

# Lexicographie assistée par ordinateur.

## Signification de "Banque" dans le vocabulaire économique

Jean Leselbaum<sup>1</sup>, Dominique Labbé<sup>2</sup>

<sup>1</sup> SIGNIFIER – 4, square Théophile Gautier – 75015 Paris – France –  
signifier.leselbaum@wanadoo.fr

<sup>2</sup> CERAT-IEP – BP 48 – F 38040 Grenoble Cedex 9 – France – [dominique.labbe@iep.upmf-grenoble.fr](mailto:dominique.labbe@iep.upmf-grenoble.fr)

### Abstract

We present a method to define the meanings of a word in an author's works or in a collection of texts (corpus). We bring to light the vocabulary associated with this word and all the synonymous which are used in the same contexts. As an example, we examine the case of "bank" in the economic French vocabulary. This word has two main meanings : "operations on financial markets" and "financial group

### Résumé

Méthode pour définir les sens précis d'un mot dans un corpus ou chez un auteur. On recherche le vocabulaire associé à ce mot (univers lexical) puis tous les synonymes potentiels : vocables de même catégorie grammaticale et employés dans des contextes semblables. La méthode est illustrée avec le mot "banque" dans le vocabulaire économique et social français contemporain. Ce mot possède deux sens principaux : opérateur sur les marchés financiers et groupe financier.

**Mots-clés** : Lexicographie - Vocabulaire économique - Français - Banque.

## 1. Introduction

Les dictionnaires de langue donnent une représentation plus ou moins complète et exacte du lexique français, mais ils ne peuvent répondre à la question de savoir quels sont le ou les sens précis qu'un mot possède chez un auteur donné ou dans un secteur particulier de la vie sociale. Nous pensons que les méthodes de la lexicologie (étude de la langue) et de la lexicographie (technique de confection des dictionnaires) peuvent aider à répondre à cette question. En effet, pour déterminer les sens possibles d'un mot en langue, les lexicologues établissent les différents contextes d'emploi possibles (paraphrases) et remplacent le mot qu'ils cherchent à définir par d'autres qui lui sont sémantiquement voisins ou opposés (recherche des synonymes et des antonymes). Les mêmes techniques peuvent être appliquées aux œuvres d'un auteur ou à des corpus représentatifs de tel ou tel types de discours.

Malheureusement, pour l'instant, l'apport de l'ordinateur à la lexicographie semble encore assez faible et se limite essentiellement à l'édition de "concordances" qui permettent de retrouver les attestations d'une graphie dans un corpus de textes. Nous proposons d'aller plus loin de deux manières. En premier lieu, la recherche ne se fera plus sur les graphies mais sur les vocables, ou entrées de dictionnaire, auxquels sont rattachés chacun des mots du corpus étudié (par exemple, le masculin singulier des adjectifs, l'infinitif des verbes...) D'autre part, nous proposons d'automatiser les procédures des lexicologues et notamment la recherche des

contextes significatifs, des synonymes et des paraphrases qui permettent de distinguer les principales significations.

Pour illustrer cette méthode, nous prendrons l'exemple du substantif féminin "banque" dans le vocabulaire de l'information économique contemporaine. Auparavant, on pourra consulter un dictionnaire de langue ou de synonymes. Par exemple, dans le dictionnaire *Robert des synonymes* : "caisse de crédit, de dépôts, établissement de crédit, comptoir". Cette courte liste correspond-elle aux emplois des locuteurs et notamment des acteurs de l'économie ?

Au préalable, nous avons rassemblé un corpus représentatif du vocabulaire de l'économie. Ce corpus, en cours de constitution, respecte, dans la sélection des textes, une série de critères a priori qui permettent de considérer qu'il est bien représentatif du domaine considéré (les informations émises par les acteurs de l'économie et celles qui sont écrites à leur propos par les journalistes spécialisés). Les textes sélectionnés ont subi une correction orthographique approfondie puis ils ont été normalisés et lemmatisés (Labbé, 1990) avant d'être intégrés dans la base. La partie du corpus utilisée pour cette expérience est constituée de 774 articles parus dans les rubriques économiques de quatre journaux (*Les Echos*, *Le Monde*, *Capital* et *L'Expansion*) entre janvier 1996 et décembre 1998, soit 1,2 millions de mots (300.000 pour chacun des quatre journaux), 47 650 formes normalisées différentes rattachées à 29 790 vocables. Dans cet ensemble, le mot "banque" apparaît 1096 fois. Avec quelles significations ?

## 2. L'univers lexical de "banque"

En premier lieu, nous avons établi l'univers lexical de ce mot (Hubert et Labbé, 1995 ; Labbé, 1997), c'est-à-dire l'ensemble des relations d'attraction ou de répulsion qu'il entretient avec les autres vocables dans le corpus. Pour repérer ces relations, on relève toutes les phrases contenant le vocable considéré. Le vocabulaire de ce sous-ensemble est comparé à celui du corpus entier. Lorsque, pour un mot donné, la fréquence relative dans cet univers dépasse significativement celle observée dans l'ensemble du corpus, l'association est dite "positive" ; à l'inverse, si elle est significativement inférieure, l'association est négative (antonymie). Pour mesurer ces liens d'attraction et de répulsion, on utilise le calcul des spécificités (Lafon, 1984) avec quelques adaptations (Labbé et Labbé, 1997). Les résultats sont présentés dans l'annexe 1.

Ce premier calcul appelle une remarque préalable. Dès que le mot atteint une fréquence relativement élevée (ce qui est le cas ici), on observe habituellement un nombre à peu près équivalent d'associations positives et négatives. Ici le déséquilibre est important : le vocable génère près de deux fois plus d'associations positives que négatives. Autrement dit, il y a peu de domaines (et de thèmes) dont les banques soient absentes ou peu impliquées...

Au cœur de cet univers, figurent les noms des principales banques françaises puis internationales. Pourtant, le premier nom de pays apparaissant dans cette liste n'est pas "France" mais "Suisse" ! En fait, ce serait même "Grande Bretagne" si les articles n'hésitaient pas entre "Angleterre", "Grande Bretagne" et "Royaume Uni". Enfin, New York est la seule ville qui apparaisse dans cet univers : la géographie financière mondiale est sans équivoque !

Autre remarque intéressante : au centre de cet univers, on s'attendrait à trouver aussi les monnaies et les chiffres, or c'est le contraire qui se produit : dans la liste des substantifs significativement sous-employés, on trouve : "milliard", "franc", "dollar" ainsi que tous les chiffres (dans les "déterminants")...

Quelle est alors l'activité des banques ? Les *affaires* (généralement au pluriel) comme l'indique le substantif le plus significativement sur-employé. Ce mot étant fortement

polysémique, la suite de l'univers de *banque* — comme les verbes : *détenir, diriger, accorder (crédit, prêt), regrouper, contrôler...* ou les substantifs : *filiale, investissement, gestion, fonds, compte, crédit...* — suggère quelques sens plus précis et un réseau sémantique cohérent que nous allons décrire plus bas.

Cependant, ces listes restent assez abstraites car les mots qui y figurent sont sortis de leur contexte. On propose donc un retour au texte, en demandant à l'ordinateur de sortir les passages les plus caractéristiques de cet univers. Pour cela, le programme relit l'ensemble du corpus et classe les phrases, où figure le mot recherché, en fonction de la densité (absolue et relative) des associations y figurant (faute de place, nous ne donnons en annexe 1 que les premières phrases).

La lecture de ces listes et de ces phrases suggère de nombreux sens spécifiques pour le mot banque. Il serait évidemment souhaitable de compléter l'analyse en procédant à la manière d'un dictionnaire, c'est-à-dire en regroupant de manière systématique les emplois attestés autour de quelques noyaux et en donnant, pour chacun d'eux, une définition et des exemples. Pour cela nous demandons à l'ordinateur de procéder comme les lexicographes : rechercher les synonymes et reconstituer les paraphrases correspondant à ces familles de sens.

### 3. A la recherche des synonymes de "banque"

Rappelons que deux ou plusieurs mots différents sont dits synonymes lorsqu'ils partagent un ou plusieurs sens. La synonymie s'établit en substituant un mot à un autre dans le même contexte. Si le sens n'en est pas affecté, les deux mots sont considérés comme synonymes. Naturellement, la synonymie est rarement absolue (en dehors des jargons ou vocabulaires techniques, deux mots ne peuvent être totalement équivalents). Elle concerne seulement certains emplois des mots considérés. Plusieurs situations sont envisageables :

— synonymie partielle : deux mots partagent un ou plusieurs sens mais divergent également pour certains emplois qui leur sont propres. Par exemple : *investissement* et *crédit*. C'est le cas le plus fréquent ;

— un mot de signification générale (hyperonyme) englobe des mots de sens plus spécifique (hyponymes). Ainsi par exemple : *groupe - banque - filiale - établissement...*

Enfin, il faut rappeler que les relations de synonymie n'existent qu'entre mots de même catégorie grammaticale et portent essentiellement sur les noms et les verbes. La recherche des synonymes par ordinateur nécessite donc que l'on ait, au préalable, rattaché chaque mot du corpus à sa catégorie grammaticale.

Rechercher les synonymes de *banque* consiste à se demander : quels substantifs peuvent lui être substitués, dans tout ou partie des phrases où il est employé, sans en changer le sens ? L'ordinateur ne peut répondre directement à cette question puisque la signification lui est inaccessible. En revanche, il peut dire quels sont, parmi les autres substantifs usuels du corpus, ceux qui partagent une partie significative de l'univers du mot pôle, en l'occurrence la *banque*, (associations positives et négatives). La démarche se déroule en deux temps.

1. On demande de rechercher les phrases où ne figure pas le mot pôle (*banque*) mais au moins un autre substantif (synonyme potentiel) ;

2. Dans cet ensemble, on recherche les phrases où se trouvent le maximum d'associations positives et le minimum d'associations négatives caractéristiques, c'est-à-dire les passages où l'on est certain qu'il est question des banques même si ce mot n'est pas employé. On utilise pour cela un indice d'association :

$$\text{indice d'association} = \frac{\text{Nombre d'associations positives} - \text{Nombre d'associations négatives}}{\text{Taille de la phrase}}$$

Par exemple, en moyenne, les phrases contenant le vocable *banque* ont un indice moyen d'association de 0.21. Autrement dit, ce vocable "aimante" plus d'un mot sur cinq dans les phrases où il apparaît. Dans la recherche des synonymes, on ne retient que les phrases — **ne contenant pas "banque"** mais un autre substantif (synonyme potentiel) — dont le taux d'association (à l'univers lexical de *banque*) est supérieur à 0.21. Les résultats de l'expérience sont donnés dans le tableau 1 ci-dessous.

Tableau 1. Les synonymes potentiels de "banque"

Synonymes potentiels :	Nombre de phrases	Taux de synonymie
Marché	154	0.334
Groupe	92	0.316
Activité	63	0.324
Actionnaire	57	0.327
Taux	51	0.347
Etablissement	48	0.401
Président	48	0.301
Opération	47	0.337
Compte	44	0.329
Intérêt	43	0.350
Filiale	43	0.323
Fonds	41	0.318
Société	39	0.324
Affaire	35	0.329
Fusion	32	0.339
Titre	32	0.331

Adjectifs	Nombre de phrases	Taux de synonymie
Français	106	0.328
Grand	94	0.313
Financier	92	0.335
Public	54	0.334
Européen	53	0.323
Bancaire	46	0.348
Important	40	0.355
Monétaire	36	0.353
International	32	0.338

Les synonymes potentiels sont classés par ordre décroissant d'importance : en deuxième colonne, le nombre de phrases où figure ce synonyme et qui dépassent le seuil choisi, dans la troisième, la valeur de l'indice présenté ci-dessus, c'est-à-dire le solde relatif moyen des associations dans les phrases retenues (plus ce taux est élevé, plus la synonymie est probable). En dessous des substantifs, nous donnons, pour information, les adjectifs pour lesquels la même démarche a été effectuée.

Ce tableau appelle trois remarques :

— si la présence de mots comme "groupe", "établissement", "filiale" ou "société" était attendue (et prouve que la méthode est bonne), il peut paraître étonnant que "marché" soit présenté comme le meilleur synonyme de *banque*. La présence, dans la liste, d'autres mots comme *activité*, *taux*, *opération* ou *compte* laisse penser que l'expérience mélange les synonymes proprement dits avec les termes associés à l'univers de la finance. Faudrait-il alors inventer des filtres plus sévères ? Cependant, on se souviendra que la recherche ne porte pas sur la langue française mais sur le vocabulaire économique contemporain — c'est-à-dire sur des usages spécifiques. On peut penser que l'équivalence banque(s)-marché(s) est une synecdoque, figure rhétorique consistant à désigner les entreprises d'un même secteur économique par leur activité. Ainsi dira-t-on "l'automobile" pour l'ensemble des entreprises qui produisent, commercialisent ou réparent ces véhicules. Dans le même ordre d'idée, il pourrait être logique que les observateurs de l'économie disent les "marchés financiers" pour désigner cet aspect du secteur bancaire...

— la courte liste des adjectifs suggère que la recherche des synonymes doit se faire également sur les groupes nominaux (du type "marché financier français" "grand établissement bancaire", "taux d'intérêt", etc.) Pour retrouver ces groupes, on utilise la méthode des syntagmes répétés (Pibarot et Labbé, 1998). Rappelons que le syntagme répété est une extension de la notion de "segment répété" (Salem, 1987), extension que permet la lemmatisation du corpus. Par exemple, en utilisant cette approche, le meilleur synonyme de *banque* devient : « activité(s) (sur, de) (le, les) marché(s) financier(s) », ce qui confirme la synecdoque suggérée ci-dessus : dans certains cas, les acteurs de l'économie disent : "les marchés financiers" lorsqu'ils parlent des banques, vues sous l'angle de leur activité...

— la plupart de ces mots sont eux-mêmes fortement polysémiques. Il est donc probable que la synonymie n'est que partielle et qu'elle concerne seulement quelques emplois spécifiques (zone de synonymie) et que ces emplois se recouvrent plus ou moins ouvrant la voie à des regroupements. On aboutit ainsi à deux sens principaux dans les usages de "banque".

### **3.1. Sens 1. Opérateur sur les marchés financiers**

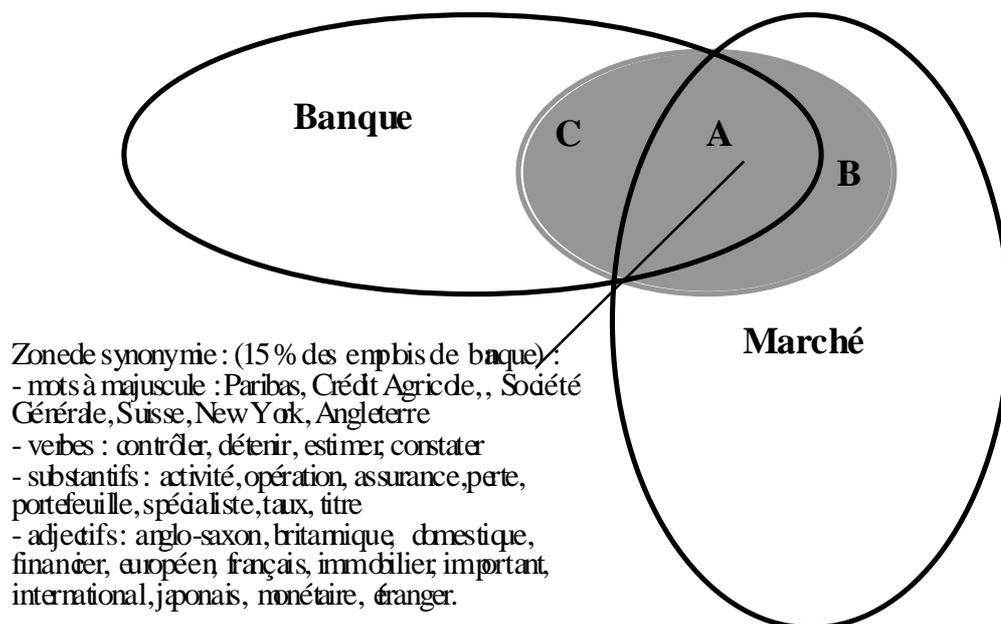
Pour identifier les zones de synonymie et mesurer leur importance, on demande à l'ordinateur de relire l'ensemble du corpus pour reconstituer le vocabulaire caractéristique partagé par "banque" et par chacun des synonymes potentiels. Le schéma ci-dessous récapitule les résultats obtenus avec "marché". La partie grise symbolise la "zone de synonymie" existant entre les deux mots. Cette zone est grossie pour les besoins de la représentation (en fait, elle ne représente que 15% de l'ensemble des emplois de "banque"). Elle correspond à trois situations différentes.

— A : 88 occurrences de *marché* (soit 3.7% du total de ses emplois), surviennent dans des phrases contenant également *banque* ;

— B : 154 des occurrences de *marché* (soit 6.5% des emplois du mot) apparaissent dans des phrases où ce vocable est employé dans des contextes semblables à ceux caractéristiques de la *banque* bien que ce mot soit absent ;

— C : 76 occurrences de *banque* (soit 7% des emplois de ce mot) figurent dans des contextes ne contenant pas *marché* mais qui sont caractéristiques de ce dernier.

Tableau 2. La zone de synonymie entre banque et marché



Au total ce sont donc à peine plus de 10% des occurrences de *marché* qui peuvent être rattachées à l'univers sémantique de *banque* et 15% de celles de *banque* qui peuvent être considérées comme synonyme de *marché*... Certes, la faiblesse relative de ce taux s'explique en partie par le choix d'un seuil élevé qui conduit à ne retenir qu'un nombre assez restreint de phrases comme étant susceptibles de fournir des contextes communs. En contrepartie, on obtient des listes brèves et une information sans ambiguïté : les marchés sont d'abord "anglo-saxons", voire *britanniques*, et le principal objet des banques ce sont des *activités* (de marchés financiers), des *opérations* sur les marchés... En effet, l'analyse des syntagmes répétés dans la zone de synonymie indique qu'il s'agit, outre des "marchés financiers", des "opération(s) (de, sur les) marché(s)", du "marché international" ou *immobilier, domestique, des titres, du taux du marché*, etc.

Pour établir un tableau complet, il faut aussi prendre en compte les vocables significativement sous-employés dans la zone grise du schéma (ils forment les antonymes de *banque* au sens d'*opérateur de marché*). La liste en est brève et éclairante :

— *Substantifs* : budget, cadre, entreprise, chaîne, chef, chiffre, chômage, emploi, formation, franc, heure, impôt, ingénieur, jour, magasin, mesure, mission, moyenne, nombre, patronat, production, salaire, salarié, site, syndicat, sécurité, temps, travail, usine.

— *Adjectifs* : humain, social.

Autrement dit, lorsque les observateurs parlent des activités de marché des banques, Ils "oublie" relativement des dimensions qui sont pourtant présentes dans leurs propos lorsqu'ils abordent d'autres thèmes.

Enfin, tout bon dictionnaire comporte des citations illustrant le sens qui vient d'être défini. Le logiciel recherche les phrases les plus caractéristiques de l'intersection des deux univers (zone A du graphique ci-dessous). Voici la phrase qui contient à la fois *marché* et *banque* accompagnés du plus grand nombre de vocables associés positivement aux deux univers et le moins de vocables significativement sous-employés dans ces mêmes univers :

*La complexité croissante des opérations financières, le développement de marchés sophistiqués et la concurrence qui pousse les banques à prendre toujours plus de risques incitent à s'interroger sur la capacité des banques centrales et des organismes internationaux à contrôler les systèmes bancaires. (Le Monde, 6 février 1996).*

Naturellement, un tel calcul "avantage" les phrases longues. On refait donc la même opération en rapportant le nombre d'associations à la longueur de la phrase. Les deux phrases suivantes sont celles qui obtiennent le meilleur score relatif :

*Les banques d'affaires anglo-saxonnes s'imposent sur le marché français. (L'Expansion, 19 décembre 1996). Les activités de marché entraînent la banque Paribas dans le rouge. (Les Echos, 29 février 1996).*

Enfin, la présence de certains synonymes potentiels de *banque*, dans la zone grise du schéma permet de rattacher à ce premier sens, une série de substantifs du tableau 1 (et donc de paraphrases) plus spécifiques : *activité, taux, opération, titre, intérêt...* Pour chacune de ces significations particulières, on procède de la même façon : recherche de l'intersection entre les vocabulaires et recherche des phrases caractéristiques.

Avec ces éléments, il sera relativement aisé de rédiger le premier paragraphe de la définition du dictionnaire du vocabulaire de la presse économique concernant les banques comme "opérateurs sur les marchés financiers". Le deuxième paragraphe concernera la banque comme "groupe financier".

### **3.2. Sens 2 : « groupe financier »**

L'encadré ci-dessous récapitule le vocabulaire caractéristique de l'intersection entre *banque* et *groupe*, et présente les phrases les plus significatives de ce second sens.

#### **Vocabulaire caractéristique des banques comme « groupes financier(s) »**

*Noms propres* : BTP, CIC, Crédit Agricole, Crédit Lyonnais, Fiat, GAN, Italie, Natexis, New York, OPA, PME, Paribas, Rivaud,

*Verbes* : contrôler, devenir, diriger, recentrer, regrouper, détenir, estimer, financer, mener, peser, prendre, venir,

*Substantifs* : acquisition, actif, actionnaire, affaire, analyste, établissement, restructuration, assurance, assureur, banquier, bilan, compte, consortium, contrôle, crise, crédit, dette, difficulté, dirigeant, donnée, entité, exploitation, filiale, activité, finance, fonds, fusion, gestion, immobilier, intérêt, investissement, métier, opération, participation, partie, perte, portefeuille, privatisation, provision, rapprochement, redressement, rentabilité, risque, spécialiste, taux, titre,

*Adjectifs* : anglo-saxon, bancaire, belge, britannique, central, chinois, suisse, commercial, direct, domestique, européen, financier, français, grand, immobilier, important, international, italien, japonais, lourd, monétaire, privé, public, régional, étranger

#### **Phrases caractéristiques :**

« Une bonne raison à cela : aux Etats-Unis et en Grande-Bretagne (et au Japon aussi, mais les fonds japonais confient leurs fonds en gestion aux banques et aux compagnies d'assurance, avec lesquelles elles ont parfois maille à partir), de très puissants fonds de pension détiennent une part considérable de la capitalisation boursière et ils ont, compte tenu des masses de fonds qu'ils gèrent, un pouvoir de discussion avec les entreprises. » (*Les Echos*, 30 janvier 1996).

« Plus pessimiste, il estime que la crise de l'industrie bancaire italienne va s'accroître, en raison d'une compétition accrue entre les banques elles-mêmes et entre les banques et les nouveaux acteurs (compagnies d'assurances, grands magasins, banque directe), de la pression de plus en plus forte des

mouvements de consommateurs, de la convergence des taux d'intérêt au niveau européen et de la stagnation de l'économie italienne. » (*Le Monde*, 29 avril 1996).

« Le partenaire idéal sera celui qui, comprenant que le groupe CIC apporte aux PME régionales, aux régions, aux métropoles françaises des services "plus" spécifiques, sans aucun doute exceptionnels, et qu'il a un fonctionnement tout à la fois efficace et décentralisé, dira : "ce groupe fonctionne, il a du potentiel, préservons son unité et ses banques régionales". » (*Les Echos*, 8 juillet 1996).

On procède de la même façon que ci-dessus : définition de la zone de synonymie — étendue et composition —, principaux syntagmes répétés puis recherche des phrases caractéristiques... A ce second sens, peuvent être rattachés une série de synonymes plus spécifiques : *actionnaire, établissement, société, fusion, fonds, filiale...*

Avec ces deux grandes entrées, nous avons présenté l'essentiel du sens que donne au mot banque le vocabulaire économique contemporain. Toutefois, la liste des synonymes potentiels présentée dans le tableau 1 n'est pas totalement épuisée. Il reste quelques vocables qui ne peuvent être rattachés à l'un ou l'autre des deux grands sens possibles que nous venons de décrire succinctement et qui fournissent des acceptions particulières que tout bon dictionnaire se doit de mentionner à la fin des principaux articles. Pour le mot "banque", les plus intéressants sont incontestablement "président" et "affaire". L'importance du premier s'explique par la synecdoque classique consistant à désigner une entreprise par son chef comme on le fait d'un pays par sa capitale. Mais ici, il ne s'agit pas du PDG, comme pour les entreprises non financières, mais du *président*, ce qui fournit un autre signe de la domination de la culture anglo-saxonne dans le domaine économique... Quant à "affaire", la majorité des emplois de ce substantif ne correspondent pas au synonyme d'activité (on disait autrefois "banque d'affaires") mais à "scandale". Voici à titre d'exemple, la phrase la plus caractéristique de ce sens que les banques ne souhaitaient sans doute pas :

*Ainsi, Barings a été la "victime" d'un opérateur de Singapour ; Daiwa a été "trompé" par le patron de sa filiale de New York ; le Crédit Lyonnais a perdu des milliards de francs dans le financement du cinéma américain par l'intermédiaire de sa filiale néerlandaise et les pertes les plus lourdes de la banque publique française proviennent, pour l'essentiel, de filiales mal ou pas contrôlées ; les déboires du Crédit Foncier sont aussi la conséquence des risques pris par des filiales...* (*Le Monde*, 6 février 1996)

#### 4. Conclusions

La signification du mot "banque" dans le vocabulaire économique contemporain est maintenant établie de manière synthétique et précise. Au fond, deux conclusions principales se dégagent : d'une part, l'importance primordiale des marchés financiers et des modèles d'organisation anglo-saxons et, d'autre part, la dématérialisation progressive des activités bancaires qui se traduit d'ailleurs par une quasi-disparition de la monnaie et des chiffres dans les articles de presse portant sur ces sujets !

Parmi les questions en suspens, nous voudrions insister sur quelques points.

En premier lieu, la fixation des seuils (pour le calcul des spécificités puis pour les zones de synonymie) demeure assez empirique. Nous avons choisi des seuils élevés afin de ne pas encombrer cette communication avec des listes trop longues. Naturellement, ce choix conduit sans doute à une présentation appauvrie par rapport à l'extraordinaire diversité d'emploi d'un des principaux vocables de la langue des "affaires" (au sens noble du terme !)

En second lieu, les résultats sont dépendants de la taille et de la représentativité du corpus. Ce genre de recherche nécessite de vastes collections de textes couvrant une période de temps assez longue. Malgré son étendue, notre corpus n'atteint peut-être pas encore cette "dimension critique". Par exemple, le lecteur aura sans doute repéré, dans les vocabulaires et les phrases caractéristiques, la trace de l'ambiance particulière des années 1996-98 marquées par les défaillances en série dans le secteur bancaire — essentiellement à cause de la crise du marché immobilier — et par les scandales (Barings, Crédit Lyonnais, Rivaud, etc.)

Naturellement, il faut que ces corpus soient exploitables : les graphies doivent être normalisées puis rattachées à leur lemme et à leur catégorie grammaticale... Certes l'opération a un coût, mais, pour parler comme les financiers, nous espérons avoir suggéré combien cet investissement pourrait être rentable.

Enfin, il est évident que l'ordinateur ne pourra jamais rédiger automatiquement un article de dictionnaire. L'exemple qui vient d'être présenté prouve simplement que la statistique textuelle apportera à l'avenir, une grande aide aux lexicographes. L'outil pourrait également trouver des applications évidentes dans de nombreuses activités allant de l'indexation et la gestion des grandes bases de données à la critique littéraire, en passant par la traduction assistée par ordinateur ou la recherche d'informations sur la toile... Bien sûr, la qualité de l'analyse dépendra toujours de l'aptitude du chercheur à poser les "bonnes questions" et à sélectionner les informations pertinentes dans les matériaux apportés par l'ordinateur.

## Références

- Dubois C. et Dubois J. (1971). *Introduction à la lexicographie*. Paris, Larousse.
- Harris Z. (1970). La structure distributionnelle. *Langages*, 20 décembre 1970.
- Harris Z. (1971). *Structures mathématiques du langage*. Paris, Dunod.
- Hubert P. et Labbé D. (1995). La structure du vocabulaire du général de Gaulle. In BOLASCO S. et al, *IIIe Giornate internazionali di analisi statistica dei dati testuali*, Rome, CISU, II, p 165-176.
- Labbé C. et Labbé D. (1994). Que mesure la spécificité du vocabulaire ?. Grenoble, CERAT. Repris dans : *Lexicometrica*, 3, 2001.
- Labbé D. (1998). Le nous du général de Gaulle. *Quaderni di studi linguistici*. 4/5, p 331-354.
- Lafon P. (1984). *Dépouillements et statistiques en lexicométrie*. Genève-Paris, Slatkine-Champion.
- Lebart L. et Salem A. (1994). *Statistique textuelle*. Paris, Dunod.
- Martinez W. (2000). Mise en évidence de rapports synonymiques par la méthode des cooccurrences. In Rajman M. et Chappelier J.-C. eds. *Actes des 5<sup>e</sup> journées internationales d'analyse des données textuelles*. Lausanne, Ecole polytechnique fédérale, vol 1, p 85-94.
- Matore G. (1973). *La méthode en lexicologie*. Paris, Didier.
- Mortureux F. (1997). *La lexicologie entre langue et discours*. Paris, SEDES.
- Muller C. (1977). *Principes et méthodes de statistique lexicale*. Paris, Hachette.
- Picoche J. (1977). *Précis de lexicologie française*. Paris, Nathan.
- Picoche J. (1986). *Structures sémantiques du lexique français*. Paris, Nathan.
- Niklas-Salminen A. (1997). *La lexicologie*. Paris, A Colin.

Rey A. (1977). *Le lexique. Images et modèles du dictionnaire à la lexicologie*. Paris, A. Colin.

Salem A. (1987). *Pratique des segments répétés. Essai de statistique textuelle*. Paris, Klincksieck.

## Annexe

Univers lexical de « banque » dans le corpus « vocabulaire de la presse économique »

### 1. Vocabulaire significativement suremployé

(Seuil de 1%, classement par catégories grammaticales et spécificités décroissantes).

**Mots à majuscule** : Paribas, Crédit Lyonnais, Rivaud, Crédit Agricole, Société Générale, Indosuez, CIC, BNP, Suisse, CDR, Italie, Lazard, Mediobanca, BTP, AFB, Natexis, Barings, JP Morgan, Crédit Mutuel, Comptoir des Entrepreneurs, Delmas, Lehman Brothers, Bâle, NatWest, Warburg, Pallas-Stern, Hokkaido Takushoku, Angleterre, Marsalet, BIP, Hongkong, Lucien, UBS, Douroux, BIANC, Goldman Sachs, GAN, Bolloré, Royaume-Uni, Crédit Foncier, OPA, Eurotunnel, Fiat, PME, New York,

**Verbes** : détenir, diriger, recentrer, demander, mener, accorder, appeler, avoir, regrouper, devenir, obliger, contrôler, imaginer, prendre, estimer, constater, peser, venir, financer,

**Substantifs** : affaire, activité, filiale, établissement, investissement, fonds, taux, compte, gestion, crédit, opération, assurance, risque, finance, perte, créance, caisse, prêt, contrôle, association, financement, agence, bilan, faillite, épargne, dette, immobilier, sauvetage, sicav, assureur, tutelle, détail, gouverneur, défaillance, commerçant, encours, défaisance, caution, abandon, rapprochement, analyste, consortium, recapitalisation, intérêt, entité, donnée, réserve, système, commerce, client, fusion, acquisition, mandat, proximité, institution, portefeuille, autorité, commission, président, provision, argent, privatisation, actif, dépôt, spécialiste, marché, ministère, département, dirigeant, actionnaire, banquier, métier, participation, rentabilité, redressement, titre, partie, union, difficulté, fédération, pouvoir, crise, dossier, organisation, exploitation, restructuration,

**Adjectifs** : grand, français, financier, central, public, bancaire, régional, international, important, commercial, britannique, vert, italien, immobilier, populaire, domestique, mutualiste, créancier, suisse, monétaire, étranger, anglo-saxon, direct, chinois, belge, privé, japonais, lourd, européen,

**Pronoms** : qui, que, celui, un,

**Adverbes** : ainsi, notamment, auprès, récemment, très, puis, enfin, tant, encore, trop, non, alors, longtemps, aussi,

**Déterminants** : le, leur, certain, deuxième,

**Conjonctions et prépositions** : de, et, par, comme, parmi,

### 2. Vocabulaire significativement sousemployé

(Seuil : 1%, classement par catégories grammaticales et spécificité décroissante)

**Noms propres** : Alain, Renault, Etats-Unis, Amérique, Américain, Français, PDG, France Télécom,

**Verbes** : affirmer, produire, représenter, falloir, dépasser, investir,

**Substantifs** : image, matière, qualité, économie, mission, industriel, cours, recette, recherche, temps, impôt, constructeur, milliard, géant, patronat, alliance, stratégie, chômage, technologie, ingénieur, salaire, croissance, firme, nombre, prix, entreprise, emploi, marque, usine, chiffre, production, virgule, jour, chaîne, salarié, succès, monde, budget, voiture, contrat, site, fabricant, magasin, jeune, travail, gamme, film, guerre, vente, moyenne, commande, formation, franc, sécurité, chef, syndicat, distributeur, lieu, consommateur, loi, cadre, dollar,

**Adjectifs** : nécessaire, beau, destiné, entier, jeune, seul, vendu, plein, unique, social, nouveau, humain, numérique,

Pronoms : tout, y, il, je, nous, vous,

Adverbes : pas, vite, bien, ne, moins,

Déterminants : tel, neuf, ce, cinq, mille, cent, deux, notre, dix, mon, quarante, cinquante, vingt, quatre, chaque, trois, même, six, huit, trente,

**Conjonctions et prépositions** : sur, contre, que, pendant, dès, à, chez, mais, en, près, pour, quand, voilà

### 3. Phrases les plus caractéristiques (en valeur absolue) :

Parmi les plus gros deals de l'année, outre la fusion record dans l'assurance, on retrouve le rapprochement entre le Crédit Local de France et le Crédit Communal de Belgique, l'acquisition de la banque Indosuez par le Crédit Agricole, et les OPA de Paribas sur la Navigation Mixte ou de la Caisse des Dépôts sur le Crédit Foncier (*L'Expansion*, 19 décembre 1996)

Le contrôle des banques devient un casse-tête pour les autorités de tutelle la multiplication des défaillances bancaires souligne la difficulté croissante, pour les banques centrales et les grands organismes internationaux, de mesurer les risques pris par les établissements financiers. (*Le Monde*, 6 février 1996)

**4. Phrases les plus caractéristiques (en valeur relative) :**

Les banques françaises ont aussi recentré leurs activités et simplifié leurs structures (*Le Monde*, 4 octobre 1997).  
Encore récemment, la Banque Commerciale Privée (BCP) et surtout la banque Pallas-Stern ont ainsi déposé leur bilan. (*Les Echos*, 23 avril 1997)

La banque Rivaud avait créé un système de recyclage d'argent sale. (*Le Monde*, 26 juin 1997)