

2^o JOURNEES IUT DE LA RECHERCHE EN SCIENCES SOCIALES ET HUMAINES

LES ORGANISATIONS AU RISQUE DE L'INFORMATION

Toulouse 17 - 18 mars 1994

Comité scientifique

Robert BOURE (Sciences de l'information et de la communication, LERASS, IUT de Toulouse), Patrick CHASKIEL (Sciences de l'information et de la communication, IUT de Nancy), Jean-Luc DALLEMAGNE (Sciences économiques, IUT du Val de Marne), Jean-Pierre DURAND (Sociologie, Université d'Evry Val d'Essonne), Claude JAMEUX (Sciences de gestion, IUT d'Annecy), Philippe JEANNIN (Sciences économiques, LERASS, IUT de Tarbes), Anne-Marie LAULAN (Sociologie, Université de Bordeaux 3), Michel MARCHESNAY (Sciences de gestion, Université de Montpellier 1), Pierre MEVELLEC (Sciences de gestion, Université de Nantes), Bernard MIEGE (Sciences de l'information et de la communication, Université de Grenoble 3, Responsable scientifique du programme PARINFO), Jean MOUCHON (Président de la Société Française des Sciences de l'Information et de la Communication, Université de Lille 3), Michel PERRIN (Anglais de spécialité, Université de Bordeaux 3), Jocelyne SIMBILLE (Droit public, LERASS, IUT de Toulouse), Christine VAILLANT (Anglais de spécialité, ERDIL, IUT de Toulouse).

Dans ce que certains appellent déjà la "société d'information et de communication", et au sein d'organisations publiques et privées, l'information est de plus en plus présentée comme une variable incontournable du changement. Ce discours, très prégnant dans le monde professionnel, est (partiellement) repris par une fraction des chercheurs en Sciences sociales et humaines.

Dans le même temps, parce qu'elle est bien souvent réduite à ses dimensions "techniques" l'information est traitée comme un adjuvant, nécessairement bénéfique, permettant aux organisations de se développer et de se réformer. Toute idée de tension, de perturbation et, *a fortiori*, de conflit, est évacuée. De plus en plus souvent, l'information est invoquée comme la solution "miracle" ou "passe-partout" (c'est selon) dès qu'apparaît une crise générale ou locale.

Des études empiriques ont pu montrer que l'action de l'information dans les organisations peut être saisie autant comme une solution que comme une source de problèmes, parfois nouveaux ou spécifiques. Il semble donc peu rigoureux de l'appréhender de façon univoque et, qui plus est, à travers les seules relations de détermination technicistes. L'environnement, les rapports de forces, les relations que les acteurs entretiennent avec elle ne doivent pas être sous-estimés.

Parce qu'elle s'étend dans les organisations publiques et privées, l'information suscite le développement de multiples dispositifs et supports dont le fonctionnement n'est pas sans incidence sur les pratiques individuelles et collectives. Mais, ici encore, des travaux de recherche ont révélé qu'outils et dispositifs liés à l'organisation du travail, à la gestion, à la communication, ne produisent pas d'effets homogènes dans l'espace et stables dans le temps.

Sur le thème *Les organisations au risque de l'information*, et dans un cadre scientifique, les 2^o Journées IUT de la Recherche en Sciences sociales et humaines se proposent d'interroger le statut et le rôle de l'information dans les organisations publiques (administrations d'Etat, collectivités territoriales....) et privées (entreprises, associations, syndicats, partis politiques...). Elles seront organisées par le LERASS, laboratoire d'accueil reconnu par la DRED, et placées sous le patronage d'un Comité scientifique.

Les communications pourront privilégier le fonctionnement global des organisations ou bien les rapports inter-individuels, se référer à des travaux empiriques (connus, inédits) ou développer des approches théoriques, s'inscrire dans le cadre d'une discipline particulière (Economie, Gestion, Psychologie sociale, Droit...) ou se placer du point de vue de l'interdisciplinarité.

Sont plus particulièrement attendues des communications dans des axes suivants:

- 1) Outils et méthodes de production/diffusion de l'information dans les organisations
- 2) Information et "performance" des organisations
- 3) Information, rapports de forces, jeux de pouvoirs et conflits

Ces Journées sont ouvertes à tout enseignant en poste ou non en IUT ainsi qu'aux chercheurs et aux professionnels.

Afin de faciliter le travail de lecture par le Comité scientifique, il est demandé à ceux qui souhaitent proposer une contribution d'envoyer, avant le 30 novembre 1993, un résumé de deux pages maximum, faisant état des hypothèses, de la problématique utilisée et de la méthodologie mise en oeuvre.

Les travaux des 2^o Journées IUT de la Recherche en Sciences Sociales et Humaines seront publiés au cours de l'année 1994 sous la forme d'Actes.

Pour toute information, s'adresser à:
Robert BOURE, LERASS
IUT, II5 Route de Narbonne
31077 Toulouse Cedex
Tel: 62 25 82 80
Fax: 62 25 80 01

**2° JOURNÉES IUT DE LA RECHERCHE EN
SCIENCES SOCIALES ET HUMAINES**

Réponse à l'appel à communications

NOM et prénom:

Qualité:

Adresse professionnelle:

Tel professionnel:

Fax:

Adresse personnelle:

J'ai l'intention de présenter une communication sous le titre de (joindre un résumé de deux pages maximum, avant le 30 novembre 1993):

Bulletin de participation aux 2° Journées IUT

(à remplir par tout participant; joindre un chèque ou un bon de commande de 300 F pour participation aux frais, incluant les deux déjeuners)

NOM et prénom:

Qualité:

Adresse professionnelle:

Tel professionnel:

Fax:

Adresse personnelle:

**A renvoyer à Viviane COUZINET, Comité d'organisation des Journées IUT,
LERASS, IUT 115 Route de Narbonne, 31077 Toulouse cedex.**

Tel: 62 25 82 80

Fax: 62 25 80 01

Mise en évidence de concepts dans le discours oral par une nouvelle méthode d'indexation sur des critères statistiques

Baldit P. (*), Retourna C. (*), Quoniam L. (*), M. Roux(**), Dou H. (*)

(*) CRRM - Faculté des Sciences de St Jérôme - 13397 Marseille Cedex 20

(**) Chargé de Mission aux Relations Industrielles - Délégation Régionale PACA - CNRS - 31 Chemin Joseph Aiguier - 13402 Marseille Cedex 20

Résumé

La problématique posée dans le cadre de cette étude vise à acquérir une meilleure connaissance du phénomène de l'innovation dans l'entreprise. Au travers d'entretiens semi-directifs menés avec un certain nombre de chefs d'entreprises, un important volume d'informations a été recueilli sous forme de discours oral.

Il s'agit alors de développer une méthodologie d'analyse statistique du discours de façon à dégager les concepts apparaissant comme pertinents dans le processus d'innovation de manière semi-automatique. Ceci passe au préalable par une nécessaire réduction de la diversité du vocabulaire rencontré sans perte d'informations importantes. En effet la principale difficulté de l'analyse du langage naturel réside dans la grande diversité du vocabulaire employé. Un certain nombre d'outils et de techniques employées ici permettent de ramener ce vocabulaire à une dimension acceptable pour une analyse statistique.

Une fois ce vocabulaire réduit, au moyen de technique statistique adaptée, on peut alors mettre en évidence un certain nombre de groupement mots. Ces groupements sont rattachés aux concepts préalablement reconnus pertinents relativement à l'innovation. Ces concepts, définis *a priori*, sont validés ou invalidés par l'analyse d'un ensemble d'entretiens qui permettra de dégager une "cartographie" des mécanismes d'innovation dans l'entreprise.

En ce qui concerne l'analyse statistique des entretiens, elle s'effectue sur des données représentant la présence ou l'absence de mots dans :

- 1 - Les réponses aux questions posées durant l'entretien.
- 2 - Les phrases contenues dans l'ensemble de l'entretien.

Notre laboratoire a développé un algorithme de sériation par blocs qui est capable d'effectuer des regroupements par classes de manière globale suivant les deux dimensions à analyser simultanément. Ces classes sont déterminées à la fois par un groupement de mots et un groupement de phrases (ou réponses). Les classes sont ensuite affectées à des concepts et lorsqu'il y a ambiguïté sur les concepts, il est nécessaire de revalider la liste des concepts en cours.

Ainsi un concept va regrouper un ensemble de mots et de réponses. Ces regroupements vont permettre une nouvelle indexation sur des critères statistiques des interviews recueillies, afin d'aboutir à une consultation de ces interviews sous forme hypertexte. L'utilisateur de cette base de donnée hypertexte va pouvoir se "déplacer" de réponses en réponses par l'intermédiaire de mots clefs symbolisant le concept à analyser.

Mise en évidence de concepts dans le discours oral par une nouvelle méthode d'indexation sur des critères statistiques

Baldit P. (*), Retourna C. (*), Quoniam L. (*), M. Roux(**), Dou H. (*)

(*) CRRM - Faculté des Sciences de St Jérôme - 13397 Marseille Cedex 20

(**) Chargé de Mission aux Relations Industrielles - Délégation Régionale PACA - CNRS - 31 Chemin Joseph Aiguier - 13402 Marseille Cedex 20

Position du problème

L'objectif fixé dans cette communication est celui de la recherche et de la mise en évidence des points clés - facteurs reproductibles - du processus d'innovation dans les petites et moyennes entreprises. Cette problématique s'inscrit dans le cadre du programme STRIDE¹ (Science et Technologie pour les Développements Régionaux en Europe) de la Communauté Européenne, dirigé en France par l'ANVAR.

S'il est de plus en plus admis que l'information est une matière première pour l'innovation dans les entreprises [1], de nombreux modèles existent pour décrire ce phénomène.

L'intérêt de l'innovation dans le processus de développement économique ne fait plus de doutes depuis les travaux de Schumpeter [2], premier à avoir attiré l'attention sur ce phénomène en situant l'innovation comme la phase qui suit immédiatement l'invention et permettant de la faire passer au stade commercial. L'innovation est vue comme génératrice de rupture, pour l'entreprise innovante, par rapport au contexte concurrentiel. L'entreprise est projetée, lors de la mise en oeuvre de l'innovation, au-delà de la concurrence, l'innovation lui procurant un avantage compétitif provisoire. Cet avantage ne prendra fin que lorsque la diffusion de la nouveauté, entraînant des entreprises imitatrices, la fera réintégrer son état antérieur.

Le schéma le plus ancien du processus d'innovation reste le classique modèle linéaire Idée - Développement - Production, ou Idée - Technique - Réalisation (Ideas - Techniques - Commodities en Anglais), correspondant en fait au schéma Recherche et Développement traditionnel.

Mais d'autres approches tendent vers une représentation plus complexe de cette activité de R&D. On peut par exemple considérer qu'à chacune des dimensions évoquées plus haut peut se rattacher une multitude de notions, domaines de connaissances ou domaines techniques, interdépendants, et décrivant les connaissances nécessaires à l'accomplissement de chaque phase. Mais ces connaissances sont également liées entre elles d'une phase à l'autre. Ainsi des techniques relativement diverses reposent sur des idées scientifiques communes, et des réalisations variées partagent des techniques communes. On aboutit par là à une représentation proche de celle d'un réseau neuronal [3].

L'innovation peut également se percevoir, sous un autre aspect, comme la résolution de problèmes scientifiques ou techniques posés pour répondre à des exigences liées à des conditions de coûts ou de marché [4].

¹Rassemblant la délégation régionale du CNRS PACA, la Chambre de Commerce et d'Industrie de Marseille (CCIM), le Centre de Recherche Rétropective de Marseille (CRRM) et le Laboratoire d'Economie et de Sociologie du Travail (LEST)

On retrouve ici la notion de recomposition de son univers intérieur par une entreprise [5]: ce sont les notions idéales de firmes "actives" et de firmes "passives". Les entreprises "actives" recherchent délibérément l'innovation et intègrent leur environnement en tant que ressource dans leur processus de conception, planification et gestion de l'innovation. Les entreprises "passives", en revanche, ne recherchent pas systématiquement l'innovation ni ne prennent en compte l'environnement extérieur, si ce n'est au travers des structures prévues à cet effet.

Cette typologie, volontairement succincte, peut être complétée par celle issue des travaux de Pavitt [6], décrivant les différences de "trajectoires technologiques" des entreprises. Elle distingue tout d'abord les entreprises dominées par les offreurs (de bien et de machines), où les innovations sont essentiellement de processus. Ensuite viennent les offreurs spécialisés, innovant plutôt vers les produits, puis les producteurs de masse, innovant aussi bien en produits qu'en procédés. Enfin on compte les entreprises basées sur la science, où les innovations sont directement liées à l'apparition de nouvelles connaissances scientifiques.

Mais par-delà ces différentes typologies, on peut tenter de déterminer les divers paramètres intervenant dans un processus d'innovation. On peut penser qu'ils ne seront pas trop éloignés des facteurs reproductibles - les sept "facteurs d'innovativité" - recensés par l'ANVAR [7] tout au long de ses années de contacts avec les entreprises : facteur lié à l'environnement, aux ressources humaines, aux ressources technologiques, aux ressources financières, à l'organisation, à la stratégie et enfin au dirigeant lui-même. La nature des entreprises considérées, PME/PMI à forte vocation de sous-traitance ou de services aux entreprises dans une zone économique en difficulté (Pourtour de l'Etang de Berre) sera de nature à souligner ou diminuer l'importance de facteurs envisagés dans une perspective globale des entreprises.

L'objectif de cette communication est d'arriver à dégager, au travers d'une source d'informations particulière qu'est un entretien semi-directif avec un dirigeant d'entreprise, les principaux thèmes abordés par ce dernier tout au long du récit de l'histoire de l'entreprise et de ses produits. Des travaux en cours et à venir viseront, à partir des thèmes isolés dans le discours du dirigeant, à rattacher ou déterminer les facteurs d'innovativité présents ou potentiels pouvant expliquer ou révéler un mécanisme d'innovation effectif ou potentiel.

Application au traitement d'interviews

Acquisition de l'information

L'information recueillie et destinée à être analysée se présente sous la forme d'un entretien, enregistré sur bandes audio, avec le dirigeant de l'entreprise ou un membre de la direction.

Les thèmes abordés lors de cet entretien semi-directif, mené par un binôme d'enquêteurs, sont essentiellement² : l'histoire de l'entreprise et de ses produits, sa situation et ses perspectives de développement, la manière dont l'entreprise reste en contact avec le marché et son évolution, la question du changement technologique dans l'activité de l'entreprise, le problème des qualifications à l'intérieur de l'entreprise et de leurs adéquations avec les projets de celle-ci, enfin la manière dont l'entreprise accède à l'information qui lui est nécessaire (contacts, réseaux, ...).

Les entretiens, enregistrés, sont retranscrits sur support informatique en vue de l'analyse statistique.

Traitement préalable

²Consignes d'interviews mises au point avec le concours du LEST, 35 avenue Jules Ferry 13626 Aix en Provence CEDEX

Un entretien sous sa forme brute ne se prête pas à une analyse statistique, aussi bien en raison de la quantité d'information à traiter que de la diversité du vocabulaire rencontré. En effet un entretien se présente en général sous la forme d'un texte (sous forme de questions-réponses) de 10 à 15 pages comprenant entre 5000 et 7000 mots (*occurrences*). Il est alors nécessaire, dans l'optique d'une analyse de la structure d'un entretien, de parvenir à réduire cette diversité de vocabulaire, et donc de ramener la quantité d'informations présentes à des proportions abordables en termes de temps de calculs.

Ce travail est réalisé en quatre étapes.

Tout d'abord, il est nécessaire d'établir le dictionnaire complet des mots (formes) présents dans un entretien, ce qui est accompli au moyen d'une des fonctionnalités d'un logiciel du commerce développé par le CNRS [8]. L'ordre de grandeur du dictionnaire pour un entretien se situe entre 1000 et 1700 formes.

Une deuxième étape consiste, à partir du dictionnaire obtenu, à lemmatiser toutes les formes pour lesquelles le dictionnaire nous indique une racine commune, ou plus précisément une forme réduite commune.

La troisième étape est la suppression de certains mots outils, telles les locutions, conjonctions de coordination ou de subordination, que l'on retrouve très souvent (à très forte *fréquence*) et dont la présence n'est pas significative. Un dénombrement des formes restantes dans l'entretien sous sa forme réduite [9] donne un aperçu du nombre (diminution de 50%) et de la répartition en fréquence des formes restantes.

La dernière étape est une analyse et sélection pour suppression de certains mots parasites ou considérés comme non porteurs de sens dans la perspective envisagée.

Le résultat final se compose d'une matrice récapitulant pour chaque phrase de l'entretien la présence ou l'absence de chaque forme du vocabulaire réduit.

		Descripteurs = Mots du vocabulaire lemmatisé						
Individus = Phrases	1	1			1	1		
		1	1	1				1
	1	1						
		1	1	1	1	1		1
	1		1		1	1		
				1		1		1

Figure 1

Méthode Statistique

Nous avons utilisé pour cette étude un outil élaboré dans notre laboratoire appelé DATABLOC qui traite des matrices de données binaires de présence absence. Suivant les travaux [10] de l'équipe de F. MARCOTORCHINO³ en analyse relationnelle des données et de J.M PROTH [11] sur les îlots de production, nous avons développé une heuristique adaptée aux données rencontrées en science de l'information. L'heuristique qui a été développée permet de traiter un grand nombre de données dans des temps acceptables sur IBM-PC. Bien

³IBM CEMAP (Centre Européen de Mathématiques Appliquées), 68-76 quai de la rapée, 75592 PARIS Cedex 12 (France)

qu'initialement prévue pour des analyses de données bibliométriques il s'est avéré intéressant de tester cette heuristique sur des entretiens semi-directifs.

L'heuristique proposée conduit à une classification automatique non hiérarchique des données. La sériation s'effectue simultanément sur les deux espaces, celui des descripteurs et celui des individus. Outre les capacités mathématiques de réduction de l'information de ce logiciel, il s'est révélé primordial d'avoir la possibilité de consulter les interviews réalisées en ayant un guide de lecture issu de l'analyse statistique. Dans cette optique une visualisation hypertexte des interviews permettant de "naviguer" entre les phrases se rapportant au même thème a été mise en place. Nous allons rappeler brièvement sans faire appel aux détails mathématiques qui ne rentrent pas dans le cadre de cette présentation, la méthode utilisée.

La donnée de base à traiter par le logiciel est une matrice binaire de présence absence que l'on peut schématiser selon la figure 1.

Les lignes de cette matrice correspondent aux individus à analyser et les descripteurs aux caractéristiques de ces individus. Ainsi à l'intersection d'une ligne i et d'une colonne j on trouvera codifiée la présence du descripteur j pour l'individu i par la notation binaire 1 et l'absence de ce descripteur sera codifiée par la notation binaire 0. On recherchera à faire apparaître des blocs diagonaux de densité de présence maximale en permutant les lignes et les colonnes de la matrice de départ (figure 2).

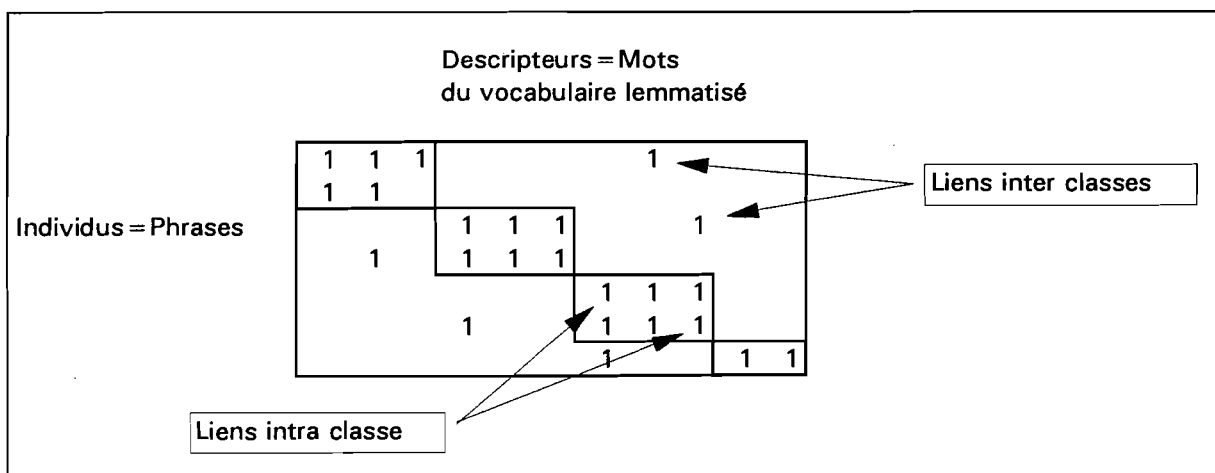


Figure 2

Nous constituons donc des classes (ici 4) qui sont composées d'individus ayant en commun certaines caractéristiques. Selon les données traitées ces regroupements peuvent avoir des significations différentes. Le logiciel permet également d'isoler les descripteurs présents dans la quasi-totalité des individus (descripteurs "triviaux"⁴) de manière automatique en calculant le seuil de séparation en fonction de la distribution statistique des données contenues dans la matrice. Un processus analogue est également utilisé afin de déterminer s'il y a lieu, les individus qui peuvent être considérés au sens statistique comme étant triviaux (possédant la quasi-totalité des descripteurs). Dans le domaine lexicographique, on pourra ainsi faire apparaître les mots qui se retrouvent dans la quasi-totalité des phrases et considérer ceux-ci comme étant des mots "triviaux" pour la matrice analysée. Les blocs formés lors du traitement statistique permettent de dégager des zones délimitées par un certain nombre de phrases et de mots, ces zones étant porteuses de significations potentielles.

⁴Le terme trivial ne sera pas pris en compte ici dans son sens strict mais au sens statistique.

Les méthodes statistiques usuelles comme l'ACP ou l'AFC [12] produisent des résultats difficiles à analyser par la multiplicité des axes de projection générés et le grand nombre d'individus traités (problèmes de représentation graphique : superposition des individus).

Nature des données traitées

Très souvent en science de l'information nous sommes confrontés à des distributions caractéristiques lors de l'étude de matrices binaires de présence absence de descripteurs sur une population donnée. En effet si on classe par ordre décroissant en fréquence les mots rencontrés dans un texte quelconque on obtient une distribution qui suit une loi de type Zipfienne [13] (figure 3).

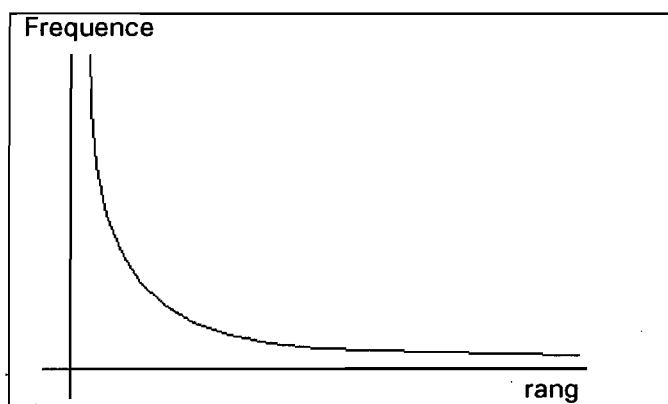


Figure 3

Les mots de forte fréquence sont les mots les plus utilisés dans la langue française et les mots de très faible fréquence correspondent à des termes très précis du texte. Suivant cette particularité, il est possible de mettre en évidence trois zones zones [14] (figure 4).

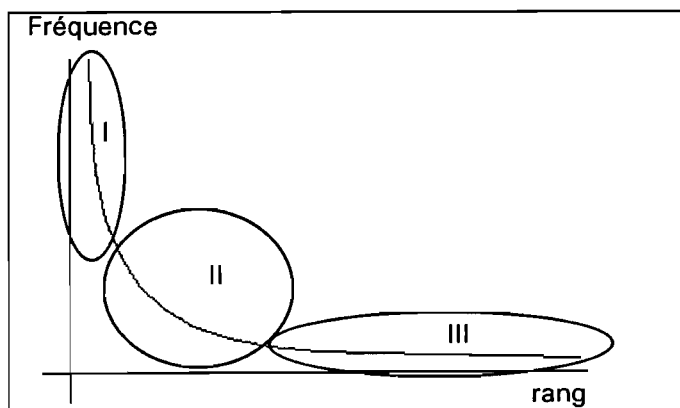


Figure 4

- Zone I : Mots triviaux, généralement termes utilisés dans le discours
- Zone II : Mots probablement porteurs d'information intéressante
- Zone III : Mots probablement peu porteurs d'information intéressante

De plus ces matrices ont la particularité d'être hypercreuses (peu de présence de descripteurs). Du fait de la structure non gaussienne de cette distribution[15], les méthodes d'analyse de données classiques bien qu'applicables sont biaisées par ce phénomène (ACP, AFC etc.). C'est la raison pour laquelle nous avons développé une nouvelle méthode permettant de

définir les différentes zones et d'analyser chaque partie séparément afin de déterminer des groupements de mots dans des secteurs de la courbe où les fréquences de ces mots sont relativement peu éloignées.

Analyse statistique

La matrice est traitée par le logiciel DATABLOC⁵ afin de faire émerger des groupes constitués de phrases et de mots clefs représentant les sujets abordés par le dirigeant d'entreprise.

Interprétations et Résultats

L'interprétation des résultats obtenus (sur un échantillon de 5 interviews analysées) peut se pratiquer selon deux niveaux, interprétation statistique et interprétation de contenu des interviews (différenciation du contenu et de la structure du discours).

Du point de vue statistique le nombre et la taille des classes apparues lors de la sériation donnent un reflet de la structure globale du discours.

Des classes riches en mots, mais ne regroupant qu'un nombre relativement restreint de phrases (classes "rectangulaires") sont significatives de phrases longues dans le texte de référence (figure 5). Ceci peut caractériser des phrases longues utilisant un vocabulaire étendu et/ou abondant de nombreux thèmes (chef d'entreprise "bavard"), mais ne nous permet pas une interprétation plus fine en terme de sens.

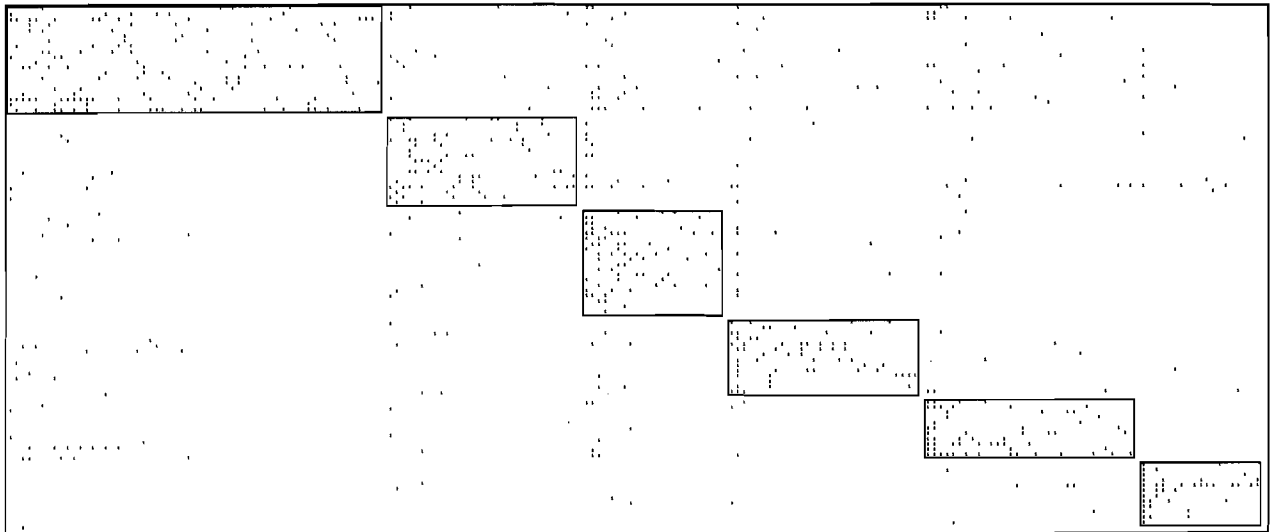


Figure 5

Il en va de même pour des classes plus nombreuses et de faibles dimensions (classes "carrées"), significatives de phrases plus concises (figure 6), utilisant un vocabulaire plus restreint et/ou ne traitant qu'un minimum de sujets (chef d'entreprise "peu loquace").

⁵logiciel développé au CRRM

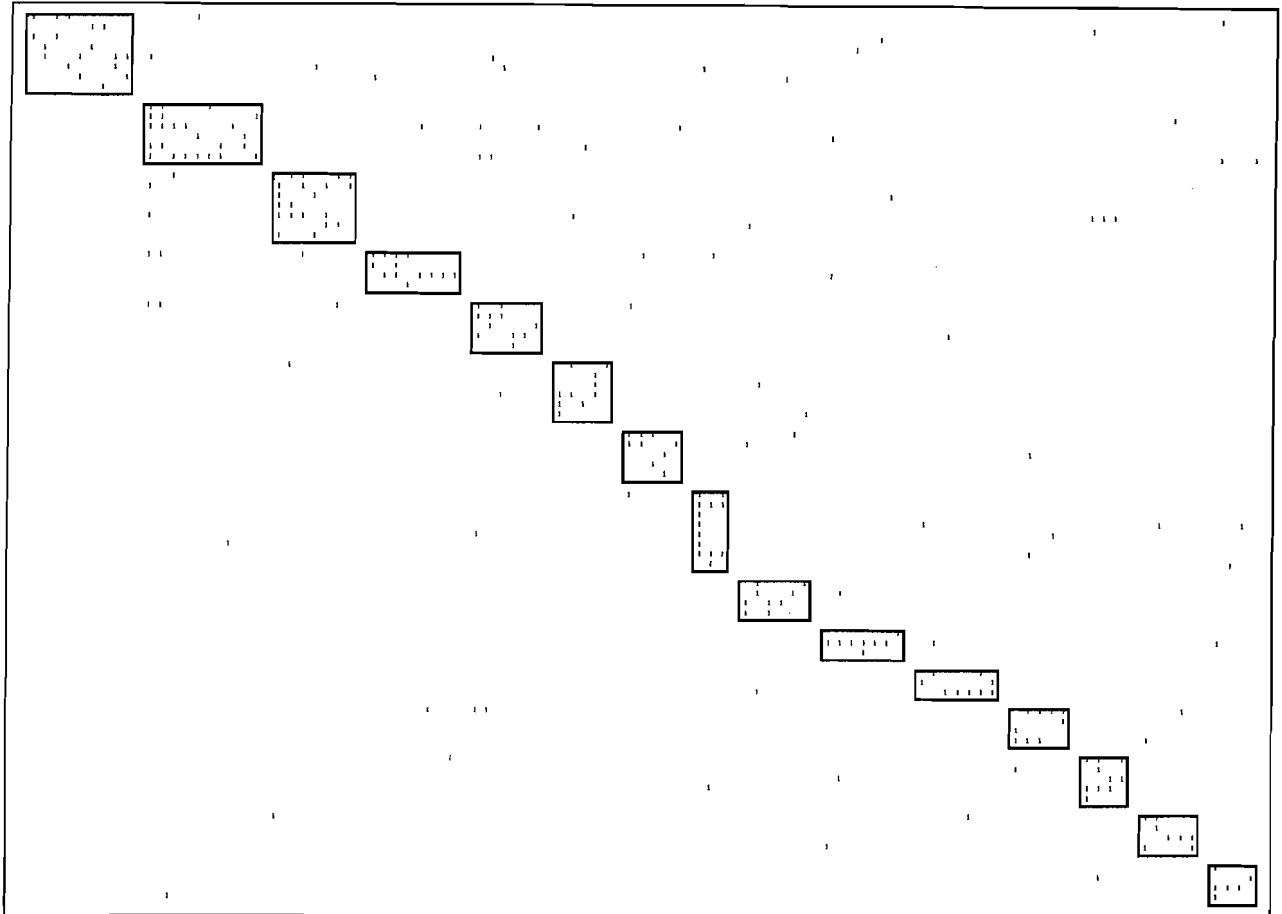


Figure 6

Ce type d'interprétation statistique est difficile à généraliser sur les interviews analysées en raison des nombreux biais qui peuvent être apportés à ce genre d'analyse. L'un des biais les plus importants pouvant être le comportement de l'intervieweur lui-même, qui peut influencer directement la disposition du chef d'entreprise à répondre ainsi que le volume et la nature de ses réponses.

Toutefois des travaux sont en cours pour déterminer des indicateurs permettant de quantifier la structuration du discours.

Une interprétation plus sémantique est en revanche permise par la visualisation sous forme hypertexte des classes obtenues par la sériation et des phrases du texte source auxquelles elles font référence.

En effet, une classe est composée d'un ensemble de mots sous leur forme réduite et des numéros des phrases dans lesquelles ils se trouvent. Ce n'est donc pas exploitable directement du point de vue de l'interprétation du discours.

Mais un outil de visualisation nous permet de consulter les classes et leur contenu en regard des phrases de référence du discours (figure 7).

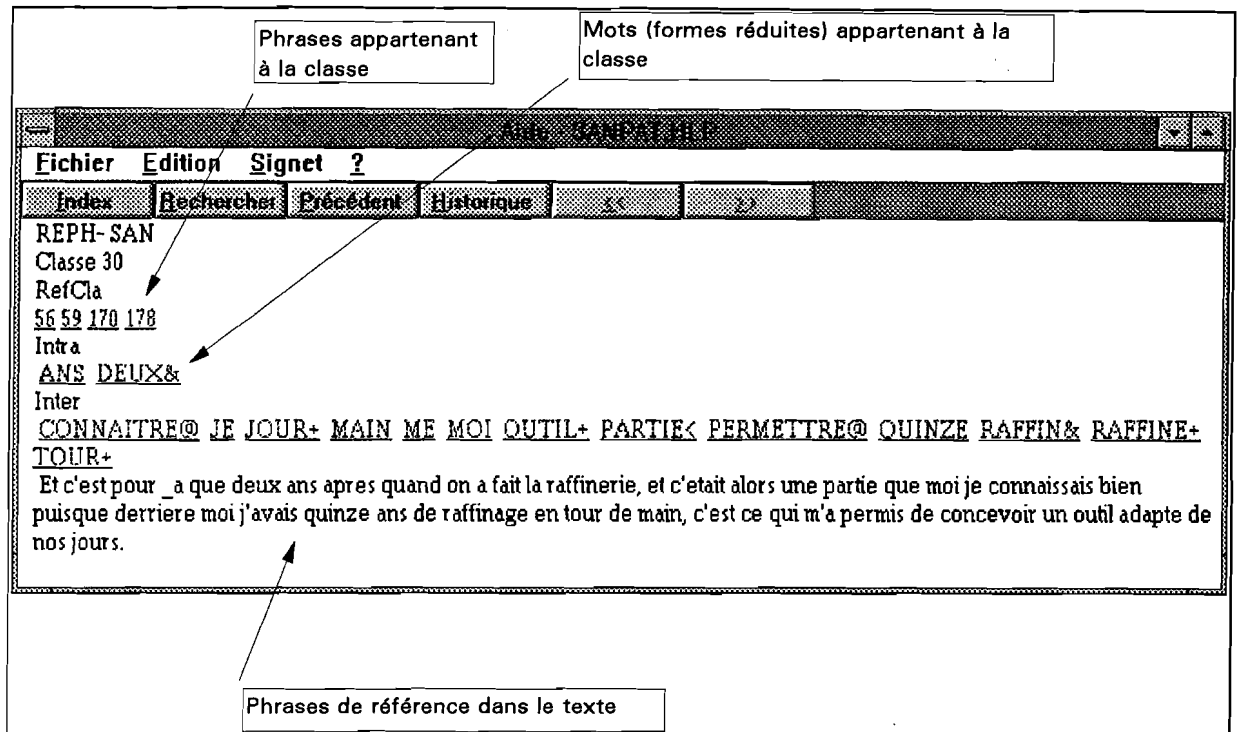


Figure 7

Le retour au document primaire, étape indispensable de toute aide à la lecture, permet ici d'établir la nature des liens entre les différentes phrases d'une même classe.

Certains de ces liens sont relatifs à la polysémie de certains mots : plusieurs phrases dans une même classe comportant un même mot prenant tour à tour un sens différent dans chacune d'entre elles, et donc établissant un lien "statistique" entre ces phrases et non "sémantique". Ceci aurait pu être évité par un traitement de désambiguïsation, impliquant l'utilisation d'un dictionnaire de synonymes informatisé et une étude linguistique morpho-syntaxique [16]. Mais la lourdeur du traitement aurait alors débordé du cadre de cette étude de faisabilité d'une chaîne de traitements relativement simple sur un micro-ordinateur (type PC - 486).

Mais d'autres liens permettent de reconstituer un cheminement d'idées présentes en plusieurs endroits du discours, et constituant les thèmes ou préoccupations importantes du texte.

Nous en avons ici un exemple (figure 8), où la première référence de la classe 9 est affichée et le reste des phrases de la classe porté plus bas :

Cette lecture au travers des groupes de phrases issus de la sériation nous permet de revenir dans la perspective de recherche des mécanismes d'innovation dans les PME/PMI.

La recherche de ces mécanismes ne se fait pas en "aveugle", mais en parallèle et en interaction avec une analyse en lecture classique des interviews. Un certain nombre de ceux-ci ont été lus, et quelques concepts ont été relevés comme pouvant être pertinents dans les processus d'innovation au sein des PME/PMI⁶. Ces concepts, ou catégories, sont en l'état actuel des travaux, plus ou moins détaillés et répartis selon "Contexte Historique", "Organisation de la Production", "Technologies et Procédés", "Produits et Savoir-Faire", "Stratégie de Marché", "Etat de la Demande", "Dépendance", "Secteur d'Activité", "Qualification", "Normes", "Discours".

Il semble alors intéressant d'essayer de retrouver, valider ou infirmer ces notions dans les phrases du texte original relatives aux classes dégagées par la sériation.

On peut ainsi prendre l'exemple de cette classe, issue d'un entretien différent de celui évoqué plus haut :

Classe 19, Réf 157

"lui, c'était plutôt un commercial, un administrateur, un gestionnaire et moi j'étais plutôt la partie technique; de sorte qu'à un moment donné on a commencé à se faire des réflexions et se dire: "Si on fait quelque chose pour mettre de l'huile en bouteille, il faut faire quelque chose qui ne soit pas comme tout le monde, mettre l'huile dans une bouteille en verre c'est pas tellement enrichissant, on va simplement copier"".

Il est possible de rattacher cette phrase (Réf 157) à la notion de "Stratégie de Marché", relevant plus précisément d'une stratégie de différenciation.

La classe suivante apporte quelques précisions sur la nature du projet mis en oeuvre :

Classe 22, Réf 173

""Mais si, le fournisseur qui m'a fourni les machines, allez voir (*autre société*), il peut faire des bouteilles", voilà le déclic du départ."

Classe 22, Réf 174

A ce moment là, je savais qu'on pouvait faire de la bouteille mais c'était les produits et (*autre société*) a été très intéressée de trouver quelqu'un qui, dans l'agro-alimentaire, voulait bien se lancer dans l'aventure parce que c'était une aventure.

On considère ici que la catégorie "Organisation de la Production" est concernée, et plus précisément la collaboration établie avec un partenaire économique dans la mise au point d'une nouvelle forme de conditionnement.

Ces deux exemples sur des travaux expérimentaux montrent qu'il est tout à fait envisageable, à condition de bien avoir au préalable cerné le domaine à étudier, de considérer la classification ici présentée comme une aide à la lecture qualitative de certains documents.

⁶Grille établie en collaboration avec le LEST

Conclusion

Il convient de préciser en conclusion ce que peut apporter cet outil d'analyse statistique en regard des inconvénients répertoriés.

Il est clair que ce qui est proposé ici constitue un outil d'aide à la lecture, qui ne dispense en aucun cas d'une lecture du document primaire [17]. Cette philosophie est d'ailleurs adoptée dans la visualisation sous forme hypertexte des classes en regard du texte original.

De plus, l'analyse fine du contenu des références au texte ne peut se faire qu'avec un expert du domaine d'étude envisagé, qui saura évaluer la pertinence des thèmes dégagés du texte. Ceci est conforme à la démarche pratiquée dans notre laboratoire dans le cadre de la veille technologique [18].

La lenteur du traitement global est en mettre en parallèle avec l'accessibilité du matériel utilisé (IBM PC standard).

Précisons que, dans le cas des interviews constituant l'échantillon de travail, les auteurs sont tout à fait conscients des aléas propres à une interview semi-directive, et aux nombreux facteurs qui peuvent influencer sur la qualité des rapports interviewer-interviewé. Cependant nous avons pu localiser au moyen de cet outil des concepts intéressants relativement à l'innovation.

Une telle analyse offre une lecture homogène reproductible d'un entretien à l'autre, ce qui n'est pas toujours le cas lors de la lecture séquentielle de plusieurs entretiens par une seule et même personne.

Une bonne idée de ce que peut être la structure interne d'un texte est également offerte. La structure et le nombre des classes reflètent la physionomie du texte étudié.

Enfin, élément important, elle permet de lier, de manière statistique, des parties de texte "physiquement" éloignées, et qui donc auraient pu passer inaperçues à la lecture séquentielle. L'analyse sémantique vient ensuite déterminer la valeur de ces liens dans le cadre du domaine étudié.

- [13] ZIPF (G.K.)
Human behaviour and the principal of least effort
Addison Wesley, 1949
- [14] QUONIAM (L.)
La bibliométrie : méthodologie *in* La veille technologique,
Dunod, Paris, 1989
- [15] HAITUN (S.D.)
Stationary scientometric distributions ,part III.: The role of the zipf distribution
Scientometrics, V4, N°3, 1982, p181-194
- [16] WARMESSON (I.)
Structuration du lexique et documentation automatique
Revue française de bibliométrie, n°9, Septembre 1991, pp324-351
- [17] DEVALAN (P.), CANDORET (J.P.)
La bibliométrie, un outil de veille technologique pour les entreprises
CETIM Information, n°116, Juin 1990
- [18] DOU (H.), DESVALS (H.)
La veille technologique
Dunod, Paris, 1992