

Richard FOURNIER

Sociologue, poète, écrivain, journaliste et haïkiste.

(1991)

"L'emploi de l'analyse de contenu
en évaluation de programme :
introduction à la logique
du modèle *ANOVEP*."

Un document produit en version numérique par Jean-Marie Tremblay, bénévole,
professeur de sociologie retraité du Cégep de Chicoutimi

Courriel: jean-marie_tremblay@uqac.ca

Site web pédagogique : <http://www.uqac.ca/jmt-sociologue/>

Dans le cadre de: "Les classiques des sciences sociales"

Une bibliothèque numérique fondée et dirigée par Jean-Marie Tremblay,
professeur de sociologie au Cégep de Chicoutimi

Site web: <http://classiques.uqac.ca/>

Une collection développée en collaboration avec la Bibliothèque
Paul-Émile-Boulet de l'Université du Québec à Chicoutimi

Site web: <http://bibliotheque.uqac.ca/>

Politique d'utilisation de la bibliothèque des Classiques

Toute reproduction et rediffusion de nos fichiers est interdite, même avec la mention de leur provenance, sans l'autorisation formelle, écrite, du fondateur des Classiques des sciences sociales, Jean-Marie Tremblay, sociologue.

Les fichiers des Classiques des sciences sociales ne peuvent sans autorisation formelle:

- être hébergés (en fichier ou page web, en totalité ou en partie) sur un serveur autre que celui des Classiques.
- servir de base de travail à un autre fichier modifié ensuite par tout autre moyen (couleur, police, mise en page, extraits, support, etc...),

Les fichiers (.html, .doc, .pdf, .rtf, .jpg, .gif) disponibles sur le site Les Classiques des sciences sociales sont la propriété des **Classiques des sciences sociales**, un organisme à but non lucratif composé exclusivement de bénévoles.

Ils sont disponibles pour une utilisation intellectuelle et personnelle et, en aucun cas, commerciale. Toute utilisation à des fins commerciales des fichiers sur ce site est strictement interdite et toute rediffusion est également strictement interdite.

L'accès à notre travail est libre et gratuit à tous les utilisateurs. C'est notre mission.

Jean-Marie Tremblay, sociologue
Fondateur et Président-directeur général,
LES CLASSIQUES DES SCIENCES SOCIALES.

Cette édition électronique a été réalisée par Jean-Marie Tremblay, bénévole, professeur de sociologie au Cégep de Chicoutimi à partir de :

Richard FOURNIER

" L'emploi de l'analyse de contenu en évaluation de programme: introduction à la logique du modèle *ANOVEP* "

Un article publié dans *L'évaluation sociale : savoirs, éthique, méthodes*, pp. 86-117. Actes du 59^e Congrès de l'ACSALF, 1991 sous la direction de Judith Légaré et Andrée Demers. Montréal : Les Éditions du Méridien, 1993, 372 pp.

[Autorisation formelle accordée par l'auteur le 27 septembre 2008 de diffuser toutes ses publications dans Les Classiques des sciences sociales.]



Courriel : ricfourn@videotron.ca

Polices de caractères utilisée : Comic Sans, 10 points.

Édition électronique réalisée avec le traitement de textes Microsoft Word 2008 pour Macintosh.

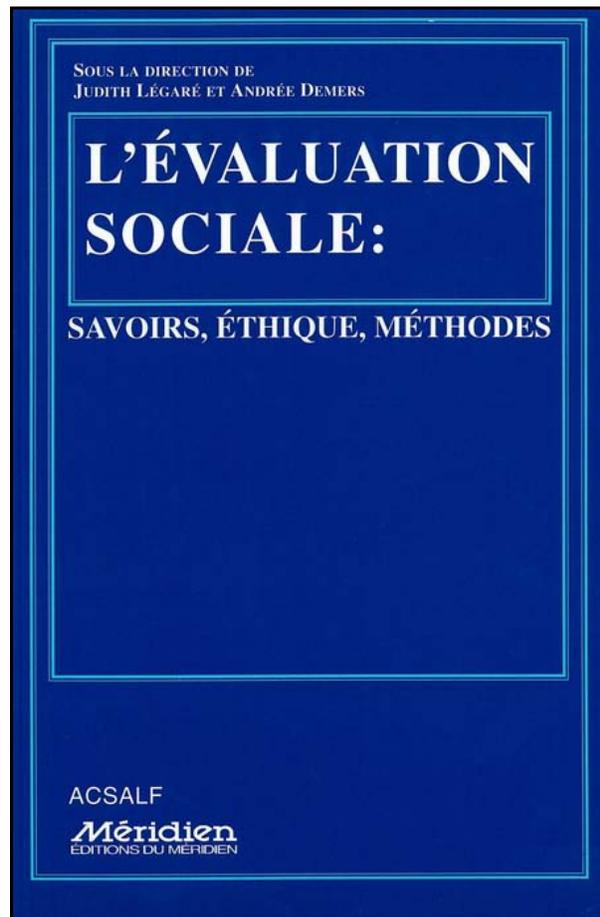
Mise en page sur papier format : LETTRE US, 8.5" x 11".

Édition numérique réalisée le 29 janvier 2013 à Chicoutimi, Ville de Saguenay, Québec.



Richard FOURNIER

"L'emploi de l'analyse de contenu en évaluation
de programme: introduction à la logique
du modèle *ANOVEP*".



Un article publié dans *L'évaluation sociale : savoirs, éthique, méthodes*, pp. 86-117. Actes du 59^e Congrès de l'ACSALF, 1991 sous la direction de Judith Légaré et Andrée Demers. Montréal : Les Éditions du Méridien, 1993, 372 pp.

Table des matières

Introduction

1. Mandat à l'origine de *MODEPAS*

1.1. *Le principe de réplication*

1.2. *Le principe de modularité*

1.3. *Le principe d'interaction avec les partenaires*

1.4. *Le principe d'économie*

1.5. *Le principe de rétroaction interne*

2. La logique du modèle *ANOVEP*

3. Incidences épistémologiques d'*ANOVEP*

3.1. *Le schématisme cognitif dans ANOVEP*

3.2. *Portée informationnelle d'ANOVEP*

3.3. *Statut de la subjectivité dans ANOVEP*

RÉFÉRENCES BIBLIOGRAPHIQUES

Richard FOURNIER

**"L'emploi de l'analyse de contenu en évaluation de programme:
introduction à la logique du modèle *ANOVEP*".**

Un article publié dans *L'évaluation sociale : savoirs, éthique, méthodes*, pp. 86-117. Actes du 59^e Congrès de l'ACSALF, 1991 sous la direction de Judith Légaré et Andrée Demers. Montréal : Les Éditions du Méridien, 1993, 372 pp.

Introduction

[Retour à la table des matières](#)

[87]

Conformément au thème de cet atelier, je voudrais partir d'un emploi de l'analyse de contenu en évaluation de programme pour réfléchir à des questions d'épistémologie et de méthode.

Il s'agira donc essentiellement d'examiner une application des mathématiques à la sociologie de la connaissance, sous la forme d'un modèle logico-mathématique de construction de la grille en analyse de contenu informatisée, le modèle *ANOVEP* (ANalyse Organique de la Variabilité en Évaluant des Proportions / ANalysis Of Variability in Evaluating Proportions). L'acronyme décrit, comme on le verra, ce qui caractérise le modèle sur le plan mathématique.

Cette application des mathématiques à la sociologie de la connaissance a été réalisée dans le but d'expérimenter une solution opératoire à un problème de méthode présenté par l'analyse de contenu non automatique à codeur unique. S'il n'y a qu'un seul codeur, en effet,

comment déterminer le biais au codage, lorsque le codeur s'écarte de la grille ?

Après qu'il eût été développé, le modèle a donné lieu, d'autre part, à une première application informatique : la production d'un logiciel d'aide à la décision en évaluation de programme, appelé *MODEPAS* (MOdélisation de Données en Éducation Populaire Autonome Subventionnée), du nom de son objectif ¹. En introduction, je [88] rappellerai donc brièvement l'origine de cette trajectoire de recherche.

La sociologie de la connaissance, d'abord. En général, le sociologue de la connaissance s'intéresse aux déterminants sociaux de la connaissance, occupé d'objets tels les idéologies, les attitudes, les valeurs, les croyances, les représentations collectives, etc. Or, voilà autant de réalités sociales qui se présentent d'abord à nos instruments de recherche sous l'angle de données qualitatives, au sens fort du terme où l'usage anglo-saxon parle de *soft data*, c'est-à-dire de données imprécises quand à leur mesure, incertaines quant à la façon de les recueillir ou vulnérables quant à leur crédibilité (Cobb et Thrall, 1981). On peut, en effet, faire de la sociologie de la connaissance à analyser des rapports d'entrevue, domaine classique des données qualitatives, aussi bien que des articles de journaux, des histoires de vie ou, comme ici, quand on procède à l'analyse de contenu de formulaires administratifs.

Remarquons que dans tous ces cas, la sociologie de la connaissance, si elle le désire, peut s'allier avec l'informatique pour rechercher des algorithmes spécialement conçus dans l'optique de traiter des données textuelles. On connaîtra ainsi, le cas échéant, soit des logiciels de gestion documentaire (*Text Retrieval Systems*) permettant repérage, archivage, résumé automatique, annotation de textes, etc. (Hudon, 1991), soit des logiciels d'analyse de contenu (*Theory-Building Software*) (Tresh, 1990) destinés, ceux-là, à décrire des objets de discours ou des structures d'argumentation. *ANOVEP*, modèle logico-

¹ *ANOVEP* a été mis au point de 1982 à 1984 au Service de la recherche et du développement de la Direction de l'éducation des adultes du MEQ et expérimenté sur ordinateur central à partir d'échantillons de corpus, programmation en SAS par Hélène Boutet et Steve Aubry, à la Direction de la planification et de la recherche du MEQ de 1985 à 1987.

mathématique de construction de données en analyse de contenu à co-
deur humain, appartient à cette dernière famille d'algorithmes.

Le système *MODEPAS*, programmé en APL par Michel Saucier d'après *ANOVEP*, a été conçu et réalisé au Service de l'éducation populaire de la Direction de l'éducation permanente du MEQ de 1988 à 1989. Le système a été développé sur Macintosh et implanté sur PC. Avec les données, l'application occupe dans les deux cas 6 Mo. L'auteur reconnaît sa dette en cours de route à l'endroit de la compétence en APL de Michel Saucier pour la programmation des calculs de *MODEPAS*.

Deuxième point. L'analyse de contenu dont nous parlons ici est une analyse traditionnelle, définie de la façon suivante : à l'origine de la méthode, un contenu fait de symboles (texte, film, image, etc.) [89] est d'abord dénombré en unités de sens par un codeur selon une grille ou système de catégories. À l'aide de ce système de catégories est alors possible, à partir d'un traitement statistique, l'analyse des significations, associations non arbitraires entre un ou des signes et un ou des interprétants (Granger, 1968). Nous employons donc ici le terme d'analyse traditionnelle pour désigner cette origine de la méthode, et la distinguer de sa pratique en linguistique. L'analyse traditionnelle, sans doute basée sur l'analyse computationnelle, reste encore conçue en effet, de notre point de vue, comme une méthode d'analyse de la communication au sens large, ainsi qu'à ses débuts, et non simplement linguistique.

Cependant, le modèle dont je présenterai un exemple d'application comporte un élément, le coefficient MRC de Mesure de Rendement du Codage (*Coding Proficiency Index*), élément caractéristique d'*ANOVEP*, qui permet de déterminer le rendement des catégories au codage dans le cas de l'analyse à codeur humain. Sur le plan théorique, il s'agit alors, doit-on noter, d'un cas de figure. À ce titre, il s'applique aussi à l'analyse automatique, lorsque l'analyste ou l'indexeur dans un logiciel d'analyse automatique, par exemple, modifie à mesure les caté-

gories ou les codes qu'il a déjà construits. Nous n'entrerons cependant pas ici dans ce domaine.

En troisième lieu, il convient de constater que l'emploi de l'analyse de contenu en information de gestion ne va pas de soi. L'information de gestion est une fonction de l'entreprise. Elle consiste à définir, recueillir, interrelier et traiter l'information apte à répondre aux besoins d'une organisation (Hurtubise, 1976 ; Poulin et Trudeau, 1982). L'analyse de contenu est surtout une pratique de recherche (Holsti, 1969). Entre ces deux paradigmes sociaux, le conduit n'est pas évident. En particulier à cause d'un problème de méthode. Rappelons-nous, en effet, qu'une organisation résolue à utiliser l'analyse de contenu comme outil de gestion, c'est-à-dire décidée à transformer l'instrument *ad hoc* de la recherche académique en une base périodique et standardisée de collecte et d'analyse de données, ne peut se payer à répétition une équipe de codeurs, essentiellement un groupe d'employés faisant strictement la même chose avec des résultats un peu différents. Cependant, la technique [90] des coefficients de fidélité, technique basée sur le contrôle du travail d'un codeur par celui de ses pairs, technique propre à l'analyse traditionnelle préoccupée de données valides, requiert précisément qu'il en soit ainsi, puisqu'elle reste encore, à ce jour, la méthode à suivre, lorsqu'on a besoin de faire produire à l'analyse de contenu des données cumulatives, donc comparables, par exemple, d'un cycle administratif à l'autre. Par défaut, la solution courante sera donc de se passer de données fiables.

Néanmoins, du point de vue de l'information de gestion, il devient entre-temps graduellement avantageux, sinon nécessaire pour l'organisation, de pouvoir se doter des moyens d'exploiter d'une façon spécifique les données dites qualitatives. Le besoin, au surplus, est facilement admis en ce qui regarde les données d'origine textuelle, vu les volumes de la paperasserie déplacés par la gestion des pièces administratives. Il y a donc ici une inclination compréhensible des organisations à penser pouvoir traiter ces données d'origine textuelle sur un pied d'égalité avec les données quantitatives, soit côte à côte dans un même fichier, comme le permet *ANOVEP*, soit en relation à un fichier quelconque. Le gain anticipé dans les deux cas est d'augmenter le ratio

information traitée / information recueillie tout en améliorant éventuellement la décision.

Dernière remarque introductive. En évaluation de programme, des facteurs de conjoncture incitent l'organisation à se munir d'un système d'information de gestion où figurent d'une façon fiable des données qualitatives (croyances, valeurs, idéologies, attitudes, etc.). Cela est singulièrement notable pour ce qui a trait aux programmes de prévention, d'éducation, de réhabilitation, d'accompagnement, etc., les plus difficiles à évaluer parce que principalement fondés sur ce genre de données (Fournier, 1985). Évoquons simplement, au titre de ces facteurs de conjoncture qui incitent les utilisateurs de systèmes d'information de gestion à se pencher sur le puits de la culture comme source de données qualitatives, l'apparition d'économies à croissance zéro, la gestion de la décroissance, la libéralisation des marchés, la remise en cause des régulations et des objectifs de l'État, la restructuration de l'entreprise. En retour, la fonction d'évaluer les programmes vient alors potentiellement soutenir l'information de [91] gestion, en se présentant, par exemple, comme un moyen utile pour raffiner un système d'indicateurs de gestion (Fournier, 1990a).

Pour ce qui est de l'application d'*ANOVEP* dont je veux vous parler, c'est au point de rencontre de ces deux approches, information de gestion et évaluation de programme, qu'on en saisit le mieux le contexte. Je présenterai donc brièvement, en première partie de mon exposé, le mandat à l'origine de l'application d'*ANOVEP* dans le logiciel *MO-DEPAS*.

Dans la partie suivante, je m'attacherai à décrire le problème de méthode qu'implique la décision d'employer l'analyse de contenu comme outil de gestion, et à expliquer l'aménagement de ce problème que propose le modèle. Interviendra alors en guise de conclusion une réflexion épistémologique.

1. Mandat à l'origine de *MODEPAS*

[Retour à la table des matières](#)

Le mandat à l'origine de l'application d'*ANOVEP* dans le logiciel *MODEPAS* provient de la volonté d'évaluer un programme de soutien financier s'adressant à des organismes communautaires d'éducation populaire et, pour ce faire, de doter l'organisation responsable de gérer le programme d'un cadre d'évaluation. Un tel cadre sera ici défini minimalement comme une base de données apte à structurer questions et réponses sur les enjeux du programme. Il viendra concrétiser la décision de mieux orienter le programme en fonction des objectifs fixés, des ressources investies, des résultats obtenus ou escomptés. Par éducation populaire, on entendra alors « l'insertion d'un adulte ou d'un groupe d'adultes dans un processus de responsabilisation » au moyen d'actions appelées « interventions éducatives » (MEQ, 1987). On accepte en outre que ces interventions se situent sur un continuum entre deux pôles opposés : scolarisation et socialisation.

D'autre part, s'ajoute une contrainte organisationnelle. Aux yeux de l'administration, évaluer le programme s'inscrit dans la perspective de pouvoir accréditer et classifier les organismes admissibles. Sous le terme de classification s'entendra donc ici une opération logico-mathématique qui assure la distribution de caractéristiques stables sur une structure d'ordre et, en conséquence, la répartition d'objets y appartenant, savoir des organismes communautaires [92] d'éducation populaire. En conséquence, l'évaluation doit nommément prendre en compte dans le processus de prise de décision la multiplicité des organismes, la diversité des activités, contenus, objectifs et résultats des projets présentés et, parmi ceux-ci, tant les projets admis à une subvention que ceux refusés (Fournier, 1989a).

Pour rencontrer ces exigences, le cadre d'évaluation où se développa le logiciel *MODEPAS* en tant qu'application d'*ANOVEP* en vint ainsi à reposer sur cinq principes (Fournier, 1989b).

1.1 Le principe de réplication

[Retour à la table des matières](#)

Les données qui servent à l'évaluation doivent être standardisées quand à leur construction (ou syntaxe), leur définition sémantique (ou sociologique) et leur aptitude à s'insérer dans un processus reconnu d'acquisition et d'exploitation de données, tel l'analyse de contenu, l'analyse multivariée, les typologies, l'analyse des correspondances, etc.

Le premier effet attendu d'une telle pétition de principe était qu'un ensemble de données ne perde pas sa fonctionnalité à se voir transféré d'unité administrative (ce fut précisément ce qui arriva en cours de projet); ensuite, question de respecter la norme communément admise, il fallait assurer la possibilité qu'un chercheur sur les mêmes données en suivant les mêmes manières de procéder aboutisse au même résultat; finalement, il était nécessaire que les données soient évidemment comparables dans le temps, puisqu'il s'agit de construire ici une base de données textuelles.

1.2 Le principe de modularité

Exigence de conception, par ailleurs, le cadre d'évaluation doit ici conserver la souplesse nécessaire pour permettre de réaliser au moment opportun les modifications désirées aux trois niveaux suivants de la gestion logique des données : les fichiers, les protocoles de requêtes, les systèmes logico-mathématiques d'exploitation. Or, l'une des façons d'y arriver était de concevoir la construction d'un système modulaire sur micro-ordinateur d'une part (*MODEPAS* fonctionne aussi comme système résidant dans un ordinateur central) et, d'autre part,

de travailler avec un modèle logico-mathématique [93] qui autorise à modifier les catégories, ce que permet précisément *ANOVEP*.

1.3 Le principe d'interaction avec les partenaires

[Retour à la table des matières](#)

En raison de la perspective de recherche-action manifeste dès l'étape du design du modèle *ANOVEP* (Fournier, 1983), il était naturel que le cadre d'évaluation où le modèle allait servir à construire et exploiter des données textuelles puisse prévoir que les partenaires du programme, organismes communautaires et institutions scolaires en particulier, participent à la définition, la construction et l'exploitation des données concourant à leur propre évaluation.

Il aura donc fallu définir des destinataires et les produits documentaires à leur intention (graphiques, tableaux, synthèses, etc.), ce qu'on retrouve dans les menus du logiciel *MODEPAS*. Il aura été nécessaire aussi, même si l'initiative d'implanter ce retour ne relève pas de l'analyste de contenu ou du spécialiste en évaluation, de prévoir les conditions informatiques, méthodologiques et administratives assurant le retour aux organismes, sous la forme des traitements pertinents, des informations qu'ils auront communiquées à l'organisation. En troisième lieu, on aura dû également établir les conditions de la participation des agents du programme, administrateurs et administrés, à la définition des objets de l'évaluation. À ce dernier impératif, la méthode traditionnelle pour construire un thésaurus en analyse de contenu devait ici servir de point de départ (Fournier, 1983, p. 53, note 42).

Par ailleurs, ce principe d'interaction entre partenaires a été formellement inscrit, sur le plan théorique, dans la possibilité que donne le système d'analyse de contenu *ANOVEP* de modifier les critères d'observation du programme ou les catégories de l'analyse et, sur le plan pratique, dans la capacité du logiciel *MODEPAS* d'être transportable sur réseau, sur PC ou sur Mac.

1.4 Le principe d'économie

[Retour à la table des matières](#)

Le cadre d'évaluation devait viser à rentabiliser au maximum le fait, pour un organisme participant, d'avoir produit l'information [94] écrite exigée des organismes communautaires aux fins de gérer le programme de financement. C'est la logique du design d'*ANOVEP* à l'origine, faire d'une pierre deux coups : construire, à des fins d'en rendre valide l'analyse de contenu, des informations qualitatives qui resteraient autrement en puissance bien que déjà recueillies, les obtenant ainsi à un coût inférieur à celui, par exemple, de rapports d'entrevues ou de questionnaires d'évaluation *ad hoc*; et construire, ce faisant, des informations qualitatives non singulières, c'est-à-dire reproductibles d'un cycle administratif à l'autre au moyen des pièces administratives (formulaires) générées automatiquement à travers les opérations courantes du programme.

Cette logique de la construction des données se complète alors de la possibilité d'envisager des échanges avec des bases de données quantitatives existant en d'autres unités administratives de l'organisation ou de l'administration publique.

1.5 Le principe de rétroaction interne

Le cadre d'évaluation devait finalement permettre d'intégrer les résultats de l'évaluation au suivi du programme et, pour autant, de garantir au système d'information de gestion dérivé d'*ANOVEP* que l'interaction nécessaire entre plusieurs niveaux de requêtes (aide à la décision, au contrôle des opérations, à l'exécution des opérations) se réalise sur la base préalable d'interactions assurées entre des variables valides.

Cette exigence se traduit dans le contenu de l'interface et des menus de *MODEPAS*: le fait de pouvoir déterminer le rendement au co-

dage des catégories construisant les variables de l'analyse permet, en effet, de jouer avec ces niveaux de requête selon le degré de description ou de simulation désiré. Un menu *ANOVEP*, intégré à la structure du logiciel, est la plaque tournante de l'implantation de ce principe de rétroaction entre des niveaux de connaissances.

En terminant sur ce principe de rétroaction interne le compte-rendu des exigences de départ de l'application d'*ANOVEP* à une situation concrète de gestion, on constatera donc finalement que c'est le cadre d'évaluation dans son ensemble qui est rejoint et influencé par la structure d'*ANOVEP*, et ce, tant sous l'aspect de la méthode, [95] savoir la détermination du biais au codage, que sous l'aspect de l'exploitation qui sera faite des données, savoir l'interrogation et la simulation des phénomènes à même une base de données pondérées (*weighted index*) selon un indice de « subjectivité » du codage.

Cependant, une conséquence apparaît aussitôt sur un autre plan, celui des incidences cognitives qu'on peut rattacher à un tel développement organisationnel. Sur ce plan, en effet, le cadre d'évaluation proposé lors de l'implantation du système *MODEPAS* se rattache à une stratégie de mise en œuvre de systèmes informatiques dits à base de connaissances, catégorie de systèmes qu'on peut imaginer appelés, dans l'avenir, à compléter, sur une base modulaire du moins, des systèmes d'informatique de gestion plus traditionnels dits à base de données. Deux courants de pensée, l'algorithmique classique et la recherche en intelligence artificielle, se rencontrent ainsi dans l'expérience d'implanter *ANOVEP*.

Il y a donc lieu de s'attarder maintenant sur la structure et la logique du modèle *ANOVEP* de ce dernier point de vue, puisque s'y retrouve également le point de vue, important pour le sociologue ou le spécialiste en science humaines, de la représentation des connaissances en analyse de contenu. Je retiendrai donc sous cet aspect du modèle les deux éléments d'*ANOVEP* les plus aptes à nourrir la réflexion épistémologique que je veux amorcer en conclusion. Ces deux éléments sont le codage sur graphes et le coefficient *MRC* de mesure du rendement des catégories au codage.

2. La logique du modèle ANOVEP

[Retour à la table des matières](#)

La logique ou la structure d'*ANOVEP* ont déjà été exposés ailleurs (Fournier, 1983, 1984, 1990d). Quant au domaine où le modèle apparaît, on sait déjà qu'il s'agit d'un programme « mou » (Fournier, 1985), c'est-à-dire à base de données qualitatives, dont on fait une évaluation au moyen de données textuelles, expression écrite de données qualitatives : valeurs, croyances, attitudes, idéologies, projets, modèles (*patterns*), etc.

La matière de l'analyse est constituée des formulaires où les gens exposent, en réponse à des questions ouvertes, le projet pour lequel un soutien financier est requis, les moyens de le réaliser et les [96] besoins auxquels il répond, selon un cycle administratif en trois étapes : production d'un projet, d'un plan d'intervention et, une fois le projet réalisé, d'un rapport d'activités.

L'objectif de l'analyse est d'en arriver, à partir de l'exploitation d'une base de données qualitatives, à pouvoir plus finement arrêter un choix de projets, évaluer les critères de sélection, établir un suivi, gérer les opérations, mesurer les impacts et, finalement, orienter le programme en connaissant mieux ce qui est financé : activités, pratiques, résultats.

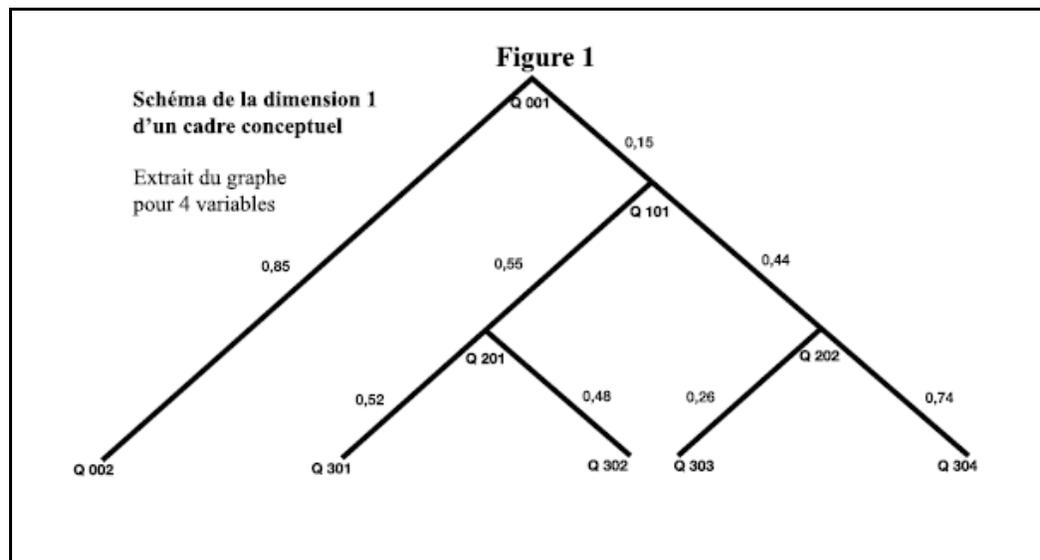
Le cadre conceptuel de l'analyse comporte trois dimensions : les modèles pédagogiques (87 variables à l'origine), les modèles de développement (31 variables) et les modèles organisationnels des organismes communautaires (15 variables). Dans la version finale de *MODEPAS*, *ANOVEP* travaille avec 188 éléments, comprenant les variables, les indicateurs d'unités de sens et les critères d'observation du programme. Touchons un mot de chacun.

Les indicateurs d'unités de sens ou descripteurs sont établis en employant la notion de domaine de référence, « ensemble des traits sémantiques applicables à l'objet » (Fournier, 1983). L'opération revient à réaliser en premier lieu un recensement pratique des indicateurs par expérience, en l'occurrence celle du chercheur et celle des responsables de la gestion et de l'animation du programme, et par réflexion, inventaire des sources et de la littérature. Le produit de ce recensement se présente sous la forme de catégories d'observation, qu'on appliquera dès lors au matériel à l'aide d'échantillons du corpus. Furent ainsi tirés trois échantillons de 97 à 100 cas chaque fois, à même l'un ou l'autre univers de deux sortes de pièces administratives. Les catégories initiales, reconstruites en cours de route à la lumière de ces codages successifs, ont ensuite été validées auprès de responsables, et la version obtenue testée sur échantillon avant d'être adoptée.

Le dépouillement du corpus à la recherche des indicateurs d'unités de sens se fait à l'arrivée des formulaires, selon la grille de codage en rangées et colonnes connue en analyse traditionnelle. Dans celle-ci, par exemple, la valeur de chaque rangée reproduit la valeur de l'unité de sens en termes de fréquence brute au dénombrement, et [97] chaque colonne la valeur du corpus, de la dimension ou de la catégorie de celui-ci. L'analyse traditionnelle propose de la sorte une mesure du corpus par fréquences simples à partir d'un jugement de présence ou d'absence. Du point de vue de l'opération de mesure, le corpus en effet peut alors être décrit par le ou les tableaux des distributions obtenues. C'est ici que l'analyse selon la méthode d'*ANOVEP* diverge de l'analyse traditionnelle.

Pour quantifier un corpus avec *ANOVEP*, en effet, on laissera tomber les fréquences brutes et la grille rectangulaire, mais ce sont ces mêmes fréquences qui, transposées en rapports de proportions ou valeurs de probabilité, sont portées sur des graphes (orientés sans circuits), après avoir fait l'hypothèse que la quantification en analyse traditionnelle obéit à la structure mathématique d'un codage par application. D'où le premier élément du modèle à considérer : le codage sur graphes.

Prenons-le dans l'abstrait et regardons la figure suivante (fig. 1). Elle décrit un ensemble de valeurs de probabilité, ou ensemble de données selon notre définition précédente, organisé par inclusion en sous-ensembles. La relation d'inclusion est représentée par une relation de parenté dans un arbre binaire. Celle-ci tient que, pour chaque sommet non terminal du graphe ou sommet c , par exemple pour Q101 et Q202, les sommets a et b successeurs de c sont les fils ou descendants véritables de c . À chaque nœud de niveau $n-1$, on retrouve donc une relation d'ordre, si on fait du graphe une lecture verticale (classes distributives), et une classe homogène, où chaque terme comprend tous les autres ou les exclut, si notre lecture est horizontale (classes ensemblistes) (Fournier, 1983, 1988). Nous voilà pour l'essentiel en présence d'un graphe *ANOVEP*, technique d'organisation des catégories en analyses de contenu.



D'où maintenant les variables viennent-elles ? Demandons-nous d'où procède le graphe. En effet, selon le postulat d'un codage par application, la phase de conceptualisation du modèle aura comporté auparavant d'organiser le thésaurus à partir d'une base de données codées en arbre exponentiel $P(E)=2^{|E|}$. Ce qui revient à définir un graphe appliqué sur un questionnaire binaire. En conséquence de quoi, le produit traditionnel d'un codage, savoir la quantification [98] d'un corpus, est re-

présenté dans *ANOVEP* par l'ensemble (oui, non)^E de la classification dichotomique des réponses à sa propre question que réalise le codeur chaque fois qu'il a sélectionné une unité de sens pour lui attribuer, dans une case de sa feuille de codage, sa place dans la grille.

Cette place est toujours un sommet de dernier niveau (ou niveau n) du graphe. Par exemple, sur la portion de graphe illustrée, partie de la dimension 1 d'un cadre conceptuel à trois dimensions, les quatre sommets pendants Q301 à Q304 à l'extrême droite sont les quatre indicateurs de deux variables dichotomiques. À chaque sommet pendant du graphe, les indicateurs renvoient ainsi à des variables rapportées à l'une ou l'autre des dimensions d'un cadre conceptuel. Un graphe *ANOVEP* compte ainsi autant de variables qu'il contient de sommets de niveau $n-1$. À chaque variable correspond une catégorie de la grille d'analyse.

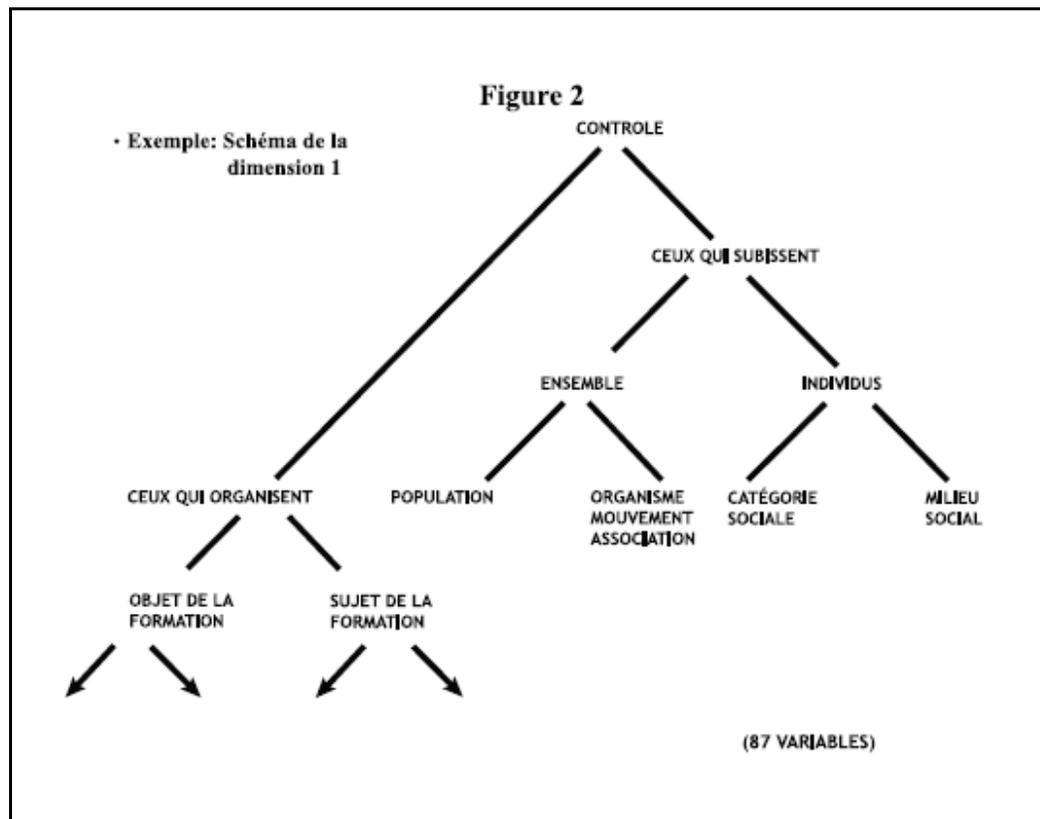
Sur le plan d'une technique de construction de la grille ou sur le plan d'une méthode de représentation des connaissances en analyse de contenu, le graphe *ANOVEP* est ainsi un graphe étiqueté.

La compréhension de cette méthode de représentation de la grille de codage en analyse de contenu appelle néanmoins, se voudrait-elle complète, deux précisions. Pour commencer, relevons en effet qu'on y retrouve ce qui distingue la façon de quantifier un corpus dans *ANOVEP*: on calculera des trajets sur des graphes, en s'aidant pour [99] ce faire de la théorie de l'information et de la théorie des graphes. L'incidence méthodologique de ce procédé de quantification repose sur le fait que la description et l'explication d'un corpus dans *ANOVEP* se pratiquent avec des indices (calculés depuis des valeurs de trajets) et non sur le décompte des fréquences brutes, comme dans la tradition.

La seconde précision est un peu plus élaborée. Sur le plan mathématique, le binarisme de l'arbre exponentiel renvoie en effet à une structure de groupe abélien ou de diagramme commutatif. Or, une telle opération de structuration peut s'appliquer dans la langue à l'opposition binaire d'une chose et son contraire, opposition propre au jugement de présence ou d'absence d'un attribut (par exemple, les couples chaud/froid, haut/bas, fort/faible, loin/proche, etc., des différenciateurs sémantiques d'Osgood). Et le résultat peut se concilier

en même temps avec l'opposition entre contraire et contradictoire, opposition propre aussi à la langue naturelle ou à la vie courante.

À titre d'illustration, considérons la figure 2, jumelle de la précédente, où la valeur de probabilité que possède une unité de sens d'être mentionnée au texte est remplacée à chaque sommet par la définition de la catégorie de la grille d'analyse à laquelle fait référence la mention.

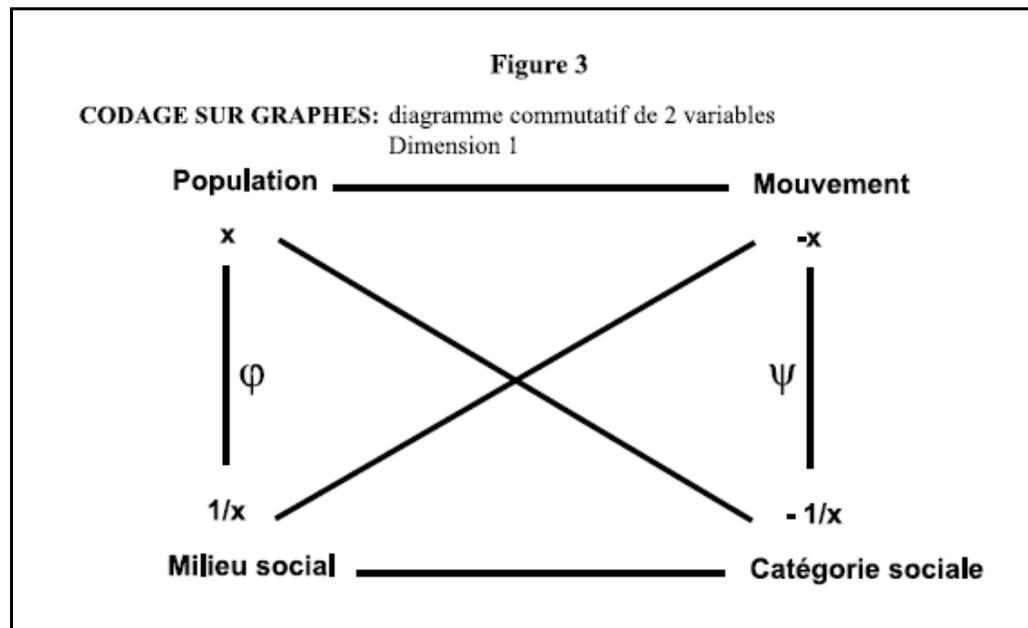


[100]

Soit la classe formée par les quatre sommets pendants, à droite. Ces quatre sommets de la classe correspondent aux indicateurs de la proposition : « La cible est soit un individu soit un ensemble ». De gauche à droite, ils se lisent respectivement ainsi : « Si la cible est un ensemble, elle s'identifie à la population en général (indicateur 1) ou à un mouvement, une organisation, une association, etc. (indicateur 2) ; si la

cible est un individu, elle est membre d'une catégorie sociale (indicateur 3) ou bien elle appartient à un milieu social (indicateur 4) ».

Or, cette proposition est susceptible d'une représentation faisant ressortir le groupe des transformations dont elle est issue. Une telle représentation est donnée en effet par la structure mathématique du groupe de Klein (Barbut, 1969). Par exemple, le groupe des transformations de notre proposition est le suivant :



On pourrait montrer en quoi cette façon apparemment binaire de distribuer un attribut à quatre valeurs à l'intérieur d'une classe d'après la structure du groupe de Klein correspond au carré d'Aristote, au carré logique de Piaget ou au carré sémiotique de Greimas (Fournier, 1983). Car, ce sont là autant de manières d'exprimer, en terme logiques ou linguistiques, ce fait banal que le langage commun ou la vie courante s'accommodent fort bien de dire que P exclut son contraire non-P, tout en acceptant l'existence de sa contradiction -1/P.

[101]

Par exemple, ainsi en va-t-il de ces quatre sommets pendants fictifs de la vie quotidienne : « chaud » (x), « froid » ($-x$), « tiède » (-

1/x), « frais » (1/x). Nous reviendrons en troisième partie sur cette place du binarisme de la langue dans le modèle *ANOVEP*.

Car, pour l'instant, outre le codage sur graphes et ses incidences, il existe au regard d'une réflexion épistémologique un tout autre élément d'*ANOVEP* dont il importe aussi de parler, puisque cet élément est complémentaire du mode de quantification par trajets sur les graphes (ou du mode de représentation par graphe étiqueté), et qu'il est non moins essentiel au modèle. Il s'agit de l'emploi de l'indice S de rendement d'un codage, emploi qui caractérise le modèle *ANOVEP* sur le plan opératoire et lui procure sa spécificité analytique.

Par rapport à l'analyse traditionnelle, en effet, dès lors que dans *ANOVEP* on prend les moyens de décrire des trajets sur un graphe, on transforme ce faisant la grille de codage en grille d'analyse. Les indices d'*ANOVEP* procurent ici ces moyens, et permettent d'organiser la gestion des unités de sens du thésaurus (ou gestion sémantique du thésaurus), objectif pragmatique de n'importe quelle méthode d'analyse de contenu. Or, sur le plan cognitif, cette opération de gérer le thésaurus dans *ANOVEP* est à double versant. Car il faut constater alors, en effet, qu'à l'instar de l'analyse traditionnelle, les règles de production de sens qui président à la gestion des unités de sens dans *ANOVEP* sont celles de l'analyste (par exemple ici : sociologie des groupes, des mouvements sociaux, de l'éducation permanente, de la participation; analyse multivariée, de correspondances, méthode typologique, etc. (Fournier. 1988)). Mais on prendra note aussitôt que, dans la conceptualisation du domaine de référence que construit *ANOVEP*, ces règles de production de sens s'exercent non sur des données, mais sur des niveaux de connaissances, comme on l'a vu (chaque niveau de variables n-1), et sur des niveaux de lecture (lecture horizontale et lecture verticale) de ces mêmes connaissances. La technique de construction de la grille et de représentation des connaissances dans *ANOVEP* apparente ainsi le modèle aux techniques usitées en intelligence artificielle (Fournier, 1988, 1989c). Le lien avec l'analyse traditionnelle est assumé par les indices fournis [102] par le calcul de trajets sur des graphes, lesquels indices jouent dès lors leur rôle d'instruments de lecture.

Le point à considérer regarde le plus important de ces trois indices, l'indice S , mesure du rendement des catégories au codage. C'est la valeur du coefficient MRC du même nom, essentiellement défini comme le rapport entre l'efficacité du graphe à construire une variable (ou indice C) et la valeur d'entropie de la variable obtenue (Fournier, 1984). L'égalité entre ces deux quantités représente alors le cas de figure de l'« objectivité parfaite » pour un canal sans bruit. Symbolisés par la lettre S , la valeur de l'indice varie de 0 à 1 bits selon le cas de figure considéré. À la valeur zéro, situation d'« objectivité parfaite », le rendement du système de catégories est maximal : ce cas correspond à la situation où la compréhension des catégories par le codeur coïnciderait parfaitement avec la définition qu'on en donne. À l'inverse, le cas où S est égal à l'unité, situation de « subjectivité parfaite », figure l'éventualité où le codeur n'aurait rien compris à rien.

Dans la réalité, les valeurs de S se distribuent entre ces deux termes : un codage de 0,09 bits, par exemple, sera réputé dans *ANOVEP* plus conforme aux significations de la grille qu'un codage de 0,29 bits. L'emploi de l'indice S constitue donc, si on peut dire, la spécificité du modèle *ANOVEP* en ce qui concerne le problème de l'objectivité (*content validity* (Zeller et Carmines, 1980)) en analyse de contenu.

Sur le plan du traitement des données, on peut utiliser de diverses façons un tel renseignement. Cela revient chaque fois à pouvoir disposer d'un fichier de données textuelles où les variables sont introduites au fichier déjà accompagnées de leurs poids d'objectivité. C'est de cette manière fondamentale qu'*ANOVEP* constitue essentiellement une technique de construction de la grille en analyse de contenu.

Prenons un exemple pratique en évaluation de programme. Séparant en terciles les variables selon un poids de plus ou moins grande « objectivité », l'analyste en effet peut se donner des versions contrastées d'un corpus suivant le calibre des variables auxquelles il fait appel. Le tableau 1 décrit une telle entreprise appliquée à un programme de financement.

[103]

Tableau 1

Rendement au codage des catégories d'une étude

Valeurs du coefficient MRC de Mesure du rendement des catégories au codage

Dans la table, la ligne horizontale divise en terciles les valeurs S de l'indice de mesure du rendement des catégories au codage, valeurs disposées par ordre d' « objectivité croissantes » de (+-) à (-+). Cet ordre est obtenu en rapportant la valeur moyenne de l'indice S à chaque nœud du graphe, variable du cadre d'analyse, à celle de son coefficient de variation.

Var.	Moy.	CV.	(+/-)	Var.	Moy.	CV.	(+/-)
S003	0,916	0,29	+ -	S305	0,229	0,41	= =
S007	0,911	0,31	+ -	S306	0,240	0,44	= =
S010	0,906	0,30	+ -	S006	0,274	0,74	= +
S103	0,874	0,31	+ -	S107	0,276	0,68	= +
S108	0,476	0,27	+ -	S211	0,219	0,56	= +
S215	0,332	0,34	+ -	S212	0,218	0,55	= +
S011	0,814	0,44	+ =	S303	0,225	0,56	= +
S102	0,420	0,40	+ =	S201	0,169	0,53	- =
S105	0,385	0,52	+ =	S208	0,176	0,50	- =
S109	0,751	0,42	+ =	S302	0,127	0,54	- =
S209	0,342	0,48	+ =	S317	0,155	0,51	- =
S210	0,410	0,45	+ =	S318	0,154	0,52	- =
S002	0,725	0,56	++	S319	0,160	0,49	- =
S009	0,379	1,15	++	S320	0,154	0,52	- =
S203	0,300	0,39	= -	S001	0,027	4,09	- +
S204	0,301	0,36	= -	S005	0,042	3,00	- +
S313	0,239	0,37	= -	S101	0,164	0,76	- +
S314	0,272	0,39	= -	S104	0,154	0,67	- +
S315	0,244	0,38	= -	S202	0,206	0,58	- +
S316	0,223	0,38	= -	S207	0,153	0,55	- +
S321	0,238	0,26	= -	S301	0,166	0,58	- +
S322	0,238	0,26	= -	S304	0,163	0,56	- +

Groupements pour un rapport de recherche

<u>tercile (*)</u>	<u>tercile (**)</u>	<u>tercile (***)</u>
Q001	Q006	Q002
Q005	Q107	Q009
Q101		
Q104		

Source : R. Fournier, 1990b.

[104]

Sur une liste, les variables d'une dimension sont ordonnées selon leur valeur croissante en objectivité. En examinant les variables du premier tercile, réputées les plus « objectives », l'évaluateur de programme ou l'analyste de contenu peut ainsi obtenir la description du déroulement du programme la plus conforme à la définition institutionnelle (légale, administrative, réglementaire, politique) de celui-ci, portrait de la réalité selon les définitions qu'on a voulu s'en donner. En pratique, l'évaluateur travaillera donc d'abord à établir ses propositions de consolidation du programme avec les variables du premier tercile. À l'inverse, s'il veut connaître par où le déroulement du programme est perméable à l'innovation, culturelle ou sociale, ou sensible à l'effet de conditions changeantes survenues sur le terrain, il examinera en premier lieu les variables du troisième tercile, les plus difficiles à coder probablement parce que la définition au codage s'écarte le plus des définitions instituées. De la même façon, avons-nous déjà indiqué ailleurs, la valeur *S* est directement utilisable avec un moteur d'inférence pour gérer l'incertitude dans un éventuel *ANOVEP-Expert* (Fournier, 1989c).

Les résultats établis jusqu'à maintenant à expérimenter l'emploi d'un coefficient *MRC* de mesure de rendement du codage avec *MODEPAS* ont surtout trait à trois ordres de travaux : 1) la possibilité de tracer, tel que suggéré ci-dessus, une description contrastée d'un programme selon sa plus ou moins grande conformité au projet de ses concepteurs ou selon le besoin d'y effectuer des changements (Fournier, 1990b); 2) la possibilité d'établir de la même façon des schèmes causals sur le fonctionnement d'un programme ou le déroulement des activités qui le constituent (Fournier, 1990c); 3) la possibilité pour l'analyste ou les partenaires engagés dans l'évaluation d'un programme de modifier les catégories de l'analyse de contenu (variables dépendantes) ou les critères d'observation (variables indépendantes) du programme (Fournier, 1990c).

Il y a d'autres possibilités d'utilisations du coefficient *MRC* en recherche : révision des catégories, formation des codeurs, enrichissement du thésaurus, mise à jour de la base de connaissance, études de

réplication, etc. Si je rappelle ici une illustration générale de l'expérimentation du modèle en évaluation de programme, c'est que [105] ce niveau de généralité de l'emploi de l'indice me semble adapté à exposer à la réflexion, sans rien limiter, l'essentiel de la logique et de la structure d'*ANOVEP* à partir des deux éléments caractéristiques qu'on lui a reconnus, le codage sur graphes et la mesure du rendement des catégories au codage.

En elle-même cependant cette illustration aurait l'heur, comme je l'ai dit, d'être généralisée. À court terme, par exemple, on pourrait ainsi signaler que les développements théoriques du modèle *ANOVEP* à partir de l'expérience de *MODEPAS* ont principalement trait à : 1) l'évaluation de la fonction du nombre Q trouvé au codage; 2) l'isomorphisme du modèle avec l'analyse de systèmes; 3) le rapport du modèle avec la Théorie classique des tests ou la Théorie des questionnaires.

À plus long terme, d'autre part, le développement théorique intéressant se situe au carrefour de la sociologie de la connaissance et de l'analyse de contenu. On aura remarqué, en effet, que le modèle réalise explicitement, dans sa façon de définir la quantification reçue de l'analyse traditionnelle, le postulat que l'interprétation en analyse de contenu est un processus stochastique, c'est-à-dire un système produisant une suite d'unités de sens selon certaines probabilités. Reste donc à évaluer le postulat ainsi concrétisé par *ANOVEP*. Travail stimulant, car du même coup sera explicité, en partant des heuristiques du sociologue, le lien entre *ANOVEP* et les techniques de représentation des connaissances que développe l'informatique cognitive sous le terme d'intelligence artificielle. Nous voilà de cette manière arrivés à considérer les incidences épistémologiques du modèle.

3. Incidences épistémologiques d'*ANOVEP*

[Retour à la table des matières](#)

Je choisirai parmi les incidences épistémologiques d'*ANOVEP* celles qui me semblent relever directement de l'emploi de l'analyse de contenu en information de gestion et, plus précisément, en évaluation de programme. Il n'est pas sans intérêt, en effet, lorsqu'il s'agit d'une pratique scientifique, de rattacher à l'horizon épistémologique auquel elles appartiennent des considérations qui seraient nées d'une [106] « utilisation administrative des sciences sociales », comme l'exprimait naguère Hubert Guindon (1960)².

Ces incidences sont au nombre de trois. La première regarde l'activité de schématisation cognitive propre à l'esprit humain; la seconde, la portée informationnelle d'*ANOVEP* et la dernière, le statut de l'intersubjectivité d'après le modèle.

3.1. *Le schématisme cognitif dans ANOVEP*

La méthode du codage sur graphe caractéristique d'*ANOVEP* provoque directement, on l'a vu, l'activité de schématisation du sujet dans son fonctionnement cognitif; elle renvoie ainsi à des questions d'ordre général, en particulier la question critique. Or, ce n'est pas là une question ignorée en analyse de contenu. Dès l'origine de la méthode de

² Dans un article intitulé « The Social Revolution of Quebec Reconsidered », Hubert Guindon a cette phrase en 1960 sur l'État québécois : « Administrative social science, which, in the final analysis, in the only kind of social science wanted by these dominant bureaucracies, will have its heyday ». Trente ans plus tard, l'expression « utilisation administrative des sciences sociales » semble rendre compte assez justement encore de l'état achevé de cette prophétie laïque.

l'analyse de contenu traditionnelle, en effet, on reliait l'activité de construire la grille d'analyse à la question critique. À tous le moins y procédait-on, même si de loin en certains cas, par exemple chez Thomas et Znaniecki (1918), par référence au paradigme de l'objectivité dans la méthode expérimentale. Et on a continué avec régularité de le faire, témoin les débats sur la question de savoir si l'on doit construire les catégories *a priori* ou *a posteriori*.

Sans doute, à travers ces débats normatifs les modèles connus d'analyse renvoient-ils à des épistémologies souvent distinctes; analyse de contenu au sens traditionnel, analyse linguistique, analyse documentaire, analyse du discours, stratégie de fouille automatique dans des bases de données, analyse structurale, etc.

Cependant, quelle que soit l'épistémologie de référence, on s'accorde à reconnaître que l'une des normes commune à toutes les modélisations de données tentées par l'analyste consiste à pouvoir en arriver à définir un type d'erreur, le biais, à partir des techniques employées. Quelquefois ce type est assimilé directement à l'erreur [107] de mesure, d'autres fois non. Mais toujours l'idée directrice veut tenir compte de l'interprétation de la communication par l'observateur.

On cherche ainsi à déterminer l'effet de l'activité de l'interprète sur les résultats obtenus. Lorsqu'on saisit cet effet, l'activité d'interprétation qui l'a produit est ordinairement perçue comme exogène au contenu, mais constitutive néanmoins de l'apparition de celui-ci dans l'analyse, puisqu'elle est supposée, entre autres, découler de la situation de communication ainsi établie. Cette activité d'interprétation constitutive de l'apparition d'un contenu dans l'analyse est alors prise comme la conséquence de la rencontre de deux savoirs plus ou moins fortement institués; le savoir social de l'analyste ou du codeur, premier sujet, et le savoir social d'un second sujet, individuel ou collectif, qu'on vise à objectiver.

Deux conséquences découlent de ce rapport entre deux sujets ou deux sièges d'activité épistémique. D'abord, même si comme dans un langage, il ne peut y avoir de significations sans le support de groupes, cette rencontre ici entre deux sujets épistémiques est souvent perçue, du fait même de la méthode employée, comme singulière. Les analystes de contenu, observait Krippendorff (*in* Mochmann, 1980), sont

ordinairement « conçues *ad hoc*, appliquées une seule fois et ensuite oubliées ». Il ajoutait que les assiste aussi en si bon chemin le fait qu'elles portent fréquemment sur des documents historiques ou qu'elles restent proches de l'herméneutique. Cependant, il y a plus : qui est intéressé, comme disait l'autre, à faire carrière dans la répliation (Krippendorf, *in* Mochmann, 1980)?

Deuxième conséquence, l'analyse de contenu en elle-même est inséparable de son propre schématisme cognitif. Celui-ci peut être emprunté à la mathématique, quand il s'agit, par exemple, de représenter ses résultats. Ainsi, on figurera les liaisons thématiques dans le journal intime de Goebbels par les distances entre des thèmes répartis en constellations, les différenciateurs sémantiques d'Osgood se traduiront sur une matrice, etc. Mais, seront également tributaires d'un schématisme cognitif les connaissances qui ont servi à obtenir ces résultats, plus ou moins structurés par analogie avec les mathématiques. Par exemple, il faut en passer par le schématisme cognitif de la langue, si on veut établir un corpus dont on regroupe les [108] syntagmes par le schématisme d'un langage institutionnel dans le cas d'un corpus scientifique, en physique par exemple (Ackerman et Zygouris, 1966), etc. Définition des données, construction de la grille, plan d'expérience, schème de codage, etc., une représentation instrumentale des connaissances est nécessaire à la méthode pour obtenir ses résultats, représentation qui comporte ses propres schématismes, y compris ceux de l'argumentation (Grize, 1984).

Relevant rapidement ces deux conséquences, savoir une soumission inévitable au singulier et une ambition des schématismes à l'universel logique, on serait donc justifié à prétendre que l'analyse de contenu se déploie, sur le plan de la méthode au moins, à l'image d'une sorte de ruban de Moebius de la connaissance, renvoyant incessamment l'idée de communauté à celle de raison et inversement. Mais, de cette façon, le souci d'« objectivité », préoccupation constitutive de l'analyse traditionnelle, nous apparaîtra-t-il en prise directe sur la sociologie de la connaissance. C'est pourquoi j'ai choisi, pour ma part, en début d'exposé, de présenter le modèle *ANOVEP* comme le produit d'une recherche se passant à la jonction de l'analyse de contenu et des mathématiques appliquées à la sociologie de la connaissance. Méthode

d'analyse de la communication, en effet, l'analyse de contenu travaille à ce titre de plein gré, me semble-t-il, à relier, tout comme la sociologie de la connaissance s'y adonne, la logique à l'idée de communauté, dans cet esprit où, à la limite, le champ de l'intersubjectivité tend à apparaître comme le nouveau lieu transcendantal.

Cela dit, un statut épistémologique inédit de la méthode se laisse voir à partir de l'expérience d'*ANOVEP*. Dans ce déplacement critique du sujet kantien, une première incidence épistémologique d'*ANOVEP* serait en effet d'attirer l'attention sur l'activité de coder un corpus, d'en offrir une première tentative, sans doute imparfaite, de formalisation.

Regardons, à titre d'exemple, ce que Francis Jacques appelle « l'alternative langagière ». C'est l'alternative pragmatique qu'on est forcé d'accepter, du moment qu'on assume qu'un mot, en pratique, ne doit désigner qu'une seule chose, auquel cas ou bien ces choses sont distinctes (homonymes) ou elles ne le sont pas (synonymes) [109] (Jacques, 1979). C'est l'alternative où se trouve le codeur. Car, nous voici alors, en effet, en plein cœur de la polysémie inhérente à un corpus quelconque, là où, placé devant un indicateur de sens, le codeur devra continuellement distinguer, afin de porter un jugement de présence ou d'absence, entre exemple et analogie (Fournier, 1989c).

Plus précisément, l'alternative pragmatique renvoie à l'activité de coder un corpus d'au moins trois façons. D'abord, par la question : qu'est-ce qu'une catégorie? La plan de l'interrogation est ici philosophique (ontologie) : dans des procédés ou des structures comme le binarisme de la langue, tel que celui dont s'est en partie inspiré *ANOVEP* (Fournier, 1983), que signifie, par exemple, classier de l'être? (Le non-être serait-il de l'être, comme disait l'autre?). Ou ne peut-on classier que de l'étant? Etc.

Par ailleurs, sur le plan psychologique, prétendre prendre en compte la décision d'un codeur qui range des unités de sens suppose une représentation de la structure de décision propre à cette opération. Dans *ANOVEP*, où l'activité de coder est vue comme un canal sans bruit à travers la théorie de l'information (Fournier, 1984), on convient, au carrefour de la théorie des graphes et de la théorie de l'information, de donner de cette activité une représentation formel-

lement fondée sur le signal plutôt que sur le signe, une représentation signalétique plutôt que signifiante, et relevant à ce titre de la cybernétique plutôt que de la linguistique. C'est à cela que se ramène finalement, en effet, l'inédit de lire un corpus à travers des indices de traitement de l'information employés pour construire des données, par exemple. Là réside la différence des indices *ANOVEP* d'avec le calcul fréquentiel de l'analyse traditionnelle.

Le résultat obtenu ainsi est une langue artificielle, i.e. à base logico-mathématique et, en ce sens, une syntaxe. Ce n'est donc que par le truchement de la sémantique du sociologue (mais ce pourrait aussi bien être celle du docimologue, du politicologue, de l'ingénieur, etc.), que la langue artificielle d'*ANOVEP* devient le moyen technique de faire passer le corpus d'un état premier, l'état de corpus clos tel que donné au codage par le décompte fréquentiel de l'information [110] de gestion (base de données), à un autre état, celui d'un corpus ouvert à l'interprétation (base de connaissances) (Fournier, 1983).

En cet état nouveau, le corpus en effet est devenu un système d'information de gestion : des éléments de savoir (ou faits) sont réunis par des relations internes (ou relations réglées) entre des propriétés dérivées d'une heuristique (en l'occurrence ici celle du sociologue analyste). Par exemple, un semblable corpus ouvert serait ici celui que fournirait l'analyse de l'éducation populaire au moyen de modèles causals obtenus avec *MODEPAS* (Fournier, 1990c).

Une troisième conséquence épistémologique d'*ANOVEP* remet au centre de sa réflexion l'activité de coder un corpus. Comment, en effet, s'articule encore avec une psychologie cognitive, mais cette fois sur le plan logico-mathématique, l'opération binaire qui consiste à coder ou faire coder un texte par référence à un questionnaire arborescent : que peuvent dire, par exemple, la structure mathématique de groupe ou le carré logique de Piaget sur l'à-propos de choisir chaque fois entre l'exemple et l'analogie pour ranger une unité de sens. Etc. (Fournier, 1983, 1989c)?

À travers ces trois questions : qu'est-ce qu'une catégorie? Qu'est-ce qu'un corpus ouvert ou une connaissance? Qu'est-ce que classifier des objets?, on aura donc remarqué que, par rapport aux langages documentaires (e.g. *SYNTOL*), il n'y a pas de différence de nature entre

le langage logico-mathématique dans lequel les données ou les résultats d'*ANOVEP* sont produits et les leurs. Mais, ce qui distingue *ANOVEP* de ces langages réside, au contraire, dans des questions de procédure logique et de procédure informatique de modélisation et de calcul, procédures essentiellement issues du postulat de coupler théorie des graphes et théorie de l'information (Berge, 1973; Shannon et Weaver, 1975).

Ce qui m'amène à considérer la seconde incidence épistémologique du modèle.

3.2. Portée informationnelle d'ANOVEP

[Retour à la table des matières](#)

Au moment où l'on se parle, la rencontre de l'analyse de contenu avec l'informatique est déjà vieille de 30 ans et des poussières.

[111]

Cependant l'idée de relier l'analyse computationnelle ou le calcul de fréquences de l'analyse traditionnelle au principe d'entropie maximum d'une distribution, comme le fait *ANOVEP*, est plus récente. Elle s'inspire en partie de la méthode d'analyse d'entropie de Darcy et Aigner sur des attributs, et de l'analyse d'attributs de Bettman, deux méthodes apparues concurremment au début des années 80 (Bettman, 1979; Darcy et Aigner, 1977).

À son tour, prenant appui sur la théorie de l'information, *ANOVEP* développe ainsi un formalisme qui le distingue de deux autres classes de logiciels s'en rapprochant, les logiciels d'analyse de texte (*DÉREDEC*) et les langages de représentation de la langue naturelle (*NETL*). La différence essentielle qui sépare *ANOVEP* de ces derniers sur le plan des procédures est une différence de sémantique (Paquin, 1990; Fahlman, 1979).

En effet, la sémantique à laquelle renvoie la formalisation logico-mathématique d'*ANOVEP*, comme on l'indiquait tantôt, est celle du sociologue ou d'une heuristique propre à l'analyste, en somme, et que ce dernier applique. Ce n'est donc jamais, si on peut ainsi dire, la sémanti-

que d'un contenu, au sens où l'on parlerait de la sémantique d'un domaine produite à travers une analyse syntaxique du vocabulaire. Par là, la langue artificielle d'*ANOVEP*, beaucoup moins élaborée qu'en analyse automatique (elle ne contient pas d'analyseur syntaxique), est une organisation de signification syntaxique en linguistique.

C'est pourquoi sa fonction, quant au contenu à analyser, n'est pas d'abord une fonction d'analyse, mais de communication : permettre, en langue naturelle, de passer d'une façon fiable du code du sujet, ici l'organisme communautaire qui présente un projet, à celui du ou des analystes. La langue artificielle d'*ANOVEP* ne produit pas l'analyse, mais en représente la condition de production.

Ainsi se déduit la parenté du formalisme d'*ANOVEP* avec celui de l'approche informationnelle. L'analyse de contenu, indiquait à l'origine Laswell, fait référence à une théorie générale de la communication. On peut ajouter qu'elle n'en fournit aucune, cependant, si ce n'est une structure d'argumentation, la sienne, plus ou moins [112] rattachée à l'idée de la rationalité. Et, là-dessus, d'autres avant nous en ont déjà signalé la faiblesse. Sans capacité d'inférer, rappelait par exemple Krippendorf, « vous n'avez qu'une technique de validation de la perception, pas une technique d'analyse » (Muchielli, 1974). Et on a déjà fait remarquer aussi, à propos de la quantification d'un corpus, qu'un nombre n'est jamais réellement un attribut.

En sorte que le donné dans le modèle *ANOVEP* n'est pas autre chose qu'une quantité signalétique. Le modèle représente, au sens strict, un cas de rencontre avec la machine à l'intérieur d'un paradigme scientifique où l'énergie est remplacée par l'information (Granger, 1967; Simondon, 1969). En ce « nouvel âge de la raison mécanique », écrit Jean-Pierre Sérès (1989, p. 430-437), les machines « ne sont plus seulement l'intermédiaire entre le monde et nous pour l'exploitation des ressources de celui-ci, mais l'intermédiaire entre lui et nous sous la forme d'un modèle (théorique) et d'un modèle sur lequel notre action est, d'entrée de jeu, possible ». Or, *ANOVEP* est une semblable machine théorique : dans *MODEPAS* par exemple, la possibilité pour l'analyste de modifier les catégories de la grille ou les critères d'observation à l'aide du coefficient MRC lui vient de la structure des

données dans la machine, non d'une méthode de construction de la grille (Fournier, 1984).

En conséquence remarquons qu'il résulte ici de ce rapport à la machine un effet pratique quant à la structure politique de nos prises de décisions sociétales. Par le biais du paradigme informationnel évoqué ci-dessus, en effet, l'application *ANOVEP* à l'information de gestion dans le logiciel *MODEPAS* se trouve à définir l'évaluation de programme comme étant, à la base, une situation de communication entre partenaires, administrateurs et administrés. Sous cet aspect convenait-il sans doute alors, en première partie de cet exposé, de rappeler la relation étroite existant entre la structure d'*ANOVEP* et le cadre d'évaluation développé pour répondre au mandat de production du logiciel *MODEPAS*.

Ce rappel opportun me permettra, au reste, de conclure brièvement avec la troisième incidence épistémologique du modèle.

[113]

3.3. Statut de la subjectivité dans ANOVEP

[Retour à la table des matières](#)

À l'origine de la méthode de l'analyse de contenu, il y avait, en effet, comme on l'a esquissé, ce désir de pouvoir atteindre et maintenir une sorte d'objectivité acquise face à l'étude du contenu de la communication. On se rappellera, par exemple, Berelson, De Sola Pool, etc. Des raisons l'expliquent, qui relèvent de l'histoire de nos disciplines : psychologie, linguistique, sociologie, etc.

On imaginait, par exemple, pouvoir dialoguer avec un interlocuteur : connaître l'émetteur, comme disait Berelson, en étudiant le message. On espérait faire dialoguer, au sens propre, un sujet, c'est-à-dire l'analyste ou, pour cette affaire, tout récepteur du message, avec un objet, c'est-à-dire avec des représentations de droit, valables pour tous les esprits, représentations de droit qu'on présume alors s'opposer axiomatiquement à des représentations de fait, contingentes celles-ci, et qui ne regardent pas la pratique scientifique, puisque, se-

lon la formule de Lachelier, on n'en peut jamais rien dire, sinon qu'on les a.

Les choses depuis ont évolué. Pour mesurer cette évolution, pourrait être alors évoquée avec profit, me semble-t-il, la pensée bien connue d'un philosophe allemand, Martin Buber. Celui-ci distinguait dans le patrimoine linguistique de l'humanité ce qu'il appelait des mots-principes, sortes de mots-souches à l'origine des visions du monde et, par là, de la construction de celui-ci. Parmi ces mots-principes, Buber réservait une place de choix au couple verbal formé des pronoms Je et Tu, puisqu'il s'agissait là, entre autres, du mot-principe fondateur de la relation. À celui-ci s'oppose, on le devine, le couple verbal Je-Cela, mot-principe d'ordre réducteur, en ce qu'il ne peut jamais être prononcé par l'être entier (Buber, 1959).

La distinction est parlante, sur soixante ans et plus d'analyse de contenu, quant au sort réservé à une certaine épistémologie de l'objectivation. Pour un peu, en effet, le rapport classique sujet-objet pourrait être lui aussi la figure (logique) d'un principe réducteur de l'être entier. Mais, une autre possibilité, la possibilité de réintroduire un sujet, est aussi offerte à l'analyse de contenu comme l'une de ses éventualités, et ce, à la faveur de cette rencontre avec les machines, du moins au sens où nous l'avons défini plus haut. Par exemple, dans [114] l'hypothèse où le modèle *ANOVEP* satisfait au souci d'obtenir des données cumulatives par analyse de contenu, on peut donc faire confiance au codeur humain, et introduire en gestion l'analyse de contenu comme technique de prise de décision.

Mais, en formalisant ainsi l'emploi d'un coefficient MRC de mesure du rendement des catégories au codage, *ANOVEP* se trouve à proposer, à sa manière, une façon de tenir compte de la place, sinon du rôle, d'une certaine subjectivité du récepteur (ou de l'analyste ou du codeur) dans le vieux modèle communicationnel (émetteur-canal-récepteur) retenu par le processus scientifique de l'objectivation.

Il y a plus. Puisque les catégories de l'analyse peuvent être modifiées sous le contrôle des intéressés, tant l'analyste que les partenaires à la décision, on peut d'autre part imaginer, sur la base de l'emploi d'un coefficient de mesure de rendement des catégories au codage rendant possible l'analyse à codeur unique, qu'il ne s'agit donc plus tel-

lement ici de réintroduire un sujet sur le plan opératoire, mais d'opérationnaliser plutôt la place nouvelle d'une relation entre sujets épistémiques, expérimenter ainsi la latitude d'un nouvel espace dialogique entre administrateur et administrés, entre évaluateurs et évalués, que construit et garantit le modèle. Prenons-nous à rêver : au moment où la rencontre empirique avec la machine introduit la bureautique dans les administrations, on pourrait alors commencer à retrouver, en effet, dans le processus dialectique de la relation de connaissance sujet-objet, la présence, à l'aide d'*ANOVEP* ou d'autres modèles semblables, d'une certaine pratique cognitive du couple verbal Je-Tu ouverte sur la sociabilité des Nous.

La méthodologie d'*ANOVEP*, en conclusion, se veut ainsi cohérente avec l'effort éventuel de transformer, si l'on peut dire, le couple verbal Je-Cela de l'idéologie technocratique en un Je-Tu politique, dialogue entre citoyens dans la Cité. On en revient ainsi, je ne sais pas si vous l'avez remarqué, aux principes d'interaction et de participation des partenaires énoncés dans le cadre d'évaluation lié à l'application d'*ANOVEP* dans *MODEPAS*. Comme quoi il n'y a jamais si loin des préoccupations épistémologiques pointues à celles de l'intervention sociale et du jugement de valeur.

[115]

Références bibliographiques

[Retour à la table des matières](#)

ACKERMANN, W. et R. ZYGOURIS (1966) *Représentation et assimilation des connaissances scientifiques*, C.E.R.P./A.F.P.A., document no 3.

BARBUT, Marc (1969) *Mathématiques des sciences humaines*, Paris, PUF.

BERGE, Claude (1973) *Graphes et hypergraphes*, Paris, Dunod.

BETTMAN, J.R. (1979) *An Information Processing Theory of Consumer Choice*, Reading, Addison-Wesley.

BUBER, Martin (1959) *La vie en dialogue*, Paris, Aubier-Montaigne.

COBB, L. et R.M. THRALL (1981) *Mathematical Frontiers of the Social and Policy Sciences*, Boulder, Westview Press.

DARCY, R. et H. AIGNER (1977) « Entropy Analysis Technique », *Journal of Marketing Research*, août, p. 416-419.

FAHLMAN, Scott E. (1979) *NETL. A System for Representing and Using Real-World Knowledge*, Cambridge, MIT Press.

FOURNIER, Richard (1983) *Analyse de contenu des demandes reçues au programme d'aide aux organismes volontaires d'éducation populaire en 1981-82 et en 1982-83, I, L'Outil d'analyse*, MEQ, Direction générale de l'éducation des adultes, 83 pages.

FOURNIER, Richard (1984) *Une mesure de rendement du codage en analyse de contenu : l'indice MRC*, MEQ, Direction générale de l'éducation des adultes, 20 pages.

FOURNIER, Richard (1985) *L'État interventionniste : l'évaluation qualitative. L'analyse de contenu au seuil d'une nouvelle pratique*, Université du Québec, Communications, Études, no 19.

FOURNIER, Richard (1988) « Schématisation et représentation des connaissances en analyse de contenu appliquée à l'information de gestion », in : A. Turmel, (éd.), *Sociologie et intelligence artificielle*, Laboratoire de recherches sociologiques, Département de sociologie, Université Laval, p. 151-178.

FOURNIER, Richard (1989a) *Proposition d'un système de classification pour l'accréditation des organismes au programme de soutien à l'éducation populaire autonome*. Dossier technique, MEQ, Direction de l'éducation permanente, mars.

[116]

FOURNIER, Richard (1989b) *Proposition d'évaluation du programme PSE-PA de soutien à l'éducation populaire autonome*. Dossier technique, MEQ, Direction de l'éducation permanente, avril.

FOURNIER, Richard (1989c) « Représentation des connaissances et gestion de l'incertitude en analyse de contenu à codeur humain : le modèle ANO-VEP », in : B. Moulin et G. Simiand, (éd.). *Informatique cognitive des organisations. Textes du Colloque ICO 89*. Limonest, L'Interdisciplinaire informatique, p. 491-501.

FOURNIER, Richard (1990a) *Aperçu d'un système d'indicateurs de gestion à la DGRTQ*, Dossier technique. Ministère de la Main-d'œuvre, de la Sé-

curité du revenu et de la Formation professionnelle du Québec, Bureau de la coordination de la refonte des systèmes.

FOURNIER, Richard (1990b) *Bilan descriptif des réalisations du programme de soutien à l'éducation populaire en 1988-1989*. Document administratif, MEQ, Direction de la coordination des réseaux, mars, 86 pages et annexes.

FOURNIER, Richard (1990c) *MODEPAS. Manuel de l'utilisateur pour Macintosh et PC ou compatibles*, MEQ, Direction de la coordination des réseaux, avril, 92 pages.

FOURNIER, Richard (1990d) « La représentation des connaissances en analyse de contenu appliquée à l'information de gestion dans le modèle ANO-VEP », *ICO Québec*, décembre, p. 116-127.

GRANGER, Gilles-Gaston (1967) *Pensée formelle et sciences de l'homme*, Paris, Aubier-Montaigne.

GRANGER, Gilles-Gaston (1968) *Essai d'une philosophie du style*, Paris, Armand Colin.

GRIZE, Jean-Blaise (1984) *Sémiologie du raisonnement*, Berne/Frankfort-sur-Maine /New-York, Peter Lang.

GUINDON, Hubert (1960) « The Social Revolution of Quebec Reconsidered », *The Canadian Journal of Economics and Political Science*, November, XXVI : 533-551.

HOLSTI, O.R. (1969) *Content Analysis for the Social Sciences and the Humanities*, Reading, Addison-Wesley.

[117]

HUDON, Yves (1991) « La recherche textuelle : un choix technique et administratif », *Documentation et bibliothèques*, avril-juin, 37(2) : 73-77.

HURTUBISE, René (1976) *Informatique et information*, Montréal, Les Éditions Agence d'Arc.

JACQUES, Francis (1979) *Dialogiques. Recherches logiques sur le dialogue*, Paris, PUF.

KRIPPENDORFF, Klaus, in Ekkehard Mochmann (Hg.) (1980) *Computerstrategien für die Kommunikationsanalyse*, Frankfurt/New-York, Campus Verlag.

MEQ (1987) *Rapport du comité provincial sur les orientations proposées pour l'Éducation populaire réalisée par les Commissions scolaires*, juin.

MUCHIELLI, R. (1974) *L'Analyse des documents et des communications. Position du problème*, Paris, Les Éditions ESF.

PAQUIN, Louis-Claude (1990) « DÉREDEC EXPERT : un outil logiciel cognitif », *ICO Québec*, septembre, p. 13-20.

POULIN, F. et G. TRUDEAU (1982) *Les conditions de productivité des délégations du Québec à l'étranger*, Québec, École nationale d'administration publique.

SÉRIS, Jean-Pierre (1989) « Machine et communication » in *Encyclopédie philosophique universelle*, I. PUF, p. 430-437.

SHANNON, C.E. et W. Weaver (1975) *Théorie mathématique de la communication*, Paris, Retz-C.B.P. L.

SIMONDON, Gilbert (1969) *Du mode d'existence des objets techniques*, Paris, Aubier-Montaigne.

TESCH, Renata (1990) *Qualitative Research : Analysis Types and Software Tools*, Desert Hot Springs, Falmer Press.

THOMAS, William I. et Florian ZNANIECKI (1918) *The Polish Peasant in Europe and America*, Chicago, University of Chicago Press.

ZELLER, Richard A. et Edward CARMINES (1980) *Measurement in the Social Sciences*, Cambridge University Press.

Fin du texte