



**HAL**  
open science

## Le dictionnaire Electronique des Synonymes du CRISCO : méthode d'utilisation et exploitation psycholinguistique

Jacques François, Jean-Luc Manguin, Françoise Cordier, Christine Ros

### ► To cite this version:

Jacques François, Jean-Luc Manguin, Françoise Cordier, Christine Ros. Le dictionnaire Electronique des Synonymes du CRISCO : méthode d'utilisation et exploitation psycholinguistique. Colloque : Études françaises, internet et multimédia II, Mar 2003, Vancouver, Canada. pp.165-200. hal-00012502

**HAL Id: hal-00012502**

**<https://hal.science/hal-00012502>**

Submitted on 24 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.

Jacques FRANÇOIS

Jean-Luc MANGUIN  
UMR 6170 CRISCO, Université de Caen

Françoise CORDIER

Christine ROS  
UMR 6096 LaCo, Université de Poitiers

## **Le dictionnaire Electronique des Synonymes du CRISCO : méthode d'utilisation et exploitation psycholinguistique**

### **Introduction**

Le laboratoire CRISCO (Centre de Recherche Inter-langues sur la Signification en COntexte) est une unité mixte de recherche de l'Université de Caen et du CNRS ; les recherches qui y sont menées relèvent de la linguistique du français et de quatre autres langues d'Europe, et de la linguistique informatique. Dans ce dernier domaine, des travaux sur la modélisation et la représentation du sens des mots ont conduit les chercheurs<sup>1</sup> à construire un dictionnaire des synonymes du français sous forme d'une base de données exploitable par des programmes informatiques.

En 1998, la création du site Internet du laboratoire s'est accompagnée de l'ouverture de ce dictionnaire à la libre consultation par les internautes ; mais ce dictionnaire qui n'était au départ qu'un support pour la recherche en sémantique s'est avéré, au contact du public, être également un outil répondant à une formidable demande. Certes son apport à la recherche a continué à se développer, mais son originalité et la demande croissante dont il fait l'objet nous ont amenés à le considérer autrement : du statut de simple ressource, il est devenu lui-même objet de la recherche, qu'elle soit mathématique ou linguistique.

En outre, à partir de 2001, dans le cadre du programme *Cognitive* du Ministère français de la Recherche, le laboratoire LaCo de psychologie cognitive (précisément Françoise Cordier et Christine Ros) s'est associé au CRISCO pour comparer l'évaluation de la distance entre synonymes fournie par le D.E.S. et par des groupes de sujets de différentes classes d'âge. La section 7 de cet article donne un aperçu des résultats de deux expérimentations psycholinguistiques sur cette question.

### **1. Elaboration du dictionnaire**

Le dictionnaire des synonymes du CRISCO (<http://www.crisco.unicaen.fr/>) est d'un genre particulier ; c'est en effet l'un des deux dictionnaires électroniques des synonymes du français en accès libre sur Internet<sup>2</sup> ; il contient 49000 entrées, chacune suivie de ses synonymes, comme par exemple :

*démarcation: distinction, délimitation, frontière, ligne, limitation, limite, lisière, marque, séparation*

---

<sup>1</sup> Sabine Ploux, alors responsable du projet, et Bernard Victorri.

<sup>2</sup> Le deuxième étant présent sur le site de l'ISC à Lyon, constitué par une ancienne version de celui du CRISCO annotée à l'ISC, et couplée à un équivalent anglais, ce qui forme un dictionnaire bilingue.

Chaque synonyme est lui-même une entrée, de sorte que l'on peut considérer notre dictionnaire comme un ensemble de 49000 mots-vedettes (entrées) reliés les uns aux autres par un réseau de relations. Il y a relation lorsque les deux vedettes sont synonymes, et l'on dénombre un peu plus de 199 000 relations pour tout le dictionnaire.

Ces relations de synonymie proviennent de sept dictionnaires classiques : deux dictionnaires analogiques (le Grand Larousse et le Grand Robert), deux dictionnaires des synonymes du 19<sup>e</sup> siècle (Lafaye et Guizot), et trois dictionnaires des synonymes du milieu et de la fin du 20<sup>e</sup> siècle (Bailly, Bénac et Du Chazaud). Toutes les informations contenues dans ces dictionnaires nous ont été fournies sous forme de fichiers informatiques par l'Institut National de la Langue Française<sup>3</sup> ; le travail effectué au laboratoire a consisté à harmoniser les mots-vedettes de ces sept fichiers, puis à fusionner toutes ces données. Le tableau ci-dessous permet de mesurer l'ampleur de ce travail<sup>4</sup> :

	<b>Données de départ</b>	<b>Conservées</b>	<b>Ajoutées</b>	<b>Données actuelles</b>
<b>Entrées</b>	63000	48500	600	49100
<b>Relations</b>	219000	192000	7000	199000

Tableau 1 : Corrections et ajouts effectués sur les données

Par ailleurs, la fusion des dictionnaires a fait disparaître les commentaires (quand ils existaient) et la structure des articles des ouvrages d'origine ; cet appauvrissement apparent, nécessaire en raison des disparités entre les différents dictionnaires de départ, constitue en fait un tremplin pour des méthodes automatiques qui, comme nous le verrons, permettent d'analyser de manière fine et objective l'espace sémantique des mots.

Tout d'abord, la fusion des données des dictionnaires d'origine s'est faite en conservant l'origine des relations mises en commun ; au moyen d'un programme élaboré au laboratoire, mais réservé à l'usage interne, il est possible de faire des comparaisons entre dictionnaires pour un mot donné. Le programme délivre un tableau de la forme suivante, où les cases grisées représentent les relations existantes, classées par nombre décroissant de dictionnaires qui les mentionnent :

<sup>3</sup> Les relations synonymiques de l'INALF ont été exploitées dans l'élaboration du *Trésor de la Langue Française*. Martin (2001) montre dans la section II. *Le filtrage du sens* (pp.91-100) du chapitre *La réduction polysémique* comment les indications syntaxiques et sémantiques du *Trésor de la Langue Française Informatisé* (en ligne sur le site de l'ATILF) peuvent être exploitées pour une réduction automatique de la polysémie lexicale. Cette démarche, qui s'appuie sur le plus vaste dictionnaire électronique de la langue française est complémentaire de la nôtre qui se fonde sur le regroupement de liens synonymiques d'origine diverse (outre les sept dictionnaires mentionnés, un ensemble important de corrections et d'ajouts).

<sup>4</sup> Les chiffres donnés ici ne concernent que la version CRISCO du dictionnaire.

vedette	synonymes	Lafaye	Guizot	Bailly	Bénac	Larousse	Robert	Du Chazaud
version	traduction							
version	interprétation							
version	leçon							
version	relation							
version	compte rendu							
version	déclaration							
version	explication							
version	rapport							
version	mouture							
version	récit							
version	histoire							
version	variante							

Tableau 2 : Comparaison des dictionnaires pour le mot *version*

Notre dictionnaire se présente donc comme un outil fournissant des réponses relativement abruptes (puisque tous les commentaires ont disparu), mais cette sécheresse apparente est compensée par le nombre élevé d'entrées, et par la grande rapidité d'obtention des résultats. D'autre part, les données de sa fréquentation sont à la base de son évolution.

## 2. Evolution du dictionnaire

Pour l'éditeur d'un dictionnaire en ligne, les traces laissées par les internautes sont une information du plus haut intérêt ; ce sont elles qui permettent de connaître le degré de satisfaction des utilisateurs, autrement dit de savoir dans quel(s) cas le dictionnaire répond aux requêtes formulées ; cette connaissance est ensuite le point de départ des enrichissements apportés aux données.

### 2.1 Les variantes orthographiques

Pour résoudre le problème des mots qui ont plusieurs graphies possibles (comme *clé* ou *clef*), nous avons adopté la solution d'un fichier annexe juxtaposé au fichier principal. Lorsque l'internaute tape une variante orthographique dans sa requête (par ex. *chelinguer*), le programme de consultation utilise le fichier des variantes pour rediriger l'interrogation vers le mot associé (*schlinguer*), et obtenir ainsi une réponse. Nous avons élargi ce concept en y incluant les entrées du dictionnaire qui contiennent des signes diacritiques, mais en les débarrassant de ceux-ci :

*chaumiere:chaumière*  
*chasse:chasse,chaussé*  
*chasse-trappe:chasse-trape*  
*chassee:chaussée*

A noter que dans le cas de *chasse*, le fichier des variantes aboutit à deux entrées du dictionnaire ; dans ce cas, la réponse affichée correspond aux synonymes du premier, mais propose à l'utilisateur une autre interprétation possible de sa requête ; il lui suffit alors de cliquer sur un lien pour obtenir les synonymes du second terme. Cette amélioration de

l'interface de consultation permet de corriger les accents oubliés, ainsi que ceux qui sont ajoutés ou transformés par erreur. Nous avons ainsi une ébauche de correction orthographique des requêtes.

## 2.2 Les formes fléchies

La poursuite des améliorations oblige à regarder quelles ont été les demandes formulées ; l'analyse des 22 millions de requêtes par un programme automatique fait apparaître un "lexique" de plus d'un million de mots. Cependant, seulement 29 000 ont été demandés plus de 100 fois (ce qui représente 81 % des requêtes). A l'intérieur de ce lexique, certains mots sont absents du dictionnaire ; par exemple, pour les 29 000 mots demandés plus de 100 fois, environ 1600 ne font pas partie de notre ressource ; cette absence s'explique par une des trois raisons suivantes :

- le mot est une forme fléchie d'un mot du dictionnaire.
- le mot contient une faute d'orthographe.
- le mot est réellement absent du dictionnaire.

Dans les 1600 mots inconnus évoqués précédemment, chacune des trois possibilités représente environ un tiers de l'effectif. Pour permettre à notre dictionnaire de prendre en compte les formes fléchies, nous avons choisi de faire à nouveau appel au "fichier des variantes", la forme fléchie à ajouter devenant une nouvelle entrée de ce fichier ; la redirection de la requête s'effectue donc vers la forme présente dans le dictionnaire.

## 2.3 La correction orthographique

Le deuxième groupe des mots inconnus est constitué, comme nous l'avons dit, des requêtes qui contiennent une faute d'orthographe. Le traitement de toutes ces demandes erronées passe une fois encore par des entrées supplémentaires au fichier des variantes, mais dans ce cas avec l'indication dans le résultat qu'une correction orthographique a été effectuée. Par exemple, le résultat d'une requête avec le mot *tranquilité* contiendra la ligne ci-dessous affichée dans la page de réponse :

*Votre requête est : "tranquilité" (CORRECTION : tranquillité).*

## 2.4 Les mots manquants

Enfin, la dernière catégorie de mots inconnus correspond à des mots existant dans la langue française, mais absents du dictionnaire ; cette absence s'explique de deux manières :

- Le mot demandé fait sans doute partie du vocabulaire terminologique ; il désigne alors sans ambiguïté quelque chose de très spécifique, et ne possède pas de synonyme. Il est donc impossible de le faire entrer dans notre dictionnaire des synonymes.
- Ou bien le mot demandé correspond à une lacune réelle de notre dictionnaire ; pour la combler, il faut cette fois réaliser un travail lexicographique. De nouvelles entrées sont ainsi périodiquement rajoutées, choisies en prenant dans l'ordre les absences les plus flagrantes ; par exemple, nous avons récemment inscrit *éradiquer*, *extrapoler*, *évolutif*, *look*, *sur mesure*, *informel*, *ubuesque*, etc.

La mise en ligne nous a ainsi permis d'évaluer notre outil en termes de satisfaction des utilisateurs. Cette satisfaction se mesure au pourcentage de requêtes recevant une réponse ; sur l'ensemble du trafic, 87 % des requêtes sont satisfaites, dont 11 % grâce à l'adjonction du fichier des variantes.

### 3. Impact du dictionnaire

Depuis l'ouverture du site en octobre 1998, le dictionnaire a reçu plus de 22 millions de requêtes, dont plus de 12 millions dans les 12 derniers mois ; la fréquentation est toujours en progression, et la croissance du trafic se situe aux environs de 100 % par an. La figure 2 montre l'augmentation du nombre moyen de requêtes par jour, durant les deux dernières années.

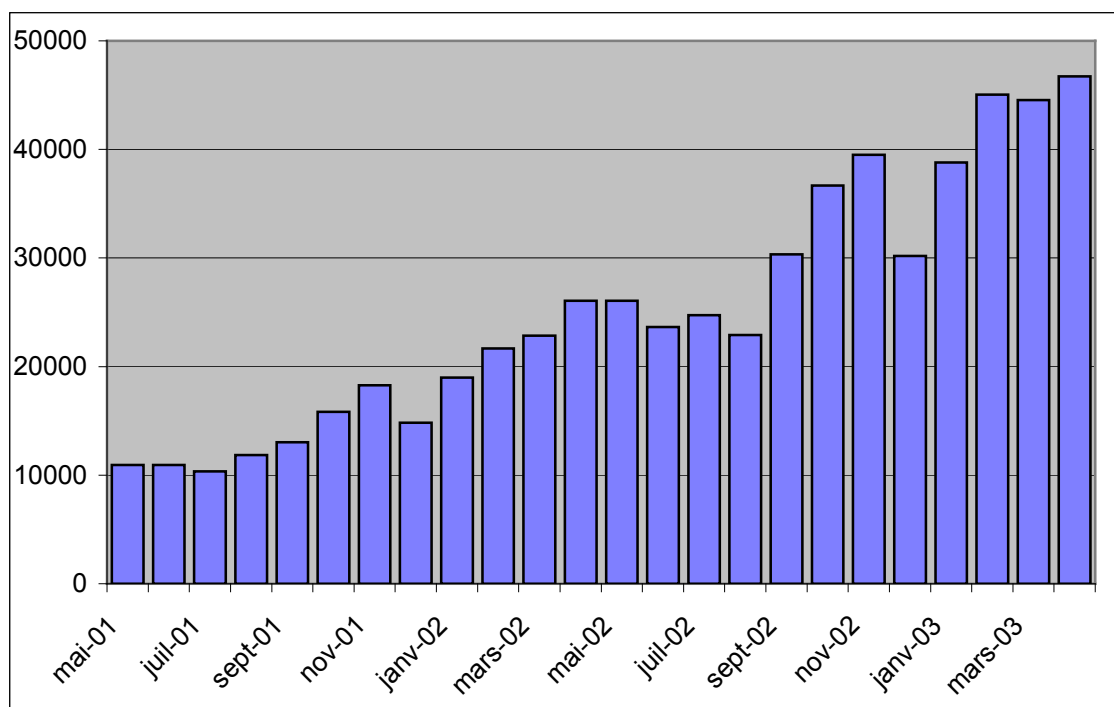


Figure 2 : Trafic du dictionnaire dans les 24 derniers mois.

La répartition des appels selon leur origine géographique révèle une très forte majorité francophone (tableau 2), même si 99 % du trafic se répartit entre 30 pays (ou super-domaines) différents, et si par ailleurs il n'est pas facile de localiser géographiquement les domaines .COM et .NET.

Super-domaine	Pays ou activité	Part de trafic (en %)
FR	France	33
CA	Canada	25
COM	(Commerce)	14
NET	(Internet)	13
BE	Belgique	5
CH	Suisse	3
EDU	Universités américaines	0,7
DE	Allemagne	0,7
	Autres	5,6

Tableau 3 : Les 10 premiers super-domaines par importance du trafic

## 4. L'accès à l'outil

Outre cela, la consultation du dictionnaire exploite les possibilités offertes par Internet, notamment l'hypernavigation qui permet aux utilisateurs de trouver rapidement le terme qui leur manque. L'interface de consultation <http://www.crisco.unicaen.fr/cherches.html> se présente sous la forme d'un formulaire contenant un champ unique, dans lequel l'utilisateur va taper le mot dont il recherche les synonymes. En réponse à sa requête, il reçoit une page contenant la liste des synonymes, et un classement fondé sur le nombre de connexions que ces synonymes forment entre eux (figure 3 en annexe).

Comme nous l'avons dit, l'interface de notre dictionnaire adopte l'hypernavigation comme moyen de déplacement dans le dictionnaire : il suffit ici de cliquer sur un terme de la liste pour obtenir ses synonymes, puisque chaque terme est lui-même un lien vers sa propre liste de synonymes.

## 5. Représentation de la polysémie lexicale

Le travail de modélisation repose sur le fait de considérer le dictionnaire des synonymes comme un graphe ; ce graphe possède 49000 sommets (les entrées du dictionnaire) et environ 400 000 arêtes (les relations de synonymie)<sup>5</sup>. Ce point de vue permet d'effectuer des analyses à l'aide d'objets issus de la théorie des graphes (voir Kahlmann, 1975).

Nous utilisons le plus souvent le sous-graphe dont les sommets sont un mot-vedette et ses synonymes, et dont les arêtes sont les relations qui existent entre ces sommets sélectionnés. La restriction ainsi effectuée nous permet d'étudier un mot en particulier, et d'aboutir à une représentation de son champ sémasiologique, autrement dit de déplier l'éventail de ses significations<sup>6</sup>.

Pour illustrer notre propos, voici une représentation simplifiée du graphe des relations synonymiques qui accompagnent le mot *baie* (figure 4).

---

<sup>5</sup> Les relations synonymiques ont en effet été symétrisées ; cette décision se justifie par les méthodes employées dans la suite ; pour la justification linguistique, on se reportera à [Kahlmann, 1975].

<sup>6</sup> L'arrière-plan théorique du D.E.S. est exposé dans Victorri & Fuchs (1996), spécialement au chapitre 3 : *Un modèle morphodynamique* (pp. 63-85). Cependant l'analyse de l'écart entre les cliques de synonymes constitutives de l'espace sémantique d'un item polysémique illustrée plus bas au §6 est purement descriptive et ne dit rien des processus de diversification du sens. Depuis Darmesteter (1887/1979, pp. 67-75), on sait qu'historiquement la diversification du sens s'effectue de manière radiale (processus de « rayonnement » selon Darmesteter) ou linéaire (processus d'« enchaînement » selon Darmesteter), entrant fréquemment en combinaison. On peut se reporter à Nyckees (1998, pp. 89-156) et Niklas-Salminen (1996, pp.159-165) sur les modalités de l'évolution du sens. La réalité cognitive de la polysémie, en particulier du type « ressemblance de famille » selon L. Wittgenstein, est sujette à controverse et l'exemple célèbre de la polysémie de l'item *game* en anglais (Wittgenstein 1961, pp. 147-8) a donné lieu à des développements intéressants à comparer de Nyckees (1998, pp. 320-7) et Touratier (2000, pp. 110).

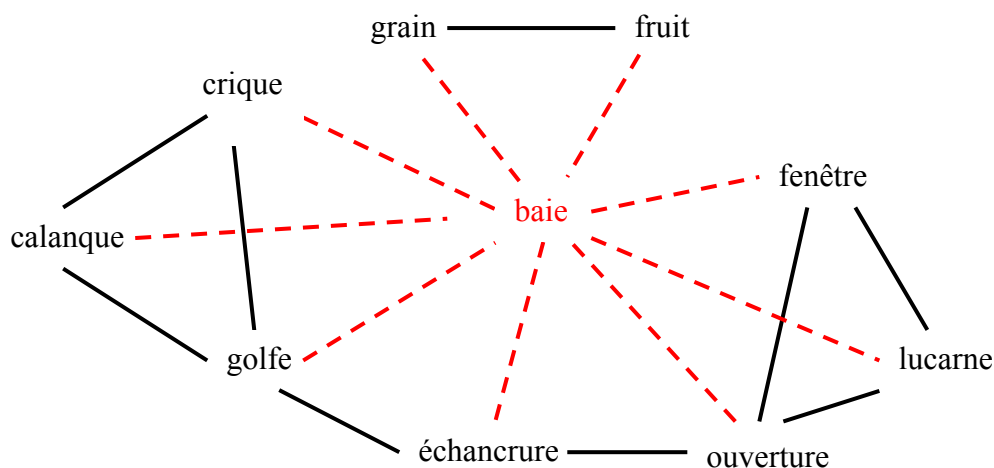


Figure 4 : Exemple de graphe de synonymie (simplifié)

Dans la théorie des graphes, les groupes où il existe un chemin de longueur quelconque entre toute paire de sommets sont appelés "composantes connexes" (voir Berge, 1967). Lorsque nous recherchons ces composantes dans le graphe lié à un mot, nous ne conservons que les relations qui n'impliquent pas le mot-vedette (dans notre exemple, celles qui sont en traits pleins). On voit alors apparaître deux composantes connexes : (*grain, fruit*) et (*calanque, crique, golfe, échancre, ouverture, fenêtre, lucarne*). Elles dissocient *baie* en deux homonymes, l'un qui concerne la baie végétale, l'autre relatif à l'idée d'une *échancre* ou d'une *ouverture*, que celle-ci soit dans une côte (*golfe*) ou dans un objet architectural (*fenêtre*) ; bien entendu, il n'y a pas de relation entre *golfe* et *fenêtre*, et si ces deux termes sont regroupés, c'est parce que le mot *échancre* joue le rôle de pivot. Dans la réalité, les composantes connexes obtenues pour *baie* sont les suivantes :

1. *abri, anse, calanque, conche, crique, croisée, enfoncement, fenêtre, golfe, havre, lucarne, ouverture, rade, vide, vue, échancre*
2. *akène, fruit, grain, graine*

L'analyse du graphe selon ses composantes connexes permet donc de dégager les homonymes d'un mot vedette ; afin de pousser plus loin l'analyse sémantique, nous devons rechercher un autre type de groupe dans le sous-graphe. Les groupes de forte cohésion, pour lesquels chaque sommet est en relation directe avec tous les autres, se nomment "cliques" dans la théorie des graphes ; dans notre exemple simplifié de la figure 4, on trouve 5 cliques : (*grain, fruit*), (*échancre, golfe*), (*échancre, ouverture*), (*golfe, crique, calanque*) et (*ouverture, fenêtre, lucarne*) ; il n'y a pas de clique (*échancre, golfe, ouverture*) car il n'y a pas de liaison entre *golfe* et *ouverture*. On voit ainsi que les cliques permettent une meilleure dissociation des sens ; d'autre part, nous pouvons ici considérer les liaisons avec le mot-vedette, car cela ne change absolument pas le nombre de cliques, mais rajoute seulement la vedette dans chacune d'entre elles. Dans la réalité, le mot *baie* possède 14 cliques. Une étude détaillée des cliques, faite pour un ensemble de mots donnés, a permis de montrer que ces ensembles à forte cohésion peuvent être assimilés aux valeurs les plus élémentaires du sens des mots, autrement dit aux nuances les plus fines que nous pouvons appréhender via notre dictionnaire. Mais l'analyse ne s'arrête pas là, et cet ensemble de cliques va subir une transformation qui nous permet de représenter de manière géométrique le champ sémasiologique du mot-vedette ; pour ce faire, nous considérons que les cliques sont des points d'un espace multidimensionnel, et que les vecteurs unitaires de cet espace sont les synonymes. Le sens du mot-vedette se trouve ainsi matérialisé sous la forme d'un nuage de



points qui se déploie dans cet espace imaginaire. Pour arriver à une représentation visible (c'est-à-dire plane) de ce nuage de points, nous effectuons sur ces coordonnées une analyse factorielle des correspondances, dont le but est de fournir les plans selon lesquels il est plus intéressant d'observer notre nuage de points. En général, les deux ou trois premiers plans sont ceux qui permettent de refléter au mieux les inter-distances entre les points (pour plus de détails, voir Ploux & Victorri, 1998). La figure 5 donne un aperçu des visualisations obtenues sur notre site par cette analyse, toujours avec le mot *baie*.

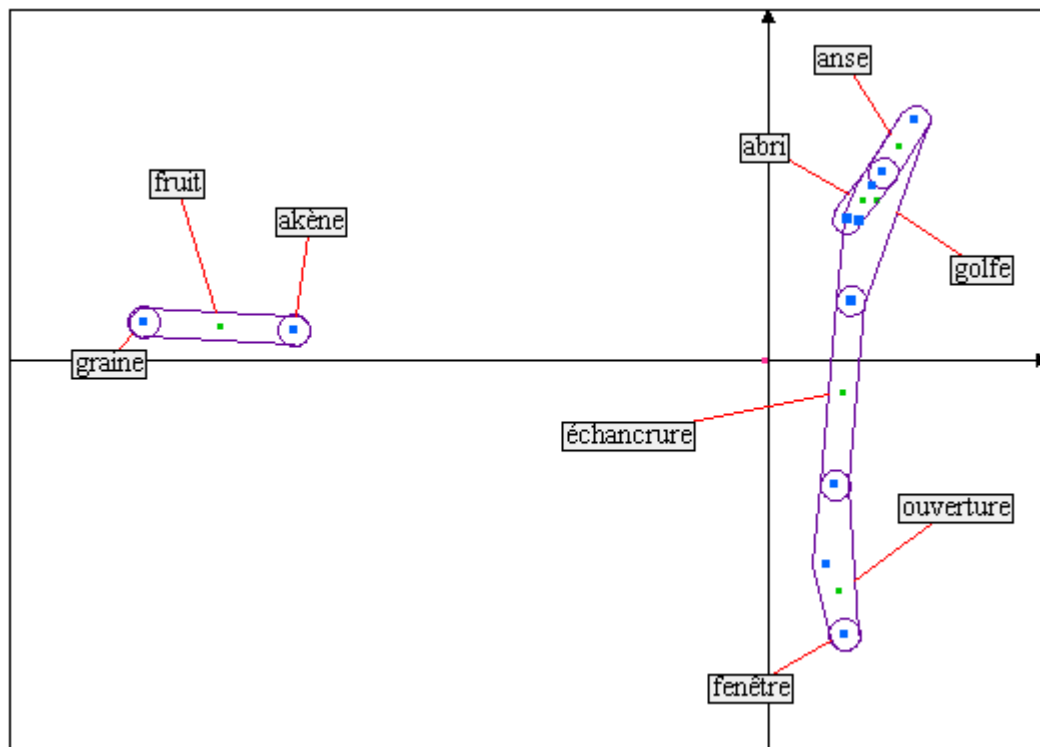


Figure 5 : Visualisation du sens par l'analyse des cliques en composantes principales.

Sur la figure, les cliques sont donc des points, et chaque synonyme est représenté par un contour qui englobe les cliques où l'on trouve ce synonyme ; on remarque sur cette représentation la dissociation entre les deux homonymes (*fruit*) et (*ouverture*, *échancre*), ainsi que l'étalement du sens de la composante qui va de *anse* à *fenêtre* dans un continuum où la liaison est assurée par le mot *échancre*<sup>7</sup>.

## 6. Exemples d'analyses de polysèmes

L'objectif principal de la première section de cet article est d'illustrer un mode d'emploi des visualisations d'espaces sémantiques délivrées par le D.E.S. Ces visualisations se présentent sous la forme de nuages de points. Plusieurs données sont à garder en mémoire afin de garder le cap au milieu du nuage !

(i) Les points du nuage représentent les cliques de synonymes introduites plus haut, par exemple pour l'item *fin* la clique {*beau*, *fin*, *perfection*}.

(ii) L'espace sémantique du terme demandé est multidimensionnel, plus exactement il présente autant de dimensions que de synonymes pris en compte. Le logiciel propose par défaut la visualisation de l'espace sémantique sur le plan défini par les deux premières dimensions (désignées comme des "axes"). C'est le plan sur lequel les cliques apparaissent le

<sup>7</sup> Si l'on exclut la relation entre *baie* et *échancre*, mentionnée par un seul dictionnaire (Le Grand Robert), on obtient alors une division de l'espace en trois homonymes.

mieux distribuées. Il est cependant possible en cliquant sur le menu « choix des axes » de faire 'tourner' la représentation. Nous donnerons plus bas deux illustrations de l'intérêt pédagogique d'un changement de plan.

(iii) Le D.E.S. n'opère pas sur des entrées indicées, c'est-à-dire que les entrées considérées dans l'un ou l'autre des dictionnaires sources comme homonymes y apparaissent rassemblées. La raison est double : d'une part les dictionnaires sources sont en désaccord sur les critères de distinction des homonymes et d'autre part, comme nous l'avons évoqué pour l'item *baie*, nous faisons l'hypothèse que la répartition des cliques dans l'espace doit fournir un critère de décision sur la dissociation de cet espace en régions séparées.

(iv) Les abscisses et ordonnées de la visualisation n'ont pas de valeur intrinsèque mais leur intersection constitue le point à partir duquel les cliques se distribuent dans l'espace sémantique de l'item demandé dans le plan sélectionné.

Nous considérerons successivement les items *fin* (nom, adjectif, adjectif substantivé), *maison* (nom), *propre* (adjectif, adjectif substantivé) et *monter* (verbe intransitif ou transitif, causatif ou non causatif) qui présentent chacun des caractéristiques différentes propres à illustrer notre méthode de lecture des visualisations du D.E.S.

### 6.1 L'item *fin* [149 synonymes, 197 cliques]

Si l'on demande la visualisation de l'item *fin*, le plan (axe) 1 - (axe) 2 délivré par défaut nous offre un spectacle curieux : d'un côté une bande compacte de cliques et à l'autre bout une clique unique, {*absolument, complètement, entièrement, fin, tout à fait*}. Les quatre synonymes de *fin* sont des adverbes, ce qui nous rappelle que dans des expressions rares (ex. je suis fin prêt) *fin* exprime le degré absolu d'une propriété et est effectivement paraphrasable par les quatre adverbes de cette clique. Cette disposition présente un inconvénient patent : même en effectuant un zoom (menu « faire un zoom ») sur la bande des cliques de synonymes non adverbiaux de *fin*, la disposition relative des cliques apparaît déformée par la priorité accordée par défaut à la distinction des emplois adverbiaux. Il est donc profitable de faire 'tourner' la visualisation en sélectionnant un plan proche, en l'occurrence 1-3 (cf. Figure 6). L'emploi adverbial y reste distinct, mais à une distance réduite, de sorte que le reste de l'espace sémantique de l'item peut se déployer et prend une forme triangulaire.

Pour lire "entre les cliques" cet espace, la méthode que nous préconisons consiste à cliquer sur les points situés aux extrémités du nuage, en l'occurrence pour un triangle sur ses trois extrémités. L'extrémité droite la plus proche de l'emploi adverbial est {*anéantissement, destruction, disparition, effondrement, fin, mort, ruine*}. Tous les synonymes sont nominaux et évoquent la *fin* comme achèvement tragique. La clique située à l'extrémité supérieure gauche est {*filiforme, fin, fuselé, grêle, maigre, mince, svelte*}. Tous les synonymes sont des adjectifs qui véhiculent un sens physique et évoquent la minceur et plus exactement la forme d'un fuseau ou d'un fil. Inversement la clique située à l'extrémité inférieure, toujours à gauche est {*fin, finaud, futé, malin, matois, narquois, roublard, rusé*}. A nouveau tous les synonymes sont des adjectifs, mais qui véhiculent un sens psychologique avec un vaste éventail d'évaluations du positif (*futé*) au négatif (*matois, roublard*). Comment passe-t-on de la finesse physique à la finesse psychologique ? En sélectionnant quelques synonymes de haut en bas, ce qui fait apparaître leur « enveloppe convexe », on voit que *rusé* partage une région avec *intelligent*, qui a son tour en partage une avec *beau*, celui-ci avec *délicat* et ce dernier avec *svelte*. C'est donc manifestement l'idée de délicatesse (entre l'ossature délicate et la remarque délicate) qui constitue le domaine de recouvrement entre les deux ordres de finesse. Et comment les emplois nominaux et adjectivaux se rejoignent-ils ? Parmi les synonymes nominaux, l'espace de *mort* empiète sur celui de *but* et celui-ci sur *perfection*, qui a son tour a une clique en commun avec *beau* : {*beau, fin, perfection*}. Mais en quoi les noms *perfection*

et *fin* ont-ils à voir avec l'adjectif *beau* ? En fait la clique concerne les emplois substantivés *le beau* et *le fin* (et non *la fin*). En effet, pensons à une histoire et à sa chute : le plus *beau* ou le *fin* de l'histoire c'est le point de *perfection* de celle-ci (dans l'ordre de l'humoristique ou du grotesque). Or nous observons que cette clique adjectivo-nominale est disposée pratiquement au point de croisement des abscisses et des ordonnées, ce qui suggère que la sélection du plan 1-3 n'altère pas l'équilibre général de l'espace de l'item *fin* entre ses emplois adverbiaux, nominaux, adjectivaux substantivés et proprement adjectivaux.

### 6.2 L'item *maison* [95 synonymes, 127 cliques]

La visualisation par défaut (plan 1-2) de l'item *maison* a une forme de banane ou de croissant de lune aux extrémités orientées vers le haut et n'est pas très parlante, dans la mesure où le sens de *maison* considéré comme central dans tous les dictionnaires, à savoir *logement* ou *foyer*, y apparaît, certes au milieu en termes d'abscisses, mais dans la région la plus basse en termes d'ordonnées. Il paraît donc souhaitable de rapprocher les cliques comportant ces deux items du centre de la représentation. On conserve l'enveloppe des deux synonymes *logement* et *foyer* en changeant les axes jusqu'à ce qu'ils se rapprochent du centre de la représentation sur le plan 3-4. Cette 'rotation' se révèle d'ailleurs fournir une clé de la polysémie de *maison*. En effet sur le plan 3-4 l'enveloppe de *logement* apparaît approximativement au centre de celle de *foyer* qui s'étend vers la gauche et surtout vers la droite. La forme globale du nuage est cette fois celle d'une croix qui suit grossièrement l'axe des abscisses et celui des ordonnées.

En commençant par le bas où la branche inférieure de la croix est la plus courte, la clique extrême est {*bouge, galetas, maison, réduit, taudis*}. Chacun de ces synonymes évoque un *logement* sans confort, voire misérable. La clique extrême de la branche gauche est {*maison, naissance, origine, race*}. Nous ne sommes plus ici dans l'ordre du *logement*, mais d'une extension de la notion de *foyer* à un *lignage* (par exemple la *maison de Savoie*). L'enveloppe de *foyer* recoupe celle de *lignée* qui elle-même empiète sur celle d'*origine*. La branche supérieure de la croix a pour clique extrême {*boîte, entreprise, firme, maison, établissement*}. Sur la Figure 7 on voit que l'enveloppe d'*entreprise* coupe celle d'*édifice* qui elle-même coupe celle de *bâtiment*. On passe donc progressivement sur l'axe des ordonnées du *logement*, de l'*habitation* (misérable, puis standard), au *bâtiment*, neutre quant à son utilisation, à l'*édifice* dont le sens figuré aboutit à *entreprise*. Enfin la branche droite de la croix se termine par la clique {*domesticité, gens, maison, serviteur*}. C'est l'item *gens* qui s'étend le plus en direction de *foyer*, relié par le mot *suite*, laquelle *suite* pouvant être un ensemble de *serviteurs* aussi bien que le prolongement d'une *lignée*. L'axe des abscisses est donc réservé à des groupes humains, dans l'ordre généalogique du côté gauche (le *lignage*) et dans l'ordre sociologique du côté droit (tous les *gens* gravitant autour d'une *maison*, en particulier les *domestiques*).

La fonction centrale de l'item *foyer* est donc confirmée : unité de dénombrement des familles dans les villages, le *foyer* est le lieu par excellence de la vie commune du clan, de la famille et des domestiques, où naît aussi la progéniture qui dans les familles 'bien nées' perpétuera le *lignage*.

### 6.3 L'item *propre* [101 synonymes, 113 cliques] <sup>8</sup>

*Propre* est un adjectif qui présente des nombreux emplois substantivés (*le propre de l'homme, c'est du propre !, mettre au propre*, etc.), en outre trois types au moins d'emplois syntaxiques, intransitif (ex. *le sens propre*), régissant un SN à référent individuel (ex. *ce qui m'est propre*) ou une construction infinitive (ex. *un argument propre à vous convaincre*) et enfin l'aptitude à se placer dans l'acception 'possessive' avant le nom qui le régit (ex. *sa*

<sup>8</sup> cf. François, J. & Manguin, J.L. (à paraître).

*propre fille ; de ses propres mains*). La visualisation par défaut sur le plan 1-2 permet de mettre en évidence le partage des deux premières propriétés par les agrégats de cliques de synonymes et inversement l'idiosyncrasie de la troisième. Encore une fois notre méthode consiste à prendre appui sur les cliques extrêmes du nuage qui a cette fois grossièrement la forme d'un croissant dont les pointes sont orientées vers le bas. La pointe inférieure droite est occupée par la clique {*entraîné, exercé, expert, habile, propre*}. Les quatre synonymes ont en commun de pouvoir régir une construction infinitive ou une nominalisation d'action (ex. *entraîné à courir, exercé au tir, expert à dénouer les fils d'une intrigue, habile à conduire un bolide*), ce qui correspond à l'un des deux emplois transitifs de *propre*. L'extrémité supérieure du croissant est occupée par la clique {*individuel, particulier, personnel, privé, propre*} où deux synonymes partagent la propriété syntaxique de pouvoir régir un complément nominal désignant un individu (*qch est particulier / personnel à qqn*). L'enveloppe de *particulier* inclut ou croise celle de *typique, spécifique, caractéristique* qui eux aussi peuvent régir ce type de complément. On est donc clairement dans la région de l'autre emploi transitif de *propre*. Comment passe-t-on de l'un à l'autre et où situer l'emploi de *propre* opposé à *malpropre* ? Sur la figure 8, on voit que l'enveloppe d'*expert* recoupe celle d'*apte*. Cette enveloppe recoupe celle de *convenable* qui elle-même partage des cliques avec celle d'*honnête*, laquelle recoupe également celles de *soigné* et de *net*. C'est donc *convenable* qui sert d'interface entre l'expression de l'aptitude (un outil *convenable* pour accomplir une tâche) et la bienséance (une jeune fille très *convenable*, c'est-à-dire respectant les convenances). La jeune fille en question a toute chance d'être *honnête* et *soignée*, « nette de sa personne », voire d'avoir du *chic* et même de faire preuve de *distinction*. Ces associations de propriétés ne sont pas insignifiantes. D'une part l'enveloppe de *net* conduit à celle de *distinct*, et ce qui distingue deux entités ou individus A de B, c'est ce qui est *propre, particulier, personnel* à A ou à B. Ainsi *net* et *distinct* assurent l'interface entre les emplois intransitifs de *propre* et les emplois où *propre* prend un complément à référent individuel. D'autre part les deux adjectifs *chic* et *caractéristique* ont un pendant substantival, *le chic* et *la caractéristique*<sup>9</sup>. Ce caractère de janus syntaxique leur assure un rôle d'intermédiaire entre les emplois de l'adjectif *propre* comme synonyme de *particulier* et ceux de l'adjectif substantivé *le propre* comme synonyme de *la particularité*. Et on constate que la clique extrême de la pointe inférieure gauche du croissant {*caractère, essence, nature, propre, propriété, qualité*} est l'exacte contrepartie nominale de la clique supérieure (gauche) {*individuel, particulier, personnel, privé, propre*}. Il se produit donc un dédoublement de la région du distinctif, selon que l'expression est adjectivale ou nominale.

Donc on ne passe pas directement d'un emploi syntaxique (*qqn est propre à faire qch*) à l'autre (*qch est propre à qqn*), c'est par l'emploi intransitif que transite l'espace sémantique. Reste la question de la position occasionnelle de *propre* devant le nom qui le régite. Les emplois 'possessifs' de *propre* n'ont guère d'équivalent, tout au plus *personnel* peut-il se substituer maladroitement à *propre* dans *sa propre fille* ou *en mains propres*. Et *personnel* ne peut pas être antéposé au nom régissant. Cette propriété de *propre* a donc un caractère idiosyncrasique.

#### 6.4 L'item *monter* [106 synonymes, 145 cliques]

Enfin nous allons terminer cette évocation des différentes classes de mots par le verbe *monter*. Deux propriétés de ce verbe nous intéressent spécifiquement. En premier lieu *monter* présente des emplois intransitifs (*le niveau monte*), pronominaux (*le prix se monte à 10*

<sup>9</sup> On peut vérifier sur la figure que l'enveloppe de l'item *caractère*, purement nominal ne recoupe celle de l'item *caractéristique*, adjectif et nom, que dans la région inférieure dévolue aux synonymes nominaux dans le champ de la particularité, dont le summum est *l'essence*.

*Euros*) et transitifs, lesquels présentent deux sous-types, les emplois transitifs causatifs, où le référent de l'objet direct subit une action accomplie par celui du sujet et a donc clairement un statut d'affecté ou de patient (*le gouvernement monte le prix de l'essence ; le porteur monte la valise dans la chambre*) et les emplois transitifs non causatifs, où le référent de l'objet est un repère locatif au long duquel s'effectue la montée (*les coureurs montent la côte*). D'autre part le dictionnaire *Robert Méthodique* introduit deux homonymes (contrairement aux autres dictionnaires *Robert*) en distinguant les emplois relatifs à l'idée d'assemblage et de montage. Nous nous posons donc une double question :

(i) la visualisation du D.E.S. permet-elle de distinguer les types d'emplois syntaxiques de *monter* et en particulier la double distinction syntaxique [+/-transitif] et sémantique [+/-causatif] ?

(ii) la présomption d'homonymie pour les emplois du type *monter un complot, un moteur, un spectacle, une entreprise* est-elle confirmée par une rupture de l'espace sémantique du verbe ?

La visualisation par défaut sur le plan 1-2 a une forme triangulaire (figure 9). A l'extrémité inférieure on trouve comme dernière clique extrême {*ascensionner, escalader, gravir, grimper, monter*}. Ces synonymes sont soit principalement intransitifs (*grimper*), soit uniquement transitifs non causatifs (*ascensionner, escalader, gravir*). La clique extrême à la pointe supérieure droite est {*augmenter, hausser, majorer, monter, rehausser, relever, renchérir, revaloriser*}. Deux de ses constituants sont des verbes à renversement (dans la terminologie de M. Rothemberg 1974) ou labiles (dans celle de M. Larjavaara 2000), *augmenter* et *renchérir*, c'est-à-dire qu'ils peuvent exprimer un changement de degré sur une échelle d'un point de vue causatif (ex. *le gouvernement augmente les taxes*) ou non causatif (ex. *les taxes augmentent*). Les autres sont transitifs causatifs. L'enveloppe d'*augmenter* recoupe à gauche celles de *dresser* et d'*exciter* qui sont causatifs (le référent de l'objet est un patient) et en bas celle de *s'élever* qui introduit un emploi pronominal, donc non causatif). Cette dernière enveloppe partage avec celle de *grimper* la clique {*grimper, monter, s'élever, se hisser*}. On voit donc que les emplois pronominaux (réfléchis ou récessifs) servent d'interface entre les emplois transitifs non causatif (*monter/grimper la côte*) et les emplois transitifs causatifs (*monter/augmenter les taxes*), si bien que les deux types sémantiques d'emplois transitifs causatifs et non causatifs sont clairement distingués.

La clique extrême supérieure gauche est {*combiner, machiner, manigancer, monter, nouer, ourdir, préparer, tisser, tramer, tisser*}. Son étendue exceptionnelle montre l'abondance des synonymes dans la région sémantique de la machination. L'enveloppe de *combiner* recoupe celle d'*assembler* qui recouvre largement celle de *dresser* vers la droite et celle d'*appareiller* vers le bas. La transition entre l'élévation exprimée dans la partie inférieure du nuage et l'assemblage exprimé dans sa partie supérieure gauche passe par l'idée d'accouplement (*l'étalon monte la jument*) qui suppose une élévation du mâle au dessus de la femelle et celle de monture (*le cavalier monte la jument*). L'enveloppe de *couvrir* s'étend aux deux types d'emplois. on peut donc en conclure que l'espace sémantique de *monter* ne présente aucune rupture justifiant l'hypothèse d'homonymie du *Robert Méthodique*.

Quant à la zone centrale, elle a une composante psychologique repérée à l'aide de l'enveloppe d'*exciter* et une composante de déplacement dans un véhicule repérée par l'enveloppe centrale de *prendre* (*monter dans / prendre un véhicule*).

## 6.5. De la visualisation de l'espace sémantique d'un item à la composition de l'article lexicographique correspondant

Les visualisations géométriques du D.E.S. établissent des relations de proximité entre les sens d'un item lexical assimilés à des cliques de synonymes, mais elles sont muettes sur les processus de diversification sémantique (par métaphore, métonymie, généralisation, spécification, etc.) qui ont produit historiquement la diversité des sens observable au début du 21<sup>ème</sup> siècle. En quoi ces visualisations peuvent-elles constituer une aide à la décision pour les lexicographes ? Ceux-ci sont confrontés à trois types de décision : (a) dissocier ou pas l'item en plusieurs homonymes, (b) regrouper les sens apparentés, c'est-à-dire établir une hiérarchie d'emplois, (c) linéariser l'article lexicographique, c'est-à-dire orienter les relations entre les sens apparentés et entre les groupes de sens. Sur le premier point, le D.E.S. peut fournir une aide précieuse par la distinction en composantes connexes, qui suggère une dissociation de l'item en homonymes. Sur le second, le tuilage local des cliques de synonymes sur les visualisations dégage clairement des groupes de sens étroitement apparentés. En revanche l'interprétation d'une affinité de sens comme une dérivation va au-delà du domaine de pertinence du D.E.S. C'est finalement un avantage, car les visualisations du D.E.S. peuvent constituer un terrain neutre dans la controverse entre sémanticiens sur la notion de dérivation de sens. La réalité historique de cette dérivation est bien cernée pour une langue comme le français où l'activité dictionnaire couvre cinq siècles (cf. Darmesteter 1897/1979). En revanche sa réalité cognitive reste sujette à débat (cf. note 6).

## 7. Exploitation psycholinguistique

L'expérimentation psycholinguistique trouve dans le D.E.S. un outil théorique et méthodologique extrêmement riche à plus d'un titre, et dont trois aspects se trouveront développés dans l'illustration expérimentale présentée ici.

- Dans la mesure où le D.E.S. fournit la liste de tous les synonymes d'un mot donné, il permet de dessiner les contours et le contenu du réseau sémantique qui s'établit autour de ce mot, réseau sémantique construit à partir d'une analyse de la langue. Du point de vue du psychologue, les données fournies par le D.E.S constituent une base de référence informant sur le volume d'éléments potentiels contenus dans une mémoire hypothétique, qui donnerait un cadre pour situer le réseau des significations qui caractérise les mémoires sémantiques individuelles. Il pourrait permettre aussi de cibler les pratiques langagières individuelles qui s'écartent, ou qui innovent, par rapport à la norme.
- Dans ses caractéristiques liées au dénombrement, le D.E.S. fournit un état objectif de la taille potentielle du réseau sémantique lié à un mot donné. Deux items peuvent donc être directement comparés sous cet aspect.
- Dans la visualisation fournie par le D.E.S., les synonymes d'un mot sont organisés en fonction de leur plus ou moins grande proximité sémantique. Les différentes significations d'un mot y apparaissent donc à la fois dans toute leur variété, et dans le cadre de sous-réseaux sémantiques, ou cliques. L'approche psycholinguistique de la polysémie y trouve une base de données, permettant de situer, entre les cliques et au sein des différentes cliques, la progression de l'appropriation du lexique.

Sur cette base, la partie psycholinguistique de cette étude consiste en une tentative pour repérer l'espace sémantique associé à 10 adjectifs dans le lexique mental de sujets francophones, et ceci pour 6 classes d'âge, correspondant à la scolarité de dernière année du primaire en France (CM2), à deux années du collège (6<sup>ème</sup> et 4<sup>ème</sup>), à une année de lycée (1<sup>ère</sup>),

ainsi qu'à deux époques des études universitaires (1<sup>ère</sup> et 3<sup>ème</sup> année). Les étudiants de 3<sup>o</sup> année à l'université (correspondant à la licence) ont une particularité: ils sont étudiants en sciences du langage, et par hypothèse, doivent jouer le rôle d'experts dans le cadre de la situation proposée ici.

Les expériences psycholinguistiques ont essentiellement deux buts: d'une part examiner la progression du lexique tout au long de la scolarité, c'est à dire à des âges où cet aspect des choses ne se trouve plus sous les projecteurs de la recherche (sur cet aspect, l'essentiel des recherches se situent juste avant, pendant ou juste après l'explosion langagière); d'autre part se focaliser sur la découverte de la polysémie des mots, donc sur un facteur d'évolution sémantique et non plus seulement lexicale.

Lorsqu'on consulte le D.E.S. pour trouver *le* mot qui convient dans *tel* contexte, une des choses que l'on apprend, s'il est besoin, c'est que les relations entre représentations sémantiques et langage ne sont pas simples. Il faut avoir présent à l'esprit que ces relations augmentent encore en complexité lorsqu'on introduit dans l'étude du langage un facteur de développement. Nous assistons en effet:

- à une **augmentation quantitative du lexique**. A l'apprentissage rapide et cumulatif des premières années (estimé à 1500 mots par an à partir de 3 ans) succède une phase d'acquisition continue, de rythme variable en fonction de l'environnement linguistique, de la scolarité, des lectures. La fréquence d'usage joue un grand rôle dans les acquisitions initiales et garde une fonction dans ces acquisitions plus tardives (Gineste & Le Ny, 2002).
- à un **enrichissement de la signification lexicale**. A l'origine, un mot peut avoir une signification très restreinte, liée à un contexte donné. Dans les premiers temps, cette signification lexicale va s'enrichir par une généralisation de l'emploi du mot à des contextes de situation variés. Nous insisterons ici, non pas sur cet aspect, mais sur la découverte de la polysémie du mot, donc sur le fait qu'un mot a rarement une signification unique. Les linguistes soulignent que, loin d'être un défaut marginal des langues, la polysémie est une qualité importante dont l'effcience reste pour une grande part à expliquer (François, 1994). Les psychologues font de plus l'hypothèse que la connaissance de la polysémie est un facteur important pour les performances cognitives (Klein & Murphy, 2001).
- à une **modification de l'organisation des significations en mémoire à long terme**, et à des changements dans la force des liens qui unissent les unités du réseau sémantique. Ces liens et leur force relative déterminent la disponibilité des mots (*i.e.* la présence des mots en mémoire à long terme, à mettre en relation avec l'évolution du lexique évoquée plus haut), et leur accessibilité, qui dépend plus directement du contexte de situation dans lequel ils sont sollicités. Un mot disponible peut, dans un certain nombre de conditions, n'être pas accessible. Sur le plan mnésique, de nombreuses recherches s'intéressent actuellement à l'âge d'acquisition des mots ou AA (Brysbaert, Van Wijnendaele, & de Deyne, 2000), âge d'acquisition qui conditionne leur temps de présence en mémoire à long terme, et qui semble avoir des répercussions sur les performances du sujet dans des tâches de nature sémantique. Nous examinerons si cette variable a un effet dans la tâche proposée ici.

## 7.1 Expérience 1

Nous rendrons compte ici d'une épreuve de production de synonymes pour dix adjectifs, tous de bonne fréquence d'usage (supérieure à 39 pour 1/1 000 000 selon la base de données Lexique (New, Pallier, Ferrand, & Matos, 2001)), et dont l'âge d'acquisition a été déterminé

préalablement à l'expérience (AA compris entre 3 ans et 6 ans), suivant la méthode préconisée par Gilhooly & Gilhooly (1980). Ces dix adjectifs sont *beau*, *curieux*, *difficile*, *doux*, *froid*, *grand*, *gros*, *propre*, *sec* et *triste*. La consultation du D.E.S. pour ces adjectifs donne un éventail de synonymes compris entre 54 (pour *curieux*) à 194 (pour *beau*). Dans la mesure où ces deux adjectifs se distinguent à la fois quant à leur fréquence d'usage (respectivement 39 et 194), leur AA (respectivement 6 ans et 3 ans), et leur "potentiel" de synonymes, nous nous focaliserons sur les données les concernant pour la tâche de production, sans pourtant limiter notre propos à ces deux adjectifs<sup>10</sup>.

La tâche de production était très simple de conception. Il s'agissait, pour les participants, d'écrire, pendant une durée de 2 minutes, tous les synonymes d'un adjectif donné venant à l'esprit. Les données recueillies avaient pour but de mettre en évidence l'étendue de l'environnement synonymique des dix adjectifs cités, ses caractéristiques statistiques (fréquence de production pour chacune des populations concernées) et ses caractéristiques sémantiques dominantes (en terme de variations des significations susceptibles d'exprimer la polysémie des mots). De manière générale, nous retiendrons deux tendances principales des taux de production qui ont été recueillis:

- Le nombre brut des adjectifs produits augmente régulièrement jusqu'au DEUG1 (1<sup>o</sup> année universitaire), et diminue pour les étudiants en licence de Sciences du Langage (3<sup>o</sup> année universitaire). Cette baisse des productions à ce niveau traduit non pas un effondrement des compétences langagières pour ces étudiants (!), mais révèle une autre stratégie de leur part: ils remplacent la quantité (tous les synonymes venant à l'esprit) par la qualité (puis-je considérer *bel* et *bien* cet adjectif comme un synonyme?). Bref, nous supposons que ces étudiants n'ont pas suivi les consignes.
- Les disparités enregistrées entre le nombre des synonymes produits selon les adjectifs-cibles restent présentes pour tous les âges, même si, pour les niveaux scolaires les plus avancés, on enregistre une modération des disparités. Ces différences dans les taux de production sont corrélées positivement avec la fréquence d'usage des cibles, leur AA, et le nombre des synonymes fournis par le D.E.S (Coefficient de corrélation  $r = 0,72$ ).

Nous trouverons ci-dessous une illustration de l'importance de ces productions en fonction des niveaux scolaires, pour les adjectifs *beau* et *curieux*.

---

<sup>10</sup> Pour avoir accès à l'ensemble des résultats sur les 10 adjectifs, vous pouvez consulter la page personnelle de Françoise Cordier au LaCo (<http://www.mshs.univ-poitiers.fr/laco/>).



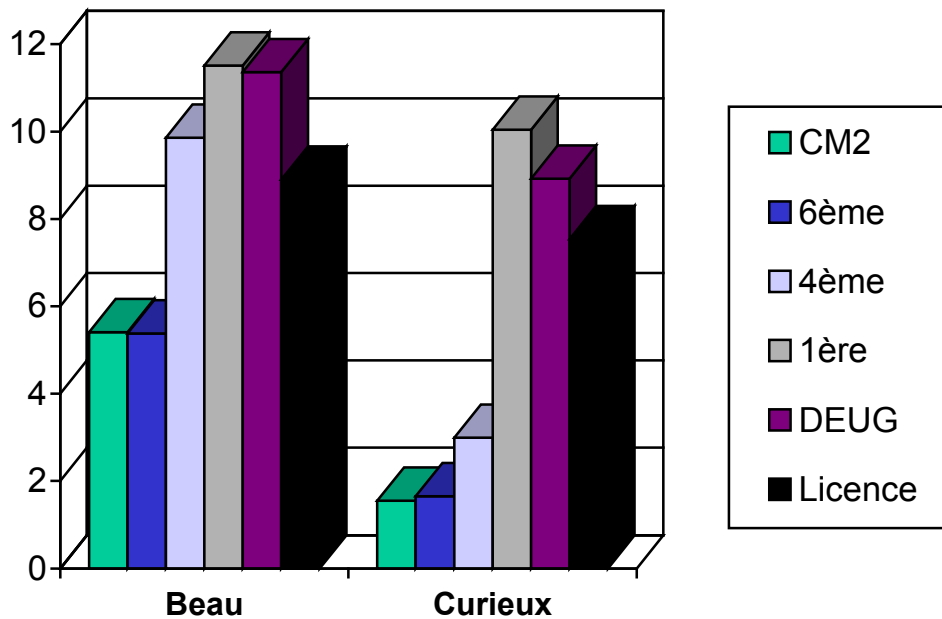


Figure 10 : Nombre brut de synonymes pour *Beau* et *Curieux* en fonction des niveaux scolaires.

Sur ce dernier point, les fortes corrélations entre productions et données fournies par le D.E.S. nous semblent donc bien être le signe que le D.E.S. peut être exploité comme point de référence par rapport au contenu d'une mémoire individuelle, non pas en terme de nombre absolu de synonymes bien évidemment, mais en terme de relation d'ordre (plus de synonymes pour tel adjectif source que pour un autre). A contrario, si on suspecte chez certaines catégories de personnes des déficits dans tel ou tel champ sémantique - cadre dans lequel se font bon nombre de recherches en neuropsychologie (Warrington & Shallice, 1984) - on peut s'attendre alors à ne plus observer chez eux ces fortes corrélations entre productions individuelles et base de données, et avoir un argument supplémentaire pour objectiver le déficit.

Si nous quittons le plan quantitatif pour un plan qualitatif, c'est à dire pour un plan sémantique, l'évolution des productions fait apparaître deux profils bien différents:

- Des synonymes, donnés fréquemment par les plus jeunes, disparaissent des productions des plus âgés. Nous illustrons cette observation ci-dessous dans le Tableau 4, pour l'adjectif *beau*, en comparant les productions de 6 synonymes. Les 3 adjectifs associés à *beau* par les plus jeunes (*cher*, *propre*, *rare*) ne sont pas mentionnés dans le D.E.S., à l'inverse de ceux cités par les scolaires les plus âgés. Ceci traduit une conformité à la langue pour les scolaires les plus âgés. Pour les plus jeunes, ce n'est pas la seule observation dont nous disposons dans notre base de données. Par exemple, à l'adjectif *propre* se trouvent associés *beau* et *joli*; et à l'adjectif *doux* à nouveau *beau* et *joli*. Toutes ces données témoignent, chez les scolaires les plus jeunes, d'une surexploitation, d'une surgénéralisation à des situations très diverses de quelques adjectifs de bonne fréquence d'usage et d'AA précoce. A l'origine de cet état de choses peuvent être invoqués les discours parentaux adressés aux petits, dont l'effet serait d'installer préférentiellement ces associations dans le

lexique: "être propre, c'est être beau". Les associations en cause, n'étant plus renforcées, disparaîtraient ensuite d'un vocabulaire moins puéril.

<i>Beau</i>	Licence	Deug	Lycée	4 <sup>ème</sup>	6 <sup>ème</sup>	CM2
<i>cher</i>				<b>13,63</b>	1,96	<b>14,63</b>
<i>propre</i>		4,25	7,14	2,27	1,96	<b>14,63</b>
<i>rare</i>		2,12	2,38	<b>20,45</b>	7,84	
<i>joli</i>	<b>97,6</b>	<b>97,87</b>	<b>88,09</b>	<b>36,36</b>	<b>70,59</b>	<b>51,22</b>
<i>magnifique</i>	<b>45,24</b>	<b>68,08</b>	<b>52,38</b>	<b>50</b>	<b>31,37</b>	<b>14,63</b>
<i>mignon</i>	<b>52,38</b>	<b>42,55</b>	<b>40,48</b>	<b>18,18</b>	<b>15,69</b>	<b>21,95</b>

Tableau 4 : Taux de production (exprimés en %) pour 6 synonymes de l'adjectif *beau*, en fonction des niveaux scolaires<sup>11</sup>.

- Des synonymes apparaissent tardivement: absents des productions des plus jeunes, ils émergent dans la période du lycée. Il s'agit de significations associées à l'adjectif-cible qui apparaissent plus tard dans son environnement synonymique. Par exemple, l'adjectif polysémique *froid* (cf. Tableau 5) est très accessible lorsqu'il renvoie à la propriété "température" : *gelé*, *glacé* sont des productions présentes dans toutes les populations testées, y compris chez les plus jeunes. Par contre, lorsqu'il traduit une caractéristique de la personnalité (*distant*, *asocial*, *antipathique*), les synonymes correspondants ne se trouvent cités qu'au lycée ou à l'université. On trouve dans les résultats de multiples exemples de cette configuration: pour l'adjectif-cible *sec*, les significations liées à l'absence d'humidité se trouvent accessibles plus tôt que celles (*sévère*, *rigide*) liées aux traits de personnalité de l'individu, ainsi que celles qui renvoient au sens de *maigre*, *décharné*.

<i>Froid</i>	Licence	Deug	Lycée	4 <sup>ème</sup>	6 <sup>ème</sup>	CM2
<i>antipathique</i>	4,76	<b>10,63</b>	7,14			
<i>asocial</i>	2,38	<b>10,63</b>				
<i>distant</i>	<b>33,33</b>	<b>23,4</b>	<b>11,9</b>			
<i>frais</i>	<b>69,04</b>	<b>25,53</b>	<b>23,8</b>	6,82	9,8	<b>12,18</b>
<i>gelé</i>	<b>40,47</b>	<b>55,32</b>	<b>40,47</b>	<b>40,99</b>	<b>64,7</b>	<b>56,09</b>
<i>glacé</i>	<b>26,19</b>	<b>46,8</b>	<b>61,9</b>	<b>45,5</b>	<b>56,9</b>	<b>43,9</b>

Tableau 5 : Taux de production (exprimés en %) pour 6 synonymes de l'adjectif *froid*, en fonction des niveaux scolaires.

<sup>11</sup> Les 3 adjectifs du second groupe sont représentés dans les productions pour toutes les classes d'âge.

Quelles sont les raisons de ces répartitions? On peut supposer que l'arrivée dans les programmes scolaires de l'étude systématisée d'œuvres littéraires, de plus en plus diversifiées, a un rôle non négligeable. On peut supposer également que les progrès dans l'apprentissage des langues étrangères s'accompagnent *de facto* d'une sensibilisation à la polysémie. Les significations multiples du même adjectif dans une langue donnée se trouvent exprimées par des mots différents dans une autre langue. Par exemple, *sec* est traduit par *dry* dans le sens d'absence d'humidité, mais par *hard* ou *cold* s'il s'agit d'une caractéristique de la personnalité. Ceci peut inciter élèves ou étudiants à faire un retour sur leur propre fonctionnement langagier, et prendre une conscience plus claire de la pluralité des significations attachées à un mot.

Il est important à ce niveau de revenir à cette configuration des résultats en faisant la relation avec le fonctionnement de la mémoire à long terme. Nous avons, pour cette première expérience, utilisé une tâche de production en temps limité. Les productions des participants correspondent donc à des synonymes qui sont tout à la fois disponibles et accessibles. Nous observons simultanément une augmentation générale des productions (cf. figure 10) et au cœur de celle-ci, une mise en place de la polysémie. C'est à dire à la fois un accroissement du lexique et un enrichissement sémantique du lexique existant.

Par rapport à la présence somme toute tardive de la polysémie dans les réponses, deux interprétations sont possibles: la première stipulerait que la découverte de la polysémie a bien lieu au cours de la scolarisation au lycée, pour les raisons évoquées plus haut, certaines significations n'étant donc pas accessibles parce que non disponibles avant. La deuxième interprétation, plus souple, soulignerait que les scolaires parmi les plus jeunes pourraient éventuellement avoir déjà présente en mémoire la polysémie de l'adjectif, mais la signification secondaire aurait une force en mémoire beaucoup trop faible pour qu'elle puisse être activée dans ce type de tâche. Les études littéraires, et les études des langues en général, auraient pour conséquence de renforcer ces associations en mémoire. De plus, les plus jeunes pourraient être cognitivement moins "flexibles" au cours de la tâche, c'est à dire demeurer dans le même schéma de signification, à la différence des plus âgés, capables de passer d'une signification à l'autre dans un temps relativement court.

Nous utilisons "signification secondaire" dans le sens qui lui est donnée dans les nombreux travaux de psycholinguistique, à savoir, pour un mot ambigu, le sens qui apparaît le moins fréquemment pour une population donnée dans une tâche d'association. Certains mots ambigus ont en effet deux significations équiprobables, alors que d'autres ont une signification dominante et une (ou des) signification(s) secondaire(s) (Marquer, Lebreton, Léveillé, & Dioniso, 1990). Il nous semble ici être dans ce dernier schéma. Si nous nous reportons au Tableau 5, nous trouvons en effet que globalement, *frais / gelé / glacé* sont beaucoup plus souvent cités à l'université que *antipathique / asocial / distant*.

Pour conclure sur cette expérience 1 dans ses relations avec le D.E.S., on mesure bien que l'interprétation qui est donnée des réponses relatives aux différentes populations scolaires est partie liée au statut des différents termes dans le D.E.S., et en particulier à l'opposition présent/absent dans l'ensemble des synonymes de tel mot.

## 7.2 Expérience 2

Nous rendons compte ici d'une tâche de jugement dont le but est de faire estimer par d'autres populations appartenant aux mêmes niveaux scolaires que précédemment la proximité sémantique de 5 synonymes par rapport à leur adjectif-cible. Nous faisons l'hypothèse que

cette proximité sémantique devrait être sensible à la fois à la progression du lexique et à la découverte ou au renforcement de la polysémie.

La procédure consistait à proposer aux participants les dix adjectifs-cibles préalablement sélectionnés (1 par feuille), accompagnés chacun de 5 de leurs synonymes, choisis autant que faire se peut de manière à ce que la polysémie de la cible se trouve illustrée. La consigne demandait aux participants, en temps libre, de sélectionner, parmi les synonymes, le plus proche par son sens de la cible, et de lui donner le rang 1. Puis de procéder de la même façon pour les 4 autres synonymes. Des comparaisons statistiques sur les classements ont ensuite été appliquées. Ces comparaisons avaient pour but de mettre en relief pour chacun des couples cible-synonyme l'évolution des jugements, ceci en fonction du niveau scolaire.

Les résultats ont mis en évidence des différences importantes selon les cibles. Ils sont illustrés ici par deux exemples, les adjectifs *propre* et *curieux*. Les jugements moyens y sont indiqués en gras lorsqu'on observe des différences statistiquement significatives<sup>12</sup> entre niveaux scolaires. Plus précisément, la différence significative est enregistrée entre les 2 niveaux scolaires pour lesquels les données sont écrites différemment.

	Licence	Deug	Lycée	4 <sup>ème</sup>	6 <sup>ème</sup>	CM2
<i>beau</i>	<b>4.50</b>	<b>4.53</b>	<b>4.37</b>	(3.95)	(3.93)	(3.65)
<i>hygiénique</i>	(2.50)	(2.21)	(2.31)	(1.98)	(2.24)	(2.54)
<i>joli</i>	<b>4.32</b>	<b>4.26</b>	<b>4.16</b>	<b>4.14</b>	<b>4.05</b>	(3.14)
<i>lavé</i>	<b>1.68</b>	<b>1.79</b>	<b>1.86</b>	<b>1.98</b>	(2.27)	(2.49)
<i>net</i>	<b>2.00</b>	<b>2.21</b>	<b>2.31</b>	(2.95)	(2.51)	<b>3.19</b>

Tableau 6: Jugements de proximité sémantique pour 5 synonymes de l'adjectif *propre*.

	Licence	Deug	Lycée	4 <sup>ème</sup>	6 <sup>ème</sup>	CM2
<i>bizarre</i>	<b>2.00</b>	<b>2.58</b>	(3.08)	(3.18)	(3.24)	(3.27)
<i>étrange</i>	<b>2.05</b>	<b>2.28</b>	<b>2.48</b>	(3.07)	(2.76)	(3.16)
<i>fouineur</i>	(2.77)	(2.56)	(2.42)	(2.40)	(2.56)	(2.22)
<i>impoli</i>	<b>4.50</b>	<b>4.65</b>	<b>4.31</b>	(3.63)	(3.88)	(3.59)
<i>intéressé</i>	<b>3.68</b>	2.93)	(2.71)	(2.72)	(2.56)	(2.76)

Tableau 7 : Jugements de proximité sémantique pour 5 synonymes de l'adjectif *curieux*

Les résultats témoignent à nouveau de deux types d'évolution:

- **Les effets d'un enrichissement du lexique sur la réorganisation de l'espace des synonymes.** C'est une modification dont on trouve surtout des exemples entre la fin de la scolarité primaire et le début de la scolarité secondaire. Dans le cas de *propre*, *joli* va ainsi voir sa distance par rapport à la cible augmenter. C'est un adjectif dont l'AA est précoce, qui a une bonne fréquence d'usage, et qui est donc bien connu, on peut le penser, des plus jeunes. Il a en fin de scolarité primaire un classement qui va se

<sup>12</sup> Les tests statistiques ont pour fonction de déterminer la probabilité d'obtenir un résultat donné. Un test dont le résultat est "significatif" nous apprend que le résultat observé a très peu de chances d'être obtenu par hasard. Par exemple, un seuil de significativité de 0,05 signifie que les différences observées ont seulement 5% de chance d'être obtenues par hasard, et ont donc une probabilité très forte de faire intervenir les niveaux scolaires (voir Vessereau, 1947).

détériorer par la suite. Lorsque d'autres mots viendront en compétition, il sera jugé plus distant. Le classement de départ de *joli* est certainement une autre conséquence de l'emploi d'une forme puérile de langage avec les enfants, déjà signalée à partir des résultats de l'expérience de production.

Ce type de remaniement est assez fréquent. On en trouve par exemple une autre illustration avec l'adjectif *difficile*. Son synonyme *complexe* voit sa distance sémantique diminuer, alors que *impossible* voit sa distance sémantique augmenter. On peut supposer là encore des habitudes langagières pour *impossible*, d'ailleurs davantage produit par les enfants les plus jeunes de notre échantillon.

- **Les effets d'un enrichissement sémantique sur l'organisation de l'espace des synonymes.** Cette modification intervient plutôt dans le secondaire, au niveau du lycée et de l'université. Pour l'adjectif *curieux*, une de ses significations (*bizarre*) se trouve jugée très distante par les scolaires alors qu'elle est jugée significativement plus proche à l'université. *Intéressé* verra par contre sa distance sémantique augmenter. Nous avons bien là mise en évidence d'un ajustement de la proximité sémantique qui prend en compte la polysémie.

L'expérience 2 exploite apparemment moins le dictionnaire des synonymes. Il serait pourtant possible d'envisager d'examiner si les jugements des sujets (en terme de relation d'ordre dans la proximité sémantique) rejoignent les données fournies par le D.E.S. quant au classement des synonymes, et en particulier si la participation d'un même synonyme à un grand nombre de cliques augmente sa proximité sémantique. Dans le cas présent, nous avons simplement trop peu de synonymes (5 par adjectif-cible) pour mener valablement une telle étude.

### 7.3 Conclusion

Nous avons ici illustré le fait que le D.E.S. est une base de données intéressante pour la conception d'expérimentations en psycholinguistique. Cette base de données permet, dans le cas présent, d'avoir une maîtrise du matériel utilisé dans le cadre d'une approche de la polysémie. Elle permet en particulier de situer dans un espace sémantique objectivé les significations qui sont maîtrisées, et ceci en fonction du niveau scolaire, dans leur progression.

Les deux expériences menées ont mis en évidence à la fois la progression du lexique, et la conquête de la polysémie des adjectifs par deux procédures différentes, une tâche de production et une tâche de jugement. Les évolutions enregistrées ne sont pas progressives, mais présentent des accélérations à certaines époques, qui semblent de façon notable liées à des apprentissages scolaires. Si l'on se borne au plan purement sémantique de l'acquisition, on peut souligner que les indicateurs objectifs que sont l'AA et la fréquence d'usage ne permettent pas de supposer un contenu plus ou moins riche, dans la mesure où le contenu sémantique de mots acquis tôt peut rester élémentaire pendant des périodes assez longues. La densification du réseau sémantique s'inscrit dans la durée.

Nous avons mené ces expérimentations sur des adjectifs, mais il n'y a aucune raison de penser que ces modifications soient liées à une catégorie grammaticale déterminée. La sémantique des noms, celle des verbes, devraient suivre la même évolution. On peut également penser que ce type d'études sur le lexique est susceptible de donner des compléments d'information quant à la compréhension et la production des métaphores par les enfants.

## 8. Bilan et perspectives

Nous avons cherché à montrer ici que le D.E.S., tel qu'il se présente en ligne sur le site du CRISCO, peut devenir un outil de découverte et d'exploration de l'espace sémantique des items hautement polysémiques du français (en particulier de ceux qui donnent lieu à des «dégrouperments homonymiques» dans certains dictionnaires, tels le Dictionnaire du Français Contemporain ou le LEXIS, mais aussi à un moindre degré le Petit Robert ou le T.L.F.). La méthode d'exploration illustrée au §6 est d'un maniement aisé, même si l'usage des techniques du zoom ou du changement d'axes demande un certain entraînement.

L'espace sémantique ainsi 'cartographié' a cependant un caractère 'totalisant', c'est-à-dire qu'il tient compte d'un inventaire de relations synonymiques bien supérieur à celui dont peuvent disposer les locuteurs francophones les plus experts. Les expériences psycholinguistiques présentées au §7 suggèrent qu'à travers les classes d'âge non seulement l'étendue du lexique disponible augmente et donc le nombre de synonymes potentiels, mais aussi — sans doute en grande partie à cause de l'institution scolaire et de l'entraînement croissant à un maniement précis de la langue française — que l'aptitude à définir et à reformuler s'affine, ce qui permet aux francophones de repérer les substitutions lexicales pertinentes, et donc à intégrer les relations synonymiques. Un pan de notre recherche consistera donc à l'avenir à élaborer un outil capable de modéliser l'espace sémantique des items polysémiques pour différentes classes d'âge.

## Bibliographie

- Bailly R. (1947), *Dictionnaire des synonymes*, Paris, Larousse.
- Bénac H. (1956), *Dictionnaire des synonymes*, Paris, Hachette.
- Berge C. (1967), *Théorie des graphes et ses applications*, Paris, Dunod.
- Bertaud du Chazaud H. (1979), *Dictionnaire des synonymes*, Paris, Editions Le Robert.
- Brybaert, M., Van Wijnendaele, I., de Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta psychologica*, 104, (215-216).
- Darmesteter, A. (1887/1979), *La vie des mots étudiée dans leurs significations*, Paris, Delagrave [1887], Champ Libre [1979].
- Dubois, J. et al. (1966), *Dictionnaire du Français Contemporain*, Paris, Larousse.
- Dubois, J. et al. (2002) *LEXIS, Dictionnaire de la langue française*, Paris, Larousse.
- François, J. (1994). La sémantique. In R. Ghiglione & J.F. Richard (Eds), *Cours de Psychologie, Mesures & Analyse*. Paris, Dunod.
- François, J. & Manguin, J.L. (à paraître) : « La polysémie adjectivale entre synonymie et sélection contextuelle : le cas de *propre* ». In : H. Bouillon (ed.) *Volume d'hommage à Jacques LEROT à l'occasion de son 65e anniversaire*, Louvain, Presses universitaires.
- François J., Victorri B. & Manguin J.L. (à paraître en 2003), Polysémie adjectivale et synonymie : l'éventail des sens de *curieux*, in O. Soutet (éd.) *La polysémie*, actes du colloque de Paris IV, novembre 2000, Presses Universitaires de la Sorbonne, pp. 157-170.
- Gilhooly, K.J., & Gilhooly, M.L.M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, 71, 105-110.

- Gineste, M.D. & le Ny, J.F. (2002). *Psychologie Cognitive du Langage*. Paris, Dunod.
- Guilbert L. et al. (1971), *Grand Larousse de la langue française*, Paris, Larousse.
- Guizot F. (1864), *Dictionnaire Universel des synonymes de la Langue Française*, Paris, Didier, (7<sup>ème</sup> édition).
- Kahlmann A. (1975), *Traitement automatique d'un dictionnaire de synonymes*, Stockholm, Université de Stockholm.
- Klein, D.E. & Murphy, G.L. (2001). The representation of polysemous words. *Journal of Memory and Language*, 45, 259-282.
- Lafaye P.-B. (1858), *Dictionnaire des synonymes de la Langue Française*, Paris, Hachette.
- Manguin J.L. & Victorri B. (1999), Représentation géométrique d'un paradigme lexical, actes de la conférence *TALN 1999*, pp. 363-368.
- Marquer, P., Lebreton, M., Léveillé, M., & Dioniso, C. (1990). A quel moment du traitement des homographes intervient la fréquence relative de leurs acceptions? *L'Année Psychologique*, 90, 489-509.
- Martin, R. (2001), *Sémantique et automate*. Paris, Presses Universitaires de France.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: Lexique. *L'Année Psychologique*, 101, 447-462.
- Niklas-Salminen, A. (1996), *La lexicologie*. Paris, Armand Colin.
- Nyckees, V. (1998), *La sémantique*. Paris, Belin.
- Ploux S. & Victorri B. (1998), Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes, *TAL*, Vol 39/1, pp. 161-182.
- Rey, A. et al. (1985), *Le Grand Robert de la langue française, Dictionnaire alphabétique et analogique de la langue française*, Paris, Editions Le Robert (2<sup>ème</sup> édition).
- Rey-Debove, J. et al. (1989), *Le Robert Méthodique*, Paris, Editions Le Robert (2<sup>ème</sup> édition).
- Touratier, Ch. (2000), *La sémantique*, Paris, Armand Colin.
- Vessereau, A. (1947), *La statistique*. Paris, P.U.F., collection « Que sais-je ? ».
- Victorri B. & Fuchs C. (1996), *La polysémie : une construction dynamique du sens*, Paris, Hermès.
- Warrington, E.K. & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107, 829-854.
- Wittgenstein, L. (1961), *Tractatus logico-philosophicus* suivi de *Investigations philosophiques*. Paris, Gallimard.

Votre requête est : "démarcation" (démarcation). Il y a 9 synonymes. [Voir leur classement](#)

### [Visualisation de l'espace sémantique](#)

#### **NOUVELLE VERSION : [Visualisation avec possibilité d'impression](#)**

Pour le mode d'emploi de la visualisation, voir la [lettre d'information numéro 8](#) (paragraphe 2)

Pour la nouvelle version, vous devez [télécharger la dernière version de Java](#) pour votre navigateur.

**démarcation** : [délimitation](#), [distinction](#), [frontière](#), [ligne](#), [limitation](#), [limite](#), [lisière](#), [marque](#), [séparation](#).

Pour faire une autre recherche :

[Classement complet des synonymes, détail des sens macroscopiques \(composantes connexes\) et microscopiques \(cliques\) de démarcation.](#)

#### **Le classement des premiers synonymes :**

limite	
frontière	
distinction	
délimitation	
séparation	
ligne	
lisière	
marque	
limitation	

Figure 3 : exemple de réponse du dictionnaire des synonymes.



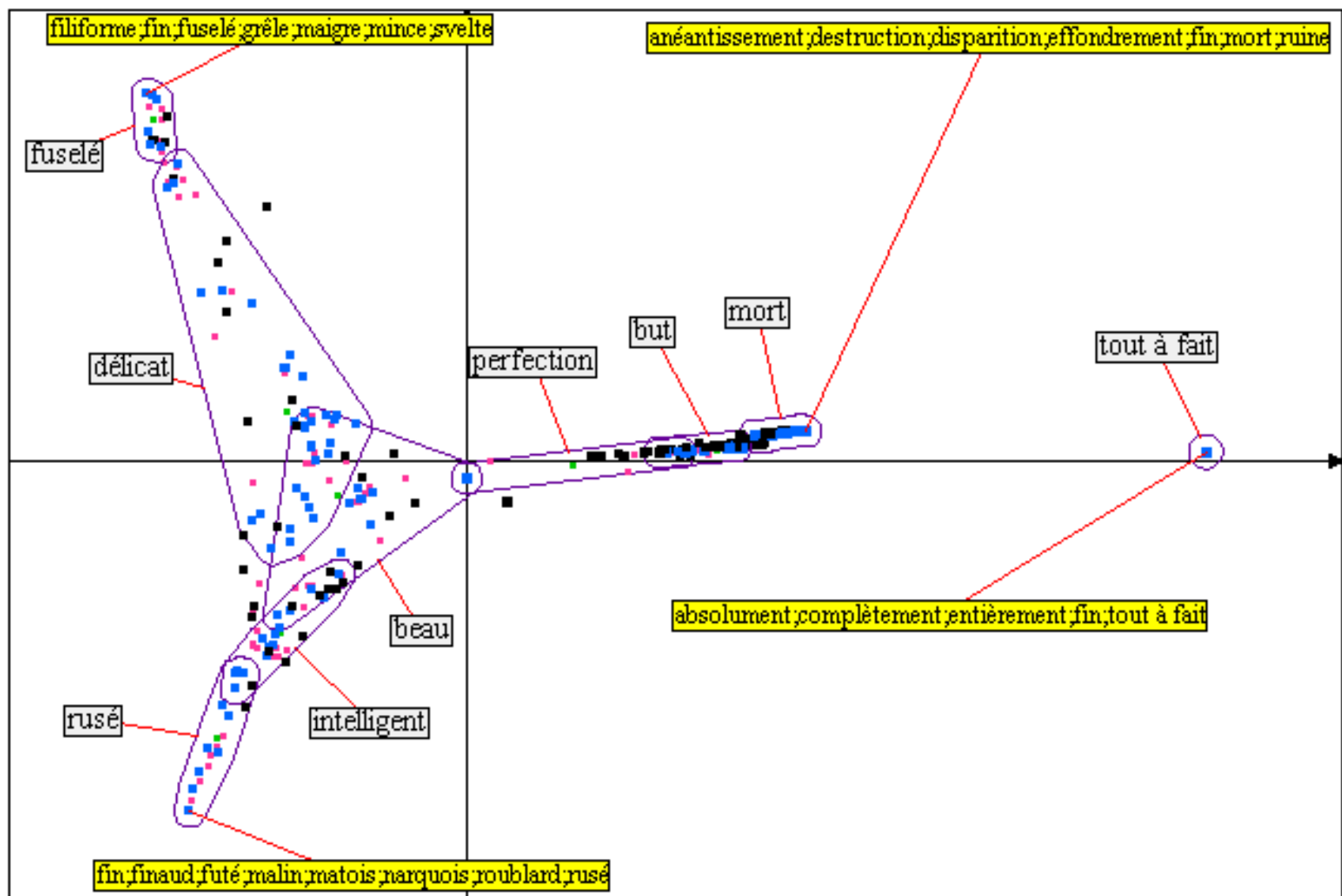


Figure 6 : Représentation du mot *fin*

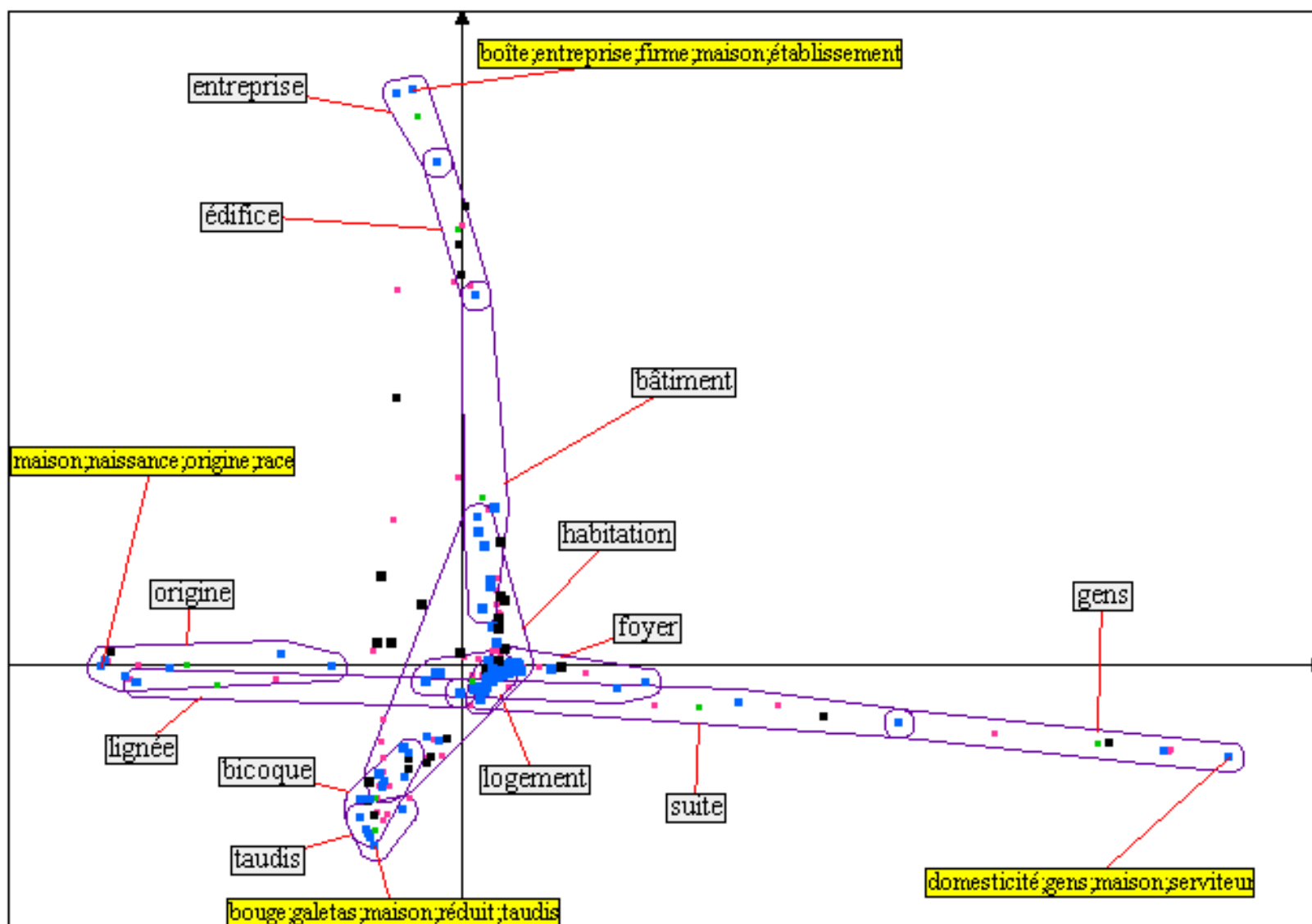


Figure 7 : Représentation du mot *maison*

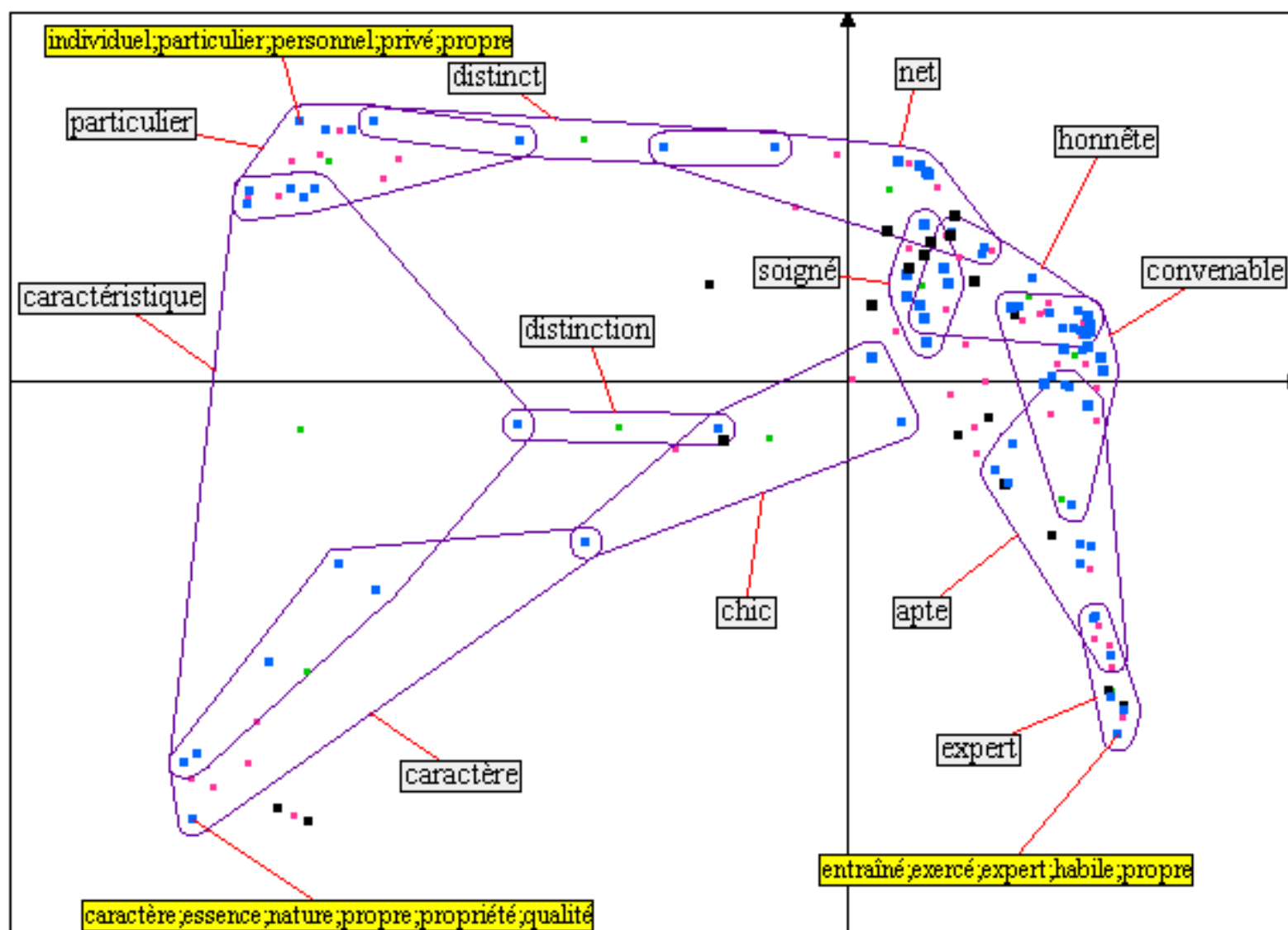


Figure 8 : Représentation du mot *propre*

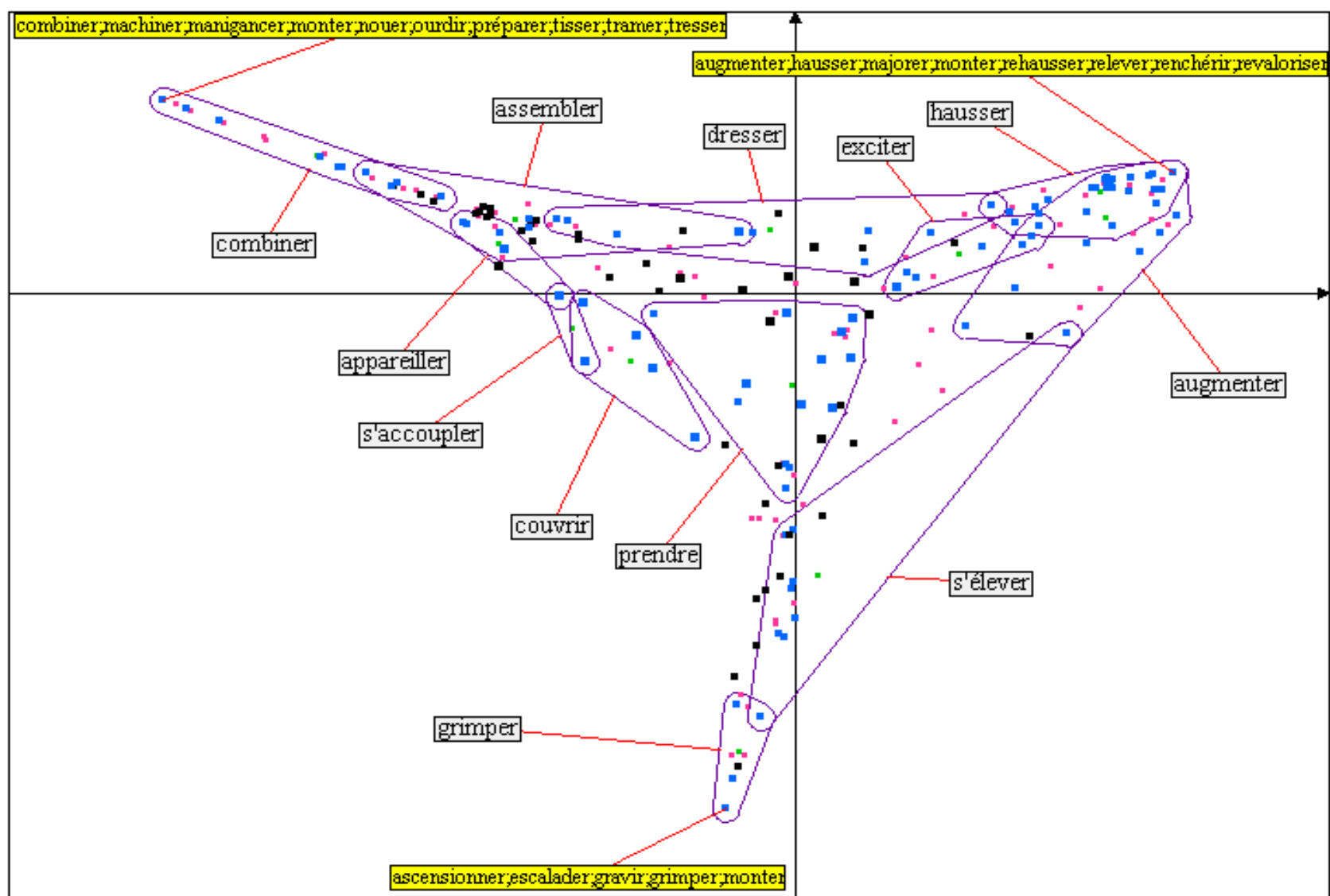


Figure 9 : Représentation du mot *monter*