

LES USAGES DE L'ANALYSE FACTORIELLE  
DANS LES REVUES DE SCIENCES SOCIALES

**Philippe BLANCHARD**  
**Charles PATOU**  
**CREDEP, Université Paris 9-Dauphine**

Comme l'a reconnu l'un de ses inventeurs et promoteurs (Benzécri, 1977), l'expression « analyse de données » était moins judicieuse que celle de « synthèse de données » pour désigner une méthode centrée sur l'analyse factorielle des correspondances, mais elle s'imposa par l'usage. Jean-Pierre Benzécri (1977) situe l'origine de cette démarche méthodologique dans un projet personnel de traduction automatique basé sur une compréhension inductive de la langue. Il s'agissait de définir un mot non plus par référence à son sens, mais simplement à l'univers de ses contextes lexicaux d'emploi réel. Au sein d'une matrice élémentaire des co-usages des noms (en lignes) et des verbes (en colonnes) de la langue française, un nom se trouve donc défini par le profil de ses fréquences de co-occurrence avec l'univers des verbes. Deux noms sont d'autant plus synonymes que leurs profils sont proches. L'analyse factorielle est la méthode qui permet de mesurer automatiquement la correspondance des noms entre eux, des verbes entre eux, et surtout des noms avec les verbes. Ce problème renvoyait au *multidimensional scaling* des psychologues américains, qui mesure le nombre de fois où un stimulus donné déclenche telle réaction. Benzécri appliquera cet algorithme pour la première fois sur le tableau des correspondances entre chinois et français.

C'est en 1963 que furent ainsi posés les jalons de l'analyse des correspondances. Elle allait devenir l'analyse des données « à la française », avec toutes les spécificités d'une école. Nous ne prétendons pas faire ici le bilan de trente ans d'existence de cette méthode, mais plutôt de passer en revue ses usages actuels. Une méthode existe à travers les applications qui en sont faites, et celles de l'analyse factorielle - envisagée ici plus largement que la seule analyse des correspondances - sont multiples.

Après un bref rappel de ce que permet l'analyse factorielle et des difficultés qu'elle pose, nous examinerons donc ce qu'en font réellement les spécialistes de sciences sociales et comment ils surmontent ces difficultés.

### *Qu'est-ce que l'analyse factorielle ?*

L'analyse factorielle est une méthode de traitement statistique visant à résumer de grands tableaux en quelques tendances synthétiques appelées facteurs (d'où l'adjectif « factorielle »). Cette méthode vise à simplifier et accélérer l'examen souvent fastidieux du détail des tris statistiques.

Une analyse factorielle peut être décrite en cinq points :

- Les données à traiter forment un tableau, soit de type individus-variables avec des données qualitatives ou quantitatives, soit de type variables-variables avec des données sous la forme de fréquences absolues, de fréquences relatives ou de coefficients de corrélation ;
- L'algorithme de traitement mathématique est basé sur le calcul vectoriel, la transformation des matrices et la géométrie multidimensionnelle, qui permet de repérer les proximités entre les lignes en fonction de leurs profils respectifs sur les différentes colonnes, et symétriquement entre colonnes en fonction de leurs profils sur les lignes. Cet algorithme fonctionne par itérations successives, chaque itération fournissant, en termes algébriques, une valeur propre qui représente un facteur explicatif. Ces facteurs sont synthétiques, c'est-à-dire qu'ils ne se matérialisent nulle part à l'état pur dans les données brutes, même s'ils sont susceptibles de se trouver après coup correspondre à des paramètres objectifs du phénomène étudié. La première itération explique le mieux possible l'information contenue dans le tableau initial (la variance du tableau), la seconde itération explique au mieux l'information résiduelle non expliquée par la première, et ainsi de suite jusqu'à épuisement de l'information contenue dans le tableau de départ ;
- Le résultat comprend en règle générale : la liste hiérarchisée des facteurs, la proportion de la variance initiale prise en charge par chaque facteur, la contribution de chacune des variables et de chaque individu à la construction de chaque axe (« inertie »), la contribution de chaque facteur à l'explication de la variance de chaque variable et de chaque individu (« contribution »), les coordonnées de chaque variable et de chaque individu sur chaque axe, les cartes factorielles en repère orthonormé à deux dimensions découlant des axes pris deux à deux, sur lesquelles sont placés les points-variables, les points-modalités et les points-individus ;
- Suivant les besoins, peuvent être ajoutées aux cartes factorielles des variables supplémentaires, qui ne participent pas au calcul des axes, mais se voient attribuer des coordonnées sur ceux-ci. Elles peuvent servir à expliquer les variables actives (ou être expliquées par elles), ou à

comparer le premier groupe d'individus avec un second, ou encore à tester les mêmes données privées d'individus en positions extrêmes qui sont susceptibles de déformer les résultats<sup>1</sup> ;

- Pour faciliter la lecture des cartes, divers affichages sélectifs sont envisageables : points les plus pertinents pour la problématique de départ, points dépassant un seuil donné d'inertie, nuages de points proches et leurs centres d'inertie, etc.

Toutes les procédures dénommées « analyse factorielle » ne comprennent pas tous ces éléments, du fait de la variété des données de départ, qui exigent des algorithmes mathématiques sensiblement différents. Nous distinguerons ici trois types<sup>2</sup> :

- l'analyse en composantes principales (en général abrégée « ACP »), adaptée aux variables quantitatives continues ;
- l'analyse des correspondances simple (« AFC »), appliquée à un tableau de contingence, ou de façon plus générale à tout tableau de nombres non négatifs, qui met l'accent sur la représentation graphique des résultats, et peut traiter des observations en petit nombre contrairement à la précédente méthode ;
- l'analyse des correspondances multiples (« ACM »), qui traite les données nominales à trois variables ou plus.

#### ***Une méthode délicate d'emploi mais séduisante***

L'analyse factorielle a la faiblesse de reposer sur des approximations. L'interprétation se limite souvent aux deux ou trois premiers facteurs. Or la littérature propose des critères divers de choix des facteurs pertinents, certains théoriques et d'autres empiriques. L'usage de ces critères varie d'un auteur à l'autre, sans toujours convaincre le lecteur.

D'autre part, les facteurs traités dégagent des tendances qui ne résument pas toute l'information du tableau. Certains aspects décisifs de la problématique peuvent être absents des premiers axes, masqués par des dimensions plus massives mais triviales ou déjà mises en évidence par une autre méthode. Les indicateurs de contribution doivent être maniés finement pour interpréter rigoureusement les axes retenus. Le danger guette de projeter des attentes sur des représentations graphiques apparemment transparentes.

En dépit de ces difficultés, l'analyse factorielle a deux grands atouts. Elle permet, suivant les cas, de gagner du temps, et/ou de mettre en évidence des phénomènes indistincts à l'œil nu. Mais son usage doit être tempéré si possible par des méthodes complémentaires. Pour P. Cibois (1990), elle peut constituer un échafaudage destiné à être démonté une fois que

l'édifice de la démonstration tient de lui-même. Elle est avant tout une méthode exploratoire globale, qui doit être complétée par des procédures confirmatoires plus précises, parfois de simples calculs, nécessaires pour prouver ce que l'analyse factorielle ébauche. L'analyse factorielle peut donc n'être qu'une étape insuffisante à elle seule car encore floue, dans le processus statistique.

A. Morineau (1993) démontre l'utilité des techniques complémentaires en reprenant une enquête menée en ex-Allemagne de l'Est en deux vagues peu après la chute du Mur de Berlin. Cette enquête voulait mesurer l'évolution des attitudes des citoyens et des consommateurs vis-à-vis de l'ouverture à l'Europe de l'ouest. Le questionnaire soumis à quelques milliers d'Allemands comprend une rubrique « signalétique » composée des variables socio-démographiques habituelles, une rubrique « politique » traitant de l'intégration et des valeurs.

L'auteur procède dans un premier temps à une analyse des correspondances multiples sur les questions de la rubrique « Politique ». Il en déduit six classes d'attitudes politiques nettement différenciées. Puis il utilise la méthode des valeurs-tests<sup>3</sup> pour caractériser finement, de manière exacte et non plus approchée, chaque classe du point de vue politique, et du point de vue de l'attitude face à l'ouverture. Les valeurs-tests appliquées dans un second temps aux variables socio-démographiques habituelles lui permettent de repérer dans chaque classe les critères qui déterminent les attentes et les craintes, l'optimisme ou le pessimisme vis-à-vis de l'ouverture. Enfin, une étude des mouvements entre classes d'attitudes d'une vague à l'autre du sondage lui permet de cerner l'évolution globale des attitudes de l'échantillon testé (et éventuellement d'en inférer l'évolution dans l'ensemble de la population ex-est-allemande).

Nous voyons ici l'intérêt de compléter l'analyse factorielle par des méthodes complémentaires. Citons parmi celles-ci :

- les classifications, par exemple la hiérarchique établie sur les bases des coordonnées calculées par analyse factorielle ; les classes peuvent être projetés sous forme de nuages sur les cartes factorielles ;
- l'analyse de similitude fournit des graphes des variables et de leurs liaisons ;
- l'analyse tabulaire (Cibois 1980, 2002) qui ajoute l'exactitude des calculs et de la représentation en aires rectangulaires à la lisibilité, favorisée par l'aspect parlant et hiérarchisé des cartes factorielles ;
- la construction de variables idéal-typiques (Cibois, 2002), pour résumer les variables et établir une typologie des individus ;

- l'analyse dite « tri-deux » développée par P. Cibois (1990) visant à ajouter sur les cartes factorielles les liens entre points, sous la forme de segments de droite dont l'épaisseur correspond à l'intensité du lien entre modalités ; c'est l'écart à l'indépendance entre deux modalités, calculé à partir du tableau croisé des deux variables, ou bien la contribution au khi-deux (écart à l'indépendance pondéré) qui mesure le lien ; on peut ainsi mettre en évidence l'intensité réelle des liens entre points, au-delà de la proximité spatiale apparente dans le plan factoriel 1-2.

L'analyse factorielle est une méthode puissante, globale et synthétique, donc séduisante. Elle donne l'impression de dispenser de l'examen minutieux des statistiques simples : tris simples, tris croisés, boîtes à moustaches, corrélations bivariées... Mais elle peut prendre un côté automatique et magique qui mène à l'erreur. Les progrès en termes de puissance de calcul informatique, d'ergonomie des logiciels et de qualité esthétique des résultats accroissent les risques d'usages abusifs. Analyse factorielle décorative, présentation superflue d'étapes secondaires de la recherche, usage non approprié au type de données, données trop complexes générant des cartes illisibles, résultats sans rapport avec la problématique<sup>4</sup> : autant de dérives qu'une bonne compréhension de la méthode permettrait d'éviter.

Cet aspect séduisant et parfois trompeur de l'analyse factorielle a-t-il limité son développement ? Ou au contraire l'analyse dite française des données s'est-elle imposée comme un canon méthodologique ? A-t-elle gagné au détriment de méthodes concurrentes, ou grâce à une collaboration avec elles ? Pourquoi a-t-elle connu le succès que l'on sait en France plus qu'ailleurs ? Nous ne répondrons naturellement pas intégralement à ces questions, tant les applications de l'analyse factorielle sont diverses et disséminées dans les divers champs de l'étude scientifique de l'homme et de la société (sociologie, science politique, psychologie, histoire, etc.) trente ans après les premiers travaux de J.-P. Benzécri. Notre enquête vise des questions plus modestes. L'analyse factorielle reste-t-elle l'apanage de spécialistes ou bien s'est-elle routinisée ? A quel type de données est-elle associée ? Quels objectifs lui sont assignés ? L'analyse factorielle intervient-elle comme méthode confirmatoire ou exploratoire ? principale ou complémentaire ? démonstrative ou illustrative ? Avec quelles autres méthodes l'analyse factorielle est-elle associée ou mise en concurrence ?

Nous verrons que ces usages témoignent d'une certaine routinisation de la méthode, mais que cette routinisation n'exclut pas une ambition didactique, comme si un écart subsistait entre des usagers familiers de l'analyse factorielle, ressentant un besoin de justification, et leurs

collègues ignorants de la chose et méfiants vis-à-vis d'elle. Puis nous montrerons comment la méthode trouve des affinités avec certaines disciplines, certains objets, certaines théories, certaines pratiques scientifiques. Ces affinités n'induisent pas nécessairement des usages non conformes de l'analyse factorielle, mais peuvent expliquer une diffusion différentielle de la méthode d'un champ intellectuel à l'autre, d'une école de pensée à l'autre.

Pour ce faire, nous dépouillerons exhaustivement trois années (1999, 2000, 2001) de neuf revues de sciences sociales et humaines choisies parmi les plus connues. Nous obtenons un échantillon de quatorze articles, répartis comme suit<sup>5</sup> :

- *Actes de la Recherche en Sciences Sociales – Mots, les langages du politique* : 4 articles par revue ;
- *Revue Française de Science Politique* : 3 ;
- *Année Psychologique – Politix – Revue Française de Sociologie* : 1 ;
- *Année Sociologique – Hermès – Quaderni* : 0.

Quatre questions s'avèrent particulièrement discriminantes pour juger des usages de l'analyse factorielle dans notre échantillon. Les premières seront l'occasion d'une exposition de la substance des articles un peu longue, mais nécessaire à la compréhension des usages concrets et divers de la méthode.

Ces questions nous permettront ensuite de distinguer des types d'usages.

#### ***Quatre critères de comparaison des usages***

*Premier critère : quels objectifs sont assignés à l'analyse factorielle ?*

Un premier groupe d'articles, très majoritaire, vise à marquer des distinctions au sein d'une population donnée : groupes sociaux, pratiques professionnelles, organisations. L'analyse factorielle est ici tout à fait conforme à sa fonction de mise en évidence de facteurs cachés et de regroupement d'éléments *a priori* dispersés aléatoirement.

Dans leur étude sur les usages de l'ordinateur au travail, M. Gollac et F. Kramarz (2000)<sup>6</sup> cherchent par exemple comment se différencient les utilisateurs d'informatique en entreprise à partir de caractères socioprofessionnels classiques (diplômes, ancienneté...) et de la perception que ces usagers ont de leur pratique. Fait-on de l'informatique par plaisir, par goût personnel, ou bien en fonction des clivages sociaux liés aux capitaux économiques et culturels ? Et à quelles conditions y trouve-t-on un bénéfice professionnel ? Toujours dans le cadre de la sociologie du travail, O. Godechot (2000) enquête sur l'espace de

concurrence formé par les salles de marché. Il souhaite mettre en évidence des catégories de *traders*. En fonction d'une batterie d'indicateurs socioprofessionnels, il montre comment ces acteurs gèrent leurs ressources et leurs compétences pour trouver la stratégie gagnante. A rebours de l'illusion du comportement rationnel de l'*homo economicus*, on s'aperçoit que les capitaux économiques et symboliques sont ici aussi déterminants.

Nous discernons les mêmes intentions de balisage d'une population dans trois autres articles, portant cette fois sur le domaine éditorial. Chez P. Bourdieu (1999), il s'agit de « construire l'espace des éditeurs » à partir de leurs caractéristiques en termes de capital économique (positionnement sur le marché éditorial...), de capital culturel (études du directeur, gestion familiale ou pas, sexe du directeur...) et capital social (emplacement géographique du siège social, dépendance vis-à-vis d'autres sociétés...). O. Godechot (1999) décrit la structure du champ des éditeurs de livres philosophiques en fonction des caractéristiques des auteurs : à un type d'auteur correspond un type de maison d'édition. Chez J. Duval (2000), l'analyse factorielle vise à décrire l'espace des journaux traitant d'économie en fonction de multiples critères économiques (composition du capital du journal, audience du média, cote sur le marché publicitaire) et journalistiques (taille de la rédaction, volume de reprises, diplômes des rédacteurs), afin de savoir s'il y a dépendance envers l'économie, ou bien si les critères professionnels garantissent une certaine autonomie aux journalistes qui se revendiquent de gauche.

Chez B. Dolez et A. Laurent (2000) et J. Chiche et aut. (2000), il s'agit également de différencier des populations, mais dans le domaine politique. Examinant les résultats de la première élection du président par tous les militants au Rassemblement pour la République (RPR) qui eut lieu en 1999, Dolez et Laurent répertorient les clivages idéologiques portés par les quatre candidats en présence. La procédure électorale, strictement calquée sur celle de l'élection à la Présidence de la République, fournit des scores par départements. L'analyse factorielle permet de classer ces unités géographiques élémentaires suivant les proportions des quatre votes, et d'en analyser les raisons en termes d'idéologies, de personnalités implantées localement, etc. A l'échelle politique réelle et non plus partisane, les élections législatives de 1997 sont l'occasion pour Chiche et aut. de repérer les évolutions idéologiques en cours. Un questionnaire détaillé, passé en sortie des urnes, permet de construire des indicateurs d'ethnocentrisme, d'attachement à l'économie ou au social, de libéralisme, etc. Les auteurs cherchent si un nouveau clivage est apparu, en plus ou en lieu et place de l'opposition entre droite

et gauche. Dans l'affirmative, si les électorats ont évolué, les offres politiques sont-elles toujours à même de le satisfaire ?

Le second type d'objectif assigné à l'analyse factorielle est également de nature classificatoire, mais procède différemment. Il consiste à expliquer une opposition déjà pressentie, voire déjà démontrée, en approfondissant le rapport entre les parties de la population entre lesquelles s'opère la distinction. Par exemple, S. Gojard (2000) se demande si les règles auxquelles se conforment les mères concernant l'alimentation de leurs enfants en bas-âge sont socialement déterminées. Si oui, y a-t-il diffusion de ces règles des catégories favorisées vers les catégories populaires, ce qui ferait des usages de ces dernières des sous-produits de ceux des précédentes ? Ou bien les deux modes d'alimentation sont-ils autonomes ? Il s'agit ici de savoir si les deux types de pratiques sont qualitativement différents, ou bien si l'un découle de l'autre. Similairement, S. Chollet et V. Duval (2000), dans leur étude sur l'évaluation gustative des vins, comparent un groupe d'experts en œnologie avec un groupe de novices. Ils traitent par analyse factorielle le corpus des mots utilisés par chaque groupe pour désigner les vins, et compare la typologie des vins établie par les experts avec celle des profanes. Il s'agit donc de mesurer la différence entre deux groupes définis *a priori*, expérimentalement, et non de définir ces groupes par l'analyse des données. Mais l'intention est identique à celle de Gojard en ce qu'on cherche si l'une des pratiques approche l'autre, ou bien si elles sont irréductibles.

Un troisième type d'objectif assigné à l'analyse factorielle est de classer les individus et de les situer dans le temps. Traitant du vocabulaire des déclarations gouvernementales belges depuis 1945, J.-C. Deroubaix (2000) recherche ainsi des césures historiques, des ruptures dans le rythmes de renouvellement du vocabulaire. De telles césures étant démontrées, il cherche si elles tiennent au clivage politique traditionnel entre gauche et droite, ou bien à des oppositions extra-politiques : thèmes dominant l'actualité au moment du changement de gouvernement, conjoncture économique, etc. La démarche est similaire chez P. Marchand et L. Monnoyer-Smith (2000) et R. Alvarez et aut. (2000), qui étudient respectivement les déclarations de politique générale françaises et espagnoles. Pour S. Bonnafous et M. Tournier (2001), il s'agit de classer des genres de paroles télévisées (interview, déclaration, prestation de campagne, débat) en fonction de leurs lexiques. Sans chercher à périodiser le corpus étudié, l'auteur met en évidence grâce à l'analyse

factorielle l'importance du critère chronologique pour distinguer les textes.

*Second critère de distinction des usages : comment sont préparées les données traitées par analyse factorielle ?*

A la diversité des objectifs de l'analyse factorielle fait écho celle des données traitées par elle. Il peut s'agir d'individus physiques : des utilisateurs d'informatique en entreprise pour Gollac et Kramarz, des *traders* pour Godechot (2000), des auteurs de philosophie chez Godechot (1999), des électeurs pour Chiche et aut., des mères d'enfants en bas-âge pour Gojard. Des textes sont aussi fréquemment analysés à travers les formes lexicales qui les constituent : descriptions des vins par les goûteurs pour Chollet et Duval, déclarations de politique gouvernementale chez Deroubaix, M et Alvarez et aut., éditoriaux syndicaux chez M. Brugidou (2000), genres relevant de diverses situations de communication chez Bonnafous et Tournier. Enfin l'analyse factorielle traite d'acteurs d'organisations : des périodiques économiques chez Duval, des éditeurs littéraires chez Bourdieu. Le cas de Dolez et Laurent est atypique en ce qu'il traite de l'agrégation géographique (départementale) d'électeurs.

S'il n'y a pas focalisation sur un type d'individus statistiques, il y a encore moins focalisation sur une taille de corpus donnée ou un type d'analyse factorielle. Les variables sont toujours de l'ordre de la dizaine, mais le nombre d'individus varie de quelques dizaines à quelques milliers. Les corpus de grande taille peuvent être traités par ACP (tableau des fréquences à partir de quelques milliers<sup>7</sup> de formes lexicales sur 7 à 38 textes chez Alvarez et aut., Deroubaix, Marchand et Monnoyer-Smith), par AFC (sans précision quantitative chez Bonnafous et Tournier) ou par ACM (questionnaires à réponses multiples passés auprès de 3000 électeurs chez Chiche et aut., auprès de 1800 mères chez Gojard). Les corpus de taille moyenne sont traités par ACP (300 éditoriaux chez Brugidou, 100 départements chez Dolez et Laurent) ou par ACM (60 à 500 entretiens ou questionnaires chez Gollac et Kramarz, Godechot 1999, Godechot 2000). Enfin, les corpus de petite taille sont traités en AFC (40 termes désignant 13 vins chez Chollet et Duval) ou ACM (40 médias chez Duval, 60 éditeurs chez Bourdieu).

Comme on le voit, le type d'analyse factorielle dépend du type de données à traiter, mais les auteurs conservent une certaine liberté. Ainsi, Chollet et Duval choisissent l'AFC au lieu de l'ACP, mais au prix d'une

transformation de la matrice : les fréquences des termes utilisés par les goûteurs pour désigner les vins sont converties en un rapport de 0 à 1. La diversité en nature et en nombre des *inputs* de l'analyse factorielle est encore renforcée par le degré de représentativité du corpus. Un tiers des articles traite de l'ensemble de la population : les suites de discours de politique générale et les départements participant à l'élection au RPR. Les autres études procèdent toutes à un échantillonnage plus ou moins justifié, plus ou moins représentatif, vis-à-vis de la population sur laquelle portent les hypothèses. Toutes les enquêtes par questionnaires, par entretiens, par codage des caractéristiques des individus ou des organisations, ou par plan expérimental sur un panel de cobayes, ressortent à cette procédure de sélection raisonnée d'individus jugés pertinents.

*Troisième critère : comment les données sont-elles traitées ?*

Ce troisième point est décisif pour notre étude. L'analyse factorielle est par excellence une méthode de traitement synthétique, globale, qui donne en quelques manipulations un résultat. Seulement, l'ensemble des opérations de préparation et de traitement des données répond à des procédures strictes, que le sociologue ne peut se dispenser de prendre en compte pour produire une interprétation juste des résultats.

L'interprétation doit intégrer une première dimension du traitement : la construction des variables. De ce point de vue, notre corpus met en évidence deux groupes d'articles. Dans le premier, le travail de construction des données est limité. C'est bien sûr le cas de l'étude du vote au sein du RPR chez Dolez et Laurent. Les résultats électoraux par départements sont traités à l'état brut, le choix de l'électeur est considéré comme un fait en soi, non modulable. Comme le veut la théorie juridique, le citoyen souverain (ici le militant RPR) vote en son âme et conscience sans en rendre compte à quiconque. Les études portant sur des textes relèvent du même traitement brut des données. Les déclarations de politique générale (chez Deroubaix, Alvarez et aut.), les éditoriaux syndicaux (chez Brugidou), les transcriptions d'émissions télévisées (chez Bonnafous et Tournier) sont traités comme des objets mathématiques simples, valant en tant que tels. Tous ces textes, pourtant hétérogènes, se réduisent du point de vue du traitement statistique à des ensembles de formes lexicales, parfaitement définies comme combinaisons uniques de signes alphabétiques. Ce n'est que lors de l'interprétation que l'auteur introduit des données étrangères au texte lui-même, données historiques, politiques, sociales, etc. De la même

manière, en aval de l'analyse factorielle sur les résultats des élections au RPR, Dolez et Laurent opèrent un travail d'interprétation important en fonction des programmes et des personnalités des 4 candidats, de la conjoncture politique, de la géographie électorale nationale, etc.

L'étude de Marchand et Monnoyer-Smith tranche avec les précédentes. La construction des variables s'affine en amont de l'analyse factorielle. Au moyen de trois logiciels spécifiques, les textes sont soumis à trois analyses complémentaires les unes des autres : une analyse lexicale, qui fournit en *inputs* à l'analyse factorielle deux grandes classes de mots, l'une abstraite et tournée vers le futur, l'autre concret et technique et faisant un bilan du passé ; une analyse logico-syntaxique, qui fournit une cinquantaine des catégories grammaticales de mots (4 catégories de verbes, 9 catégories de joncteurs, etc.) ; enfin une analyse sémantique, qui regroupe les noms communs et propres dans des réseaux sémantiques. Ce triple décryptage permet de caractériser plus finement les textes. Mais il signifie aussi que les auteurs projettent sur les textes une théorie du langage, projection que l'interprétation devra intégrer. La complexité de la préparation des données contraint la suite de la procédure.

D'autres études procèdent à une construction élaborée des données. Toutes les variables à choix multiples destinées à être traités par ACM (Bourdieu, Chiche et aut., Duval, Godechot 1999, Godechot 2000, Gojard, Gollac et Kramarz) subissent un travail de construction du questionnaire, de codage et recodage des réponses, de fixation de seuils pour les variables ordonnées ou d'intervalle, d'homogénéisation et de pondération des variables hétérogènes. Dans son étude des périodiques économiques, après une étape de choix des publications et des variables à intégrer, Duval homogénéise des variables aussi diverses que la composition des rédactions, la quantité de publicité dans les périodiques, leur structure financière, ou le nombre de lecteurs. Dans son enquête sur les mères, Gojard sélectionne des variables sociologiques, démographiques, professionnelles et éducatives.

Seule étude expérimentale de notre échantillon, la comparaison des goûteurs de vin experts et profanes menée par Chollet et Duval repose naturellement sur un protocole construit par les auteurs. L'artefact majeur se situe au niveau de la sélection des qualificatifs pertinents parmi le corpus des mots recensés dans les descriptions libres (question ouverte) des vins par les sujets. L'interprétation de l'analyse factorielle dépend de cette sélection, au détriment des mots jugés moins significatifs par les auteurs, et au détriment d'autres dimensions linguistiques et rhétoriques qui eussent pu être traités au même titre : enchaînements syntaxiques, réseaux sémantiques... Ici encore, les résultats ne pourront

être interprétés, généralisés, universalisés que sous la contrainte des choix de construction des variables.

Le traitement des données intègre une seconde dimension : la place de l'analyse factorielle en tant que telle dans le processus de recherche, et son doublement, ou non, par d'autres méthodes statistiques. De ce point de vue, trois groupes se distinguent.

L'analyse factorielle est exclusive de toute autre méthode statistique pour six de nos articles. Chez Duval, Godechot (1999), Gollac et Kramarz, Godechot (2000), tout l'effort est concentré sur une ACM de grande taille, débouchant sur des cartes factorielles à très nombreuses modalités (ex. : les professions, les diplômes, l'usage de l'informatique pour Gollac et Kramarz, le revenu et le type de méthode que les *traders* utilisent pour investir pour Godechot 2000). Puis éventuellement, des cartes factorielles des individus sont tracées, des variables supplémentaires projetées. Par exemple, chez Duval, les cartes des sous-rubriques au sein des pages économiques confirment l'orientation idéologique des rédactions, suivant qu'elles restent strictement économiques, ou bien traduisent des préoccupations sociales. L'interprétation fait intervenir une enquête complémentaire par entretiens approfondis chez Gollac et Kramarz, mais dans toutes ces études, l'essentiel du commentaire est consacré à l'explicitation de l'espace des individus construit par l'analyse factorielle.

L'article de Chollet et Duval fait la transition avec le deuxième groupe, où l'analyse factorielle est centrale mais non exclusive. L'étude sur les goûteurs de vin de Chollet et Duval comprend trois moments de recueil d'information, expériences ou entretiens. L'analyse factorielle est utilisée à l'issue de la seconde, et est la seule procédure statistique lourde. Elle est accompagnée de quelques statistiques de corrélation et de chi-deux visant à comparer les experts et les profanes d'une part, les goûteurs avec les caractéristiques physiques objectives des vins de l'autre.

Dans ce second groupe, où l'analyse factorielle est complétée par des statistiques élémentaires et des méthodes post-factorielles, se trouve l'étude de Chiche et aut. Les cartes factorielles, construites sur la base des réponses au questionnaire, comprennent aussi les quatre nuages des individus suivant leur vote pour l'un des grands partis, avec les points moyens de ces nuages. Une analyse de la variance interne et externe des nuages permet d'évaluer la capacité explicative du vote par rapport à la diversité des opinions politiques. De la même manière, l'ACP de Dolez et Laurent sur le premier tour des élections du RPR est précédée des tris simples par modalités, d'une analyse de la variance et d'une analyse de la répartition géographique des votes. Toutes ces statistiques combinées

permettent de dégager des explications de ce vote et du report d'un tour à l'autre. Enfin, nous ne revenons pas sur l'étude des déclarations de politique générale de Marchand et Monnoyer-Smith, dont nous avons vu que les trois analyses factorielles suivaient autant d'analyses de statistique linguistique.

Dans un troisième groupe, l'analyse factorielle remplit une fonction exploratoire, parfois mineure, et est complétée par des méthodes statistiques confirmatoires parfois élaborées. Chez Gojard, l'exploration de l'espace des pratiques d'élevage est complétée par des retours aux tris croisés. Elle est suivie de trois régressions logistiques à vocation inférentielle, qui visent à distinguer l'influence de la situation sociale, de l'origine, de l'expérience de la mère, etc. sur le rapport des mères à l'allaitement. Les analyses factorielles sont soit précédées (Brugidou), soit suivies (Alvarez et al., Bourdieu, Deroubaix) par une classification ascendante hiérarchique : les deux méthodes fournissent une redondance idéale pour certifier la qualité d'une typologie. Les études de textes (Alvarez et al., Bonnafous et Tournier) procèdent aussi à des mesures de richesse lexicale.

L'étude de Deroubaix est spécifique par son ambition et sa rigueur méthodologique. Elle propose une méthode générale de périodisation d'un corpus de textes se succédant dans le temps. Le plan des facteurs 1 et 2 sert à mettre en évidence une courbe en forme de S couché, le long de laquelle les 38 déclarations belges de politique générale s'alignent. Cette suite remarquable de points est ajustée par la méthode des moindres carrés. L'aspect chronologique est approfondi par une classification ascendante hiérarchique sur le lexique. Six classes lexicales se dégagent, qui s'avèrent se succéder dans le temps, parfois en se chevauchant, sous forme de vagues. L'auteur propose alors une méthode de délimitation rigoureuse de ces « vagues » de vocabulaires spécialisés. L'analyse factorielle joue ici un rôle essentiel, mais introductif à une méthode plus élaborée de traitement d'un certain type de corpus textuel, les corpus chronologiques.

La place prise dans le protocole de recherche global s'avère donc très discriminante parmi les textes de notre échantillon : l'analyse factorielle peut se suffire à elle-même ou être complétée d'une multitude de procédures plus ou moins élaborées.

*Quatrième critère : la méthode est-elle discutée en tant que telle ?*

L'analyse factorielle a-t-elle atteint un stade de routine où les évolutions se font sous forme d'améliorations à la marge ? Ou bien conserve-t-elle un caractère de nouveauté, ce qui nécessiterait, en plus de l'énoncé des

paramètres utilisés et des résultats de l'analyse, une explication purement méthodologique, voire un travail pédagogique ? Ces questions sont essentielles pour cerner la méthode dans son évolution scientifique.

Un premier groupe d'articles consacre peu de place à la méthode elle-même, se limitant aux précisions minimales sur les paramètres de l'analyse factorielle dans le corps du texte, ou en notes de bas de page. C'est le cas chez Alvarez et aut., Brugidou, Deroubaix, Marchand et Monnoyer-Smith : ces auteurs tiennent l'analyse factorielle pour connue. Elle se réduit soit à un rapide préalable à un travail de périodisation, ou bien à un débouché pour des traitements lexicométriques plus complexes et novateurs. Les *outputs* statistiques sont limités aux cartes et aux pourcentages de variance expliquée.

L'explication méthodologique est réduite chez Gojard aussi, en dépit de données plus complexes : les sources des données sont en notes, les précisions techniques dans le texte. Mais les annexes complètent l'article : les variables y sont décrites avec leurs fréquences, et deux cartes reconduisent la même analyse factorielle sur deux sous-populations afin de vérifier l'influence éventuelle d'un paramètre non traité dans le corps du texte. En revanche, Godechot (2000) présente une carte factorielle peu détaillée, sans préciser la distinction entre gros et petits points. Le commentaire ne mentionne ni les poids ni les contributions des axes, et aucune annexe ne vient clarifier la méthode. Cette sobriété déçoit au vu de la complexité de la problématique et des données mobilisées.

Les autres articles détaillent plus la méthode. Ils recourent soit à des encarts, soit à des annexes conséquentes. L'article de Chiche et aut. comprend un encart pour chaque étape de l'ACM, et une explication détaillée des procédures de construction des cartes, tandis que le questionnaire est intégralement reproduit en annexe. Duval, Bourdieu, Gollac et Kramarz présentent tous trois un long encart sur la population étudiée, la construction des indicateurs, en renvoyant à d'autres ouvrages pour la méthode elle-même. Chollet et Duval insère des précisions méthodologiques sous les cartes, Godechot (1999) sous forme de petits encarts et de légendes sous les cartes.

Au-delà de la précision méthodologique, le lecteur bénéficie parfois d'une intention didactique. Chiche et aut., par exemple, souhaitent promouvoir une « analyse géométrique des données ». Ils mettent l'accent sur les cartes, utilisent symboles et couleurs pour faciliter la lecture. Ils justifient l'usage des taux d'inertie modifiés préconisés par Benzécri, exposent les fondements mathématiques de l'ACM par petits encarts, et mettent à disposition les logiciels utilisés. On observe également une présentation assez didactique, bien que succincte, de ce

qu'est l'analyse factorielle chez Bonnafous et Tournier. Duval, Gollac et Kramarz proposent des encarts détaillés, quoique sans excès didactique. L'essentiel de l'article, comme chez Duval, est centré sur la description de la carte.

En revanche, la volonté pédagogique est faible ou nulle dans les autres articles. Ceci va plus de soi lorsque cette méthode est mineure dans la démonstration (chez Dolez et Laurent, Marchand et Monnoyer-Smith) que lorsqu'elle est centrale (Deroubaix, Gojard, Alvarez et aut., , Godechot 2000, Godechot 1999). Ces derniers auteurs ne ressentent pas la même nécessité d'expliquer leur démarche, comme s'ils ne s'adressaient pas au même public, ou comme s'ils ne faisaient pas le même constat sur l'état de la diffusion des méthodes.

Notre seconde partie entend préciser ces hypothèses sur les déterminants des types d'usages.

### *Usages synthétiques et analytiques des méthodes de factorisation*

Nous avons constaté à plusieurs reprises que les articles publiés dans un même titre présentent des airs de famille dans leurs usages de la méthode. Ces ressemblances méritent d'être qualifiées et mesurées. Plus encore, il semble intéressant de savoir si des articles proches entre eux dans la façon d'utiliser la méthode prennent en charge de façon similaire, d'une part l'approximation spécifique aux méthodes de factorisation, et d'autre part l'alternative entre analyse et synthèse mentionnée en introduction.

Pour mettre en abyme notre propos, nous cherchons en fait si un facteur caché explique les affinités entre articles. Deux pôles émergent, entre lesquels nos quatorze textes se répartiront. Ce classement des textes recouvre pour partie leur appartenance à une discipline, et plus encore leur diffusion dans une revue. Au-delà de la seule homogénéité entre articles d'un même titre, cette correspondance entre usages de l'analyse factorielle et appartenance disciplinaire indique que la construction d'une analyse factorielle n'est pas seulement dépendante des données traitées et des objectifs scientifiques, mais aussi des attentes des auteurs envers leurs outils. Posée en termes plus opératoires, cette correspondance recouvre une ambivalence de la méthode : s'agit-il d'une *analyse* de données ou d'une *synthèse* de données ? D'autres questions plus précises peuvent nous amener à apporter des éléments de réponse à cette question. Quelle place l'auteur accorde-t-il aux résultats ? L'analyse est-elle exploratoire ou confirmatoire ? Ses résultats sont-ils repris, commentés, mis à l'épreuve par d'autres données ou d'autres méthodes ? Sont-ils abandonnés si leur faiblesse est démontrée ? En bref : quelle est

la situation de l'analyse factorielle dans le processus scientifique, de l'intention à l'utilisation, en passant par le moment du calcul ?

Ces quelques éléments transversaux aux disciplines voudraient aider l'utilisateur des analyses de données à clarifier sa position par rapport à son objet et à utiliser au mieux, et quand cela est utile, l'analyse factorielle.

*Premier groupe d'articles : un usage qualificatif et synthétique des méthodes de factorisation*

A l'évidence, les articles publiés dans les *Actes de la Recherche en Sciences Sociales (ARSS)* montrent des usages extrêmement ressemblants de l'analyse de données. De plus, les contributeurs de cette revue utilisent plus que d'autres cette méthode : quatre de nos quatorze articles en sont issus, à égalité avec *Mots, les langages du politique*. Ce n'est sans doute pas un hasard car, comme la revue de P. Bourdieu, celle du laboratoire de lexicologie politique de Fontenay-Saint Cloud est également très homogène dans ses usages de la factorisation.

Concernant les articles tirés des *ARSS*, notons également que deux articles sont issus d'un numéro double portant sur « les Editeurs et les Editions » (Bourdieu, Godechot 1999). Cette proximité thématique des articles concourt peut-être à faire converger les usages de l'analyse factorielle. Il reste que ces quatre articles présentent une forte homogénéité sur des caractéristiques qui sont étrangères aux objets d'intérêt. Un certain contexte scientifique génère donc un certain usage, standardisé, de la méthode.

En fait de méthodes de factorisation, les contributeurs des *ARSS* n'utilisent qu'un type de factorisation : l'analyse des correspondances multiples (ACM). Celle-ci porte, on l'a vu, non pas sur des notes attribuées à des individus, ni sur des fréquences d'individus, mais sur les modalités des réponses apportées à un questionnaire. Les variables qualitatives et/ou quantitatives sont réduites à un codage catégoriel, qui ignore la continuité et l'ordonnement des modalités.

Dans les quatre articles issus des *ARSS*, le travail de construction des variables est très détaillé et décrit d'ailleurs par les auteurs comme un élément essentiel de leur démonstration. Ce travail de construction est plus important dans ces articles que dans d'autres. Cette focalisation sur

la construction des variables s'accompagne de précisions méthodologiques assez volumineuses.

La présentation des passages méthodologiques prend dans les *ARSS* une forme standard. Charte éditoriale, tradition disciplinaire, culture scientifique partagée ou séminaires de travail communs concourent à ce que le travail statistique soit présenté sous forme de longs encarts, et quasiment absent du corps du texte. Cette présentation se fait en deux temps. De manière parfaitement orthodoxe pour une revue scientifique, les auteurs évoquent d'abord l'objet (journalisme, édition, informatique) dans les termes de la discipline et balisent le champ d'étude avec des critères spécifiques. Ensuite pour chacun des objets (éditeurs littéraires, éditeurs de livres philosophiques, journaux économiques, pratiques informatiques) les auteurs énoncent les critères sur lesquels portera l'analyse et proposent des indicateurs correspondant à ces critères. Par exemple, le capital social dans le milieu littéraire et philosophique est mesuré chez Bourdieu par la localisation des maisons d'éditions ; la proportion de journalistes issus des grandes écoles et de l'Université mesure chez Duval le capital culturel des rédaction économiques. Ces données sont ensuite intégrées à la factorisation sous forme de variables discrètes.

La création des modalités discrètes des variables est sans conteste l'étape méthodologie essentielle des analyses de correspondances multiples. Si ce processus est peu discuté méthodologiquement par les contributeurs des *ARSS*, ils est toutefois présenté. Les auteurs usent de diverses techniques, entretiens, questionnaires ou données provenant de sources documentaires jugées objectives pour recueillir le plus possible de données exploitables. Des critères de différentes natures comme la catégorie socioprofessionnelle, le capital financier, le sexe, l'âge, la localisation, l'habileté, le résultat, la pratique ou la formation deviennent des critères objectifs d'analyse d'une population. La multidimensionalité des critères retenus n'est sans doute pas exclusive des articles des *ARSS*, elle est en tout cas exemplaire. A ce titre, l'article publié dans une autre revue qui se rapproche le plus de cet usage de la factorisation est celui de Godechot dans *Politix*. Il est aussi un des contributeurs de la revue *ARSS*. Gojard (*Revue Française de Sociologie*), elle aussi mêle dans son enquête des variables très diverses, sociologiques, démographiques, professionnelles, familiales, etc.

Tout fonctionne comme si, pour ces auteurs, il s'agissait de prendre en compte un grand nombre de variables, quelles soient homogènes ou pas. Nous pouvons voir une homologie entre cette méthode et l'objet, à savoir le matériau sociologique lui aussi multidimensionnel et hétérogène. La

société n'est-elle pas de façon primordiale diverse, voire confuse ? L'analyse factorielle elle-même n'est-elle pas particulièrement adaptée à faire émerger, en conformité avec la tradition marxiste, les déterminants cachés de la société ?

Quoiqu'il en soit, les auteurs des ARSS, mais aussi Godechot (2000) et Gojard<sup>8</sup> se distinguent par l'emploi d'un nombre important de variables et de modalités, variant de 30 à 50, pour des corpus plutôt faible, allant de 40 à 135 individus. Le ratio nombre d'observations/nombre de variables pour chacun des auteurs ne vise clairement pas à assurer aux résultats la meilleure représentativité, mais plutôt la meilleure « étendue » au sens d'une prise en compte de nombreux critères.

Les objectifs de l'analyse des correspondances, tels qu'ils sont exposés par les auteurs, sont exemplaires de cette tendance. Ils cherchent tous à dégager la « structure des espaces » où des populations (des éditeurs, des journalistes, des utilisateurs de l'informatique, des éditeurs de philosophie) sont réparties et qu'elles polarisent. Ainsi, Bourdieu cherche à « construire l'espace des éditeurs », Godechot veut « préciser les liens entre la structure du champ éditorial et celle du champ des producteurs philosophiques » et Duval « dégager la structure du sous espace du journalisme économique ». Dans la présentation que font les auteurs de leurs objets et de leurs méthodes, la notion de « *champ* » est omniprésente, simple synonyme d'espace isotopique comme un champ agricole, ou peut-être d'espace structuré, polarisé, comme un champ magnétique. Les auteurs utilisent tous l'analyse des correspondances pour « rendre compte », « représenter » ou « synthétiser » des asymétries spécifiques du champ concerné. Dans le langage de Bourdieu, les asymétries tiennent aux différences de « capital », ce que mesurent les variables. Les individus occupent des « places » dans l'espace qu'ils décrivent (Gollac et Kramarz, p. 11).

En ce sens, les usages de la factorisation seraient homologues à la notion centrale de l'œuvre de Bourdieu, ou en tout cas lui correspondraient parfaitement. Discipline et méthode se correspondent, la seconde traduisant, ou, dans le vocabulaire des auteurs, « représentant » la première. En effet, il s'agit pour chacun d'entre eux de représenter un espace en lui-même et non pas seulement, comme c'est l'objectif canonique de l'analyse factorielle, de représenter une population dans un espace. La métaphore spatiale est dans ces articles largement filée, avec les notions d'espace, de champ, d'univers. Dans ce contexte, le recours aux « oppositions » entre groupe d'individus ou de variables apparaît comme une traduction spatiale des asymétries de capital au sein d'un champ. La carte factorielle est la représentation providentielle du champ

étudié. Elle révèle par l'opposition des groupes sur les axes les asymétries de capital des populations. A ce titre, les cartes occupent dans les *ARSS* une place prépondérante, sur deux pleines pages souvent et en couleurs.

La place accordée aux cartes et l'absence de tests confirmatoires est ici caractéristique. La logique est descriptive et les résultats sont tenus pour définitifs. Il faut noter à ce sujet que la logique descriptive qui vise à représenter, préciser, dégager se distingue d'une logique exploratoire. En effet, aucun des auteurs n'infirme ses hypothèses. Pourtant, ce sont des hypothèses ambitieuses qui sont examinées. Il s'agit de mesurer les asymétries dans la répartition d'un « capital » au sein d'une population. En ce sens, l'analyse de données n'a pas pour objectif d'analyser une population, mais de la qualifier. Dans leurs articles, les contributeurs des *ARSS* distinguent, avant même l'analyse de données, des types, des groupes entre les salariés ayant un bagage scolaire important et ceux ayant été formés sur le tas (pour les usages de l'informatique), entre les éditeurs qui publient des Nobel et ceux qui publient des traductions (pour les éditeurs littéraires), etc. L'analyse de données repère ces groupes. Sa plus-value est ici de pouvoir les qualifier plus précisément qu'ils ne l'auraient été dans une simple description. Les asymétries sont définies plus systématiquement, elles ne sont plus fondées sur un ou deux critères, mais sur une multiplicité, et l'analyse de données permet de voir comment les asymétries se cumulent. Ceux qui sont en haut de l'axe pour tel critère, le sont aussi pour tel autre.

Brugidou est ici exemplaire de cette logique descriptive. Avant de procéder à une factorisation sur les mots utilisés dans des éditoriaux syndicaux, il procède à une classification hiérarchique et détermine des « groupes » de textes. De ce fait, la factorisation n'a plus pour objectif que de qualifier ces groupes et non plus de les repérer. La logique est ici clairement celle de la synthèse qui vise à résumer et rassembler des informations hétérogènes à propos d'une population.

A l'opposé de cet usage synthétique et qualificatif de l'analyse de données, un autre usage est repérable qui se fonde sur l'analyse et la catégorisation.

*Deuxième groupe : un usage exploratoire et analytique de l'analyse de données*

Comme les auteurs du « modèle synthétique », les contributeurs de la revue *Mots* font un usage important des méthodes de factorisation. Les études de lexicométrie en sont par tradition de grandes utilisatrices.

N'est elle pas « la méthode automatique la plus synthétique et la plus efficace en lexicométrie pour comparer quantitativement des textes » ? (Tournier et Bonnafous). A ces quatre articles, nous en agrégeons trois autres de la *Revue Française de Science Politique* et de la *Revue Française de Sociologie* (Gojard, Chiche et aut., Dolez et Laurent).

Ce groupe d'articles est caractérisé par une préparation minimale des données. La construction des variables est relativement simple, comparée en tout cas aux articles publiés dans les *ARSS*. Il s'agit essentiellement de discriminer des seuils de fréquence d'occurrences des formes lexicales à partir desquels elles seront incluses dans l'analyse. Les auteurs sont moins généreux en explication concernant la manière dont le corpus textuel est « préparé », c'est-à-dire segmenté ou lemmatisé.

Une des raisons qui permet, peut-être, d'expliquer, ce type d'usage de l'analyse tient dans la taille des corpus, allant jusqu'à plusieurs milliers d'occurrences croisées avec de l'ordre d'une dizaine de variables, les textes. Ici, le ratio nombre d'individus/nombre de variables est largement supérieur aux précédents. L'accent est donc mis non pas sur les dimensions d'analyse et leur fidélité avec le matériau, mais sur le volume traité.

Les articles de la *RFSP* fournissent de bons exemples de ce choix méthodologique. Deux d'entre eux portent sur des pratiques de vote, lors des élections du président du RPR en 1999 et lors d'élections nationales en 1997. Chiche et aut. travaillent d'ailleurs sur un échantillon représentatif de l'électorat français de 3000 électeurs. Ici, les variables dépendantes sont constituées par les votes pour l'un des quatre grands partis.

Les articles que nous qualifions d'« analytiques » laissent paraître une intégration de l'analyse de données dans la construction de l'article nettement différente de celle dominante chez les auteurs « qualificatifs ». C'est toute la distinction entre un usage fondé sur des représentations fortes de l'objet chez les seconds (la répartition du capital social est nécessairement asymétrique et les résultats des regroupements sont généralement peu éloignés de la représentation qu'on se faisait antérieurement de la population) et un usage fondé sur des représentations moins précises *a priori* de l'objet chez les premiers. Pour les auteurs de ce second type d'usage, représentés par *Mots*, l'enjeu de la factorisation est de repérer des groupes *ex nihilo* plus que de les qualifier. Ils accordent par conséquent une plus grande attention à la nature ambiguë des résultats des analyses de données et tentent de lever cette ambiguïté.

Pour mener cette tâche de classification à bien, ces utilisateurs prudents s'aident d'autres méthodes, souvent post-factorielles, et notamment de la classification hiérarchique. Ces classifications hiérarchiques aident à tester une hypothèse. Le plus souvent, les autres méthodes sont postérieures à l'analyse de données et indiquent alors une recherche de confirmation de la part des auteurs. Placées en revanche avant l'analyse factorielle, comme chez Brugidou, elles permettent non plus de confirmer mais de préparer l'analyse des données et fonctionnent comme une variable supplémentaire qui apparierait chaque individu à une classe. Marchand et Monnoyer-Smith se contentent certes des analyses factorielles, mais ils préparent longuement les données à factoriser.

Un autre usage des méthodes complémentaires est repérable. Il vise à compléter plutôt qu'à confirmer les résultats de la factorisation. Deroubaix est exemplaire de cet tendance. Il reformule ses hypothèses et renforce ses résultats en multipliant les méthodes et les tests. Dans une moindre mesure, les auteurs travaillant sur les déclarations de politique générale usent des méthodes complémentaires sur ce même mode. « C'est une césure gauche/droite [...] que nous pensions *a priori* retrouver dans l'étude lexicométrique des déclarations gouvernementales. Mais l'analyse du vocabulaire [...] nous a surtout permis de mettre en évidence la conjonction de deux sources de structuration des choix lexicaux : une évolution chronologique du vocabulaire et une certaine opposition entre un discours volontariste et un discours de gestion. L'aspect chronologique est dominant et se marque dans les analyses factorielles des correspondances (AFC) du tableau des formes. » C'est en ces termes que Deroubaix (p. 67) termine l'introduction de son article. L'analyse de données est donc susceptible d'infirmer une hypothèse, ou du moins une représentation, elle crée des catégories inattendues, nouvelles. C'est ce que n'avaient pas laissé entendre les auteurs des ARSS, ni plus généralement aucun auteur faisant un usage « qualificatif » de l'analyse de données. Ce second type d'usage est fondé, non plus sur une volonté de qualifier des groupes déjà existants ou supposés tels, ou de définir dans le détail des contours des groupes déjà connus dans leurs grandes lignes, mais sur ce qui peut apparaître comme une réelle pratique de classification et d'analyse d'une population d'individus. Dans l'ensemble, les auteurs « analytiques » adressent des questions précises à leurs données : « est-ce que le clivage idéologique est plus structurant que l'évolution chronologique dans le choix des termes ? » ; « les genres d'interventions publiques déterminent-ils le choix d'un vocabulaire ? ».

De plus, ces auteurs n'hésitent pas à reformuler leurs questions en prenant acte des résultats de différentes méthodes complémentaires à la factorisation. Sur ce point, les contributeurs de *Mots* sont rejoints par ceux de la *RFSP*. Dolez et Laurent notamment examinent diverses explications à la répartition des votes pour la présidence du RPR et y consacrent trois analyses factorielles, portant sur des critères différents (idéologiques, géographiques, politiques). La logique est ici celle du test qui examine avec les mêmes critères différentes hypothèses. La seule description des cartes factorielles ne suffit pas. Tout fonctionne comme si, chez les auteurs « analytiques », le rapport à l'objet était un rapport interrogatif et exploratoire qui les amènerait à examiner cet objet sous différentes faces, la factorisation étant pour eux un examen parmi d'autres.

A ce titre, certains articles occupent une position intermédiaire entre les deux grands modèles d'usage des méthodes de factorisation. C'est le cas de Chollet et Duval, seuls auteurs à procéder à « trois expériences différentes » pour déterminer les représentations que se font deux groupes de goûteurs d'une série de vins. Dans leur article, l'analyse factorielle ne constitue qu'une étape et témoigne d'un usage maîtrisé de l'analyse de données.

L'analyse factorielle a conquis ses lettres de noblesse, et est utilisée de manière routinière et non didactique par certains auteurs. En revanche, d'autres détaillent toujours les principes de la méthode.

Chez les auteurs qualitatifs, l'analyse factorielle est utilisée pour automatiser et faciliter le traitement de données hétérogènes, comme un outil, bien davantage que comme une méthode. Au sens de l'outil, l'analyse automatique de données n'intervient dans le processus scientifique ni pour tester une hypothèse ni pour examiner un objet. A *contrario*, chez les auteurs analytiques, l'analyse automatique de données constitue une véritable méthode qui permet de tester une hypothèse et de participer, avec d'autres méthodes, à la définition d'un objet. Ces deux usages ne s'opposent pas mais témoignent de l'ambivalence inhérente aux méthodes de factorisation. Tandis que dans le premier groupe d'articles, la méthode fait écho à une théorie commune, et se voit assigner, à l'exclusion de toute autre méthode, une fonction de description de l'objet, de confirmation et d'affinement d'une théorie générale, dans le second groupe, elle reste un outil parmi d'autres, intégré dans une démarche hypothético-déductive, au pouvoir limité, et bien compris comme tel. Outil ou méthode, chacun des usages présente un certain nombre d'avantages ou d'inconvénients. Pour autant

nous avons pu montrer un certain nombre d'éléments indiquant que l'analyse factorielle est utilisée moins en fonction de ces avantages et de ces inconvénients qu'en fonction de traditions disciplinaires. Il semble alors qu'au fil du temps, les usages de l'analyse factorielle se soit affinés, non pas par un développement et une meilleure maîtrise des potentialités et des limites de la méthode, mais par une relative standardisation des usages.

Ceci est dû non seulement à la compétence technique des chercheurs, mais aussi et surtout aux affinités entre l'outil et la théorie. Théorie et méthode vont donc de pair, tantôt routinières et auto-reproductives, tantôt évolutives et créatrices.

#### **NOTES**

<sup>1</sup> Cf. Lorenzi-Cioldi (1983, p. 378) pour des exemples d'usages des éléments supplémentaires.

<sup>2</sup> Ces distinctions sont schématiques, elles ne tiennent pas compte des différentes méthodes de factorisation et des différents paramétrages. Pour une présentation synthétique des différences entre analyse des correspondances, analyse en composantes principales et analyse en composantes multiples, on pourra lire : Saporta, 1993.

<sup>3</sup> La valeur-test mesure la pertinence avec laquelle une modalité donnée définit une classe d'individus. Elle est calculée à partir de la fréquence dans la classe, de la fréquence dans la population, et du degré d'inclusion de la classe dans la modalité.

<sup>4</sup> Il est par exemple recommandé, pour éviter ces pièges, de commencer par décomposer les variables en domaines à analyser séparément. Quelles que soient les méthodes utilisées, il est de toute manière utile d'alterner visions générales et gros plans, et donc de conserver en vue les tris de base.

<sup>5</sup> Cf. bibliographie des articles examinés en fin d'article.

<sup>6</sup> Les dates de publication des articles ne seront mentionnées qu'à leur première occurrence, hormis pour les deux articles d'O. Godechot, susceptibles d'être confondus.

<sup>7</sup> Les effectifs ont été arrondis pour fixer les idées.

<sup>8</sup> Gojard se distingue sur ce point en ce qu'elle analyse une population très importante de 3600 femmes.

## REFERENCES

### Articles analysés

ALVAREZ Ramon, BECUE Monica et LANERO Juan José, « Le vocabulaire gouvernemental espagnol (1979-1996) », *Mots*, n°62, mars 2000, p. 30-64.

BONNAFOUS Simone, TOURNIER Michel, « Discours et gestes télévisés : quelles méthodes ? », *Mots*, n°67, décembre 2001, p. 110-128.

BOURDIEU Pierre, « Une révolution conservatrice dans l'édition », *Actes de la Recherche en Sciences Sociales*, n°130, décembre 1999, p. 4-28.

BRUGIDOU Mathieu, « Les discours de la revendication et de l'action dans la presse syndicale (1996-1998) », *Revue Française de Science Politique*, vol. 50, n°6, décembre 2000, p. 967-992.

CHICHE Jean, LE ROUX Brigitte, PERRINEAU Pascal, ROUANET Henry, « L'espace politique des électeurs français à la fin des années 1990. Nouveaux et anciens clivages, hétérogénéité des électors », *Revue Française de Science Politique*, vol. 50, n°3, juin 2000, p. 463-488.

CHOLLET Sylvie, DUVAL Valentin, « Le degré d'expertise a-t-il une influence sur la perception olfactive ? Quelques éléments de réponse dans le domaine du vin », *L'Année Psychologique*, n°100, 2000, p. 11-36.

DEROUBAIX Jean-Claude, « Les déclarations gouvernementales se suivent et se ressemblent. Exploration d'une chronique textuelle », *Mots*, n°62, mars 2000, p. 65-93.

DOLEZ Bernard, LAURENT Annie, « Quand les militants du RPR élisent leur prédécesseur (20 novembre-4 décembre 1999) », *Revue Française de Science Politique*, vol. 50, n°1, février 2000, p. 125-146.

DUVAL Julien, « Concessions et conversions à l'économie. Le journalisme économique en France depuis les années 80 », *Actes de la Recherche en Sciences Sociales*, n° 131-132, mars 2000, p. 56-75.

GODECHOT Olivier, « Le bazar de la rationalité. Vers une sociologie des formes concrètes de raisonnement », *Politix*, vol. 13, n°52, 2000, p. 17-56.

GODECHOT Olivier, « Le marché du livre philosophique. », *Actes de la Recherche en Sciences Sociales*, n°130, décembre 1999, p. 11-28.

GOJARD Séverine, « L'alimentation dans la prime enfance. Diffusion et réception des normes de puériculture », *Revue Française de Sociologie*, vol.41, n°3, 2000, p. 475-512.

GOLLAC Michel, KRAMARZ Francis, « L'informatique comme pratique et comme croyance », *Actes de la Recherche en Sciences Sociales*, n° 134, septembre 2000, p. 4-21.

MARCHAND Pascal, MONNOYER-SMITH Laurence, « Les "discours de politique générale" français : la fin des clivages politiques ? », *Mots*, n°62, mars 2000, p. 13-29.

### Autres références

BENZECRI Jean-Paul :

- *L'analyse des données – 2. L'Analyse des correspondances*, Laboratoire de statistique de l'Université de Paris VI, Paris, Bruxelles, Montréal, Dunod, 1973.

- « Histoire et préhistoire de l'analyse de données », *Cahiers de l'analyse des données*, vol. II, n°1, 1977, p. 9-40.

- *Histoire et préhistoire de l'analyse des données*, Dunod, 1982.

CIBOIS Philippe :

- *La Représentation factorielle des tableaux croisés et des données d'enquête: étude de méthodologie sociologique*, CNRS, Laboratoire d'informatique pour les sciences de l'homme, 1980.

- *L'analyse des données en sociologie*, Presses Universitaires de France, 1990 (1984).

- « La bonne volonté scolaire. Expliquer la carrière scolaire d'élèves issus de l'immigration », in BLANCHARD P., RIBEMONT T., *Méthodes et outils des sciences sociales. Innovation et renouvellement*, L'Harmattan, coll. « Cahiers Politiques », 2002.

LORENZI-CIOLDI Fabio, « L'analyse factorielle des correspondances dans les sciences sociales », *Revue Suisse de Sociologie*, n°2, 1983, p. 365-390.

MORINEAU Alain, « Le "Thémascope" » ou analyse structurelle des données d'enquête », in GRANGE Duval., LEBART Ludovic, *Traitements statistiques des enquêtes*, Dunod, 1993, p. 135-160.

SAPORTA Gilbert, « Notions sur les méthodes factorielles », in GRANGE Duval., LEBART Ludovic. *Traitements statistiques des enquêtes*, Dunod, 1993, p. 75-90.

### **RESUME**

L'analyse factorielle est une méthode puissante, séduisante, mais délicate d'emploi. L'examen exhaustif de ses usages dans neuf revues de sciences sociales sur une période de trois ans rend compte de la diversité des données traitées, des tailles des corpus analysées, des objectifs visés. Les méthodes utilisées en complément de l'analyse factorielle sont également diverses, tout comme la place accordée à la discussion proprement méthodologique. Deux approches se dégagent de cette revue des usages, déterminées par les frontières disciplinaires et les habitudes des revues. L'une privilégie le travail de synthèse de variables hétérogènes et de qualification des catégories socio-économiques. L'autre, plus exploratoire et analytique, intègre l'analyse factorielle dans un processus hypothético-déductif où plusieurs autres méthodes sont mises en jeu.

### **ABSTRACT**

Factor analysis is a powerful and seducing method, but it has to be handled with care. We review here all the articles dealing with this method in nine major French social sciences reviews during three years. The kind of datas processed by factor analysis prove to be very diverse, as well as the size of corpuses, the aims assigned to the analysis, the joined statistical methods and the amount of discussion about the method in itself. We bring out two distinct approaches, depending on disciplinary borders and on reviews' habits. In one, factor analysis is used to synthesize heterogeneous variables and to qualify social and economical categories. The other approach is more exploratory and analytical. In it, factor analysis is part of a set of complementary methods devoted to scientific demonstration. The method has more or less standardized itself with time, each approach bringing to the fore certain aspects of factor analysis.