

Chapitre 6

**SCHÉMATISATION ET REPRÉSENTATION DES
CONNAISSANCES EN ANALYSE DE CONTENU
APPLIQUÉE À L'INFORMATION DE GESTION**

Richard Fournier

L'objet de ce travail est de faire connaissance avec la notion de schématisation développée par Grize (Borel *et al.*, 1983; Grize, 1984). L'occasion en est une application pratique à l'analyse de contenu lorsque celle-ci est employée comme technique de gestion dans une organisation. Sous cet angle, notre sujet se rattache au domaine de ce que nous pourrions appeler l'intelligence artificielle.

L'exemple d'analyse de contenu adopté, proche de l'analyse documentaire, relèverait, pour sa part, de l'ingénierie des connaissances dans les organisations plutôt que de l'analyse du langage ou de celle du discours, de l'usage de l'informatique dans les organisations (Hurtubise, 1976) plutôt qu'en linguistique. À développer cet exemple, par ailleurs, on verra qu'il n'est pas dans notre intention de s'en servir pour comparer l'analyse de contenu avec l'analyse automatique. Il s'en tient en effet à une variété d'analyse où l'ordinateur ne lit pas un texte mais parcourt, dans une langue de support ou un langage intermédiaire, le produit d'un codage effectué par un codeur humain. On pourrait avancer plusieurs raisons à l'origine de cette façon de procéder. L'une vient du fait que, de cette manière, à l'emploi de l'analyse de contenu comme technique de gestion se greffe un enrichissement de certains postes de travail clérical.

1. LA SCHÉMATISATION ET SA NATURE

Figurons-nous donc un codeur humain parcourant un texte pour y relever des indicateurs de contenu par jugements de présence ou d'absence. On peut dire que l'objet de son *modus operandi*

oscille continuellement de l'analogie à l'exemple, puisque chaque unité de sens qui démarque un indicateur le fait plus ou moins, selon ce qu'il en décide, soit à titre d'analogie, soit à titre d'exemple.

Si je rappelle cette entrée en matière, c'est qu'elle nous permet d'aborder directement la notion de schématisation à partir de l'étude qu'en fait Miéville dans le livre de Borel: *Essai de logique naturelle* (1983). Nous verrons en même temps comment ici se construit la connaissance.

Partons de la proposition de Miéville (Borel, 1983: 168) qu' "engendrer ou modifier une connaissance, c'est [...] reconstruire la signification". On peut supposer que cette opération se fait de façon logique en raison de la nécessité d'aboutir à un résultat cumulatif, puisque le savoir, ici, est cumulatif (savoir des générations, savoir scientifique, etc.). La logique sera donc un système d'opérations qui crée un monde d'objets ou de signes. Comment, à l'intérieur de ce système de signes, se fait la modification des connaissances selon Miéville? À la façon, comme dans le schéma classique de la communication, d'une interaction entre un émetteur et un récepteur liés par le recouvrement sémantique. Ainsi, un organisme volontaire d'éducation populaire (OVEP) qui complète une demande de subvention et, d'autre part, le codeur qui la transcrit et, ce faisant, fabrique les données de l'analyse sont dans un rapport de correspondance au temps T1. Ce qui se produit est qu'une classe-objet de discours, par exemple un nom, deviendra, en T2, autre chose: les constituants de la classe changent pour le même nom. Au terme de quoi, dans notre exemple, à travers des mots de passe une stratégie, la stratégie OVEP pouvons-nous dire, a rejoint le codeur, tout comme l'interprétation du codeur a rejoint, par un truchement semblable, l'analyste gouvernemental.

Intervient donc à la faveur de la communication une procédure discursive qui modifie, chez l'interlocuteur, l'état initial de l'objet (e.g. changer la signification en conservant le mot). On verra ainsi deux choses à l'oeuvre dans la modification des con-

naissances: 1) les déterminants sociopolitiques de la communication: nous les appellerons stratégies, idéologies, enjeux, perceptions des acteurs, sous-cultures, représentations sociales, préconstruits culturels ou significations différentielles, habitus, aptitudes langagières, classes sociales, etc., comme on voudra; 2) ce en quoi se définit pratiquement la "subjectivité" de l'interprétation, à savoir comment une procédure discursive bouge à l'intérieur du recouvrement et modifie l'objet. Miéville écrit:

"La modification d'un état de connaissance peut alors être caractérisée comme suit: il existe une opération qui transforme une classe-objet en une autre classe-objet de même nom. Le nom subsiste mais la classe change: ses constituants sont modifiés par l'intervention d'une procédure particulière." (Miéville, 1983: 170.)

Du côté de l'interlocuté, un objet A, qui est la classe collective de X, dit par exemple Miéville, différera de A au temps T2 dans le sens d'un nouveau recouvrement dû à l'intervention de la procédure du locuteur.

"Intuitivement parlant, la classe modifiée s'enrichit si l'objet s'inscrit dans une intervention didactique réussie; elle subira d'autres modifications, d'autres déplacements si l'objet s'inscrit dans la perspective d'une intervention plutôt polémique." (Miéville, 1983: 171.)

La modification n'est donc pas aléatoire mais dépend, en somme, de la définition de la situation dont les contraintes, entre autres, imposent le choix d'une procédure discursive.

On peut donc retenir deux choses. Que "la modification des classes collectives peut [...] rendre compte de celles des états de connaissance", selon l'expression de l'auteur (Miéville, 1983: 171),

mais aussi que la modification des connaissances est tributaire à tout coup de la procédure employée, que communication et argumentation sont synonymes.

Il en ressort pour notre propos que la mesure de la part d'interprétation qu'affiche le rendement d'un codeur est la mesure de l'argumentation qu'il emploie, peut-on dire, en convenant que cette argumentation relève du discours ou du langage institutionnel auquel il se rattache, comme dans le cas d'une École pour un chercheur par exemple, et aussi de son équation sociopersonnelle (Boudon, 1982: 396). Un codage est donc l'expression d'une argumentation; reste à voir comment cette argumentation procède en logique naturelle.

La caractéristique d'une grille, en analyse de contenu la plus traditionnelle, est, en effet, d'être un monde créé (Cartwright, 1959; Holsti, 1969) au sens même où le terme est reçu en logique (Wittgenstein, 1961). La logique d'une grille de codage en analyse de contenu peut être mathématique ou naturelle, voire prétendre à une combinaison des deux, comme dans un certain emploi de l'algèbre de Boole; mais, si l'on se fie à ce que viennent de dire les auteurs, elle est incontestablement naturelle dès qu'il y a interlocuteur, comme dans notre exemple l'analyste ou l'institution.

Regardons donc, en guise d'illustration, ce qui se passe dans la relation d'inclusion, opération qui fait la comparaison entre les éléments d'un ensemble. L'opérateur "partie de" est ici le terme primitif. Soit la relation: "Québécois partie de Canadien". Une telle relation est vérifiée, par exemple, pour l'intersection Québécois, Écossais, Inuit. On peut transformer la relation en classe distributive en l'écrivant comme ceci: "Canadien {Québécois, Écossais, Inuit}". C'est la définition qu'on rencontrerait dans un thésaurus ou dans un dictionnaire analogique ou logique. Une classe distributive exprime ainsi l'extension d'un concept. Par exemple, elle contient ici, par hypothèse sur la composition du peuple canadien, trois éléments et rien d'autre. La partie est alors nécessairement

distincte du tout, et ce dernier ne sera jamais donné par la recension des parties puisque la relation de partie à tout ne coïncide pas avec celle d'inclusion.

Mais on peut aussi exprimer la relation comme ceci:

A _____ B _____ C _____ D

où AD = canadien et AB, BC, CD aussi; mais en même temps: AC, CD et AB, BD. L'opérateur "partie de" signifie alors "est élément de", au sens où le tout est aussi élément de lui-même. Nous voici en face d'une classe collective ou méréologique, classe qui, elle, désigne une totalité réelle. C'est à celle-là, par exemple, que Miéville réserve le soin de décrire toute "espèce d'entité individuelle structurée par la relation de partie à tout" (Miéville, 1983, 169). Elle ressemble à un sac d'épicerie où toutes les combinaisons sont à la rigueur possibles: l'armée faite d'officiers, de sous-officiers, de simples soldats, par exemple, mais aussi de sièges de véhicules blindés, de manuels de règlement, de tenues de combat et d'étiquettes de buanderie, de poètes, de policiers, de rations alimentaires et de cartes à jouer, etc. (Fortin, 1981: 279). Ainsi les officiers sont-ils éléments de l'armée, mais on ne peut supposer que tout officier est l'armée comme tout Québécois est Canadien, car l'armée est une infinité de choses, en fait, une totalité réelle.

La classe méréologique revêt, pour notre propos, deux caractéristiques intéressantes. D'abord, on peut y entrer par n'importe où (Miéville, 1983: 170). Ce qui peut correspondre à la polysémie d'un corpus de la même façon, par exemple, dont le mythe, pour l'analyse, peut avoir plusieurs portes d'entrée (Fournier, 1983: 87). L'intérêt, ici, est que, si un codage est l'expression d'une argumentation, l'analyse reçoit de même l'opportunité de circuler de partout dans le corpus. Ce qui, techniquement, revient, par exemple, à dire que chaque variable d'une grille d'analyse est une relation réflexive (Fournier, 1983). La relation "est élément de", pour reprendre les termes de Miéville, s'entendra alors strictement

“au sens large où le tout est aussi élément de lui-même” (1983: 169).

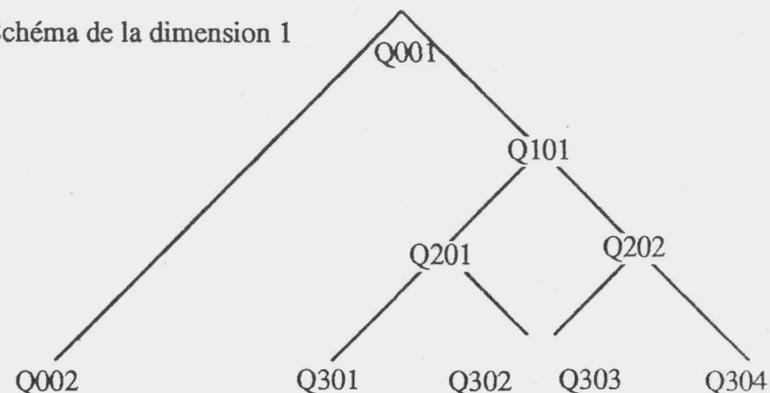
L'autre caractéristique utile à notre propos est que la classe méréologique ou collective définit en principe la classe vide comme une classe pleine (Miéville, 1983: 169). En théorie des ensembles, cette relation est exprimée par le théorème: $|E|=0, (|E|=\emptyset) \Rightarrow P(E)=1$ (Barbut, 1969). Or, on peut utiliser ce théorème comme algorithme pour construire une grille d'analyse (Fournier, 1983). Le cadre conceptuel d'une telle analyse prend alors la forme d'un arbre logique de formule $P(E) = 2^{|E|}$, arbre exponentiel répondant au schéma d'un codeur binaire en codage de Fano (Roubine, 1970). Le théorème ci-dessus agit alors comme primitive et représente le cadre conceptuel au moment de sa fondation, *i.e.* n'importe quel terme à compréhension maximum et extension nulle (Fournier, 1983: 66). La structure de grille de codage qui en résulte, l'arbre $P(E) = 2^{|E|}$ des parties de E, traduit techniquement le fait qu'un concept perd en compréhension ce qu'il gagne en extension (Fournier, 1983: 70) et il s'ensuit que l'argumentation qu'exprime le codage peut se représenter n'importe où sur l'arbre, créant un corpus ouvert.

Pour ces deux raisons, la nature spécifique de la classe méréologique permet donc de circuler à travers la polysémie du corpus en marquant, selon les niveaux de l'arbre logique, nos repères. Du fait qu'on recourt, pour quantifier des unités de sens, à la théorie des graphes (Fournier, 1983: 40), une valeur numérique peut alors s'attacher à chaque noeud, en guise de mesure, et autoriser de fabriquer des indices, selon l'exemple de grille présenté à la figure 1. Des fréquences sont transposées en trajets d'après lesquels on calcule différents indices rattachés à des variables.

Mais s'agit-il vraiment d'un avantage? Pourquoi l'analyse de contenu devrait-elle tenir ici expressément compte du jeu entre compréhension et extension en attachant cette importance à la classe méréologique? La réponse à cette question passe par l'activité du codeur. Elle réside dans la différence entre l'analogie et la classifi-

cation et renvoie à nouveau au mécanisme de la schématisation selon Miéville.

Schéma de la dimension 1



Valeur des indices C, Sk et S extraits du graphe de la dimension 1 (19 variables) pour les quatre variables de l'extrait (échantillon de 97 cas)

	Q001	Q101	Q201	Q202
C	0,99	1,15	1,23	1,22
Sk	0,99	0,57	0,41	0,40
S	0,61	0,39	0,18	0,41

Figure 1 (Fournier, 1985b: 20)

Le dynamisme par lequel se réalise la modification des connaissances dans la communication, fait en effet remarquer notre auteur, suppose encore que les objets du discours (les mots, les signes) ne signifient pas d'une manière uniforme mais selon l'une ou l'autre de deux modalités: les faisceaux et les cas:

“La première est de [...] considérer [un objet de discours] du point de vue [de ses] propriétés ou des transformations dont il y a sens à se demander si

elles peuvent [...] s'appliquer à l'objet. Nous parlons dans ce cas du «faisceau» de l'objet. Certains éléments du faisceau de l'objet sont construits, d'autres sont préconstruits, d'autres sont transformés ou construits dans le discours. Soit, par exemple, l'objet «deux». Dans un discours pédagogique [...] cet objet aura dans son faisceau les éléments suivants: appartenir à l'ensemble des entiers, être plus grand que un, être pair, être premier, etc. Dans un autre contexte, le même objet aura un autre faisceau: avoir quatre lettres, être un mot invariable, etc.

“Une deuxième façon de rapporter l'objet à ce qui le détermine revient, de façon duale, à l'intégrer au «champ» d'un prédicat. On le considère alors comme élément d'un ensemble d'objets auxquels telle propriété peut être attribuée, ou telle transformation appliquée. Soit, par exemple, le prédicat «être un nombre». Son champ comprend, entre autres, le nombre deux. La différence essentielle entre les deux points de vue est que l'objet «deux», pris avec son faisceau, est saisi comme une entité individuelle, alors que, vu dans le champ d'un prédicat, il est un cas parmi d'autres, support possible d'une généralisation.” (Miéville, 1983: 161.)

Autrement dit, sur le plan de l'axiomatisation, la différence entre classe méréologique et classe distributive recoupe la différence, pourrait-on proposer au sens de Granger (1968), entre une causalité hétérogène, *i.e.* issue de l'action de produire le mot, dans la première façon, et une causalité homogène, *i.e.* modelée sur le virtuel, dans la seconde. De la distinction faisceau-cas, on devra donc retenir encore que chaque mot a son faisceau propre, qu'un tel faisceau n'est pas une donnée stable mais varie suivant la perception qu'on peut en avoir et, surtout, que le jugement d'analogie consiste en la mise en rapport de deux faisceaux. De ce dernier trait,

en effet, vient qu'il y a différence substantielle entre analogie et classification, la même qu'entre analogie et exemple, les deux pôles vers lesquels oscille chaque fois, pour le codeur humain en analyse de contenu, chaque unité de sens qu'il croit reconnaître.

Cette différence entre l'analogie et l'exemple, Miéville l'illustre, au texte, à partir de la phrase suivante de Chomsky:

“La fonction du langage est la communication comme la fonction du coeur est de pomper le sang.”
(Miéville, 1983: 162.)

On peut faire là-dessus, propose notre auteur, deux types d'analyse. Le premier est une étude de l'analogie: une propriété globale attribuée aux objets, à savoir la propriété “posséder une structure et une fonction”, permet la mise en rapport des organisations des deux faisceaux. Miéville écrit au sujet de cette propriété:

“Leur convenant à tous deux, elle a pour effet de sélectionner une zone spécifique dans chacun des deux faisceaux, de sorte que l'on peut parler de fonction de communication pour le langage et de fonction de pompage pour le coeur.

“Un échange est alors possible entre les deux zones sélectionnées, ce qui permet de séparer la structure de la fonction. Le discours sur le coeur, objet supposé mieux connu du lecteur que le langage, permet cependant d'affirmer «qu'il y a interaction entre la structure et la fonction» et de laisser entendre que ceci vaut aussi pour le langage.

[..]

“Dans la procédure que décrit cette première analyse, l'objet «coeur» fonctionne comme un modèle à partir duquel il est possible de concevoir l'objet «langage». Le modèle autorise des manipulations

(mentales) de l'objet [...] le moins bien connu, mais à condition que celles-ci lui soient accommodées, ou soient incorporées à son propre faisceau." (Miéville, 1983: 163.)

La seconde analyse est une étude de la classe. Elle relève d'un examen de la notion d'inclusion dans la classe "avoir une structure et une fonction" — classe plus ou moins intuitive ou implicite. Structure et fonction, dira-t-on alors, sont des propriétés parmi un ensemble de propriétés communes à plusieurs objets: coeur, langage, système de transport urbain, onde porteuse, ménisque du genou, arc-boutant, électro-aimant, etc.:

"«Coeur» et «langage» peuvent en effet aussi apparaître comme deux éléments d'une classe d'objets déterminés par un ensemble de propriétés communes. Mais la séquence cesserait d'être de type analogique: les deux objets auraient même structure et même fonction, et toutes deux seraient inséparables. On aurait une procédure classificatoire." (Miéville, 1983: 167.)

Cette différence est la frontière, pour Miéville, entre l'analogie et l'exemple. Si l'un des deux objets seulement est intelligible dans sa singularité, tel l'objet "coeur" dans la phrase de Chomsky (dans le cas où l'objet "langage" se présente au sens littéral d'un autre organe), on sort de l'analogie ou de la relation horizontale entre deux singuliers et on se trouve dans le monde de l'exemple (Miéville, 1983: 199). Dans ce monde, la circulation se fait verticalement entre des singuliers et un universel par déduction ou induction d'une propriété caractéristique d'un ensemble de cas, que tantôt l'exemple illustre, tantôt il permet de construire. Ainsi, un alcaloïde est un corps organique formé de carbone, d'hydrogène, d'oxygène et d'azote (Riou-Delorme, 1957: 549): la quinine, la strychnine, l'opium, la morphine, la nicotine en sont. Si tous mes interlocuteurs comprennent qu'il s'agit d'un poison violent, c'est

une déduction; s'ils comprennent tous ce qu'est un alcaloïde, c'est une induction. Dans les deux cas, la circulation est verticale.

L'intérêt, pour notre propos, de la classe collective ou mé-réologique (telle que définie à la condition 2 dans: Fournier, 1983: 87) est donc ici de nous situer, d'après cet exemple, dans le dynamisme de la communication, à un endroit précis où l'argumentation agit sur les significations et leur construction, *i.e.* au coeur même de la coexistence entre analogie et classification. Cet intérêt nous renvoie ainsi au rôle de la classe dans la modification des connaissances et à une définition plus formelle, en terminant cette partie de l'exposé, de ce qu'est la schématisation.

Nous avons vu avec Miéville que les mots peuvent être deux choses en même temps: des faisceaux de signification (*i.e.* chaque fois une entité individuelle avec l'ensemble de ses propriétés singulières plus ou moins globalisantes et permettant la réunion d'ensembles) et des champs de prédicats ou espaces d'attributs, où les mots sont des cas interchangeables dans une classe: le champ du prédicat. Soit le mot "deux". Je puis penser $2 = a$, valeur d'une constante (ce que je fais probablement dans la vie ordinaire en pensant au mot "deux" si j'ai une réunion hebdomadaire à 14h), mais je pourrais aussi écrire $2 = a + bi$, nombre complexe, où un champ cartésien fait que le mot "deux" est interchangeable avec n'importe quel nombre réel pourvu que i soit la racine carrée de -1 (Warusfel, 1966: 46). Dans le premier cas, la constante "deux" décrit des heures comme elle peut décrire un couple, un miroir, etc.: le terme "deux" est considéré comme une entité individuelle munie de l'ensemble de ses propriétés (où d'être un nombre réel n'est que l'une de celles-là), donc une entité apte, comme telle, à se réunir (par intersection: c'est le recouvrement sémantique) avec une autre entité individuelle. Nous avons proposé de remarquer qu'il s'agit alors, dans la façon d'aborder la question, d'attacher la production du mot à une causalité hétérogène, *i.e.* issue de l'action. Mais si le terme "deux" désigne un nombre complexe (tout réel que soit ce nombre), nous voici en face "d'un cas parmi d'autres en support

d'une généralisation", comme dirait Miéville, en l'occurrence en support de la définition $z = a + bi$ d'un nombre à partie imaginaire. Nous avons affaire, disions-nous de ces situations, au modèle d'une causalité homogène, *i.e.* caractérisée par l'emprise du virtuel. C'est le cas, par exemple, de la notion d'espace d'attributs en sociologie (Barton, 1955).

Dans la pratique, les deux se confondent ou risquent d'être confondues, ce que rappellent, par exemple, Lelouche (1986) et Winston (1979) sur le maniement de l'opérateur "isa". La raison en est le dynamisme de la communication. Et, si on s'intéresse alors aux opérations logiques de base qui s'expriment à la faveur de ce dynamisme, le tableau ouvert par Miéville apparaît comme un cas particulier de schématisation au sens de Grize (1983: 99).

Pour Grize, le processus de recouvrement sémantique et, pour autant, le modèle de la communication reposent d'abord, comme il l'explique, "davantage sur l'idée de résonance ou d'induction physique que sur celle de transmission d'unités d'information" (Grize, 1983: 99). Les deux perspectives, doit-on faire observer cependant, ne s'excluent pas nécessairement (Fournier, 1985b). Grize, à tout événement, propose de son modèle la représentation suivante, dans laquelle "im(T)" veut dire "image du thème (de la communication)" (Grize, 1983: 99).

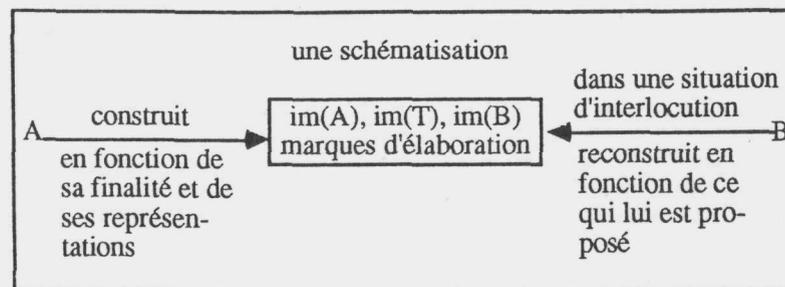


Figure 2 (Grize, 1983: 99)

Grize ajoute à la description ce qui suit:

"On a donc affaire à des activités logico-discursives aussi bien du côté de A que de B et rien ne permet de supposer que la schématisation proposée par A et celle reconstruite par B sont rigoureusement isomorphes. Toutefois, l'expérience montre qu'elles sont en général suffisamment proches l'une de l'autre pour que l'entente soit possible. De plus, A et B peuvent en principe échanger leurs rôles, de sorte que l'isomorphisme peut procéder par approximations successives." (Grize, 1983: 100.)

Cette définition de la schématisation profite de la séparation entre logique mathématique et logique naturelle. "La logique naturelle est l'étude de ces opérations de schématisation", écrit-il en effet de son modèle (Grize, 1983: 100). Mais encore faut-il "distinguer deux niveaux: celui des opérations en jeu dans toute activité discursive et celui propre à certaines procédures spécifiques" (Grize, 1983: 100). Nous avons vu deux de celles-ci avec Miéville: l'exemplification et l'analogie. Grize nous propose d'articuler son modèle sur les opérations plus générales de premier niveau qu'il appelle les opérations élémentaires. Il rejettera d'emblée, à ce moment, l'exclusivité de la logique mathématique:

"Les objets de la logique mathématique n'ont pas d'autre propriété que celle d'appartenir à leur catégorie et les prédicats d'appartenir aux leurs. Il n'en va pas de même en logique naturelle. Ici, objets et prédicats restent tout imprégnés du préconstruit culturel qui les détermine toujours en partie et ils ne cessent d'être modifiés par les activités discursives qui portent sur eux." (Grize, 1983: 100.)

Cette distance le conduit à introduire six concepts de base au lieu des deux primitives (l'objet et la proposition) de la logique

mathématique. Car si, d'un côté, l'objet, en effet, est imprégné du préconstruit culturel, le sort de la proposition, quant à elle, est entre les mains du locuteur:

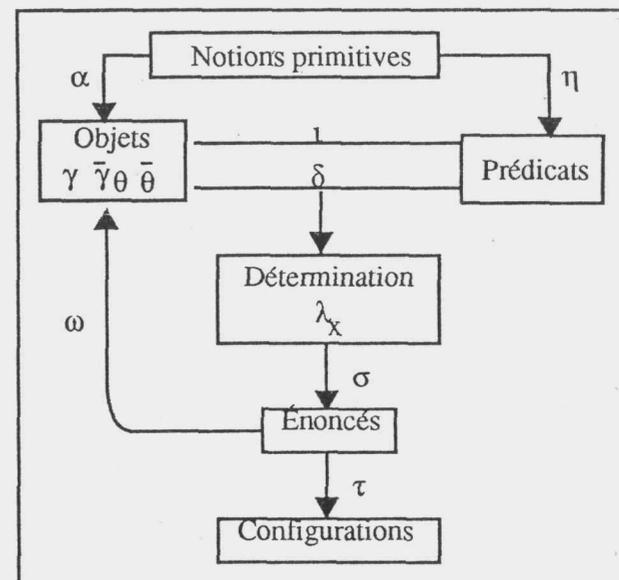
“Quant aux propositions de la logique mathématique, et même lorsqu'elles sont modales, elles ne sont à proprement parler assertées par personne et celles qui importent, c'est-à-dire celles auxquelles s'appliquent les règles de transformation, ont toutes l'unique statut de théorème. Tout au contraire, les propositions de la logique naturelle, parce qu'elle est logique de la schématisation, sont toujours prises en charge par quelque sujet énonciateur ou locuteur et leurs statuts sont fort divers.” (Grize, 1983: 101.)

Pour cette raison Grize appellera ces propositions des énoncés.

Dans la figure 3, il résume donc les opérations que permettent les concepts qu'il avance pour rendre compte, en logique naturelle, de la schématisation. À ceux-ci, localisés dans des boîtes, s'ajoutent les six opérations qu'il décrit à la suite:

Contentons-nous de remarquer que l'opération gamma (gamma barrée) revient à introduire dans la construction du discours la notion de classe collective ou méréologique (Grize, 1983: 102) à la suite des travaux de Lesniewski (cités dans Grize, 1983: 110). Cette observation suffit à notre propos et nous permet de conclure.

Si l'activité de schématisation se déroule bien telle que le supposent Grize et Miéville, *i.e.* par approximations successives obéissant, à l'intérieur d'un processus de recouvrement sémantique, à des procédures d'argumentation, que pourrons-nous dire alors de la modification des connaissances en relation avec l'une de ces activités de reconstruction: le codage en analyse de contenu? Car l'affirmation minimale qu'on a avancée jusqu'ici en réponse à cette question est, en somme, que la classe collective ou méréologique



Certaines opérations font passer d'un concept à un autre: α , η , ι , σ , ω et τ . L'une, δ associe un concept à deux autres. Les opérations γ , $\bar{\gamma}$, θ , $\bar{\theta}$ sont internes et λ_x est une famille d'opérations internes.

Figure 3 (Grize, 1983: 104)

permet de traiter du rapport entre l'analogie et l'exemple d'une façon plus satisfaisante que la seule classe distributive. Reste à axiomatiser la proposition. C'est l'objet de la seconde partie de l'exposé portant sur la représentation des connaissances en analyse de contenu.

2. LA REPRÉSENTATION DES CONNAISSANCES EN ANALYSE DE CONTENU

La perspective sociologique ici utilisée est celle de la méthode des indicateurs sociaux (Land et Spilerman, 1975). L'idée est que l'analyse de contenu, technique d'analyse de la communication, peut aussi s'employer en gestion lorsqu'il s'agit d'évaluer l'action interventionniste de l'État pour ce qui a trait, en particulier, à l'évaluation de programmes dits mous.

Nous groupons sous ce terme (Fournier, 1984) des programmes publics n'obéissant pas à des nécessités politiques évidentes pour tous et, surtout, difficiles, par ailleurs, à évaluer, étant fondés, pour l'essentiel, sur des données qualitatives: projets, opinions, attitudes, croyances, valeurs, idéologies, etc. On y reconnaît le programme de financement des groupes volontaires engagés dans l'éducation populaire (OVEP) du Ministère de l'éducation du Québec, dont nous parlions au début. Mais on pourrait tout aussi bien penser à des programmes mous à propos d'un programme de formation en industrie, de prévention routière ou de développement communautaire, etc. En contexte de société à somme nulle (Thurrow, 1981), de tels programmes publics acquièrent, en outre, une caractéristique supplémentaire: on en parle souvent comme s'ils devaient être les premiers à faire les frais des mesures visant à retailler l'intervention de l'État, sans que l'on soit mieux fixé sur leur raison d'être au moment de les abolir faute de pouvoir profiter d'un mode d'évaluation permettant le savoir accumulé. C'est à ce besoin que prétend venir répondre ici, aux côtés d'autres techniques, l'analyse de contenu.

Pour cette dernière, la difficulté à surmonter sur le plan de la méthode est évidemment celle de l'analyse à codeur unique. Il est téméraire de penser, en effet, que l'organisation puisse affecter des ressources humaines à l'analyse de contenu au-delà d'un certain seuil de rentabilité. Ce qui, entre autres lorsqu'il est question de

s'assurer de la validité des données, exclut, à toutes fins pratiques, de s'appuyer sur des équipes de codeurs, ainsi qu'en recherche le requièrent l'un ou l'autre des coefficients de fidélité traditionnels basés sur le contrôle du codeur par ses pairs. L'analyse de contenu à codeur humain gagne donc sa crédibilité comme technique de gestion dans les organisations si elle parvient à contourner cette difficulté constitutive.

“Sur cette base, notre idée pour mesurer le rendement d'un système de catégories fut de partir de la quantité moyenne de choix exercée par le codeur pour classer une unité de sens, ou de la notion d'entropie d'une variable. La méthode d'analyse que nous avons mise au point en 1983 faisait déjà reposer la mesure ou quantification du contenu sur des propriétés de la théorie des graphes: il suffisait d'y rapporter la théorie de l'information d'où provient ici la notion d'entropie.

“Il est apparu ainsi qu'on pouvait mesurer le biais au codage ou la «subjectivité» du codeur par la différence entre le coût du codage (ou le degré de complexité de construction d'une donnée) en bits, unité de mesure de l'information, et l'entropie, avec une situation d'«objectivité parfaite» dans le cas de l'égalité. Un codage de 0,07 bits, par exemple, est ainsi prétendu plus conforme à la grille, donc d'une «objectivité» plus grande, qu'un codage de 0,17 bits.” (Fournier, 1984: 3.)

La méthode a donné naissance au modèle ANOVEP d'analyse sur document du vécu d'un programme afin d'en tirer des indicateurs de performance. Par la technique appropriée, à savoir l'indice MRC (Mesure de Rendement du Codage), le modèle détermine ainsi un biais au codage en bits, ce qui permet de doter chaque variable, à son entrée dans le fichier, d'un indice de pondé-

ration qui la suit pour le reste des manipulations qu'en commande l'analyse. On voit en quoi la méthode est prête à s'apparenter aux systèmes experts employant des coefficients de certitude ou de vraisemblance. Dans le cas d'ANOVEP, par exemple, le modèle procède au départ en permettant à l'analyste de créer des typologies selon la méthode traditionnelle du sociologue (Fournier, 1985a). Le programme de traitement des données emploie les fonctions SAS, et un utilitaire est prévu avec SYMPHONY de façon à pouvoir, dans une seconde phase de la recherche-action, en déplacer l'utilisation, actuellement dévolue aux chercheurs, vers les gestionnaires du Ministère et, sur le terrain, vers les participants au programme.

Sur le plan technique, l'analyse de contenu à l'intérieur d'ANOVEP procède, pour quantifier les unités de sens, à l'aide des deux premiers indices, C et Sk, décrits à la figure 1. Ce sont ces valeurs, basées sur les décomptes fréquentiels, que l'analyste met en rapport dans les typologies. Elles changent, entre autres, selon des seuils de "subjectivité" attribuables à l'interprétation au codage, seuils que représentent les valeurs du troisième indice, noté S à la figure 1. Ces seuils ou valeurs de S sont déterminés par la méthode de l'indice MRC (Fournier, 1985b) et choisis par l'analyste en fonction des besoins: sélectionner ou former des codeurs, analyser des données, appliquer des résultats, réviser le système de catégories, affiner ou enrichir le thésaurus, corriger le protocole d'accès au fichier, mettre à jour la base de connaissances, procéder à des études de réplique.

Un exemple d'utilisation des résultats peut être de demander au fichier OVEP de classer les organismes volontaires par régions et domaines d'intervention selon le militantisme de la demande, la nature de l'information constituant les contenus de formation, l'origine du besoin ou encore l'objet de la formation. On peut aussi assimiler à un résultat utile le fait de pouvoir connaître l'allure du programme si l'on met en rapport image des partenaires et activités de formation ou, d'autre part, création d'instruments pédagogiques et stratégies de développement, etc. Par ailleurs, la

mise à jour de la base de connaissances à partir du fichier OVEP est évidemment l'opération la plus importante du point de vue de la recherche-action car elle fait intervenir gestionnaires, chercheurs et clients du programme.

Ce dernier trait achève, pour l'essentiel, de décrire l'environnement technique et organisationnel où s'emploie notre application de l'analyse de contenu à l'information de gestion. Nous en arrivons ainsi à répondre à la question qui pourrait être ici l'objet premier de curiosité: quelle parenté y a-t-il entre cette application du modèle ANOVEP et le domaine de l'intelligence artificielle (IA)?

Nous répondrons qu'une telle parenté, si elle existe, est à prendre à partir de la notion de réseau sémantique, l'une des techniques utilisées en IA pour représenter des faits. Un fait, dans un réseau sémantique, c'est une relation avec ses propriétés. Or, la formule de construction $P(E) = 2^{Ei}$ de la grille de codage reproduite à la figure 1 se propose comme un tel support, obtenu par bijection (Fournier, 1983), d'une relation avec ses propriétés. Elle peut en tout cas se présenter comme un réseau sémantique, en l'occurrence un graphe étiqueté, dans une approche descriptive de la notion de réseau sémantique. Nous prendrons donc quelques instants pour d'abord préciser cette approche, et décrire ensuite la grille de codage comme un graphe étiqueté.

La notion de réseau sémantique employée pour représenter des connaissances n'est pas étrangère à ceux qu'intéresse l'informatique. Le fait qu'elle puisse servir, par exemple, à élaborer des systèmes (Ince, 1984) rappelle que l'une de ses plus anciennes formulations, en effet, est l'emploi du diagramme logique (Macquart, 1968) pour représenter la structure d'un programme. Pour évaluer une variable numérique, par exemple, l'ordinateur fait la liste, puis compare les valeurs (instructions IF et ELSE en PL/1 ou PL/M, instructions FOR et WHILE en Basic, etc.). Une telle approche de la représentation des connaissances peut être dite procédurale

(Lelouche, 1986; Harmon, 1985; Winston, 1979) par référence au fait qu'on doit lire le programme pour savoir ce qu'il accomplit.

Dans le calcul des prédicats ou l'approche par équivalence (Lelouche, 1986; Harmon, 1985), d'autre part, des règles classent des faits. L'approche est déclaratoire, dirions-nous à la suite de Lelouche (1986), dans la mesure où le programme nous donne sur demande ses règles et ses hypothèses (faits). Dans une démonstration de PROLOG II, par exemple, l'algorithme "X est le parent de Y si X est le père de Y ou si X est la mère de Y" parcourt un graphe en arbre généalogique. Le programme dispose d'une base relationnelle, en l'occurrence la liste de personnes en tête de page, que l'algorithme exploite en s'aidant de règles.

Finalement, un réseau pourra s'essayer à une description précise des liens en langue naturelle, lieu des représentations, entre des objets, des acteurs, des événements ou autant d'éléments de réalité acceptés par le discours comme perçus. Il s'agit donc d'une description acceptée et en ce sens dite précise: on convient des termes. C'est l'approche descriptive (Lelouche, 1986) formulée à la façon du sociologue. Un graphe étiqueté (par des valeurs) sur des noeuds et des arcs est un réseau de cette sorte lorsqu'il y a transitivité complète: c'est le cas de l'arbre logique constitué par héritage dans l'exemple de PROLOG II et le cas du graphe rencontré à la figure 1. Le noeud, en ce cas, est un concept et l'arc, la relation entre des concepts.

La formule de lecture d'un graphe étiqueté (Lelouche, 1986) s'écrit:

$$N \xrightarrow{a_1} N_1$$

En tant que graphe étiqueté, la figure 1 se lit donc ainsi:

1. La valeur d'un arc a_1 est son noeud extrémité. La valeur des noeuds est donnée par les indices C, Sk.

2. Le cas (*slot*) d'un noeud N est chacun des noms de l'arc, ici catégories ou classes, susceptible d'en partir. Les indicateurs servant au codage, *i.e.* les variables de dernier niveau, sont des cas (ou *slots*); chaque variable servant d'indicateur de la variable supérieure est un cas. La formule des cas (*slots*) est $P(E) = 2^{|E|}$.

La procédure pour trouver la valeur de Sk(N), *i.e.* la valeur numérique attachée au nom de chaque catégorie dans la grille, est celle de l'héritage par valeur directe: on calcule des trajets sur le graphe (Fournier, 1983). La sortie d'imprimante se présente sous la forme d'une liste de noeuds et des valeurs trouvées directement qui leur correspondent. L'ordinateur calcule là-dessus les indices pour une liste des variables par niveau. Le programme de calcul en SAS a été fait par Hélène Boutet depuis Fournier (1983; 1985b)

Il n'est pas exclu d'ajouter l'héritage par démons au calcul de Sk(N), *i.e.* un ou des sous-programmes au besoin. Ce serait le cas, ainsi, si je voulais reconstruire les catégories autrement (les recombinaison ou les combiner en d'autres sous-ensembles) en conservant les mêmes indicateurs au codage. Je partirais, en ce cas, d'informations existantes. La situation n'est pourtant pas si différente dans l'héritage par défaut, où je partirais, par hypothèse, de propriétés vraisemblables ou probables pour réviser les catégories.

On peut donc observer en terminant qu'il y a possibilité de combiner les trois types d'héritage, selon les besoins de l'analyste ou l'état de sa connaissance du domaine. Par exemple, ici, le démon (améliorer les catégories) vient après l'héritage par valeur directe (qui donne, entre autres, les premières typologies), et la procédure par défaut vient après le démon (une fois les catégories révisées pour de bon, etc.). La règle qui engendre alors les métarègles, comme elle le fait pour tout le reste de l'analyse, est celle de l'indice MRC, selon qu'il varie de 0 à 1.

On remarquera que la combinaison graduelle de trois procédures pour trouver la valeur de $Sk(N)$ nous donne ainsi le moyen d'explicitier à mesure les éléments pertinents de la représentation, ce qui, en soi, est un critère de représentation puisqu'il permet, par exemple, de rendre celle-ci plus complète.

En résumé, on aura finalement construit par réseaux sémantiques (tel le réseau en figure 1) la structure d'une base de données et d'une base de comportements. Les données, ce sont les valeurs des classes ou les faits; la base de comportement, les typologies obtenues par chaînage avant ou arrière à l'aide de l'indice MRC. Base de comportements et base de données constituent alors, en termes de représentation, une base de connaissances (BDC).

Quel jugement porter sur cette BDC? Dans quelle mesure est-elle apte à servir, hypothèse implicite de l'exposé, à illustrer ce qu'est la schématisation des connaissances en IA? Notre réponse, fort brève, tiendra lieu de conclusion à l'ensemble de ce travail.

CONCLUSION

Dans un article, Ganascia (1985: 1143) indique qu'il y a trois phases d'étude ou trois niveaux de connaissances supposés dans la construction d'un système expert.

Le premier est le niveau **structurant**, *i.e.* celui du mécanisme qui engendre la certitude et soutient, par exemple, le processus de validation des connaissances. On peut reconnaître ce niveau dans la technique de l'analyse de contenu et le programme SAS: quantification (graphes, trajets) et représentation par réseaux. Un second est appelé **conceptuel**: il renferme l'ensemble des concepts opératoires ou des "noeuds du raisonnement" (Ganascia, 1985: 1142) de l'expert. C'est ce que constituent, en somme, les théories

sociologiques qui président chez l'analyste aux choix et à l'étude des typologies. À ces concepts opératoires, ajoute l'auteur, on associe un descripteur élémentaire, qui s'exprime dans le formalisme de premier niveau. L'indice MRC peut jouer ce rôle sur les grilles, comme on s'est efforcé de le montrer. Finalement, un niveau **cognitif**, dans la schématisation propre à un système expert, comprend l'ensemble des connaissances mises en oeuvre par l'expert: il s'agit ici, par exemple, de la sociologie des associations ou des mouvements sociaux, de la sociologie de la connaissance qui a servi à l'élaboration du cadre conceptuel, de l'expérience du terrain, etc. Les connaissances de ce niveau doivent pouvoir se traduire dans le langage défini par les deux autres niveaux: le formalisme de représentation du niveau structurant (*i.e.* le programme SAS, fait pour ça, et la structure $P(E) = 2^{E^E}$), et les descripteurs existants du niveau conceptuel, *i.e.* les typologies.

Bien que l'application en question n'en soit qu'à ses débuts, on rencontrerait donc néanmoins, avec cet exemple concret de schématisation des connaissances dans la construction d'une grille en analyse de contenu, une application à l'information de gestion des trois stades de conception d'un système expert. C'est le jugement, du moins, que je soumets à votre attention en le prenant comme conclusion de cet exposé.

BIBLIOGRAPHIE

- BARBUT, Marc, *Mathématiques des sciences humaines*, tome 1, Paris, PUF, 1969.
- BARTON, Allen H., "The concept of property-space in social research", dans: Paul. A. LAZARSFELD et MORRIS ROSENBERG, *The Language of Social Research*, Glencoe, The Free Press, 1955: 41-54.

- BOREL, Marie-Jeanne, Jean-Blaise GRIZE et Denis MIÉVILLE, *Essai de logique naturelle*, Berne/Francfort-sur-Maine/New York, Peter Lang, 1983.
- BOUDON, Raymond et François BOURRICAUD, *Dictionnaire critique de la sociologie*, Paris, PUF, 1982.
- CARTWRIGHT, Dorwin, "L'Analyse du matériel qualitatif", dans: Léon FESTINGER et Daniel KATZ, *Les méthodes de recherche dans les sciences sociales*, tome 2, Paris, PUF, 1959: 481-538.
- FORTIN, Andrée, *Mode de connaissance et organisation sociale*, Montréal, (CIDAR), Département de sociologie, Université de Montréal, 1981.
- FOURNIER, Richard, *Analyse de contenu des demandes reçues au programme d'aide aux organismes volontaires d'éducation populaire (OVEP) en 1981-82 et en 1982-83*, 1. *L'Outil d'analyse*, Québec, Ministère de l'éducation du Québec, 1983.
- FOURNIER, Richard, *L'État interventionniste: l'évaluation qualitative. L'analyse de contenu au seuil d'une nouvelle pratique*, Québec, Université du Québec, Vice-présidence aux Communications, 1984. ("Études", 19.)
- FOURNIER, Richard, *Application du modèle ANOPEP d'analyse de contenu à l'évaluation de programmes*, Québec, Ministère de l'éducation du Québec, 1985a.
- FOURNIER, Richard, *Une mesure de rendement du codage en analyse de contenu: l'indice MRC*, Québec, Ministère de l'éducation du Québec, 1985b.

- GANASCIA, Jean-Gabriel, "La conception des systèmes experts", *La recherche*, vol. 16, n° 170, octobre 1985: 1142-1151.
- GRANGER, Gilles-Gaston, *Pensée formelle et sciences de l'homme*, Paris, Aubier-Montaigne, 1968.
- GRIZE, Jean-Blaise, *Sémiologie du raisonnement*, Berne/Francfort-sur-Maine/New York, Peter Lang, 1984.
- HARMON, Paul et David KING, *Expert Systems*, New York, Wiley, 1985.
- HOLSTI, Ole R., *Content Analysis for the Social Sciences and Humanities*, Reading (Mass.), Addison-Wesley, 1969.
- HURTUBISE, Rolland, *Informatique et information*, Montréal, Agence d'Arc, 1976.
- INCE, D.C., "A Source code control system based on semiotics nets", *Software Practice and Experience*, vol. 14, n° 12, décembre 1984: 1159-1168.
- LAND, Kenneth et Seymour SPILERMAN, *Social Indicators Models*, New York, Russel Sage Foundation, 1975.
- LELOUCHE, Ruddy, *Représentation des connaissances et systèmes experts*, Québec, Département d'informatique, Université Laval, 1986 (ronéotypé).
- MACQUART, Serge et Jean-Marie POULIN, *Éléments de langage P/11*, Québec, Centre de traitement de l'information, Université Laval, 1969.
- MIÉVILLE, Denis, "Analogie et exemple", dans: Marie-Jeanne BOREL, *op. cit.*: 149-225.

RIOU, Paul et Joachim DELORME, *Traité de chimie*, Montréal, Beauchemin, 1957.

ROUBINE, E., *Théorie de l'information*, Paris, Masson, 1970.

THUROW, Lester C., *The Zero-Sum Society*, New York, Penguin Books, 1981.

WARUSFEL, André, *Dictionnaire raisonné de mathématiques*, Paris, Seuil, 1966.

WINSTON, Patrick H., *Artificial Intelligence*, Reading (Mass.), Addison-Wesley, 1979.

WITTGENSTEIN, Ludwig, *Tractatus logico-philosophicus*, Paris, Gallimard, 1961.