



HAL
open science

La dictionnaire Internet : l'exemple du dictionnaire des synonymes du CRISCO

Jean-Luc Manguin

► **To cite this version:**

Jean-Luc Manguin. La dictionnaire Internet : l'exemple du dictionnaire des synonymes du CRISCO. Colloque annuel de l'AFLS - Le français aujourd'hui : Problèmes et méthodes, Sep 2003, Tours, France. [14 p.], 10.4000/corela.1130 . hal-00012219

HAL Id: hal-00012219

<https://hal.science/hal-00012219v1>

Submitted on 19 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.

Jean-Luc MANGUIN

CNRS - Université de Caen – CRISCO (UMR 6170)

**LA DICTIONNAIRIQUE INTERNET :
L'EXEMPLE DU DICTIONNAIRE DES SYNONYMES
DU CRISCO**

Résumé : D'une manière générale, la dictionnaire peut s'envisager comme l'élaboration du « contenant », tandis que le « contenu » relève de la lexicographie. A travers notre expérience d'un dictionnaire des synonymes sur Internet, nous montrerons que cette forme de dictionnaire, apparue récemment, présente un intérêt à la fois pour l'utilisateur et pour le concepteur. Une ressource en ligne est en effet intéressante pour ses utilisateurs car ils bénéficient de l'évolution continue du produit ; pour les concepteurs, le retour d'information donné par la fréquentation du site permet d'orienter cette évolution. Il est en effet relativement aisé d'analyser les requêtes formulées par les utilisateurs et de repérer ainsi les lacunes de la base de données. En plus de leur intérêt lié à leur disponibilité, les ressources en ligne peuvent constituer un puissant outil d'apprentissage ; par exemple, notre dictionnaire des synonymes permet de construire des représentations visuelles et interactives du sens qui forment un support de choix pour l'enseignement du vocabulaire.

Abstract : Generally, we can understand dictionaries as the building of the « container », while the « contained » is a matter of lexicography. Through our experience of an on-line dictionary of synonyms, we show that this type of dictionary, which has recently appeared, is interesting both for its user and for its author and designer. An on-line resource is interesting for its users indeed because they benefit from the continuous evolution of the product ; for the authors of dictionaries, the information feedback which is given by the log files of the website can direct this evolution. It is quite easy, indeed, to analyse the requests set out by the users, and then to find the shortages of the database. In addition to their interest coming from their availability, on-line resources may become a powerful learning tool ; for instance, it is possible, from our dictionary of synonyms, to build interactive visualizations of meaning which form a choice aid for the teaching of vocabulary.

Mots-clefs : dictionnaire, dictionnaire, Internet, représentation du sens, synonymes.

Introduction

Si la notion de *dictionnaire*, définie par Bernard Quemada, vise à circonscrire tous les aspects de l'élaboration d'un dictionnaire dans le cadre d'un objectif, elle se démarque ainsi de la *lexicographie*, qui se charge alors de « l'analyse des mots de leurs significations » (Pruvost, 2002) ; néanmoins, comme le rappelle aussi Jean Pruvost dans le même ouvrage, ces deux notions sont complémentaires et intimement liées. Sans entrer dans le détail des problèmes de dictionnaire qui se posent dans le cas des ouvrages imprimés, ou des lexiques électroniques sur CD-Rom, nous pouvons cependant affirmer que l'apparition des dictionnaires consultables en ligne sur Internet a contribué, du fait des spécificités de ce média, au renouvellement des questions et des réponses de la dictionnaire.

Dans cet article, nous montrerons que l'expérience de notre dictionnaire des synonymes du français, présent sur Internet depuis l'automne 1998¹, nous a permis de résoudre de manière particulière des questions de dictionnaire traditionnelles comme l'adaptation de l'ouvrage à l'usage envisagé, ou encore son évolution en termes de mots-vedettes à ajouter, voire à ôter.

1. Le dictionnaire des synonymes du CRISCO

Le dictionnaire des synonymes du CRISCO est essentiellement constitué d'un fichier dans lequel chaque ligne contient un mot-vedette suivi de la liste de ses synonymes, comme par exemple ci-dessous pour le mot *démarcation* :

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

Ces mots-vedettes et leurs synonymes proviennent de la fusion de sept fichiers contenant les relations présentes dans les dictionnaires qui ont servi de base à cette construction (voir les références de ces dictionnaires à la fin de l'article), et qui nous ont été fournis par l'Institut National de la Langue Française. Il faut tout d'abord signaler que lors de l'élaboration du fichier final, il était impossible de conserver les distinctions et divisions d'articles faites dans certains dictionnaires, en raison de l'hétérogénéité des classifications opérées suivant les lexicographes (voir Ploux, 1997). Le fichier résultant de la fusion présente donc la liste des synonymes « en vrac ». Cette idée de fusionner plusieurs dictionnaires se trouve déjà dans la thèse d'André Kahlmann (1975), dans laquelle on rencontre également des arguments en faveur de la symétrisation de la relation synonymique. On verra dans la suite de cet

¹ Ce dictionnaire est consultable gratuitement à l'adresse suivante : <http://www.crisco.unicaen.fr/>

article, que la symétrisation (que nous avons effectuée par traitement automatique) et la disparition des hiérarchies (qu'il était impossible de conserver) constitue un tremplin pour l'application de nouveaux modèles.

A l'issue de cette fusion et de cette symétrisation, il était nécessaire de regrouper certaines entrées et certaines relations ; par exemple, le verbe *émouvoir* possède parmi ses synonymes *exciter*, mais aussi, selon certains auteurs, *exciter un sentiment* et *exciter une passion*. Il nous semble que ces deux derniers constituent une redite, et qu'il n'est pas nécessaire de les intégrer. Ainsi, de 63000 entrées nous sommes passés à 49000, tandis que le nombre de relations passait de 220 000 à 200 000. Les relations erronées provenaient d'erreurs de saisie dans les fichiers d'origine, comme par exemple des synonymes de verbes pronominaux attribués à la forme non pronominale. Enfin, il faut signaler que ce travail de correction se poursuit continuellement, mais dans un but d'enrichissement du dictionnaire (voir les paragraphes 3 et 4 ci-après).

2. Impact et utilisation du dictionnaire

Une des questions essentielles de dictionnaire réside dans l'élaboration d'un cahier des charges nécessaire à la confection d'un dictionnaire ; ce cahier des charges prend avant tout en compte le type de public visé, ainsi que le type de dictionnaire envisagé (général ou encyclopédique, par exemple). Dans notre cas, les contraintes éditoriales ne se sont pas posées, puisque le dictionnaire des synonymes a été conçu au départ comme un support aux travaux du laboratoire, et sa mise en ligne comme une vitrine de l'activité de l'unité. Néanmoins, nous avons repris la problématique à rebours, grâce aux traces laissées sur le serveur Web par les internautes dans ce que l'on appelle en informatique les « fichiers de logs ». Techniquement, l'analyse statistique de ces fichiers permet de recueillir toutes les informations relatives à l'utilisation de notre ressource, et la comparaison avec des données qui concernent d'autres serveurs fournit une appréciation du placement du site sur une échelle qualitative allant du ludique au professionnel.

Sur les cinq années écoulées depuis l'ouverture du site, nous observons une croissance continue de forte valeur, ce qui indique que le site ne subit pas le moindre désintérêt de la part du public des internautes ; cet attrait peut s'expliquer grâce à l'hypertextualité de la ressource : la réponse fournie contient les synonymes du mot demandé, mais ceux-ci sont des liens vers leurs propres synonymes ; la navigation dans le dictionnaire devient ainsi très rapide, et constitue une raison de son engouement :

Année	Trafic (en millions de requêtes)	Progression
1999	0,48	
2000	1,91	298 %
2001	4,60	140 %
2002	9,85	114 %
2003	16,15	64 %

Tableau 1 : trafic annuel du dictionnaire des synonymes

Il est cependant plus intéressant de connaître la répartition du trafic, entre autres suivant les jours de la semaine :

Jour	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
Part de trafic %	16,7	17,9	18,7	18,3	15,6	5,8	7,1

Tableau 2 : répartition hebdomadaire du trafic

On remarque déjà dans cette répartition la faible valeur du trafic en fin de semaine, ce qui laisse deviner une utilisation professionnelle du dictionnaire des synonymes. Un outil de mesure d'audience, tel que le Weboscope®, délivre des analyses de trafic plus poussées, comme l'histogramme journalier en fonction de l'heure d'appel du client (c'est-à-dire l'heure qu'il est chez la personne qui consulte le site) ; il faut en effet tenir compte du fait que presque un tiers de notre trafic provient de la communauté francophone d'outre-Atlantique.

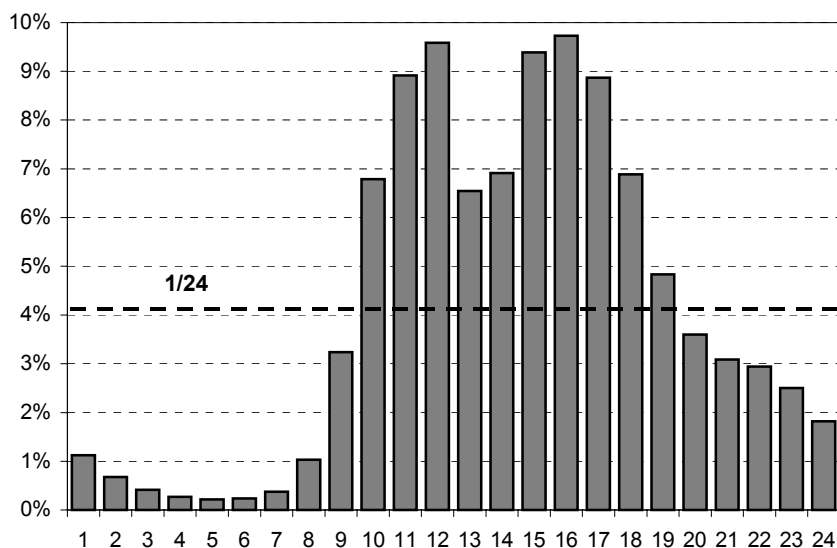


Figure 1 : histogramme journalier du trafic suivant l'heure client

Il est clair sur cette figure que l'essentiel du trafic est concentré aux heures de bureau ; on peut même y observer la pause-déjeuner. La conclusion qu'il faut en tirer, c'est que notre dictionnaire est perçu comme un outil bureautique supplémentaire. Ceci provient de son caractère « cumulatif » et non « distinctif », pour reprendre la typologie de Bernard Quemada (Quemada, 1968) ; notre dictionnaire fournit ainsi aux rédacteurs des mots qu'il connaissent en « compréhension », mais qu'ils ont à l'instant même du mal à retrouver en « production » ; enfin, si l'on compare notre histogramme à celui d'un site encyclopédique, cette image est encore plus évidente, et nous conduit à penser que notre ressource est certainement satisfaisante pour retrouver un mot que l'on connaît, mais qu'elle n'est pas idéale pour l'apprentissage, au contraire d'un dictionnaire « distinctif » qui explique les nuances d'emploi entre les différents synonymes d'un mot.

3. Enrichissement du dictionnaire

Une question tout aussi essentielle dans la vie d'un dictionnaire est l'évolution de son contenu au fur et à mesure de ses éditions successives ; les lexicographes doivent en effet faire évoluer les entrées en fonction des évolutions de la langue (il va de soi que les définitions évoluent elles aussi). Dans un cas comme le nôtre, la solution à ce problème est grandement facilitée par les traces des demandes formulées par les internautes. Nous avons construit un programme de traitement qui récupère et classe par leur fréquence les mots les plus demandés par les utilisateurs. A titre d'exemple, à la fin de l'année 2003, sur 33 millions de requêtes, les mots les plus demandés étaient :

Mot	Score
permettre	42898
amour	38968
faire	31015
important	30545
maison	28639
utiliser	25914
travail	24609
donner	24296
présenter	23336
dire	21699

Tableau 3 : les 10 mots les plus demandés

Il est facile, avec un dépouillement supplémentaire, de savoir à quels mots de cet ensemble le dictionnaire n'a pas répondu, et, après analyse,

de connaître la raison du mutisme de notre ressource. Nous donnons ci-après la partie supérieure du résultat de cette analyse en mars 2002, et les conséquences qui en ont découlé dans le développement de notre dictionnaire.

Mot	Score	Type d'erreur
pédestre	1505	D
entrez un mot ici	1116	A'
focus	664	A
travaux	648	B
synonyme	597	C
enjeux	541	B
#	496	A
allo	490	A'
sex	480	C
cela	455	D
interactif	450	D
mysogine	441	C
éponyme	417	A'
palindrome	404	A'
amont	362	D
france	353	A'
mature	349	D
anticonstitutionnellement	347	A'

Tableau 4 : les mots sans réponse les plus fréquents

Comme on le voit sur ce tableau, les requêtes peuvent être classées en plusieurs catégories, dont voici le détail ci-après :

3.1 Type A et A' : impossibilités

On trouve dans cette catégorie des requêtes auxquelles un dictionnaire des synonymes, aussi parfait soit-il, ne pourrait répondre ; en effet, nous avons parmi ces requêtes soit des absurdités (comme #), soit des mots qui n'ont pas de synonyme (comme *anticonstitutionnellement*).

3.2 Type B : formes fléchies

Ce second type correspond à l'utilisation d'une forme fléchie lors de la requête (par ex. adjectif au féminin, nom au pluriel, verbe conjugué) ; étant donné que ce ne sont pas là des « fautes », il est important de répondre à ce type de demande. Pour cela, nous avons ajouté à notre ressource un dictionnaire des formes fléchies qui va rediriger en cas de

besoin la requête vers la « forme canonique » qui se trouve dans le dictionnaire de base.

3.3 Type C : fautes d'orthographe

Depuis le début de la mise en ligne de notre dictionnaire, nous avons déjà installé une correction des signes diacritiques sur les requêtes, afin de ne pas pénaliser les utilisateurs étrangers qui ne disposent pas toujours de clavier avec accents. Nous y avons ajouté la correction des fautes les plus fréquentes ; techniquement, la correction consiste à rediriger la requête erronée vers une entrée standard du dictionnaire, comme pour les formes fléchies. Il existe cependant des cas où la requête peut conduire à plusieurs interprétations, comme par exemple *tache* ; le fichier de redirection va conduire à la première interprétation possible, tout en proposant à l'utilisateur des liens vers les autres options. Pour *tache*, le fichier de redirection contient la ligne suivante :

tache : *tache, tâche, tacher, tâcher*

Quand une redirection par correction orthographique est faite automatiquement, l'utilisateur en est averti, puisqu'il voit sur son écran apparaître un avertissement de ce type :

Votre requête est : "synonyme" (CORRECTION : synonyme).

Actuellement, dans un lexique de requêtes de 1,48 millions de « mots », nous corrigeons environ 1200 mots, ce qui représente 0,6 % des requêtes du dictionnaire. Ce « lexique des fautes corrigées » est désormais assez large pour commencer à établir une typologie, afin d'enrichir encore le correcteur, mais cette fois sans avoir à consulter le fichier des requêtes sans réponse.

3.4 Type D : lacunes

Ce dernier type d'erreur reflète les lacunes du dictionnaire ; on y trouve en effet des mots auxquels il serait relativement facile de donner des synonymes, comme par exemple *mature* qui peut recevoir comme synonymes *adulte, mûr* et *sage*. Le constat de ces imperfections entraîne donc un travail de lexicographie qui fait entrer de nouveaux mots-vedettes dans le dictionnaire ; ceux-ci sont bien entendu pourvus de synonymes faisant ou non déjà partie du dictionnaire.

L'effet de tous ces ajouts et corrections peut se mesurer au niveau du taux de réponse du dictionnaire (figure 2) ; il est toutefois important de nuancer ce taux brut par le détail des tranches de fréquence des mots ; ainsi, les mots demandés plus de 1000 fois (sur 33 millions de requêtes) représentent plus de 54% des requêtes, et dans cette tranche, le taux de réponse du dictionnaire est de 99,9 % (96,5 % de réponses directes, plus 3,4 % grâce aux corrections).

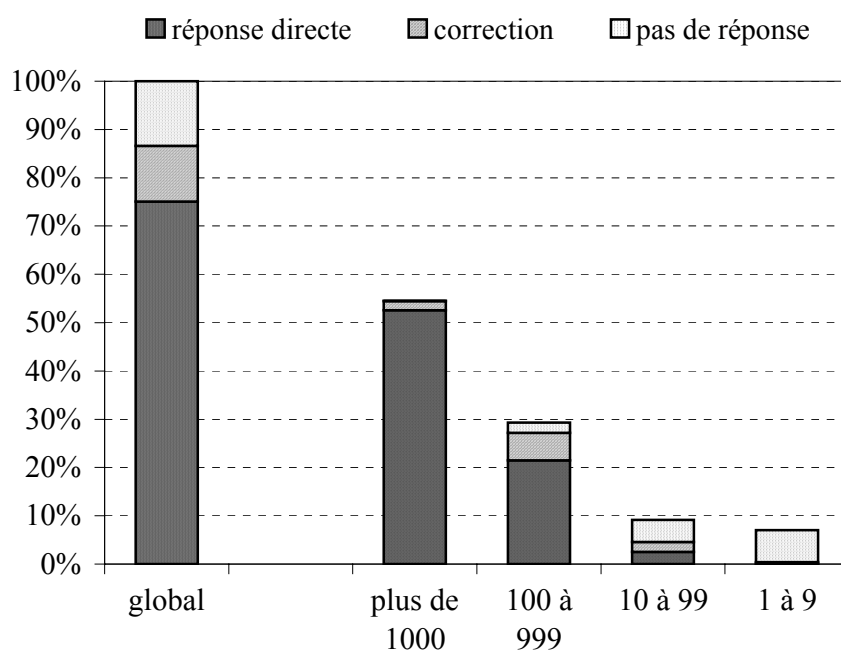


Figure 2 : taux de réponse du dictionnaire

4. Evolution de la présentation

A l'origine, le dictionnaire a été conçu comme un outil de recherche, et sa mise en ligne relevait au départ d'une exposition des travaux du laboratoire ; il s'est rapidement avéré que cet outil répondait à un besoin particulièrement fort, mais comme nous l'avons signalé, il satisfait un usage rapide et fonctionnel. Pour le faire évoluer vers des aspects et des fonctionnalités plus didactiques, nous avons profité des travaux de modélisation effectués, et nous les avons adaptés à l'usage en ligne. Enfin, nous développons actuellement des modules dont le résultat encore plus synthétique devrait accroître son intérêt pédagogique.

4.1 Le modèle en graphe

Le fondement de nos travaux réside dans la modélisation du dictionnaire des synonymes en un graphe dont les sommets sont les mots-vedettes du dictionnaire, et dont les arêtes sont les relations synonymiques. Puisqu'il est question d'arêtes cela suppose que la relation de synonymie est symétrique (voir Berge, 1967). Ce modèle s'inspire, comme nous l'avons dit, de la thèse de Kahlmann (op. cit.), à ceci près que chez ce dernier, les relations ne sont pas symétrisées ; cela dit, les arguments en faveur de la symétrisation se trouvent aussi chez cet auteur. Lors de l'étude d'une unité lexicale appartenant aux entrées du dictionnaire, nous restreignons notre matériau d'étude au sous-graphe qui contient seulement le mots étudié et ses synonymes, ainsi que les

relations existant entre ces unités. L'exploitation de ce sous-graphe se fait alors en énumérant les ensembles à cohésion lâche (composantes connexes) et les ensembles à cohésion forte (cliques). Prenons l'exemple du mot *baie*, dont le graphe de synonymie simplifié est le suivant :

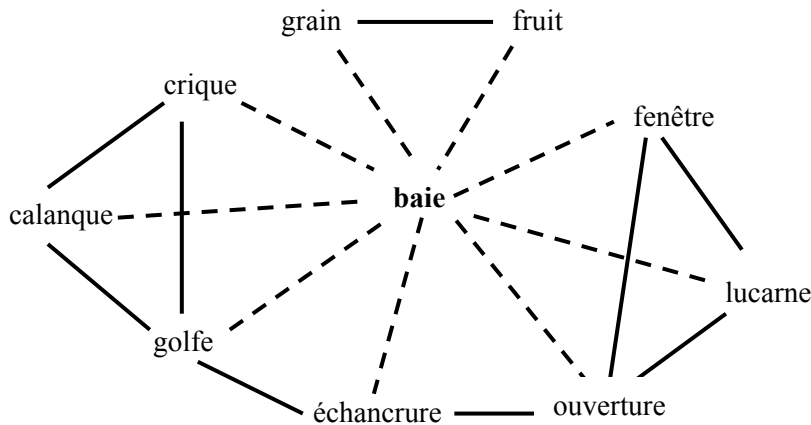


Figure 3 : graphe de synonymie simplifié du mot *baie*

L'énumération des composantes se fait en ne tenant compte que des liaisons qui n'impliquent pas le mot étudié, ici *baie* ; les composantes connexes (pour la terminologie, voir Berge, op. cit.) sont celles où il existe une chaîne de longueur quelconque entre deux sommets ; on en voit ici deux, l'une qui concerne la *baie* dans le sens végétal (avec comme sommets *grain* et *fruit*), l'autre reflète le sens de *baie* en tant qu'*échancre*, dans une côte ou dans une façade. Les composantes connexes sont donc utiles pour séparer ce qui correspond généralement à des homonymes.

Les cliques sont des sous-ensembles où chaque sommet est en relation avec tous les autres ; par exemple, {*calanque*, *crique*, *golfe*} constitue une clique, mais on ne peut ajouter à cet ensemble le mot *échancre*, car ce dernier n'est pas en relation avec *crique* ni avec *calanque*. Ces cliques constituent un ensemble dont le détail donne une idée précise des sens élémentaires pris par le mot étudié.

4.2 Représentation du sens à partir de l'ensemble des cliques

Selon une méthode décrite par ailleurs (voir Ploux, 1997 et Ploux & Victorri, 1998), il est possible d'utiliser l'ensemble des cliques d'une unité lexicale pour construire un espace géométrique multidimensionnel assimilable à l'espace sémantique de l'unité étudiée ; dans cet espace, le sens de l'unité est représenté par un nuage de points, chaque point correspondant à une clique, donc à un sens élémentaire. La répartition de ces points dans l'espace donne visuellement une idée de la proximité ou de l'éloignement entre ces différents sens. Dans cette représentation, les

synonymes sont des portions de l'espace, la zone d'influence de chaque synonyme englobant les cliques où celui-ci est représenté. Ces représentations sont disponibles sur notre site, et pour l'exemple de *baie*, l'espace sémantique obtenu a l'allure suivante² :

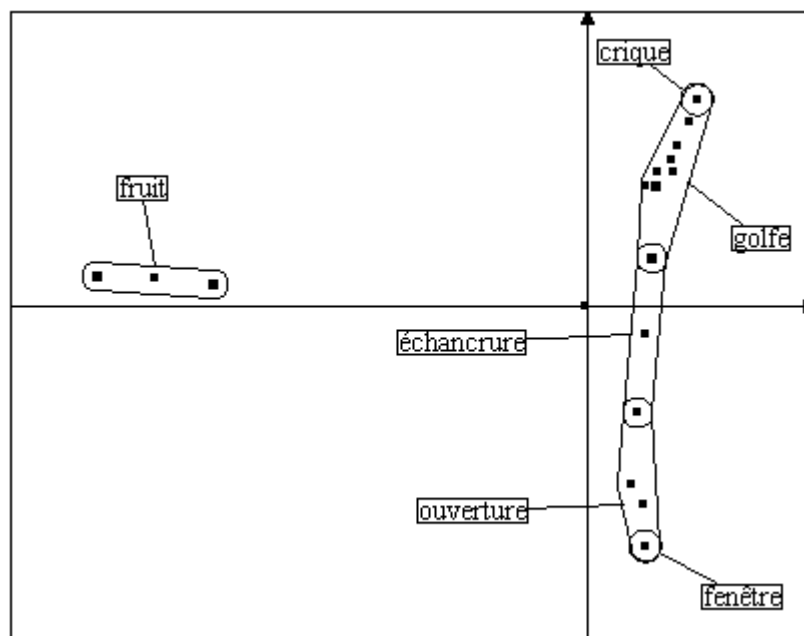


Figure 4 : espace sémantique du mot *baie*

On peut remarquer sur cette figure que les deux composantes de *baie* sont nettement séparées, ce qui confirme la présence des deux homonymes ; d'autre part, la partie droite montre un continuum qui va de *fenêtre* à *crique*, en passant par *échancrure* ; si ce dernier mot n'avait pas de liaison synonymique avec *golfe*, par exemple, un dégroupement homonymique apparaîtrait avec d'une part la composante architecturale, et d'autre part la composante maritime. L'application de cette méthode à l'étude des unités polysémiques du français permet de mieux appréhender les champs sémantiques, notamment parce qu'elle peut servir à examiner simultanément deux unités relativement proches (voir par exemple Manguin, François & Victorri, à paraître).

4.3 Les indices de similitude

On peut cependant remarquer que la représentation du sens dans un espace géométrique implique la connaissance de la distance entre les différents sens d'un mot ; nous pouvons considérer que, si l'on se limite aux seules relations dans le sous-graphe du mot étudié, les différents sens sont portés par les différents synonymes. Dès lors, le calcul d'une

² Sur le site Web du laboratoire, ces représentations sont en couleurs, et interactives, ce qui en améliore la perception.

distance entre ces synonymes revient à trouver un moyen d'attribuer une valeur aux arêtes du graphe de synonymie, et de se dispenser de l'énumération des cliques qui est assez coûteuse en temps de calcul. Nous avons choisi de calculer la valeur affectée à chaque relation par les indices de communauté (ou indices de Jaccard). Ainsi, pour chaque paire de sommets du graphe considéré, l'indice de similitude sera défini par le nombre de sommets communs, divisé par le nombre de sommets en relation avec l'un des deux membres de la paire (voir entre autres Legendre & Legendre, 1998, pp. 253-258) :

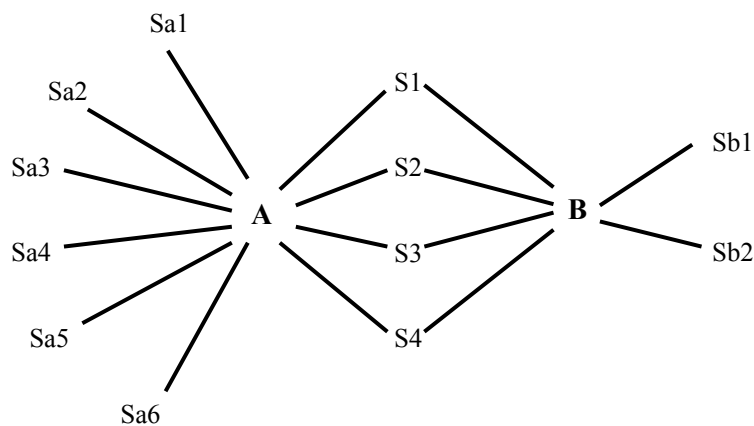


Figure 5 : exemple de calcul de similitude entre deux sommets A et B

Dans l'exemple de la figure, la similitude entre A et B vaut $4/12$, puisqu'il y a 4 sommets communs sur un ensemble de 12 sommets en relation avec A ou B. L'établissement d'une relation évaluée entre le mot étudié et ses synonymes, mais aussi entre ces synonymes, permet de quantifier la polysémie d'une unité lexicale, et également de réaliser des groupements de sens pertinents qui rassemblent les synonymes entre eux (Manguin, à paraître A).

4.2 Représentation du sens à partir des indices de similitude

La relation évaluée que nous définissons à partir des indices de similitude est assimilable à une inter-distance entre les différents synonymes (y compris le mot étudié). Nous pouvons alors, comme nous l'avons déjà montré (Manguin, à paraître B), construire une représentation dans un espace géométrique de dimension faible (2 voire 3 pour les unités très polysémiques) par échelonnement multidimensionnel (Multi-dimensional scaling, en abrégé MDS ; voir Kruskal et Wish, 1978). La représentation diffère de celle que l'on obtient par l'analyse des cliques par le fait que les points sont ici les synonymes eux-mêmes, et qu'ils sont situés dans un espace de dimension ordinaire (en général un

plan) ; la figure suivante reprend l'exemple de *baie* avec ce type de visualisation :

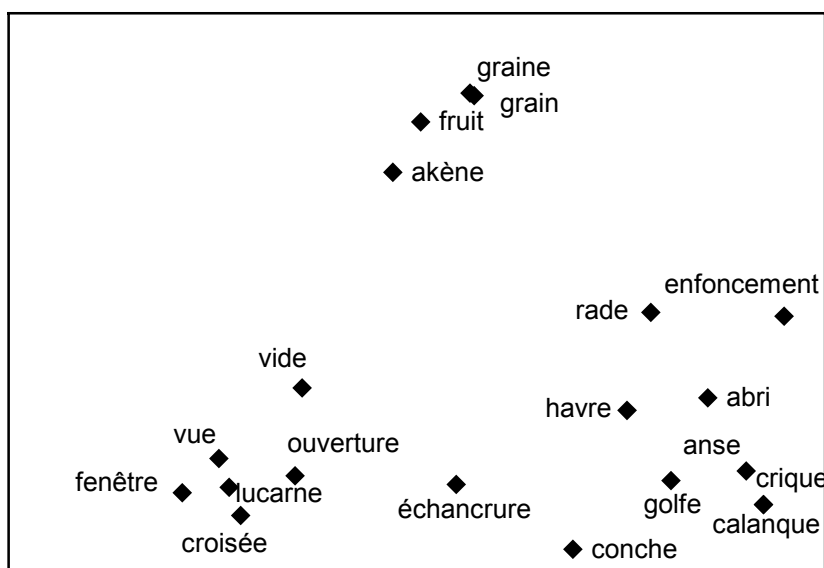


Figure 6 : espace sémantique du mot *baie* (seconde méthode)

Sur une représentation de ce type, nous avons déjà réalisé et présenté (Manguin, à paraître B) une adjonction des contextes, dans le cas où l'unité lexicale étudiée est un adjectif. Nous avons choisi d'examiner un adjectif polysémique épithète dans certains contextes nominaux. Après récupération de cooccurrences dans un corpus catégorisé (Frantext) et traitement de cet ensemble de données, les noms dont l'adjectif étudié est épithète peuvent être placés sur la représentation plane de l'espace sémantique. On obtient ainsi une représentation qui devient interactive si elle est placée sur un support informatique dynamique (comme c'est le cas pour les visualisations disponibles sur le site Internet du CRISCO) ; ce type de représentation permet non seulement les distinctions de sens, mais aussi les distinctions d'emploi, en d'autres termes les solidarités lexicales consacrées par l'usage.

Conclusion et perspectives

Comme nous venons de le montrer, Internet a largement contribué au renouvellement des questions de dictionnaire ; en effet, la structure même des sites Web permet de garder des traces de la « lecture » effectuée par les internautes, et cela apporte au concepteur de la ressource en ligne des réponses rapides et tangibles en matière de présentation et d'évolution du contenu. Les contraintes éditoriales étant faibles, l'évolution même de la ressource en ligne devient continue, et transparente pour l'utilisateur. Mais au-delà de la forme particulière que prennent les réponses aux questions dictionnaires, et même au-delà de

la réorganisation du processus d'élaboration d'une ressource lexicale (les questions ne se posant plus forcément en amont de la finalisation du produit), c'est avant tout l'émergence de questions nouvelles suscitées par l'accès aux ressources qui caractérise à notre avis la dictionnaire Internet ; nous en donnons ci-dessous deux exemples.

Premièrement, la gratuité et la simplicité d'accès à certains dictionnaires (comme le nôtre) rendent possible la réalisation d'intericiels (pour reprendre la terminologie québécoise), petits programmes bureautiques qui vont, de manière totalement imperceptible par l'utilisateur, consulter des ressources gratuites du Web pour rapporter dans l'application courante les données requises. Cela dit, cette possibilité alléchante peut, entre des mains malveillantes, se révéler être un instrument de pillage à des fins inavouables ; comment, dès lors, mettre à disposition des ressources linguistiques dans l'intérêt de la communauté ?

Deuxièmement, l'arrivée sur Internet des corpora étiquetés apporte aux internautes avertis (et férus de sciences du langage) une grande richesse de données ; puisque le matériel ne paraît plus imposer de limite, comment créer désormais une synergie entre les données paradigmatiques que constituent les dictionnaires, et les données syntagmatiques que sont les corpora ? Nous sommes convaincus qu'un tel couplage peut faire émerger un véritable renouvellement des dictionnaires, en apportant d'abord aux définitions ou aux choix de synonymes des exemples certifiés et datés, et ensuite en permettant l'accès immédiat à des représentations du sens qui intègrent la dimension contextuelle ; mais il importe dans ce travail de tenir compte de l'aspect diachronique de la langue, car, comme le rappelle Jean Pruvost (op. cit.), l'intérêt historique des éditions successives des dictionnaires est d'avoir fixé l'état d'une langue à la date de leur parution.

Dictionnaires sources

BAILLY, René (1946) : Dictionnaire des synonymes. Paris, Larousse.

BÉNAC, Henri (1956) : *Dictionnaire des synonymes*. Paris, Hachette.

BERTAUD DU CHAZAUD, Henri (1971) : *Nouveau dictionnaire des synonymes*. Paris, Robert.

GUIZOT, François (1864) : *Dictionnaire Universel des synonymes de la Langue Française*. Paris, Didier (7ème édition).

LAFAYE, Pierre-Benjamin (1858) : *Dictionnaire des synonymes de la Langue Française*. Paris, Hachette.

Grand Larousse de la Langue Française (1971) : Paris, Larousse.

Le Grand Robert, dictionnaire de la langue française (1985) : sous la dir. d'A. REY. Paris, Robert.

Références bibliographiques

- BERGE C. (1967), *Théorie des graphes et ses applications*, Paris, Dunod.
- KAHLMANN A. (1975), *Traitement automatique d'un dictionnaire de synonymes*, Stockholm, Université de Stockholm.
- KRUSKAL, J. B. & WISH M. (1978) : *Multidimensional scaling*, Beverly Hills, Sage Publications.
- LEGENDRE P. & LEGENDRE L. (1998), *Numerical Ecology*, Amsterdam, Elsevier.
- MANGUIN, J.L. (à paraître A) : « Regroupements de synonymes par indices de similitude : exemple avec l'adjectif *ancien* ». Communication au colloque *Les adjectifs non prédicatifs*, 28-29 novembre 2002, Université de Paris-Nord Villetaneuse. A paraître dans *Les Cahiers de Lexicologie*.
- MANGUIN J.L. (à paraître B) : « Utilisation d'un corpus catégorisé pour l'étude et la représentation de la synonymie en contexte ». 3^{èmes} *journées de linguistique de corpus*, 11-13 septembre 2003, Université de Bretagne Sud, Lorient.
- MANGUIN J.L., FRANÇOIS J. & VICTORRI B. (à paraître), « Polysémie adjectivale et rection nominale : quand *gros* et *gras* sont synonymes », *Actes du colloque « L'adjectif en français et à travers les langues »*, Caen, 28-30 juin 2001, Presses Universitaires de Caen.
- PLOUX S. (1997). « Modélisation et traitement informatique de la synonymie ». *Linguisticae Investigationes*, XXI (1), Amsterdam, John Benjamins.
- PLOUX S. & VICTORRI B.(1998), « Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes », *Traitement Automatique des Langues*, Vol 39/1, pp. 161-182.
- PRUVOST, Jean (2002) : *Les dictionnaires de langue française*. Paris, PUF, Collection « Que sais-je ? ».
- QUEMADA, Bernard (1968), *Les Dictionnaires du français moderne (1539-1863). Étude sur leur histoire, leurs types et leurs méthodes*, Paris, Didier.