

# Presses universitaires de Liège

---

Mesures de l'écrivain | Björn-Olav Dozo

---

## 3. L'analyse factorielle des correspondances

p. 23-47

### Texte intégral

- 1 L'analyse factorielle des correspondances est un outil statistique de présentation et d'analyse de tableaux de données. L'objectif est de classer un nombre d'individus en fonction de variables qui les décrivent. L'analyse factorielle permet d'afficher les données d'un tableau sous la forme graphique d'un nuage de points répartis sur un plan. Ce graphique n'est finalement que le résumé des résultats de l'analyse des données : il permet de présenter les traits saillants de l'analyse d'un tableau qui serait impossible à lire tel quel, c'est-à-dire sans réorganisation, classement et traduction visuelle.

### 3.1. USAGE DE L'ANALYSE FACTORIELLE DES CORRESPONDANCES MULTIPLES PAR BOURDIEU

- 2 L'analyse factorielle permet de faire surgir la structure des données, la façon dont les modalités de chaque variable se situent les unes par rapport aux autres, de manière différentielle et relationnelle. La sociologie structurale de type bourdieusien, et Bourdieu lui-même, en ont fait un outil de représentation puissant : l'outil permet de mettre au jour la structure multidimensionnelle et relationnelle du champ étudié.
- 3 Ainsi, dans le deuxième chapitre de *La Distinction*<sup>20</sup>, intitulé « L'espace social et ses transformations », Bourdieu recourt à l'analyse factorielle des correspondances multiples pour expliciter sa conception de l'espace social. Quand il explique ce qu'entraîne la création de classes d'individus, il raisonne à partir d'un modèle fondé sur la description d'individus par des variables<sup>21</sup>, et se réfère à J.-P. Benzécri<sup>22</sup>, le fondateur et le promoteur de l'analyse factorielle des correspondances en France<sup>23</sup>. Il utilise l'analyse factorielle pour critiquer les dérives qu'entraîne l'usage de certaines catégories en statistique. Il est intéressant de constater qu'il mobilise de la sorte un outil statistique pour mettre en cause certaines catégorisations qui étaient largement utilisées par d'autres statistiques. Par exemple, il interroge la classification des variables, réparties traditionnellement en variables dépendantes — celles qui sont liées au phénomène que l'on cherche à expliquer — et variables dites indépendantes, qui ne dépendraient donc pas de ce phénomène.

Les relations singulières entre une variable dépendante (telle l'opinion politique) et des variables dites indépendantes comme le sexe, l'âge et la religion, ou même le niveau d'instruction, les revenus et la profession, tendent à dissimuler le système complet des

relations qui constituent le véritable principe de la force et de la forme spécifiques des effets enregistrés dans telle corrélation particulière. La plus indépendante des variables « indépendantes » cache tout un réseau de relations statistiques qui sont présentes, souterrainement, dans la relation qu'elle entretient avec telle opinion ou pratique. Là encore, au lieu de demander à la technologie statistique de résoudre un problème qu'elle ne peut que déplacer, il faut, par une analyse des divisions et des variations qu'introduisent, au sein de la classe découpée par la variable principale, les différentes variables secondaires (sexe, âge, etc.), s'interroger sur tout ce qui, présent dans la définition réelle de la classe, n'est pas consciemment pris en compte dans la définition nominale, celle que résume le nom employé pour la désigner et, par conséquent, dans l'interprétation des relations dans lesquelles on la fait entrer (BOURDIEU, 1979, p. 114-115).

- 4 On retrouve donc dans cet extrait la variable principale qui permet de définir la classe et que Benzécri appelait  $\alpha$ , et les variables secondaires, notées  $\beta$  et  $\gamma$  par Benzécri et qui pourraient définir des sous-classes. L'AFC a permis à Bourdieu de mettre au jour tout le « réseau de relations statistiques » qui existe entre les variables dépendantes et indépendantes. L'usage qu'il fait de l'AFC montre en quoi les classifications construites *a priori* présagent déjà d'un découpage du monde qui masque des relations « souterraines », des corrélations entre variables *a priori* indépendantes mais qui, dans les faits, apparaissent liées, ce que met en évidence l'AFC. Ce refus de réduire le monde social à des classes préconstruites fut une des grandes prises de position de Pierre Bourdieu à travers son usage des statistiques. Son utilisation de l'analyse factorielle des correspondances multiples fut créatrice et l'aida dans sa critique de la statistique « officielle » (dans le sens où les institutions officielles utilisaient cette statistique pour justifier des décisions politiques par exemple) jusque-là en vigueur<sup>24</sup>. Il dut d'ailleurs essayer plusieurs critiques à ce sujet. Ainsi, le graphique bien connu de *La Distinction*, qui représente l'espace social dans les années 1970<sup>25</sup>, n'est pas une analyse factorielle au sens propre : il s'agit d'une reconstruction manuelle à partir d'un ensemble d'analyses factorielles partielles. La principale critique (outre le repérage d'erreurs ponctuelles) fut qu'il n'utilisait la statistique que pour confirmer des idées qu'il s'était déjà forgées. Nous suivons Michel Gollac lorsqu'il prend la défense de Bourdieu en disant que « la statistique sans idées [est] généralement sans résultat » et en soulignant que « les ACM menées postérieurement à la publication de son graphique par Bourdieu [ont] systématiquement confirmé la structure qu'il pronostiquait »<sup>26</sup>.
- 5 L'usage qu'a fait Bourdieu des graphiques de l'analyse factorielle offre surtout un excellent résumé d'un travail de recherche, résumé visuel et efficace, qui permet l'appréhension quasi immédiate par le lecteur — après un travail d'interprétation, comme tout graphique — d'un grand nombre de données et surtout des relations qu'elles entretiennent. Il s'agit là d'un des avantages principaux de la méthode, que Bourdieu a su très bien exploiter : le graphique fait apparaître « une concentration de l'ensemble des possibles »<sup>27</sup> d'un espace social particulier, en soulignant que chaque pratique ne prend sens que par rapport aux autres<sup>28</sup>. Finalement, la meilleure expression pour qualifier l'utilisation par Bourdieu de l'analyse factorielle est que celle-ci lui permettait de raisonner « toutes choses égales réunies » (par opposition à d'autres méthodes, par exemple les régressions linéaires, qui fonctionnent « toutes choses égales par ailleurs » — c'est-à-dire des méthodes où l'on teste la validité de l'explication d'un phénomène à partir de la modalisation de ce phénomène et de la permutation une à la fois de chaque modalité d'une des variables explicatives du phénomène).

### 3.2. APERÇU HISTORIQUE DE L'UTILISATION DE L'ACM EN SOCIOLOGIE

- 6 Il existe donc, en sociologie, un lien très fort entre l'analyse factorielle des correspondances et la sociologie structurale. L'article de Paul-André Rosental, publié vingt ans après *La Distinction*, parle même de « laïcisation » de ce type d'analyse en constatant la multiplicité des usages et des disciplines qui y eurent recours depuis les premiers graphiques de Bourdieu. Rosental met en évidence comment et pourquoi le marquage de l'AFC par la sociologie « structurale » a pris forme au départ :

En reprenant cette citation de J.-P. Benzécri dès les premiers développements de ses considérations sur l'espace social, Pierre Bourdieu explicitait le lien qui, en sciences sociales, allait pendant plusieurs années unir analyse des correspondances et approches

« structurales ». L'outil mathématique venait apporter son appui démonstratif à la théorie sociologique, tandis que celle-ci identifiait dans le fonctionnement de la société une incarnation concrète de la vision multidimensionnelle proposée par l'analyse factorielle (ROSENTAL, 1997, p. 185).

- 7 Ce « marquage structural » de l'outil et son appropriation privilégiée par cette école sociologique ont suscité de longues polémiques sur l'usage même qu'il était possible de faire de l'AFC : pouvait-elle être utilisée en dehors d'une interprétation structurale ? L'adéquation postulée un peu rapidement par certains bourdieusiens entre le « champ » et le produit de l'AFC (le graphique) ne relève-t-elle pas de la surinterprétation ? La diffusion de la méthode dans d'autres sphères de recherche a quelque peu pâti du développement de cet a priori à son sujet. Cependant, la laïcisation qu'évoque en boutade Rosental avait pourtant bien eu lieu, au sens propre cette fois, lors de cette appropriation de l'outil par Bourdieu. En effet, le créateur de l'AFC, J.-P. Benzécri, soutenait quant à lui, et cela dès les premiers développements de cette technique, que les facteurs qu'elle mettait en évidence renvoyaient à la vraie nature des choses créées par Dieu. L'analyse factorielle pouvait ainsi dégager « de la gangue des données le pur diamant de la véridique nature »<sup>29</sup>. Dans leur article sur l'histoire de l'analyse des correspondances en sociologie en France, Karl M. van Meter, Marie-Ange Schiltz, Philippe Cibois et Lise Mounier soulignent les implications de cette position :

This philosophical and religious thought has very important practical consequences. For Benzécri, what is important is not the data whose analysis gives an approximate image, but the factors which reflect the essence of reality. Thus one must eliminate any return to the data to interpret results, any use of post-factorial analysis which tries to find in the original data the results of the correspondence analysis which are only an approximation of the data themselves (VAN METER, SCHILTZ, CIBOIS et MOUNIER, 1994, p. 129).

- 8 Bourdieu, lorsqu'il utilise la technique, la débarrasse de cette croyance en la nature divine des facteurs mis au jour par l'analyse<sup>30</sup>. Il va utiliser toutes sortes de moyens de contrôle des résultats graphiques, du cosinus carré des points projetés à la contribution des individus ou des modalités aux axes. Nous reviendrons sur ces moyens lorsque nous expliquerons en détail le fonctionnement de la méthode (voir section 3.5, p. 45-47). Pour le moment, il est important d'insister sur le fait que, pour Bourdieu, les résultats graphiques de l'analyse factorielle ne sont lisibles qu'accompagnés des tableaux de données de contrôle que l'analyse produit.
- 9 L'analyse factorielle fut donc investie de croyance dès le départ, puis utilisée en lien avec une théorie particulière, ce qui eut de sérieuses conséquences sur sa diffusion et son utilisation. La puissance présumée de l'outil, investi de grandes espérances par son créateur, puis la « récupération » par Bourdieu et son école, corrélées aux oppositions dans le champ de la sociologie qui naquirent de la montée en puissance de cette école, sont autant d'étapes qu'il faut garder à l'esprit lors de l'utilisation et de la diffusion d'analyses factorielles en sciences humaines.
- 10 Si l'on veut caractériser avec précision l'usage que Bourdieu fait de l'analyse factorielle des correspondances, on peut parler d'usage « interprétatif ». Nous entendons par « interprétatif » le fait de produire du sens social à partir de l'organisation, de la structure, de la proximité relative des variables. Dans le graphique de *La Distinction*<sup>31</sup>, l'étiquetage des axes en, respectivement, « capital économique » et « capital culturel » n'est pas un donné de l'outil mathématique, mais résulte d'une interprétation des regroupements de variables que l'analyse a opérés. Chaque variable renvoie peu ou prou à la possession ou non d'une forme de capital, et la proximité relative de variables permet au chercheur de « lire » la structure sociale à partir du graphique<sup>32</sup>.
- 11 Gisèle Sapiro, par exemple, explique peu ce « saut » entre la représentation graphique du résultat mathématique et la représentation graphique de l'interprétation bourdieusienne du champ littéraire. Pourtant il est bel et bien présent. Dans un article de la revue *Poetics*, elle pose une question qui permet de mettre en évidence sa conception de l'utilisation de l'analyse factorielle :

Could such a structure [la structure du champ littéraire] be formalized through Multiple Correspondence Analysis (MCA), as Bourdieu did for the academic world (BOURDIEU, 1984)? [...] a geometric representation of the literary field had never be tried. [...] I shall first present the statistical inquiry and the structure of the literary field as displayed by the MCA, then an

interpretation of the first and second axes of inertia, and finally a historical interpretation of the results of the MCA. (SAPIRO, 2002, p. 388)<sup>33</sup>

- 12 La question concerne donc la possibilité d'une représentation de la structure du champ littéraire sous forme graphique, mais la formulation « the structure of the literary field as displayed by the MCA » suppose une adéquation directe entre la structure du champ littéraire et la représentation graphique offerte par l'analyse factorielle. Or cette adéquation entre l'outil mathématique et la théorie des champs ne va pas de soi : Bourdieu a imposé ce lien au fil de ses recherches, mais une analyse factorielle n'est finalement rien d'autre qu'une opération mathématique. Le surcroît de sens que les théoriciens du champ lui accordent, comme mode de représentation privilégié de la structure d'un champ à partir d'un ensemble de variables bien choisies, est lié à la théorie elle-même : en soi, l'analyse factorielle permet seulement diverses opérations mathématiques favorisant l'exploration d'un corpus de données. Cette remarque n'enlève rien à l'efficacité et à la qualité de l'usage qu'en fait Bourdieu ; cependant, cette explicitation nous semblait nécessaire, dans la mesure où elle nous prémunit d'une lecture trop essentialisante des données visualisées.
- 13 Nous aurons également recours dans ce travail à cette technique consistant à donner du sens à une configuration de variables en l'identifiant au champ littéraire. Cette manière de procéder permet des interprétations fortes si le choix des variables est intéressant. Néanmoins, l'ACM ne nous servira pas qu'à représenter un champ : sa fonction première, l'exploration et le classement d'un corpus de données, nous sera également d'une grande utilité. Félicité des Nétumières souligne la pertinence de cet usage plus basique de l'ACM :

[L']objectif premier [de l'analyse factorielle] est de permettre au chercheur d'appréhender le plus simplement possible la masse de données dont il dispose dans son enquête et d'en extraire les informations pertinentes. Pour reprendre une expression consacrée, l'analyse factorielle est « un aveu d'ignorance », un « radar tourné vers le brouillard » qui « sert avant tout à dépendre à grands traits les dimensions les plus importantes d'une variation dans un nouveau champ de recherche ».

C'est donc, en premier lieu, lors de la phase exploratoire des données, que l'analyse factorielle se présente comme un outil particulièrement utile. Au commencement de toute étude, le chercheur consacre toujours un temps qui peut se révéler très long à « sentir les données », c'est-à-dire tout d'abord à découvrir la population de son fichier, à la décrire à l'aide de ses principales caractéristiques, puis à sélectionner les variables dont il peut supposer qu'elles auront quelque chose à voir avec le sujet de son étude. [...] L'ACM [analyse des correspondances multiples] [...] permet en quelque sorte d'automatiser cette étape, sans qu'il soit nécessaire d'émettre la moindre hypothèse préalable concernant les associations éventuelles entre les différentes variables (DES NÉTUMIÈRES, 1997, p. 276).

- 14 L'usage que l'auteure prône de l'analyse factorielle est bien différent de celui des bourdieusiens : il ne s'agit pour elle que d'un outil d'exploration de corpus, qui « permet de se faire une idée ». S'il est exact qu'il ne faut pas trop d'hypothèses préalables concernant les associations et corrélations entre variables — et ainsi éviter de forcer l'interprétation de l'ACM par une lecture conditionnée —, cela ne veut pas dire que tout se fait automatiquement et que le sens « jaillit » du graphique : des hypothèses ont dû être posées au moment de la construction des variables, dont on suppose la pertinence pour décrire l'objet étudié, et d'autres prennent forme lors de l'opération de lecture du graphique, qui met en évidence les corrélations entre les variables utilisées.

### 3.3. ANALYSE FACTORIELLE DES CORRESPONDANCES (AFC) : PRINCIPE GÉNÉRAL

- 15 Après avoir posé les enjeux de l'utilisation de la méthode, il faut en comprendre le fonctionnement. Nous n'aurons pas la prétention de présenter ici une démonstration mathématique des différentes étapes de l'analyse factorielle. Il ne nous semble en effet pas pertinent de nous livrer à un exposé complexe, fondé sur des calculs matriciels dont il faut connaître l'algèbre, pour expliquer le mécanisme détaillé de ce qui doit rester un outil. Différents ouvrages, dus à des statisticiens, exposent cet aspect de l'AFC. Le lecteur intéressé pourra s'y reporter avec fruit<sup>34</sup>.
- 16 Le reproche traditionnel fait aux utilisateurs de l'analyse factorielle des correspondances en sciences sociales est d'utiliser l'outil comme une boîte noire, qui « mouline » un tableau des données et en fournit un résultat qu'ils commentent de manière parfois impressionniste. Pour éviter cet écueil, il nous semble utile de présenter les grandes lignes du processus de

- transformation d'un tableau de données en un graphique représentant un plan factoriel. Cela devrait aussi familiariser le lecteur non spécialiste au principe général de la méthode.
- 17 La représentation graphique finale de l'analyse se présente sous la forme d'un nuage de points répartis sur un plan. Pour passer du tableau à ce plan, plusieurs opérations sont nécessaires.
- 18 Tout d'abord, il faut présenter les données sous la forme d'un tableau croisé (appelé table de contingence). Il s'agit dans notre cas du croisement des modalités de deux variables descriptives des individus. Dans notre exemple, les deux variables sont la génération et les professions, deux catégories pouvant décrire nos individus, et les modalités sont les différentes valeurs que ces variables peuvent prendre<sup>35</sup>. L'intersection de deux modalités comprend l'effectif partagé par ces deux modalités (par exemple 6 hommes politiques dans la troisième génération).

	Homme politique	écrivain	avocat et associés	journaliste	fonctionnaire et associés	enseignant et associés	commerçant et associés	rentier	religieux	secrétaire particulier	cadre supérieur	directeur d'école	médecin et associés	TOTAL
1 <sup>re</sup> génération	1	3	5	5	6	2	0	0	0	0	0	1	0	23
2 <sup>e</sup> génération	2	0	3	6	4	2	1	1	0	0	0	0	2	21
3 <sup>e</sup> génération	6	4	10	1	2	3	2	0	1	0	1	0	0	30
4 <sup>e</sup> génération	0	4	3	5	2	7	0	0	1	2	0	1	1	26
5 <sup>e</sup> génération	0	1	1	0	4	5	0	0	1	0	1	0	1	14
6 <sup>e</sup> génération	0	1	0	1	0	3	0	0	0	0	0	0	0	5
TOTAL	9	13	22	18	18	22	3	1	3	2	2	2	4	119

Tableau 1 – Exemple de table de contingence

- 19 Ensuite, il faut calculer le profil du tableau, c'est-à-dire exprimer chaque effectif en pourcentage, pour permettre leur comparaison.
- 20 Le profil de chaque ligne du tableau précédent est :

	homme politique	écrivain	avocat et associés	journaliste	fonctionnaire et associés	enseignant et associés	commerçant et associés	rentier	religieux	secrétaire particulier	cadre supérieur	directeur d'école	médecin et associés	TOTAL
1 <sup>re</sup> génération	4,35	13,04	21,74	21,74	26,09	8,70	0,00	0,00	0,00	0,00	0,00	4,35	0,00	100,00
2 <sup>e</sup> génération	9,52	0,00	14,29	28,57	19,05	9,52	4,76	4,76	0,00	0,00	0,00	0,00	9,52	100,00
3 <sup>e</sup> génération	20,00	13,33	33,33	3,33	6,67	10,00	6,67	0,00	3,33	0,00	3,33	0,00	0,00	100,00
4 <sup>e</sup> génération	0,00	15,38	11,54	19,23	7,69	26,92	0,00	0,00	3,85	7,69	0,00	3,85	3,85	100,00
5 <sup>e</sup> génération	0,00	7,14	7,14	0,00	28,57	35,71	0,00	0,00	7,14	0,00	7,14	0,00	7,14	100,00
6 <sup>e</sup> génération	0,00	20,00	0,00	20,00	0,00	60,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00
Moyenne	5,65	11,48	14,67	15,48	14,68	25,14	1,90	0,79	2,39	1,28	1,75	1,37	3,42	100,00

Tableau 2 – Profils des lignes de la table de contingence

- 21 On peut considérer que chaque profil est en quelque sorte un vecteur, au sens mathématique du terme, c'est-à-dire que le point « première génération » pourrait être situé dans un espace à 13 dimensions (le nombre de modalités de la variable « professions ») par le vecteur (0,0435 ; 0,1304 ; 0,2174 ; 0,2174 ; 0,2609 ; 0,087 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0,0435 ; 0). On comprend aisément que l'esprit humain est incapable de se représenter un tel espace. Tout le principe de l'analyse factorielle va donc consister en la réduction du nombre de dimensions (de 13, on va passer à un espace en deux dimensions, celui d'un plan) grâce à des projections successives de ces vecteurs.
- 22 Avant cela, il faut considérer un autre concept fondamental de l'analyse factorielle : la masse, ou poids, qui va être associée à chaque profil. On l'a vu dans le premier tableau, les effectifs de chaque génération ne sont pas les mêmes. Il a fallu les ramener à 100 pour pouvoir les comparer. Mais on ne peut négliger leur disparité : la première génération comptait un effectif total de 23, tandis que la sixième par exemple, n'en comptait que 5. La masse associée à chaque profil est le moyen par lequel on tient compte de cette disparité. Cette masse est calculée de la sorte : elle est égale à l'effectif de la ligne dans le tableau de départ divisé par l'effectif total du tableau de départ.

	poids
1 <sup>re</sup> génération	23/119=0,193
2 <sup>e</sup> génération	21/119=0,176
3 <sup>e</sup> génération	30/119=0,252

4e génération	26/119=0,218
5e génération	14/119=0,118
6e génération	5/119=0,042

Tableau 3 – Poids des lignes de la table de contingence

- 23 Ensuite, un troisième concept fondamental est celui de distance entre les points de cet espace multidimensionnel. La distance à laquelle nous sommes accoutumés dans un espace euclidien en trois dimensions est appelée distance pythagoricienne. Soient un point A qui a pour coordonnées  $(x_1, y_1, z_1)$  et un point b dont les coordonnées sont  $(x_2, y_2, z_2)$  ; la distance pythagoricienne se calcule selon la formule :

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

- 24 En analyse factorielle, cependant, on utilise une variante de la distance pythagoricienne, appelée distance pondérée. Cette distance est utilisée pour mesurer la distance entre deux profils de points. Cette pondération n'est pas celle qui correspond à la masse du profil ; elle renvoie au poids différentiel de chaque dimension de l'espace considéré. L'effet de la pondération de cette distance est justement un des aspects très utiles de l'analyse factorielle. Greenacre le décrit en ces termes:

In practice this weighting has the effect that response options which occurs less frequently are made to contribute more highly to the interprofile distance, while those that occur more frequently are made to contribute less (GREENACRE, 2004a, p. 11).

- 25 C'est pour cela que l'analyse factorielle met en évidence les écarts de certains points représentant certaines modalités peu répandues par rapport aux autres modalités qui caractérisent tous les individus. On prendra toute la mesure de ce principe lors de l'interprétation.
- 26 La pondération est calculée en divisant, à l'intérieur de la racine, chaque carré de la différence entre les coordonnées de deux profils de points par la moyenne entre les coordonnées normalisées de la dimension dont dépendent ces coordonnées. Par exemple, pour les avocats, cette moyenne équivaut à 0,1467. Cette pondération a pour effet que les modalités (par exemple les professions) qui sont le moins souvent mobilisées pour décrire les individus vont contribuer beaucoup au calcul de la distance entre deux individus, et celles dont les occurrences sont plus nombreuses vont contribuer moins au calcul de cette distance.
- 27 Pour prendre un exemple, on peut calculer la distance entre la première et la sixième génération. Pour ce faire, prenons leurs deux vecteurs

première génération (0,0435 ; 0,1304 ; 0,2174 ; 0,2174 ; 0,2609 ; 0,087 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0,0435 ; 0)  
sixième génération (0 ; 0,2 ; 0 ; 0,2 ; 0 ; 0,6 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0)

- 28 et appliquons la formule :

$\sqrt{\frac{\sum (x_i - y_i)^2}{m_i}}$ , où  $x_i$  correspond à chaque coordonnée du premier vecteur,  $y_i$  à chaque coordonnée du second et  $m_i$  à la masse (au « poids ») de chaque dimension dans cet espace. Cette masse est calculée comme la moyenne des coordonnées des profils de points (ou vecteurs) pour chaque dimension de l'espace. Cela donne pour les deux vecteurs pris comme exemple :

$$\sqrt{\frac{\frac{(0-0,0435)^2}{0,0565} + \frac{(0,2-0,1304)^2}{0,1148} + \frac{(0-0,2174)^2}{0,1467} + \frac{(0,2-0,2174)^2}{0,1548} + \frac{(0-0,2609)^2}{0,1468} + \frac{(0,6-0,087)^2}{0,2514} + \frac{(0-0)^2}{0,019} + \frac{(0-0)^2}{0,0079} + \frac{(0-0)^2}{0,0239} + \frac{(0-0)^2}{0,0128} + \frac{(0-0)^2}{0,0175} + \frac{(0-0,0435)^2}{0,0137} + \frac{(0-0)^2}{0,0342}}{2}} = 0,139$$

- 29 Greenacre explique pourquoi cette distance est appelée distance  $\chi^2$  :

Because of the analogy with the chi-squared concept of calculating squared differences between proportions relative to their expected, or mean, values, this distance is known as the *chi-square distance* (GREENACRE, 2004a, p. 11).

Le quatrième et dernier concept nécessaire à la compréhension des principes généraux de l'analyse factorielle est celui d'inertie, ou plus spécifiquement de moment d'inertie. Ce

principe est emprunté à la mécanique. Chaque objet a un centre de gravité. Chaque particule constituant cet objet à une certaine masse (m) et se trouve à une certaine distance (d) de ce centre de gravité. Le moment d'inertie de l'objet est défini comme la somme des  $md^2$  pour chaque particule, c'est-à-dire  $\sum_r m_r d_i^2$

- 30 Dans le cas de l'AFC, on a un ensemble de points dont la masse totale vaut 1 (ou 100 %). Ces profils de points ont un centre de gravité (leur profil moyen) et on peut mesurer la distance (la distance  $\chi^2$ ) entre ces profils de points. L'inertie de cet ensemble de points peut donc être calculée. Chaque profil de point contribue dans une certaine mesure à cette inertie totale. Par exemple, la contribution de la première génération au total général de l'inertie s'élève à :

$$0,193 \times \left( \frac{(0,0435-0,0565)^2}{0,0565} + \frac{(0,1304-0,1148)^2}{0,1148} + \frac{(0,2174-0,1467)^2}{0,1467} + \frac{(0,2774-0,1548)^2}{0,1548} + \frac{(0,2609-0,1468)^2}{0,1468} + \frac{(0,087-0,2514)^2}{0,2514} + \frac{(0-0,019)^2}{0,019} + \frac{(0-0,0079)^2}{0,0079} + \frac{(0-0,0239)^2}{0,0239} + \frac{(0-0,0128)^2}{0,0128} + \frac{(0-0,0175)^2}{0,0175} + \frac{(0,0435-0,0137)^2}{0,0137} + \frac{(0-0,0342)^2}{0,0342} \right) = 0,068$$

- 31 Dans ce calcul, 0,193 représente la masse du point ligne de la première génération, et la racine carrée élevée au carré renvoie à la distance  $\chi^2$  au carré de ce point ligne par rapport au centre de gravité de l'ensemble des points lignes. Les coordonnées du centre de gravité sont la moyenne des coordonnées des profils des points lignes.
- 32 L'inertie totale des points lignes est calculée en additionnant les inerties de chaque point ligne. L'inertie totale vaut ainsi 0,649. L'inertie a une interprétation géométrique très importante : sur le graphique de résultat de l'AFC, plus l'inertie des points est grande, plus leur dispersion est importante. On peut donc dire que l'inertie est la mesure de la dispersion des profils de points dans l'espace multidimensionnel.
- 33 Maintenant que nous avons travaillé sur la matrice de départ pour pouvoir en donner une représentation multidimensionnelle, il est nécessaire de transformer cette représentation multidimensionnelle, qui est de peu d'intérêt pour l'esprit humain, en une représentation en deux dimensions. Pour cela, on procède par projection et réduction. On va donc arriver à un plan qui offre une représentation la plus proche possible du contenu de la matrice de départ, mais dans un espace réduit.
- 34 La démarche est analogue à celle effectuée dans une régression linéaire : il s'agit, dans le cas particulier d'un nuage de points sur un plan, de rechercher la droite qui passe au plus près de tous les points.
- 35 Pour l'analyse factorielle, Greenacre explique :

In CA we have an analogous situation in that we are looking for a low-dimensional space, usually a plane, which reflects as accurately as possible the chi-square distances between the profiles. In CA this turns out to be equivalent to looking for the plane which is in some sense 'closest' to all the point (GREENACRE, 2004a, p. 15).

- 36 C'est cette proximité et comment elle est calculée qu'il s'agit maintenant d'explicitier. En fait, pour trouver quel plan satisfait à cette condition de proximité, il faut trouver un plan qui réduit au minimum la distance  $\chi^2$  entre ce plan et chaque point. Le point dans ce plan qui est le plus proche d'un profil de point est appelé *projection* du profil de point sur le plan.

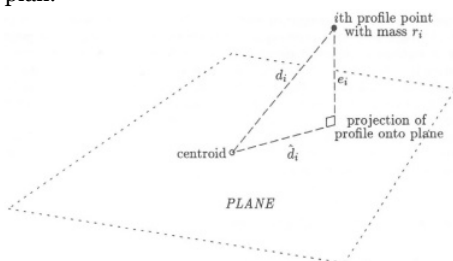


Figure 1 – Projection d'un point-ligne sur le plan factoriel<sup>36</sup>

Il s'agit d'une application classique du théorème de Pythagore :  $d_i^2 = \hat{d}_i^2 + e_i^2$

Si l'on considère maintenant l'inertie totale des profils de points, on a vu qu'on pouvait l'exprimer comme la somme des multiplications de chaque masse du profil du point par la

distance élevée au carré entre la particule et le centre de gravité, soit  $\sum_i r_i d_i^2$ , où  $r$  est la masse du profil du point, et  $d$  la distance de ce point au centre de gravité.

- 37 Cette inertie est décomposable en deux constituants, qui sont l'inertie comprise dans le plan par la projection et l'inertie résiduelle, soit l'équation :

$$\sum_i r_i d_i^2 = \sum_i r_i \widehat{d}_i^2 + \sum_i r_i e_i^2$$

- 38 On peut lire cette équation ainsi : inertie totale = inertie dans le plan + inertie résiduelle.  
 39 L'inertie résiduelle permet de mesurer combien d'inertie a été perdue en réduisant l'espace des profils de points en un espace à deux dimensions. C'est le plan qui permet de conserver le maximum d'inertie qui est retenu pour exprimer les données en deux dimensions. Il s'agit en fait des deux premiers vecteurs de la base vectorielle de la matrice considérée.  
 40 Les vecteurs propres permettent de décrire un espace de coordonnées : chaque point à situer est décrit à partir d'un vecteur multiple de ceux-ci. Par exemple, dans un espace classique en trois dimensions, les vecteurs propres sont (1,0,0), (0,1,0) et (0,0,1). Chaque vecteur étant linéairement indépendant, il n'est pas possible de réduire davantage le nombre de vecteurs nécessaires pour décrire l'espace tridimensionnel. En revanche, dans notre exemple de départ, le tableau des points lignes, qui est de dimension 6x13, peut être décomposé en cinq vecteurs propres. Une réduction du nombre de vecteurs nécessaires pour situer les points dans l'espace multidimensionnel est possible, car les vecteurs coordonnées des points lignes ne sont pas tous linéairement indépendants, c'est-à-dire que pour ces vecteurs  $a_1, \dots, a_6$ , il existe des nombres  $\lambda_1, \dots, \lambda_6$  non tous nuls tels que

$$\sum_{j=1}^6 \lambda_j a_j = \lambda_1 a_1 + \dots + \lambda_6 a_6 = 0$$

- 41 La relation linéaire entre ces vecteurs implique donc qu'ils puissent être exprimés grâce à une base vectorielle réduite, fondée sur ces cinq vecteurs propres. De 6 vecteurs coordonnées (ou profils) des points lignes, on passe à une base vectorielle de 5 vecteurs. Il faut donc compléter ce que nous disions sur les projections des profils de points en vue d'une réduction du nombre de dimensions (pour arriver à un plan) en insistant sur le fait qu'avant la projection pour la réduction, les points sont à exprimer dans la base vectorielle constituée des vecteurs propres linéairement indépendants. Dans notre exemple, pour conserver toute l'information contenue dans le Tableau 1 et la représenter dans l'espace, il faudrait donc un espace en cinq dimensions qui permettraient de situer chaque point. En sélectionnant les deux premiers axes (totalisant à eux seuls 41,9 % de l'inertie totale), on réduit les trois autres dimensions par projection, projection qui apporte 22,79 % de l'inertie totale (en plus de ces 41,9 %). L'inertie du plan factoriel est ainsi portée à 64,69 % de l'inertie totale du tableau de départ.  
 42 Le pourcentage de chaque axe, que l'on retrouve sur les graphiques, donne l'inertie rendue par chaque axe et le pourcentage total du graphique exprime l'inertie rendue du graphique.  
 43 Ainsi, le Tableau 1 donne le graphique page suivante (voir figure 2). Ce graphique exprime 64,69 % de l'inertie totale des profils de points, le premier axe 38,74 % et le deuxième 25,95 % (c'est-à-dire l'inertie propre à chacun des deux premiers axes, plus la projection sur ceux-ci de l'information apportée par les trois autres axes). L'inertie résiduelle est de 35,31 %. Au stade actuel de l'explication de la méthode, nous ne commenterons pas ce graphique en détail<sup>37</sup>.



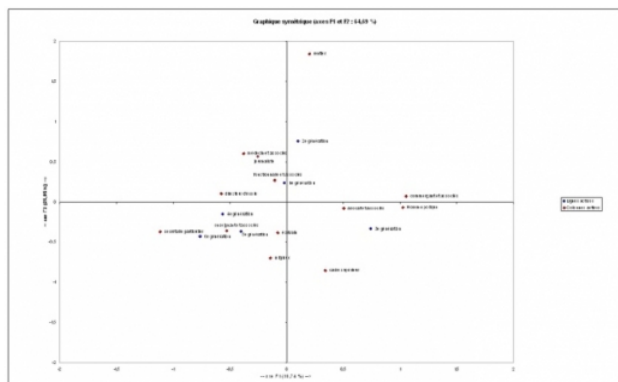


Figure 2 – Analyse factorielle des correspondances du tableau 1

- 44 Les développements suivants nous ont ainsi permis de comprendre comment, d'un tableau de données, on passe successivement :

à des profils de points (points lignes ou points colonnes) — ce qui permet de s'en servir comme coordonnées d'un espace multidimensionnel ;  
 puis à la masse de ces profils de points — ce qui permet d'évaluer leur importance les uns par rapport aux autres, et leur contribution à la masse totale ;  
 puis à la distance de ces profils de points par rapport au centre géométrique de cet espace, en définissant le centre géométrique comme la moyenne des coordonnées de tous les profils de points ;  
 enfin à l'inertie de ces profils de points — ce qui permet de rendre compte de l'équilibre général de l'espace multidimensionnel, en mesurant la contribution de chaque profil de point à l'inertie totale de l'espace multidimensionnel (cette mesure est obtenue par la pondération de la distance de chaque profil de point par rapport au centre géométrique).

- 45 Les points-lignes et les points-colonnes ne se décomposent pas en deux plans factoriels identiques : les deux premiers vecteurs des bases vectorielles ne sont pas les mêmes, et les bases vectorielles elles-mêmes ne sont pas les mêmes. Néanmoins, il est possible de projeter les points-colonnes sur l'espace vectoriel des points-lignes et vice-versa. Il faut cependant se méfier des interprétations des proximités entre ces points de différentes natures. Les « qualités<sup>38</sup> » des projections de ces différents points ne sont pas homogènes, et le fait qu'ils ne viennent pas du même espace de départ (même si ceux-ci sont liés) est encore une source d'erreur supplémentaire.
- 46 L'analyse factorielle des correspondances est très utile pour comparer deux variables décrivant des individus. Mais si l'on a plusieurs variables à comparer, il faut un autre outil, qui peut être conçu comme une généralisation du premier.

### 3.4. ANALYSE FACTORIELLE DES CORRESPONDANCES MULTIPLES (ACM) : PRINCIPE GENERAL

- 47 La différence principale entre l'AFC et l'ACM concerne les données qui sont analysées. L'ACM analyse des tableaux constitués d'individus (au sens général) et de modalités qui décrivent ces individus. L'intersection entre la ligne de chaque individu et la colonne de chaque modalité contient soit un « 0 », si l'individu ne peut être décrit par cette modalité, soit un « 1 », s'il le peut. Par exemple, prenons les écrivains de l'Académie royale de Langue et Littérature françaises de Belgique<sup>39</sup> depuis sa fondation jusqu'en 1995 comme individus et différentes variables composées de plusieurs modalités pour les décrire. On obtient le tableau suivant (Tableau 4), que l'on va nommer Z, et qui est une table de contingence.

Nom	Généralistes						Partisans d'origine						Militants						de doctrine déclinée						Universels						Grands types de productions						Genres pratiqués									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
Ernest Allouart	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46

Tableau 4 – Table de contingence du personnel littéraire de l'ARLLF

48 Ensuite, ce tableau est transformé en un autre tableau, appelé tableau de Burt, sur lequel est produite l'AFC. Ce tableau de Burt est en fait le résultat de la multiplication de la matrice transposée de Z par Z. On appelle matrice transposée d'une matrice Z (et on la note Z<sup>t</sup>) la matrice obtenue en prenant l'élément de Z en ligne i, colonne j et en le plaçant en ligne j et colonne i. On a donc Z<sup>t</sup><sub>(j,i)</sub> = Z<sub>(i,j)</sub> avec i, j = 1,..., N. Le tableau de Burt (B) est donc construit selon la formule B = Z<sup>t</sup> Z. Il s'agit d'une matrice nommée matrice d'adjacence, c'est-à-dire symétrique, avec en lignes et en colonnes les différentes modalités descriptives des individus, et à l'intersection de chaque ligne et de chaque colonne, l'effectif commun aux deux modalités.

Nom	Généralistes						Partisans d'origine						Militants						de doctrine déclinée						Universels						Grands types de productions						Genres pratiqués									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20 <th>21</th> <th>22</th> <th>23</th> <th>24</th> <th>25</th> <th>26 <th>27</th><th>28</th><th>29</th><th>30</th><th>31</th><th>32</th><th>33</th><th>34</th><th>35</th><th>36 <th>37</th><th>38</th><th>39</th><th>40</th><th>41</th><th>42</th><th>43</th><th>44</th><th>45</th><th>46 </th></th></th>	21	22	23	24	25	26 <th>27</th> <th>28</th> <th>29</th> <th>30</th> <th>31</th> <th>32</th> <th>33</th> <th>34</th> <th>35</th> <th>36 <th>37</th><th>38</th><th>39</th><th>40</th><th>41</th><th>42</th><th>43</th><th>44</th><th>45</th><th>46 </th></th>	27	28	29	30	31	32	33	34	35	36 <th>37</th> <th>38</th> <th>39</th> <th>40</th> <th>41</th> <th>42</th> <th>43</th> <th>44</th> <th>45</th> <th>46 </th>	37	38	39	40	41	42	43	44	45	46
Ernest Anouï	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46

49 Le tableau suivant (tableau 5) est similaire à un tableau sur lequel l'AFC peut fonctionner. À partir de celui-ci, on applique la série de principes expliqués pour l'AFC, et l'on calcule les profils, les masses, l'inertie de chaque ligne. On repère les vecteurs linéairement indépendants, ce qui permet de constituer une base de vecteurs propres, puis on projette les profils des vecteurs lignes obtenus sur le plan factoriel formés des deux premiers vecteurs propres. Une des seules différences est que le pourcentage d'inertie totale n'a plus de signification statistique : il est « pollué » par la présence de la première diagonale dans le tableau de Burt. Cette diagonale a une grande inertie due à sa nature particulière et vient parasiter les calculs d'inertie. Nous ne nous attarderons pas sur cette difficulté, car les dernières versions des logiciels de calcul incluent des ajustements permettant de ne plus obtenir de résultats aberrants concernant l'inertie<sup>40</sup>.

50 Une autre différence avec l'AFC subsiste néanmoins ; elle est inhérente au nombre de facteurs : l'inertie du premier plan factoriel dans une ACM est nécessairement faible<sup>41</sup> car plus le nombre de dimensions de l'espace de départ est élevé (c'est-à-dire plus il y a de modalités descriptives des individus), plus la projection de cet espace sur un plan (espace à deux dimensions) entraîne une perte d'informations. Cela a des conséquences sur la qualité de la projection de l'espace, mais la méthode de l'analyse factorielle permet néanmoins de garder les traits les plus saillants de la structure des données, lesquels permettront une interprétation.

	première	deuxième	troisième	quatrième	cinquième	sixième	Namur	Flandre orientale	...	prose courte
première	16	0	0	0	0	0	1	1	...	8
deuxième	0	15	0	0	0	0	0	1	...	7
troisième	0	0	16	0	0	0	0	0	...	6
quatrième	0	0	0	17	0	0	1	1	...	4
cinquième	0	0	0	0	10	0	0	1	...	5
sixième	0	0	0	0	0	3	0	0	...	0
Namur	1	0	0	1	0	0	2	0	...	0
Flandre orientale	1	1	0	1	1	0	0	4	...	1
Bruxelles	4	8	4	7	4	3	0	1	...	12
Flandre occidentale	1	0	0	0	0	0	0	0	...	0
Brabant	2	0	2	1	0	0	0	0	...	0
Anvers	2	3	1	3	1	0	0	0	...	4
Hainaut	1	2	5	2	1	0	0	0	...	7
Liège	3	2	5	2	2	0	0	0	...	8
Limbourg	1	1	0	0	0	0	0	0	...	0
Luxembourg	0	0	1	0	1	0	0	0	...	1
philologie classique	1	1	0	0	0	0	1	0	...	1
droit	9	9	10	8	3	2	1	2	...	14
philosophie et lettres	2	1	1	0	0	0	0	0	...	1
arrêtés avant majorité ou humanités	2	2	3	4	2	0	0	1	...	7
école militaire	1	1	0	0	0	0	0	0	...	1
médecine	0	2	0	0	0	0	0	0	...	1
école normale	1	0	1	2	1	0	0	0	...	3
sciences politiques et sociales / psychologie	0	1	0	0	0	0	0	0	...	1
philologie romane	0	0	0	2	1	1	0	0	...	1
histoire	0	0	0	0	1	1	0	1	...	0
histoire de l'art	0	0	0	0	1	0	0	0	...	0
psychologie	0	0	0	0	1	0	0	0	...	1
théologie	0	0	1	1	1	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...	...
prose courte	8	7	6	4	5	0	0	1	...	30

Tableau 5 – Extrait du tableau de Burt (B) des modalités décrivant le personnel littéraire de l'ARLLF

- 51 Enfin, pour projeter les individus sur le même plan factoriel que les modalités descriptives, il faut procéder au même type d'analyse sur la matrice des individus : recréer une matrice carrée en multipliant Z par sa transposée, mais cette fois à droite :  $B' = ZZ'$ . La matrice obtenue,  $B'$ , est de dimension  $i$ , et peut subir le même traitement que la matrice B. On trouve les vecteurs propres linéairement indépendants, et on projette les profils de vecteurs lignes de la matrice  $B'$  sur le plan factoriel. Cependant, il faut rester attentif au fait que les proximités détectées sur le plan factoriel entre les vecteurs issus de la matrice B et ceux issus de la matrice  $B'$  sont trompeuses. De là vient la difficulté d'interprétation : la « qualité » de la projection n'est pas nécessairement la même pour tous les points des deux matrices, et si des points issus du même espace (de la même matrice) ont déjà des qualités de projection différentes, il faut encore redoubler de prudence dans les interprétations de proximité à partir de deux plans factoriels différents, entre des points provenant d'espaces différents (constitués soit par les lignes en fonction des colonnes, soit par les colonnes en fonction des lignes), même si ceux-ci sont liés.

### 3.5. LECTURE ET INTERPRÉTATION DES RÉSULTATS

- 52 Après cette explication des principes de l'analyse, on peut légitimement se demander à quoi une telle analyse peut servir. Pour en voir l'intérêt, il faut passer par une interprétation du plan factoriel obtenu. Différents principes de lecture doivent accompagner celui qui se livre à une interprétation. L'objectif est finalement de donner sens aux oppositions mises en relief par l'analyse sur le graphique. Ces conclusions ne peuvent être tirées que grâce à une bonne connaissance de l'univers analysé par le graphique. Comme l'écrit Philippe Cibois, « il s'agit d'un exercice où l'on fait intervenir de nombreuses connaissances qui ne sont pas entrées en tant que telles dans l'analyse »<sup>42</sup>.
- 53 Il est bon de garder en mémoire les quelques principes de base suivants :

il faut avant tout vérifier la pertinence du plan par rapport aux oppositions saillantes qu'il met en évidence. En général, il est logique de le former à partir des deux premiers axes factoriels, qui mettent au jour les grandes oppositions structurant les données. Ces deux premiers axes sont les deux premiers vecteurs propres de la base qui permettent d'exprimer l'espace multidimensionnel créé par tous les vecteurs du

tableau de départ. Ces deux premiers vecteurs propres reprennent la plus grande part d'inertie pouvant être représentée sur un plan ; de plus, ils opposent fortement les traits distinctifs des modalités, car l'inertie est calculée à partir de la distance  $\chi^2$ , comme on l'a vu, et cette distance est pondérée grâce à la division par la masse de la coordonnée dans l'espace multidimensionnel : plus cette masse sera élevée, moins sa contribution à la distance au centre sera grande. Cette propriété entraîne une plus grande dispersion dans le plan factoriel des profils de points dont le rapport distance – masse de la coordonnée sera élevé, c'est-à-dire dont la masse de la coordonnée sera peu élevée. Néanmoins, il est parfois utile de sélectionner le deuxième et le troisième axes factoriels à la place des deux premiers, si le premier ne met au jour qu'une opposition peu pertinente pour différentes raisons, par exemple parce qu'elle serait trop évidente.

Il faut ensuite trouver un sens aux axes factoriels. Quelles sont les variables qu'ils opposent ? Quelle interprétation peut-on en tirer ?

Il est nécessaire d'attirer l'attention sur les modalités qui contribuent le plus à la formation des axes, que l'on repère grâce à la consultation des tableaux de contribution aux axes calculés par le programme construisant l'AFC. On doit corrélérer ce point au précédent, car les deux permettent une interprétation plus fine des oppositions que les axes dessinent.

Enfin, il est recommandé de considérer uniquement les modalités qui sont « bien » représentées sur le plan. Pour cela, il faut recourir au cosinus carré de la projection du point représentant la modalité sur le plan. Pour chaque modalité, on mesure ainsi la qualité de la représentation de celle-ci sur le plan grâce au cosinus de l'angle formé par le plan principal et le vecteur (le profil de point) dont l'ensemble des coordonnées situent cette modalité dans l'espace multidimensionnel de départ. On recourt donc à la projection (voir la figure 1, p. 35). Lorsqu'une modalité est dans le plan factoriel, l'angle défini est nul, ce qui entraîne un  $\cos^2$  égal à 1. Au contraire, une modalité orthogonale (c'est-à-dire dont le vecteur allant du centre jusqu'au point le représentant dans l'espace multidimensionnel est perpendiculaire aux axes factoriels) à chacun des axes du plan factoriel aura un  $\cos^2$  nul.

Tous les individus [ou modalités] occupant des positions intermédiaires entre ces deux extrémités auront un  $\cos^2$  compris entre 0 et 1, d'autant plus proche de 1 que l'individu est bien représenté par sa projection sur le plan.

En règle générale, les individus contribuant bien aux axes du plan y sont bien représentés (puisque ce sont ceux qui forment les axes). L'indicateur  $\cos^2$  est principalement à utiliser quand on veut interpréter les points centraux (LECOUTRE, 1998).

- 54 Cette démarche d'interprétation ne donne qu'un canevas de base pour permettre au lecteur du résultat de l'analyse factorielle, c'est-à-dire à celui qui ne voit que le plan, de se faire une idée des oppositions que celui-ci met en évidence.
- 55 Enfin, l'un des principaux avantages d'un graphique d'analyse factorielle synthétique, c'est de permettre des rapprochements rapides et une vision d'ensemble efficace. Comme le dit très bien Michel Gollac,

Le plaisir ressenti à la lecture des graphiques est à la fois celui de la connaissance et celui de la liberté. L'analyse statistique permet ici de se dégager des contraintes que l'exposé linéaire fait peser sur la pensée sociologique. Le style de Bourdieu, s'efforçant, dans chaque paragraphe et chaque phrase, de conserver la plus grande part possible de l'ensemble des relations qui constituent l'objet dont il parle, doit être apprécié (entre autres) en regard des difficultés de l'écriture à rendre compte d'une réalité plus immédiatement accessible à des formes plus visuelles d'expression (GOLLAC, 2005, p. 60).

- 56 Gollac insiste aussi sur la possibilité, pour le lecteur du plan factoriel, de « se promener dans l'espace social », ou, pour reprendre l'expression de Bourdieu, d'avoir un court instant le plaisir « de vivre toutes les vies ». Évidemment, la quantité de données à collecter pour s'offrir ce genre de petits plaisirs est énorme, mais que l'on songe combien les sociologues et même (et c'est peut-être ce qui est le plus important) les non-sociologues ont été marqués par la représentation de l'espace social offert dans *La Distinction...* Cette pédagogie du graphique vaut la peine, à notre sens, d'utiliser un outil peut-être délicat et difficile à manipuler, mais surtout puissant et éclairant.
- 57 Il est utile de répéter que l'analyse factorielle permet de prendre en compte l'individu

singulier dans une démarche sociologique qui paraît l'abolir dans de grands ensembles.

Par leur pratique de l'analyse factorielle, Bourdieu et ses disciples ont réconcilié l'attention portée aux individus singuliers avec la démarche quantitative. Nous disposons ainsi de graphiques où figurent des personnages connus par ailleurs, et connus des lecteurs : grands patrons, écrivains, éditeurs. Quels bénéfices en tirons-nous ? Le premier relève du poppérisme ludique : nous, lecteurs, pouvons vérifier en nous amusant que nos connaissances n'invalident pas l'interprétation proposée. Mieux, le « graphique des individus » permet aux lecteurs d'insérer dans l'analyse des variables qui auraient été omises par l'auteur (GOLLAC, 2005, p. 62–63).

- 58 Dans notre cas, l'analyse factorielle tend donc à réconcilier le point de vue individuel de l'auteur avec son inscription sociale. Elle est un puissant outil dans une étude prosopographique telle que la nôtre, fondée sur la recension et l'analyse de données qui permettent de décrire les écrivains. Car l'objectif de l'étude prosopographique est finalement de mettre en évidence la structure du champ littéraire en fonction des attributs qui décrivent les agents.

## Notes

20. BOURDIEU, 1979, p. 109-187.

21. « [...] les individus rassemblés dans une classe qui est construite sous un rapport particulier mais particulièrement déterminant apportent toujours avec eux, outre les propriétés pertinentes qui sont au principe de leur classement, des *propriétés secondaires* qui sont ainsi introduites en contrebande dans le modèle explicatif » (BOURDIEU, 1979, p. 113).

22. « Ce que dit bien Jean Benzécri : "Soit des individus  $\alpha\beta_1\gamma_1$ ,  $\alpha\beta_2\gamma_2, \dots$ ,  $\alpha\beta_n\gamma_n$ , décrits chacun comme possédant trois traits (ou caractères). En faisant abstraction des deux derniers éléments de chaque description, on dira que tous ces individus rentrent dans une seule espèce définie par le trait  $\alpha$ , et qu'on pourra appeler, en bref, l'espèce  $\alpha$ . Mais même si le trait  $\alpha$  permet de définir cette espèce et d'en reconnaître les individus, on ne peut étudier celle-là sans égard aux traits  $\beta$ ,  $\gamma$  de ceux-ci. De ce point de vue, si on note B l'ensemble des modalités  $\beta$  que peut revêtir le deuxième caractère, et C l'ensemble des modalités  $\gamma$  du caractère trois, étudier l'espèce  $\alpha$  ce sera étudier  $\alpha BC$ , c'est-à-dire outre le premier trait qui est fixé, tout ce que peut être le deuxième (B) ou le troisième (C) ; et de plus les associations permises entre ces derniers (de tel  $\beta$  avec  $\gamma$  plutôt qu'avec  $\gamma'$  ou  $\gamma''$ ) » (BENZÉCRI, 1978, cité par BOURDIEU, 1979, p. 113, note 7).

23. On peut remarquer une homologie structurale dans la trajectoire de Bourdieu et Benzécri : promoteurs chacun d'un outil puissant (la théorie des champs pour le premier et l'analyse factorielle pour le second), ils ont fondé chacun leur revue (*Actes de la Recherche en Sciences sociales*, 1975, et *Les Cahiers de l'Analyse des Données*, 1976) et en font un lieu de défense et d'illustration de leur théorie, lieu où ils publient eux-mêmes abondamment.

24. Pour être technique, rappelons que Bourdieu utilise les variables correspondant aux catégories de pensée traditionnelles (profession, etc.) comme variables supplémentaires, « projetées » dans la représentation d'un espace social construit à partir des variables que le sociologue a lui-même créées, en fonction des différences de modes de vie des agents qu'il a observés et interrogés. « In the Correspondence Analyses of *La distinction*, the status variables (age, father's profession, education level, income) are set out (with the exception of social fractions) as *supplementary elements*, that is, they do not participate to the construction of the social space, but they are represented as points in the space of properties. This was done, Bourdieu points out, in order to "give its *full strength*" to the demonstration that the differences in lifestyle are interpretable in terms of status. This means that Bourdieu was deliberately making an *explanatory use of CA* (going against the misbelief that CA is a "merely exploratory" procedure » (ROUANET, ACKERMANN et LE ROUX, 2000, p. 6).

25. BOURDIEU, 1979, p. 140–141.

26. GOLLAC, 2005, p. 59, note 63.

27. GOLLAC, 2005, p. 60.

28. D'où le qualificatif de « relationnel » que Bourdieu emploie tant. Il faut néanmoins anticiper en signalant déjà que l'on retrouvera cet adjectif plus loin dans ce travail, à propos de l'analyse structurale des relations sociales cette fois, mais que les deux sens ne sont pas identiques (la relation n'étant pas de même nature dans les deux cas).

29. BENZÉCRI, 1976, p. 144. Benzécri expose sa position (et sa croyance en son outil) dans plusieurs articles : voir notamment BENZÉCRI 1969 et 1973.

30. Stephen Jay Gould, dans un article très éclairant sur l'usage historique de l'AFC en psychologie, démonte admirablement cette « essentialisation » des facteurs mise en évidence par l'analyse. Voir GOULD, 1997, p. 273–362.

31. BOURDIEU, 1979, p. 114–115.

32. On reviendra sur les modalités de cette lecture (voir section 3.5, p. 45–47). Ce qui est important ici, c'est de souligner cette opération d'interprétation du chercheur qui donne du sens à la configuration de variables qu'il observe.

33. Le fait que l'auteure développe cette question dans une revue anglaise vient probablement du manque de familiarité des lecteurs de la revue avec la méthode employée : par exemple, il n'existe pas d'explication nette de l'adéquation entre concept et outil dans l'article qu'elle publie dans les *Actes de la recherches en sciences sociales* en 1996 (SAPIRO, 1996, p. 3–35).
34. Voir notamment les ouvrages suivants, écrits pour expliquer la méthode aux chercheurs intéressés, mais non spécialistes des mathématiques : CIBOIS, 1983 et GREENACRE et BLASIUS, 1994.
35. Cet exemple est repris de notre mémoire de DEA (DOZO, 2004). Pour la construction de ces variables, voir le mémoire p. 69–71 pour les générations et p. 72–73 et p. 84–87 pour les professions.
36. Figure extraite de GREENACRE, 2004a, p. 16.
37. Pour le lecteur curieux, on peut en donner une interprétation rapide : le graphique permet de repérer trois « groupes » de modalités issus du croisement des deux variables (professions et générations). Le premier est celui autour de la partie supérieure du deuxième axe (première et deuxième générations, fonctionnaires, journalistes, médecins). Le deuxième est à droite : troisième génération, hommes politiques, cadres. Le troisième est en bas à gauche : quatrième, cinquième et sixième générations, enseignants.
38. Pour la définition de cette notion, voir plus bas (voir section 3.5, p. 46).
39. En abrégé, ARLLF.
40. Pour plus de détails, se reporter aux deux articles de GREENACRE, 2004b et 2006.
41. Voir par exemple l'article de SAPIRO, 2002. Dans cet article, elle explique qu'il y a 236 modalités actives pour son ACM, et que les trois premiers facteurs contribuent respectivement pour 5,6 %, 3,9 % et 3,2 % à l'inertie totale. Elle écrit en note : «the low percentage is due to the large number of variables» (p. 389).
42. CIBOIS, 1983, p. 90.

© Presses universitaires de Liège, 2011

Conditions d'utilisation : <http://www.openedition.org/6540>

#### *Référence électronique du chapitre*

DOZO, Björn-Olav. 3. *L'analyse factorielle des correspondances* In : *Mesures de l'écrivain : Profil socio-littéraire et capital relationnel dans l'entre-deux-guerres en Belgique francophone* [en ligne]. Liège : Presses universitaires de Liège, 2011 (généré le 21 octobre 2019). Disponible sur Internet : <<http://books.openedition.org/pulg/2253>>. ISBN : 9782821896338. DOI : 10.4000/books.pulg.2253.

#### *Référence électronique du livre*

DOZO, Björn-Olav. *Mesures de l'écrivain : Profil socio-littéraire et capital relationnel dans l'entre-deux-guerres en Belgique francophone*. Nouvelle édition [en ligne]. Liège : Presses universitaires de Liège, 2011 (généré le 21 octobre 2019). Disponible sur Internet : <<http://books.openedition.org/pulg/2241>>. ISBN : 9782821896338. DOI : 10.4000/books.pulg.2241.

Compatible avec Zotero