le graphe.

Les éléments de G vérifient $\forall i \leq n, \forall j \leq n, G_{ij} = \frac{A_{ij}}{\sum_{0<k\leq n} A_{ik}}$, ce qui peut s'interpréter également comme une mesure de proximité entre les voisins à l'ordre 1. Les

4. En fait le dictionnaire doit préalablement être lemmatisé. Pour plus de détail sur la procédure de construction d'un graphe à partir d'un dictionnaire voir [Gaume *et al.*, 2002]

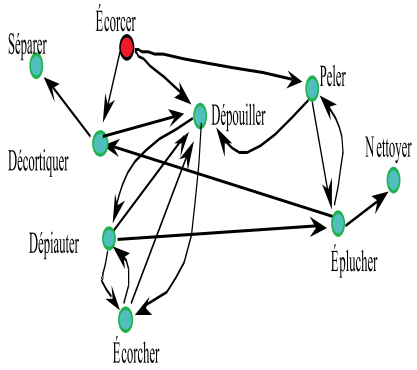

FIG. 2 – *Verbes voisins d'ordre 1, 2 et 3 de Ecorcer.*

méthodes mentionnées ci-dessus et qui reposent aussi sur la matrice associée, utilisent également la proximité à l'ordre 1.

Cependant, il est possible d'étendre le propos en prenant en compte un ordre supérieur. Cela permet de définir la proximité entre deux sommets non pas par la simple existence d'un arc entre ces deux sommets mais par l'existence d'un nombre de sommets auxquels ils sont indirectement reliés (voir, e.g., [Gaume, 2003b]). Notons que la construction de similarités ou dissimilarités à partir de la matrice associée et d'indices tels que celui de Dice par exemple, permet de prendre en compte l'ordre 2, (voir, e.g. [Jouve *et al.*, 2001]), mais, en général, pas au delà.

Dans notre cas, prendre en compte l'ordre p revient à s'intéresser à la situation de notre marche aléatoire à l'instant p: on obtient alors un tableau de proximité en considérant la matrice de transition au $p$ième instant qui est donc la matrice G^p.

La représentation du graphe s'obtient par ACP du tableau G^p, p pouvant être alors considéré comme un paramètre servant à fixer un "diamètre d'influence". Nous avons choisi ici une ACP standard [5] et on représente la proximité entre sommets à partir de la distance euclidienne calculée sur les lignes du tableau des Proxémies G^p. Cependant, d'autres choix de métriques sont bien sûr possibles (notamment celle du χ^2 qui conduit à l'AFC) et c'est la raison pour laquelle nous développons ci-dessous notre propos dans un cadre plus général.

5. Cela se justifie de par la nature de nos graphes qui sont de type "petit monde à échelle libre" ([Newman, 2003], [Gaume, 2003a]). En effet, selon [Watts, 1999] la plupart des grands graphes de terrains peu denses se caractérisent au moyen de deux indices : C, qui mesure son taux d'agrégation et L qui est la moyenne des plus courts chemins. Les graphes de type "petit monde" se caractérisent par des taux d'agrégation forts et des moyennes des plus courts chemins faibles. De façon schématique, on peut dire qu'ils répondent au principe de "mes amis ont tendance à être amis entre eux".

3 ACP granulaire

Le tableau G^p contient une information calculée sur l'ensemble du graphe qu'il serait possible de représenter, comme indiqué ci-dessus, au moyen d'une ACP de G^p. Le terme ACP doit être pris ici dans un sens large, c'est dire comme la décomposition spectrale d'une matrice avec des métriques convenables. Notamment, pour $p = 1$, cela recouvre l' analyse spectrale du Laplacien ou l'analyse de contiguïté. Cependant, une telle représentation va s'avérer peu efficace dans la mesure où il sera bien difficile d'interpréter quoique ce soit de la représentation de plusieurs dizaines de milliers de sommets[6] en, disons, 3 dimensions.

De plus, notre intention est de représenter localement le graphe. En effet, notre objectif est d'obtenir "la forme" d'une région sémantique à partir de laquelle on peut naviguer dans le graphe : on veut ainsi obtenir partir d'une proximité calculée sur l'ensemble du graphe, une représentation locale "autour" de l'un de ses sommets préalablement sélectionné.

Pour ce faire, nous proposons d'utiliser une version pondérée de l'ACP de G^p et, pour obtenir une représentation autour du mot m, on considère la diagonalisation d'une matrice de la forme

$$Gp'M_1D_mGpM_2$$

où D_m est une matrice diagonale de poids et M_1 et M_2 sont des matrices associées respectivement aux espaces des colonnes et des lignes de Gp.

Un choix "naturel" de D_m consiste à sélectionner un "diamètre Proxémique autour" du sommet choisi en posant

$$D_m = diag((\mathbf{1}_{[\pi,1]}((G^p)_{m,i}))_{i=1,\ldots,n}),$$

où $\pi \in [0,1]$ est une valeur qui permet de choisir le diamètre Proxémique autour de m. Ce paramètre permet d'obtenir un continuum d'analyses, de l'analyse globale pour $\pi = 0$ à la plus locale pour $\pi = 1$,en passant, pour les valeurs intermédiaires, par la prise en compte des seuls mots dont la probabilité d'être atteints à l'instant p et partant de m est supérieure à π.

Cette approche permet d'obtenir une représentation "locale objective" du graphe à partir d'une information globale[7]. En effet, si en ultime étape, on procède à une sélection de sommets, il est clair que les Proxémies sont elles mesurées sur l'ensemble du graphe : on effectue en quelque sorte un "zoom" sur une partie privilégiée du graphe, autour d'un mot privilégié.

Bien sûr, il est également possible d'effectuer d'autres choix pour D_m et notamment de chercher à représenter l'ensemble du graphe mais "vu à partir" d'un mot préalablement choisi, depuis le "point de vue de m" : en somme, une représentation

6. l'ordre des graphes linguistiques que nous étudions oscille entre 10 000 et 100 000 sommets suivant la couverture lexicographique du dictionnaire étudié.

7. c'est là que l'on voit bien que notre approche est Gestaltiste, voir par exemple [Victorri et Fuchs, 1996]

"locale subjective". Il suffit alors de substituer aux indicatrices de véritables poids décroissants en fonction de la Proxémie des sommets au sommet élu.

Par exemple, on peut prendre, en notant, $G^p_{m.}$ la ligne de G^p correspondant au mot m:

$$D_m = diag(G^p_{m.}).$$

Cela revient à pondérer chaque sommet du graphe par sa probabilité d'être atteint, dans la marche aléatoire, à l'instant p en partant de m.

On peut aussi "zoomer" non plus comme ci-dessus autour d'un seul sommet (objectivement ou subjectivement) mais en se focalisant sur plusieurs régions éloignées, afin d'en pouvoir étudier en détail les zones de sens articulatoires, par exemple : c'est le cas si l'on remplace D_m par $\frac{1}{2}(D_r + D_s)$ dans ce qui précède, où r et s sont deux sommets quelconques.

Nous avons jusqu'à maintenant uniquement exploité des poids directement dérivés de la Proxémie sur les sommets du graphe, mais on peut également envisager des poids construits sur la base de l'ACP, c'est à dire à partir de la distance euclidienne - bien sûr, si notre graphe est de type petit monde, les résultats obtenus par l'une ou l'autre approche seront assez voisins. Pour ce faire, nous considérons une fonction positive décroissante K et nous posons

$$D = diag(K(\|v_1 - m\|_{M_2}),...,K(\|v_n - m\|_{M_2})).$$

Par exemple, on peut prendre pour K la fonction qui à t associe $e^{-\lambda t^2}$, ou celle qui à t associe $\mathbf{1}_{[x<\lambda]}(t)$. De telles fonctions sont déjà exploitées dans le domaine de l'Analyse en Composante Principale standard puisque, poser $D = diag(\mathbf{1}_{[x<\lambda]}(\|v_1 - \overline{v}\|),...,$ $\mathbf{1}_{[x<\lambda]}(\|v_n - \overline{v}\|))$, où $\overline{v} = \frac{1}{n}\sum_{v\in V} G^p_v$ est la "moyenne" des sommets, revient à faire de l'Analyse en Composante Robuste en éliminant les individus qui se trouvent à une distance à la moyenne supérieure à un λ donné. Notre approche est donc une façon de faire une analyse "robuste" autour d'un sommet privilégié. Par ailleurs, en prenant K un noyau unitaire et en posant, $D = diag(\frac{1}{nh}\sum_j K(\|v_1 - v_j\|),...,\frac{1}{nh}\sum_j K(\|v_n - v_j\|))$, on obtient une version "non paramétrique" de l'ACP. Mais, notre propos est ici plus général puisqu'il englobe plusieurs analyses. Ainsi, si M_1 est l'identité et M_2 est la matrice diagonales des inverses des sommes des colonnes de Gp, on retrouve l'AFC de cette dernière. Remarquons pour conclure que celle-ci s'obtient aussi simplement en prenant M_1 et M_2 toutes deux égales à l'identité, mais en remplacant Gp par la matrice dont les éléments sont les racines carrées des éléments de Gp [Comenges et Volles, 1979].

4 Application aux dictionnaires de synonymes

Nous illustrons notre approche sur un Graphe de verbes construit à partir d'un dictionnaire de synonymes constitué de sept dictionnaires classiques (Bailly, Benac, Du Chazaud, Guizot, Lafaye, Larousse et Robert) dont ont été extraites les relations synonymiques. Ce premier travail de fusion, effectué à l'Institut National de la Langue Française (aujourd'hui ATILF: http://zeus.inalf.fr/synonymes.htm) a produit une série

de fichiers ; les données de ceux-ci ont été regroupées et homogénéisées au sein du laboratoire CRISCO par un important travail de correction sur le fichier final. C'est ce dictionnaire qui sert de base aux travaux dans [Ploux et Victorri, 1998] sur les cliques tels que ceux de Bernard Victorri et Jean-Luc Manguin (http://elsap1.unicaen.fr/dicosyn.html) ou Sabine Ploux (http://dico.isc.cnrs.fr/dico/fr/chercher).

Dans un souci de clarté et de concision nous nous limiterons au graphe DicoSyn_Verbes qui est le graphe des verbes extrait du graphe ci-dessus. Dicosyn_verbe a 9 043 sommets, il a 50 948 arêtes, sur sa plus grande partie connexe (8 835 sommets) son L est égal à 4,1694 et son C est égal à 0,3186, c'est typiquement un petit monde. La Figure3 représente la courbe de distribution des incidences de ses sommets celle-ci est caractéristique des réseaux petits mondes hiérarchiques [Ravasz et Barabási, 2003] (en log-log elle forme approximativement une droite), voir aussi [Duvignau et Gaume, 2004].

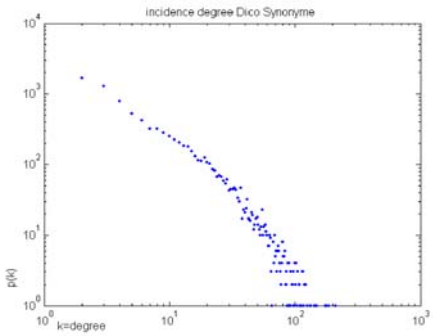

FIG. 3 – *Distribution des incidences.*

Exemple 1 : La Proxémie du verbe "écorcer" et sa visualisation

Dans DicoSyn_Verbes, le verbe "écorcer" a 9 synonymes : {baguer, décortiquer, démascler, dépouiller, gemmer, inciser, peler, tondre}.

Dans la liste ci-dessous, le nombre qui précède chaque verbe est son rang en fonction de sa Proxémie (pour p=6) à ÉCORCER.

[1 ECORCER, 2 DEPOUILLER, 3 PELER, 4 TONDRE, 5 OTER, 6 EPLUCHER, 7 RASER, 8 DEMUNIR, 9 DECORTIQUER, 10 EGORGER, 11 ECORCHER, 12 ECALER, 13 VOLER, 14 TAILLER, 15 RAPER, 16 PLUMER, 17 GRATTER, 18 ENLEVER, 19 DESOSSER, 20 DEPOSSEDER, 21 COUPER, 22 BRETAUDER, 23 INCISER, 24 GEMMER, 25 DEMASCLER, 26 BAGUER, 27 EVINCER, 28 ETRILLER, 29 ETRANGLER, 30 EPURER, 31 EMONDER, 32 ECAILLER, 33 EBRANCHER, 34 EBOURRER, 35 EBARBER, 36 TAMISER, 37 TAILLADER, 38 SPOLIER, 39 SEVRER, 40 SCRUTER, 41 SCARIFIER, 42 SALER, 43 SAIGNER, 44 S'EPOILER, 45 REVOQUER, 46 RUINER, 47 RETOURNER, 48 RETIRER, 49 RANÇONNER, 50 RAISONNER, 51 QUITTER, 52 PRIVER, 53 PILLER, 54 PERDRE, 55 OUVRIR, 56 NETTOYER, 57 MONDER, 58 MARQUER, 59 LIRE, 60 ISOLER, 61 GRUGER, 62 FUSILLER, 63 FRUSTRER, 64 FOUILLER, 65 FILOUTER, 66 FAUFILER, 67 FAUCHER, 68 EXPROPRIER, 69 EXAMINER, 70 ESTAMPER, 71 ESCROQUER, 72 ENTAMER, 73 ENTAILLER, 74 EFFEUILLER, 75 DEVETIR, 76 DEVELOPPER, 77 DEVASTER, 78 DEVALISER, 79 DETRONER, 80 DETROUSSER, 81 DESHERITER, 82 DESHABILLER, 83 DESENVELOPPER, 84 DESENCOMBRER, 85 DESAVANTAGER, 86 DEROBER, 87 DEPOURVOIR, 88 DEPIAUTER, 89 DEPECER, 90 DENUER, 91 DENUDER, 92 DENANTIR, 93 DEMONETISER, 94

DEGARNIR, 95 DEGAGER, 96 DEFEUILLER, 97 DEFAIRE, 98 DECEREBRER, 99 DECOURONNER, 100 DECHAUSSER]

On peut remarquer que le verbe DÉPOUILLER qui apparaît en tête de liste (le plus Prox de ÉCORCER) est en relation hyperonymique avec le verbe ÉCORCER.

La Proxémie organise dans un continuum les notions d'hyperonymie, de cohyponymie intradomaine (par les sommets les plus Prox) et de cohyponymie interdomaines (par les sommets un peu moins proches). La notion de " proxémie " qui recouvre ces trois notions permet de souligner le glissement de sens continu qu'il y a d'un mot en relation synonymique (cohyponyme intradomaine) vers un mot en relation métaphorique (cohyponyme inter domaine) au fur et à mesure que la Proxémie au mot de référence diminue ([Duvignau, 2002]).

De plus, les proxémies ainsi calculées entre les mots sont en accord avec les " approximations sémantiques par analogie " produites par les jeunes enfants. Par exemple, l'énoncé spontané " je déshabille l'arbre " pour [j'écorce l'arbre], produit à 2 ans et demi, manifeste un rapprochement entre verbes qui valide notre modèle : DÉSHABILLER est Prox de ÉCORCER (c'est le 82e mot le plus Prox de ÉCORCER - voir ci-dessus). Ce type d'analyse[8] mené à partir d'un corpus de 230 " approximations sémantiques par analogie " produites par des jeunes enfants (1;8 ans à 4;2 ans) montre que la moyenne du rang du verbe dit par l'enfant (comme DÉSHABILLER) dans la proxémie du mot " juste " (comme ÉCORCER) est de 239, (ce qui est peu sur les 9043 verbes présents dans notre graphe) [Duvignau, 2002], [Duvignau et Gaume, 2003], [Duvignau, 2002] .

Ci-dessous un extrait du corpus recueilli par Karine Duvignau :

Extrait du Corpus enfant :

(1) " Tata, ton coeur y *clignote* " (Cyrielle 2 ans) / la tête sur la poitrine de sa Tata

(2) " je *déshabille* l'arbre " (Camille 2 ;8 ans) / enlevant un bout d'écorce à un tronc d'arbre

(3) " les arbres y *pondent* des fruits " (Julie 2 ;11 ans) / à propos de l'image d'un pommier

(4) " allez *allumes, allumes* tes yeux ! " (Joane 3 ans) / à un adulte qui a les yeux fermés

Extrait du Corpus adulte à visées didactiques :

(5) "Faudra-t-il chercher à *racommoder* les principes *ébréchés*?" (Poincaré, *La valeur de la science*)

(6) "Notre faiblesse nous oblige à *découper* en tranche l'univers" (Poincaré, *Science et méthode*)

(7) "Les voyelles ... sont plus aptes à '*aimanter*' l'attention de l'enfant" (Boysson-Bardiés, *Comment la parole vient aux enfants*)

(8) "c'est en utilisant à son profit la formidable puissance du hasard que l'embryon *bricole*, de manière aveugle, son système immunitaire" (Jean_Claude Ameinsen, *La sculture du vivant*)

La Figure 4 donne la visualisation des 100 sommets les plus proches de "écorcer" pour $p = 6$. Pour qu'il y ait 100 sommets affichés, $D_{\text{écorcer}}$ est tel que :

8. travaux partiellement financés par le Projet " Structure et Acquisition du Lexique, L'analogie pour l'apprentissage des verbes " Programme CNRS Ecole & Sciences cognitives, http://dilan.irit.fr/

$$\forall i,\, D_{\text{écorcer}^{ii}} = \mathbf{1}_{[0.001,1]}((G^p)_{\text{écorcer},i}).$$

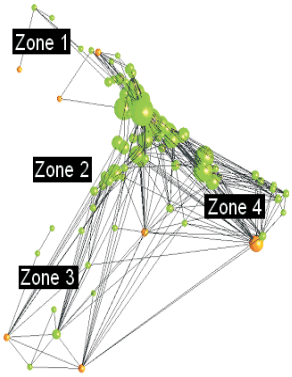

FIG. 4 – *Projection des 100 verbes les plus proches d'Ecorcer.*

La Figure5 est un"zoom" permettant d'identifier les verbes appartenant aux quatre zones autour d'écorcer.

Exemple 2 : La forme conceptuelle du verbe "jouer" et sa polysémie

En Figure6 nous donnons la représentation des voisins de "jouer", c'est à dire qu'ici $D_{jouer} = A_{jouer}$ qui est la ligne de jouer dans la matrice associée A du graphe DicoSyn_Verbes, (ici on ne veut afficher que les voisins de "jouer"). On peut remarquer en Figure7 que la forme ainsi représenté est en accord avec la structure conceptuelle du verbe "jouer" et que les quatre pôles principaux reflètent la structure polysémique du verbe "jouer".

Références

[Berge, 1958] C. Berge. Théorie des Graphes et ses applications, DUNOD Paris, 1958.

[Benzecri *et al.*, 1973] J. P.Benzecri, F. Benzecri, A. Birou et S. Blumenthal. L'analyse des données : Vol. 2 : L'analyse des correspondances, Paris/Bruxelles/Montréal : Dunod , vi, 1973.

[Chung, 1997] F.R.K. Chung. Spectral Graph Theory. Regional Conference Series in Mathematics, 92, American Mathematical Society, 1997.

[Comenges et Volles, 1979] D. Comenges et M. Volles. Analyse factorielle sphérique : une exploration. Annales de l'INSEE, 35, 1979.

[Duvignau et Gaume, 2004] K. Duvignau et B. Gaume. Pour une ergonomie cognitive des dictionnaires électroniques, in Fouille de textes et organisation de document, numéro spécial de la revue Doccument Numérique, Hermes (à paraître), 2004.

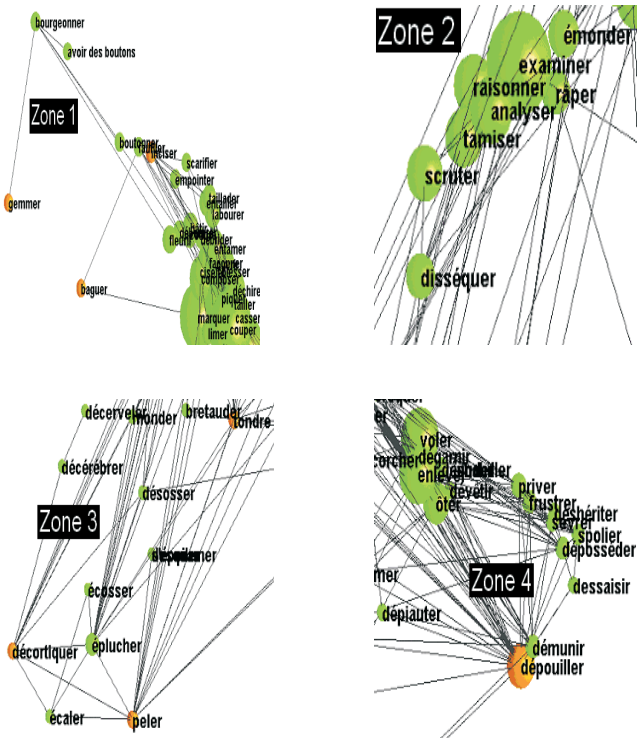

FIG. 5 – *Visualisation des zones sémantiques autour d'Ecorcer.*

[Duvignau, 2002] K. Duvignau. La métaphore, berceau et enfant de la langue, Doctorat en Sciences du Langage, Université Toulouse-Le Mirail, 2002.

[Duvignau et Gaume, 2003] K. Duvignau et B. Gaume. For Early Verb-Learning Based on Analogy, Cognitive Systems, à paraître, 2003.

[Duvignau, 2002] K. Duvignau. Métaphore verbale et approximation, In Regards Croisé sur l'Analogie Duvignau K., Gasquet O., Gaume B. (éditeurs), Revue d'Intelligence Artificielle, vol. 17, n° 5-6. Hermès-Lavoisier, 2003.

[Ferré et Jouve, 2002] L. Ferré et B. Jouve. Vertex partitioning of a class of digraphs. Mathématiques Informatique et Sciences Humaines, 158, 59 - 77, 2002.

[Gaume *et al.*, 2002] B. Gaume, K. Duvignau, O. Gasquet, M-D. Gineste . Forms of Meaning, Meaning of Forms, Journal of Experimental and Theoretical Artificial Intelligence, 14(1):61-74, 2002.

[Gaume, 2003a] B. Gaume. Analogie et Proxémie dans les réseaux petits mondes, Regards Croisés sur l'Analogie, Actes des Journées Analogie et Cognition, JAC'02, Toulouse, 24-25 Mai 2002, In: Revue d'Intelligence Artificielle, n°5-6, Hermès Sciences, Duvignau K., Gaume B., Gasquet O (éditeurs) (2003) (à paraître),2003.

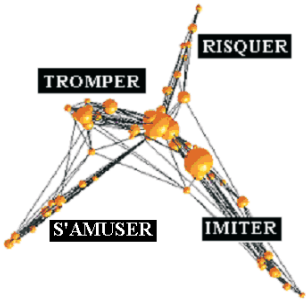

FIG. 6 – *Visualisation des "zones" sémantiques de Jouer.*

[Gaume, 2003b] B. Gaume. Random walk in lexical small world, soumis à publication.

[Jouve *et al.*, 2001] B. Jouve , P. Kuntz, E. Velin E. Extraction de structures macroscopiques dans des grands graphes par une approche spectrale. In Extraction des Connaissances et Apprentissage, vol 1, n°4, Hermes (eds), Paris, 2001.

[Jouve *et al.* 1998] B. Jouve, P. Rosenstiehl, M. Imbert M. A mathematical approach to the connectivity between the cortical visual areas of the macaque monkey, Cerebral Cortex (8), 28-39, 1998.

[Kuntz *et al.*, 2001] P. Kuntz, F. Velin, H. Briand. Iterative geometric representations for multi-way partitioning, In proc. of the 5th WSES/IEEEworld Multiconf. on circuits, Systems, Communications and Computer in Advance in Scientific Computiting, Computational Intelligence and Applications, WSEAS Press, 108-113, 2001.

[Lebart, 2001] L. Lebart. Classification et Analyse de Contiguïté, La revue de modulad, 27, 1-22, 2001.

[Mohar, 1991] B. Mohar. The Laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, Ollermann and A Schwenk, Editors, Graph Theory, Combinatronics and Applications, 871-898, New-York, 1991. John Wiley and Son, 1991.

[Newman, 2003] M. E. J. Newman. The structure and fonction of complex networks", Rapport technique disponible à l'adresse http://www.santafe.edu/~mark/recentpubs.html, 2003.

[Ploux et Victorri, 1998] S. Ploux et B. Victorri. Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes, Traitement automatique des langues,39, 161-182,1998.

[Ravasz et Barabási, 2003] E.Ravasz et A.L. Barabási. Hierarchical Organization in Complex Networks, Phys. Rev. E 67, 026112, 2003.

[Victorri et Fuchs, 1996] B. et C.Fuchs. La polysémie - Construction dynamique du sens, Paris, Hermès, 1996.

[Watts, 1999] D.J. Watts. Small words: the dynamic of networks between order and randomness. Princeton: Princeton University Press, 1999.

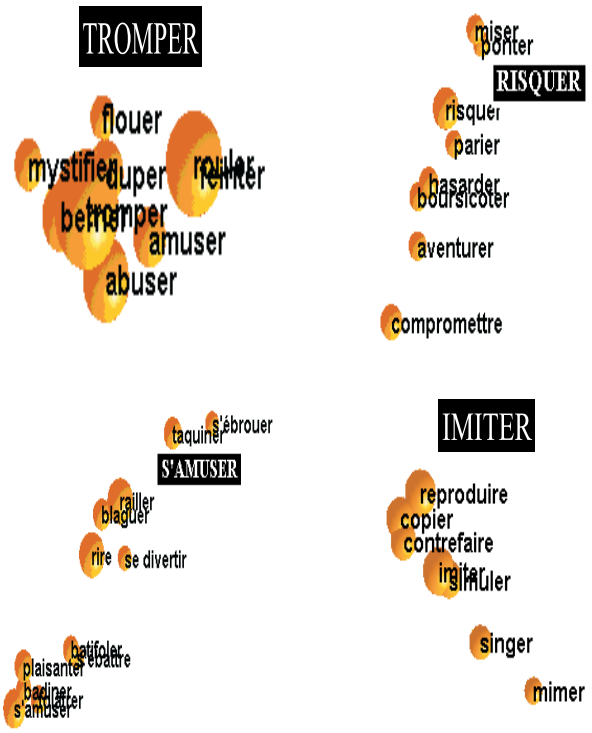

FIG. 7 – *Visualisation avec labels des quatre zones sémantiques autour du verbe Jouer.*

Summary

Extracting information from large graphs often relies on their representation onto reduced dimension spaces. Embedding into an Euclidian space is performed by means of say PCA, Principal Coordinates Analysis or Multidimensional scaling applied to dissimilarities computed from the adjacent matrix or alternatively by discrete Laplacian methods. If they are relevant for revealing the global structures of the graph, those methods fail when local aspects are of interest. However, in many situations, it is relevant to investigate local visualisation for instance by representing the graph from the point of view one particular vertex. Then, we present a method of Granular PCA which consists in applying a filtered PCA to a proximity table. Our approach provides the visualization of the graph from the point of view of one or several selected vertices. We give an example of application concerning the representation of a lexical graph, reached by computing an original measure of proximity.