

Mónica Bécue-Bertaut¹, Jérôme Pagès²

¹ EIO–UPC– c/ Pau Gargallo, 5 – 08028 Barcelona – Spain

²ENSA/ INSFA.– 65 rue de Saint-Brieuc, CS 84215 – F35042 Rennes cedex – France

In socio-economic surveys, some topics are tackled using jointly open-ended and closed questions. Here, an original methodology is proposed to analyse together closed and open-ended questions. So, Multiple Factor Analysis is extended in such a way that both contingency tables and indicator variables tables can be jointly analysed. An example illustrates this methodology.

Dans les enquêtes socio-économiques, il est utile d'aborder certains thèmes au moyen d'une batterie de questions fermées complétée par une ou deux questions ouvertes. Dans ce travail, nous proposons une méthodologie pour analyser l'ensemble de ces questions comme actives. Pour cela, l'Analyse Factorielle Multiple est étendue afin d'inclure des tableaux de contingence et des tableaux de variables indicatrices. La méthodologie est présentée et illustrée par un exemple.

Analyse de tableaux multiples, Tableaux de contingence multiples, Analyse Factorielle Multiple, Questions ouvertes

Multiple tables analyse, Multiple contingency tables, Multiple Factorial Analysis, Open-ended questions

Pour aborder certains thèmes complexes, il est utile de compléter une batterie de questions fermées par une ou deux questions ouvertes.

Le traitement statistique des questions ouvertes et des questions fermées (voir Lebart et Salem, 1994 pour une présentation générale des travaux) part de la codification des réponses en tableaux Individus \times Variables (ou Modalités). Les réponses fermées sont codées dans un tableau disjonctif complet. Chacune des questions ouvertes, limitées aux mots suffisamment fréquents, est codée comme un tableau de contingence dans lequel la case (i, k) contient la fréquence du mot k dans la réponse de l'individu i .

L'approche classique (Benzécri, 1981 ; Lebart, 1982 ; Lebart et Salem, 1994) ne permet pas de traiter conjointement les deux types de tableaux (i.e. les deux types de questions) tout en leur donnant un rôle équivalent. Elles privilégient l'un d'entre eux, mettent en évidence sa structure, puis illustrent les résultats obtenus (plans factoriels ou typologies) par l'information contenue dans les autres tableaux.

Nous proposons dans ce travail une méthode originale qui permet de traiter conjointement un ensemble de tableaux juxtaposés de différentes natures en donnant à chacun d'eux une importance équivalente. Cette méthode permet d'analyser globalement les individus à partir de

l'ensemble des caractéristiques choisies, de mettre en évidence la structure globale et aussi de comparer les structures induites par chacun des tableaux pris séparément pour déterminer les traits communs à l'ensemble ou, au contraire, propres à certains d'entre eux. Appliquée aux données textuelles, cette méthode permet de traiter conjointement des questions ouvertes et des questions fermées.

La section 2 décrit l'exemple utilisé, *Los niños y la lectura*, une enquête par questionnaire effectuée à Barcelone auprès d'enfants de 5^{ème} année d'enseignement général. La section 3 précise le tableau multiple à analyser et la notation employée. On rappelle ensuite les résultats obtenus avec l'approche classique (section 4). La section 5 présente L'Analyse Factorielle Multiple de tableaux comportant des sous-tableaux de contingence et des sous-tableaux de variables indicatrices, après avoir rappelé brièvement les principes de l'Analyse Factorielle Multiple (AFM) et de son extension aux Tableaux de Contingences Multiples (AFMTC). Finalement, la section 6 offre les résultats obtenus sur l'exemple tout en montrant les aides à l'interprétation disponibles.

Une vaste étude a été conduite dans une ville de banlieue de Barcelone (Rajadell, 1990) afin de mieux connaître les pratiques et les attitudes envers la lecture des enfants. En particulier, 895 enfants de 5^{ème} année d'Enseignement Général de Base (10-11 ans) de cette ville répondent à un questionnaire fermé, comportant une batterie de questions sur les habitudes de lecture (voir tableau 2), et complètent les deux phrases suivantes :

- *Para mí, leer es...* (pour moi, lire est...)
- *Creo que leer es importante porque...*(Je crois que lire est important parce que...)

Le tableau 1 montre quelques-unes des réponses fournies par les enfants.

Les enfants ont aussi répondu à des tests d'aptitude. En outre, on dispose des notes scolaires et de quelques informations sur la famille.

Tableau 1. Quelques exemples de réponses

<p>----A101 <i>fantástico! a mí me gusta leer mucho para luego contar lo que he leído. si cada día leo, cuando me toque leer en clase sabré leer rápido</i> ++++</p> <p><i>sino luego al final de curso no sabré leer bien</i> ----A122 <i>una diversión porque así no se pierde el tiempo</i> ++++</p> <p><i>después te encuentras algo peligroso como perro peligroso en una fabrica y lo ponen en un cartel y es bueno para el estudio y la imaginación</i> ----A130 <i>es un rollo porque hay muchas letras</i> ++++</p> <p><i>si no sabes leer no sabes escribir</i></p>

3.1. Tableau multiple à analyser

Le tableau 2 schématise le tableau multiple construit à partir des données. Aux questions fermées actives (cf. tableau 3), on associe un tableau disjonctif complet. A chaque question ouverte, on associe un tableau de contingence qui contient la fréquence avec laquelle chacun des mots est employé par chacun des individus pour répondre à la question.

Tableau 2. Tableau multiple à analyser

Individus	Variables indicatrices $K_1=23$ <i>K₁ colonnes-modalité</i>	Fréquence des mots dans les réponses à <i>Para mí, leer es...</i> $K_2=91$ <i>K₂ colonnes-mot</i>	Fréquence des mots dans les réponses à <i>Creo que leer es importante porque</i> $K_3=92$ <i>K₃ colonnes-mot</i>
-----------	--	---	--

3. . Notations

Pour le tableau *individus* × *variables indicatrices*, x_{ikq} indique si le répondant i présente ($x_{ikq}=1$) ou non ($x_{ikq}=0$) la modalité-colonne k de la question q . Plus généralement, les variables qualitatives peuvent être structurées selon J_q groupes.

Pour les J_c tableaux de contingence f_{ikj} indique la fréquence avec laquelle la réponse-ligne i emploie le mot-colonne k pour répondre à la question correspondant au tableau j .

Tableau 3. Questions fermées actives

1 . <i>A l'école, nous lisons</i>	(un peu, assez, beaucoup)
2 . <i>Chez moi, nous avons</i>	(peu de livres, assez, beaucoup)
3 . <i>Je lis</i>	(un peu, moyennement, beaucoup)
4 . <i>Je lis avec</i>	(facilité, un peu de difficulté, grandes difficultés)
5 . <i>Les livres d'école donnés par l'instituteur</i>	(me plaisent, ne me plaisent pas)
6 . <i>Je lis quand</i>	(j'en ai envie, quand je travaille, les deux)
7 . <i>Je préfère lire</i>	(en silence, à voix haute, des deux façons)
8 . <i>Lire les manuels scolaires</i>	(me plaît, ne me plaît pas, me plaît quelquefois)

Les méthodes classiques d'analyse factorielle permettent d'analyser un tableau de données dont les colonnes sont de même nature (Analyse de correspondances pour un tableau de contingence, Analyse en Composantes Principales pour un tableau de variables continues, Analyse de Correspondances Multiples pour un tableau disjonctif complet). Ces méthodes mettent en évidence la structure du tableau ; les lignes et les colonnes sont qualifiées d'actives;

ensuite, il est possible d'illustrer ensuite les résultats en employant des colonnes (ou des lignes) de tableaux supplémentaires, éventuellement de nature différente.

L'application de ces méthodes aux données textuelles offre deux possibilités. On peut donner aux modalités des questions fermées relatives à la lecture le statut de modalités actives, réaliser une Analyse des Correspondances Multiples (ACM) sur le tableau disjonctif correspondant, puis illustrer les plans factoriels en projetant les colonnes-mots, correspondant aux tableaux de contingence créés à partir des questions ouvertes, comme colonnes supplémentaires. On peut aussi analyser le tableau de contingence correspondant à l'une des questions ouvertes au moyen d'une Analyse Factorielle des Correspondances (AFC) pour ensuite projeter les variables indicatrices des différentes opinions sur la lecture comme éléments supplémentaires.

Nous montrons aux sections 4.1 et 4.2 les résultats fournis par ces deux démarches.

4.1. Questions fermées actives

On réalise une ACM en prenant comme actives les questions relatives à l'attitude envers la lecture (cf. tableau 3). Les mots employés dans les réponses ouvertes sont considérés comme des colonnes-fréquences supplémentaires. La figure 1 présente le premier plan factoriel obtenu.

4.1.1. Lecture du premier plan factoriel : les modalités actives

On choisit d'interpréter les bissectrices ce qui n'est pas en contradiction avec les pourcentages d'inertie assez proches des deux axes.

La seconde bissectrice suit la progression de l'intensité de lecture : *je lis peu, je lis assez, je lis beaucoup*. Les élèves qui déclarent *je lis peu*, rencontrent de grandes ou moyennes difficultés à lire (*je lis peu* occupe une position intermédiaire entre les modalités *je lis avec difficulté* et *je lis avec quelques difficultés*). Par contre, les « gros » lecteurs (*je lis beaucoup*) lisent avec facilité.

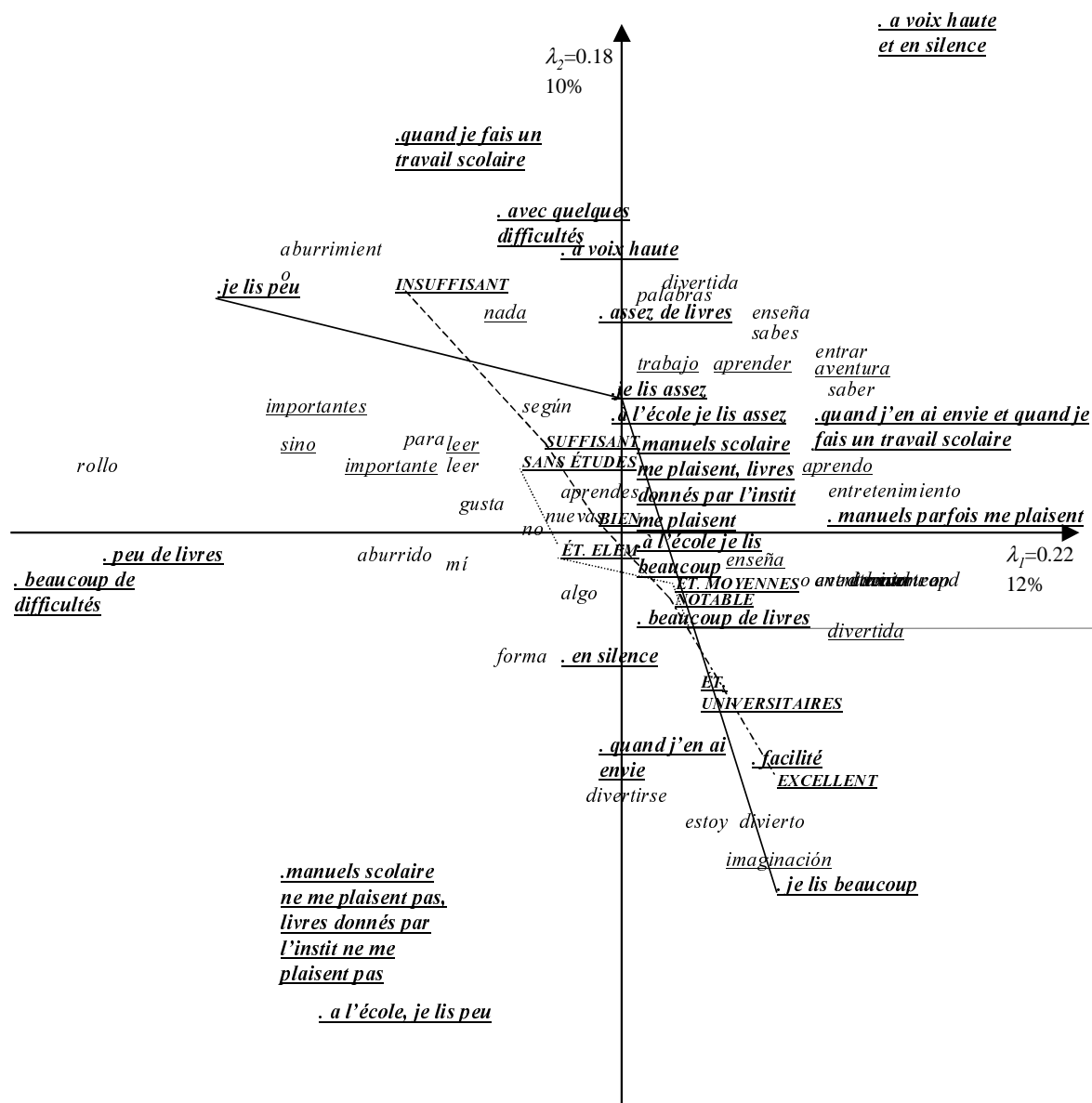
La première bissectrice concerne l'intensité de lecture scolaire : elle oppose la modalité *à l'école je lis peu* aux modalités *à l'école je lis assez* et *à l'école je lis beaucoup*. Les élèves qui déclarent lire peu à l'école sont souvent des enfants qui déclarent *lire les manuels scolaires ne me plaît pas* et *les livres donnés par l'instituteur ne me plaisent pas* ; tandis que ceux qui lisent assez ou beaucoup à l'école ont une opinion plus positive sur les textes de lecture scolaire : ils déclarent fréquemment *lire les manuels scolaires me plaît* et *les livres donnés par l'instituteur me plaisent*.

Les trois modalités de la variable *je lis quand* (*quand je fais un travail scolaire, quand j'en ai envie, les deux*) forment un triangle. Les petits lecteurs ont tendance à ne lire que dans le cadre scolaire. Parmi les moyens et gros lecteurs on trouve des enfants qui associent clairement la lecture à une activité de loisir (*je lis quand j'en ai envie*) et d'autres qui considèrent la lecture comme un loisir mais aussi comme liée au travail scolaire (*les deux*).

4.1.2. Lecture du premier plan factoriel : les modalités fermées illustratives

Les notes scolaires, projetées en tant que colonnes supplémentaires sur ce premier plan factoriel, augmentent le long de la deuxième bissectrice, de *insuffisant* à *excellent* : l'intensité déclarée de lecture et les résultats scolaires sont très liés, alors que ceux-ci ne sont guère associés avec la lecture scolaire. Il est intéressant de noter que le niveau d'études du père progresse parallèlement aux notes scolaires.

Les bons élèves, dont le père est souvent plus diplômé que la moyenne, ont tendance à se déclarer gros lecteurs ; mais que signifie lire pour eux ? Et pour les autres ? Les réponses fermées ne traduisent-elles pas trop une vision scolaire de la lecture ? C'est ce que l'on cherche à connaître à travers les réponses aux questions ouvertes.



On ne s'étonne pas du fait que beaucoup de petits lecteurs qualifient la lecture de très ennuyeuse (avec les mots *rollo*, *aburrimiento*). Quand on les interroge sur l'importance de la lecture, ces enfants emploient fréquemment le mot *sino* : *porque sino no sabemos leer, sino serias analfabeto, sino no tendrás trabajo*, etc.

Les petits lecteurs scolaires (quadrant inférieur gauche) ne sont guère caractérisés par les mots qu'ils emploient : en fait, ils ne sont pas très nombreux (86 enfants choisissent la modalité à l'école je lis peu).

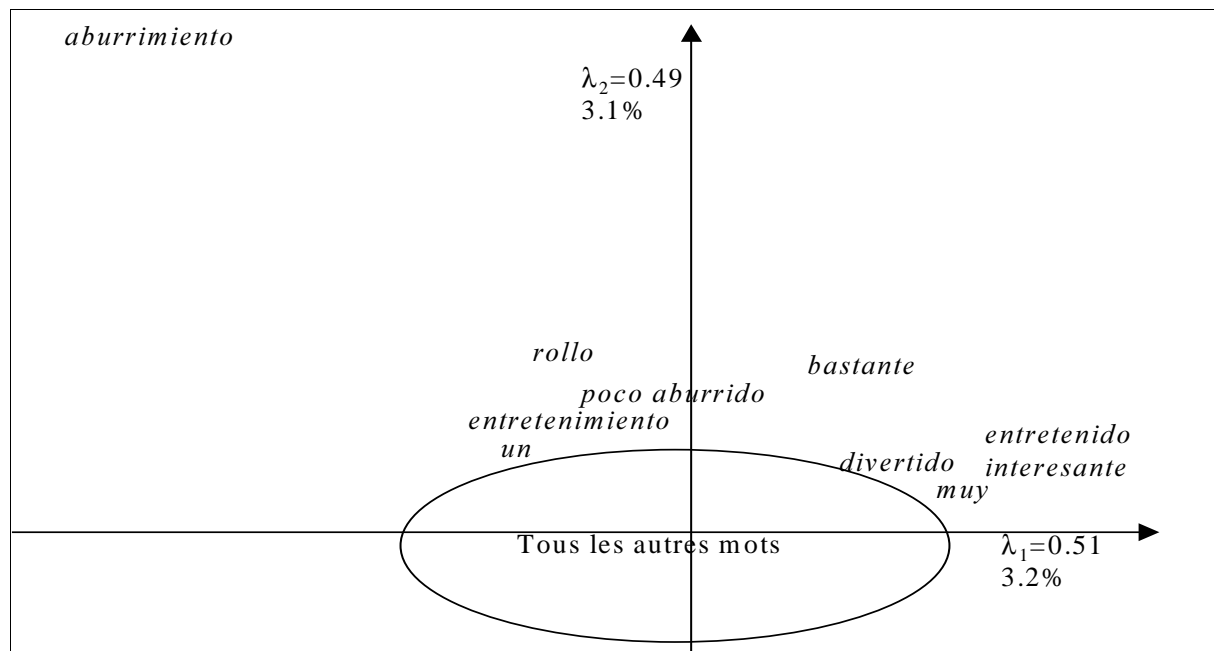
Dans la partie positive du premier axe, zone des moyens et gros lecteurs, on trouve des mots qui expriment bien une opinion positive sur la lecture : *divertido*, *divertida*, *bonito entretenimiento*, *entretenido*, *fantasía*, *aventura*, *imaginación*, *entrar*, *mundo*, etc -. Néanmoins, ces mots se trouvent mélangés avec d'autres :

- qui montrent un intérêt beaucoup plus modéré pour la lecture (*pasar rato*, de l'expression *pasar el rato*),
- qui insistent sur le rôle de la lecture dans l'apprentissage scolaire (*aprendo*, *aprender*, etc.).

En outre, les mots qui semblent mieux traduire un réel goût pour la lecture (*fantasía*, *aventura*, *imaginación*, *entrar*, *mundo*) occupent des positions moyennes sur la deuxième bissectrice, qui constitue un axe d'intensité de lecture.

4. . Questions ouvertes actives

Dans une deuxième analyse, on choisit de prendre la première question ouverte comme question active. Le tableau analysé par une analyse des correspondances simples est le tableau de contingence croisant les réponses individuelles et les mots prononcés au moins 10 fois dans l'ensemble des réponses à cette question. La figure 2 reproduit le premier plan factoriel issu de cette analyse.



INCORPORER

Figure 2. Premier plan factoriel issu de l'analyse de correspondances du tableau de contingence correspondant à la première question ouverte.

Dans cet exemple, le très petit groupe d'individus (sept enfants) qui donnent comme réponse *un aburrimiento* s'oppose à l'ensemble de l'échantillon sur les premier et deuxième axes ; cette opposition est due au fait qu'ils emploient exactement le même mot (*aburrimiento*) et que ce mot est très peu employé par les autres répondants (seulement 3 fois).

Si on déclare le mot *aburrimiento* comme mot supplémentaire, un autre mot tend à jouer un rôle équivalent. Ce problème est fréquent lors de l'analyse directe des réponses : lorsque les réponses sont courtes, il est habituel d'obtenir des premiers axes dus à de petits noyaux d'individus trop influents. Ceci nous a conduit à rejeter cette méthodologie pour cette étude.

Pour réaliser une analyse factorielle qui tienne à la fois compte des questions fermées et des questions ouvertes, il est nécessaire d'équilibrer l'influence a priori des tableaux de diverses natures dans la construction des axes.

Ceci nous conduit à étendre l'Analyse Factorielle Multiple (AFM) (Escofier et Pagès, 1988, 1994, 1998), brièvement résumée à la section 5.1, aux tableaux multiples comprenant des tableaux de contingence et des tableaux disjonctifs complets. Pour introduire les tableaux de contingence, on reprend la démarche proposée dans *l'Analyse Factorielle Multiple pour Tableaux de Contingence (AFMTC)* (Bécue & Pagès, 1999) (section 5.2). Ceci permet de considérer dans une même analyse des tableaux disjonctifs complets et des tableaux de contingence (section 5.3).

5.1. Brève présentation de l'AFM

L'Analyse Factorielle Multiple (AFM) permet d'analyser un tableau multiple : un ensemble d'individus est décrit par J groupes de variables, quantitatives ou qualitatives. L'AFM réalise une analyse factorielle du tableau juxtaposé, en surpondérant les poids des variables de façon à équilibrer l'importance de chacun des groupes.

Cette méthode offre des résultats :

- analogues à ceux de l'Analyse de Correspondances Multiples (ACM), principalement une représentation globale des lignes-individus et des colonnes- modalités.
- spécifiques des tableaux multiples, principalement la représentation superposée des structures induites par chacun des groupes de variables sur l'ensemble des individus (ou catégories d'individus) –appelées structures partielles– et la représentation des facteurs obtenus lors de analyses séparées.

5. . Introduction des tables de contingence (AFMTC)

Une méthodologie pour l'analyse simultanée d'un ensemble de tableaux de contingence a été proposée par Bécue & Pagès (1999). Elle reprend les principes de l'Analyse Factorielle de Correspondances Intra-Tableaux (AFCIT) (Escofier & Drouet 1983 ; analyse étendue aux tableaux partitionnés, sous le nom d'Analyse de Correspondances Internes par Cazes *et al.*, 1988), et élimine ainsi les différences entre les marges en lignes, et ceux de l'AFM, ce qui permet d'équilibrer l'influence des différents tableaux et fournit des représentations graphiques spécifiques de la structure en sous-tableaux.

5.3. Tableau multiple comportant des sous-tableaux de contingence et des sous-tableaux de variables indicatrices

Pour introduire à la fois des tableaux de contingence et des tableaux de variables indicatrices dans une même analyse, on considère les tableaux de contingence comme en AFCIT et les tableaux de variables indicatrices comme en ACM.

Dans toutes les analyses séparées (AFC ou ACM, selon le type de tableau), les lignes ont le même poids, calculé à partir de la marge en ligne des tableaux de contingence juxtaposés. On reprend ainsi la solution proposée par Abdessemed & Escofier (1996) pour analyser conjointement des tableaux de variables qualitatives et un tableau de fréquences (ou plusieurs tableaux de fréquences, mais aux marges égales).

Ensuite, on suit la démarche de l'AFM pour équilibrer l'influence de chaque tableau.

Cette analyse globale offre des résultats :

- analogues à ceux de l'AFC appliquée aux tableaux juxtaposés (principalement une représentation globale des lignes-individus et des colonnes-modalités et colonnes-mots) ;
- spécifiques des tableaux multiples (principalement une représentation superposée des structures induites par chacune des questions ouvertes et par chacun des groupes de questions fermées –appelées structures partielles– et la représentation des facteurs des analyses séparées).

La lecture des résultats est facilitée par les nombreuses aides à l'interprétation de l'AFM.

6.1. Analyses séparées

Dans les analyse séparées, les valeurs-propres valent : 0.22 (groupe des questions fermées), 0.51 (groupe de la première question ouverte) et 0.49 (groupe de la deuxième question ouverte). La différence entre ces valeurs justifie la nécessité d'équilibrer l'influence des groupes.

6. . Analyse globale

La figure 3 montre un extrait de la représentation des colonnes sur le premier plan factoriel. Les mots soulignés correspondent à la seconde question ouverte.

Les trois groupes de colonnes contribuent à la construction du premier axe avec, respectivement, 46%, 31% et 23% de l'inertie. Il s'agit d'une direction de dispersion commune aux trois groupes, non confondue avec les premiers axes des analyses séparées ; en effet, la première valeur propre de l'analyse globale ne vaut que 1.4, loin de la valeur maximum 3, qui indiquerait que les premiers axes des analyses séparées correspondent à un axe commun.

L'apport du groupe 1, «questions fermées», à la construction du deuxième axe est faible (18.5%), les groupes 2 et surtout le 3, « questions ouvertes », apportent une contribution plus importante (respectivement 32% et 49%).

6.2.1. Lecture du premier plan factoriel : les modalités et mots actifs

Le premier axe un axe de niveau de lecture. Sur la gauche de cet axe, les mots les plus extrêmes indiquent un fort goût pour la lecture ; même venant de la deuxième question ouverte, qui traite de l'importance de la lecture, on trouve des mots qui correspondent au plaisir de la lecture. D'autre part, les modalités de lecture *Je lis peu*, *Je lis assez* et *Je lis beaucoup* se retrouvent ordonnées de droite à gauche sur cet axe. Si l'on compare avec la figure 1, on peut

voir que dans cette deuxième projection les modalités *Je lis assez* et *Je lis beaucoup* se trouvent plus proches ; cela est dû au fait que les profils moyens des réponses ouvertes de ces deux groupes sont assez similaires et que ces profils interviennent maintenant dans la détermination des plans de représentation. Pour cette même raison, les associations entre mots sont mieux mises en évidence sur le premier plan issu de l'AFM : par exemple, *entrar* et *mundo* (souvent employés dans l'expression *entrar en un mundo*) occupent des positions proches sur la figure 3 alors que ce n'était pas le cas lorsque leur rôle était illustratif (figure 1). De même, les synonymes tendent à occuper des positions voisines, car ils sont souvent employés dans un même contexte, et ceci d'autant plus qu'ils sont employés par des catégories similaires puisque les réponses aux questions fermées sont aussi prises en compte.

On peut noter que, sur ce premier axe, les catégories fermées occupent des positions plus centrales que les mots : les enfants qui présentent les mêmes caractéristiques fermées usent un vocabulaire variable.

Le deuxième axe, principalement dû à la deuxième question ouverte, oppose une vision scolaire à la lecture vue comme un plaisir.

6.2.2. Lecture du premier plan factoriel : les modalités fermées illustratives

Après avoir décrit les attitudes envers la lecture, en tenant compte à la fois des réponses fermées et ouvertes, toute l'information complémentaire sur les enfants peut être illustrer les résultats et faciliter leur interprétation.

Ainsi, la progression des notes scolaires suit le premier axe, tout en occupant des positions plus centrales que le niveau de lecture.

On remarque aussi que *Père universitaire* et *Classe sociale élevée* (auto-déclaration, 3 niveaux : basse, moyenne, élevée) se détachent vers le haut du deuxième axe qui était peu caractérisé par les modalités fermées actives. Les enfants qui possèdent ces caractéristiques ont tendance à privilégier l'utilité de la lecture pour l'apprentissage scolaire.

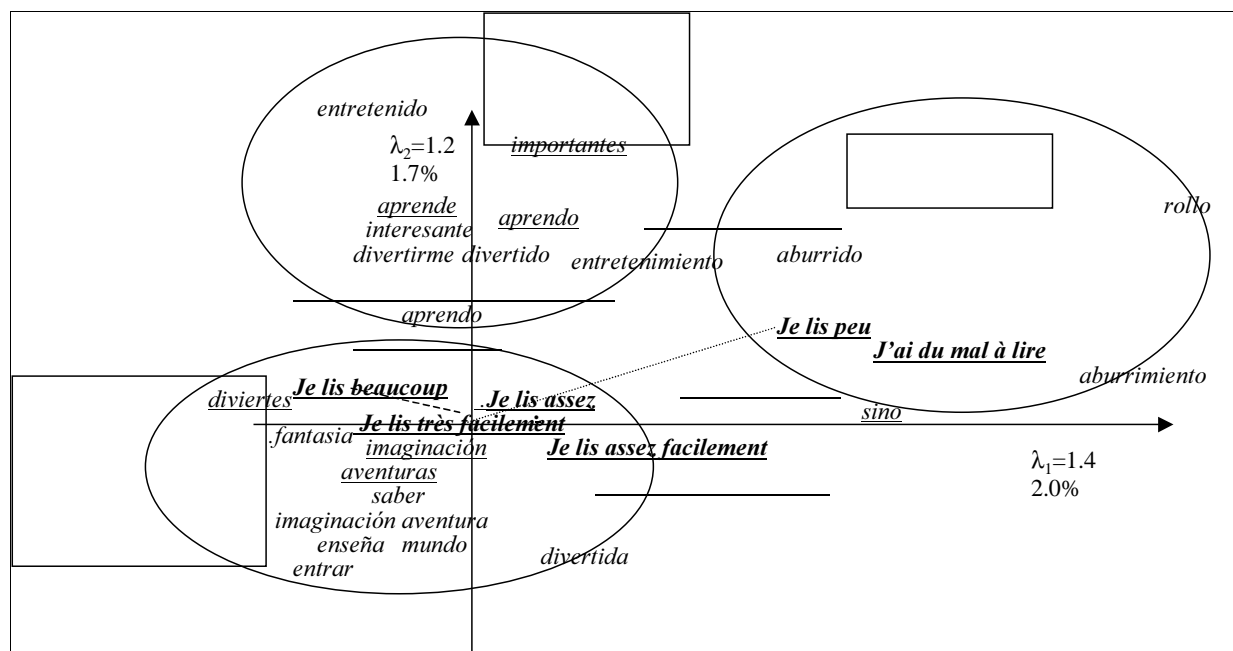


Figure 3. Premier plan issu de l'AFM du tableau multiple

6.3. Corrélations entre les facteurs des analyses séparées

Les corrélations entre les premiers axes des trois analyses séparées sont faibles : ces analyses ne peuvent pas être facilement comparées. L'apport relativement équilibré de chaque groupe au premier axe global de l'AFMTC montre que cet axe correspond à une direction de dispersion commune aux trois groupes, direction qui ne pourrait pas être mise en évidence par les analyses séparées.

6.4. Comparaison des structures partielles

6.4.1. Représentation superposée des nuages partiels

Pour comparer les trois structures, on utilise une représentation (figure 4) qui superpose la description globale des individus (ou des catégories d'individus) et celles induites par chaque tableau (on parle alors de catégories partielles). Cela permet de mettre en évidence les convergences et les divergences entre les groupes de questions.

La lecture de cette représentation superposée s'intéresse tout particulièrement aux catégories d'individus qui ont un comportement très différencié selon les questions, comme les bons élèves (note globale excellente) et les enfants de père universitaire.

Ainsi les bons élèves sont de bons lecteurs selon les questions fermées (groupe 1), alors qu'ils occupent une position plus centrale selon les questions ouvertes (groupes 2 et 3): en fait, ce groupe est hétérogène et comprend des enfants enthousiastes de la lecture et des enfants nettement moins intéressés. Face au questionnaire fermé, ils adoptent une attitude scolaire, induite par leur statut de « bon élève » et tendent à donner les « bonnes réponses ». Les réponses ouvertes, plus spontanées, traduisent davantage leur véritable goût.

La catégorie "père universitaire" occupe une position extrême sur le second axe lorsqu'elle est décrite par la deuxième question ouverte (groupe 3) : la plupart de ces enfants a bien intégré l'importance de la lecture dans les enjeux scolaires, ce qui peut être dû à l'environnement familial, alors que leurs réponses fermées (groupe 1) ne les situent que comme lecteurs modérés et que leurs réponses à la première question ouverte (qui concerne le plaisir de la lecture) les positionnent près du centre de gravité, loin des enthousiastes de la lecture.

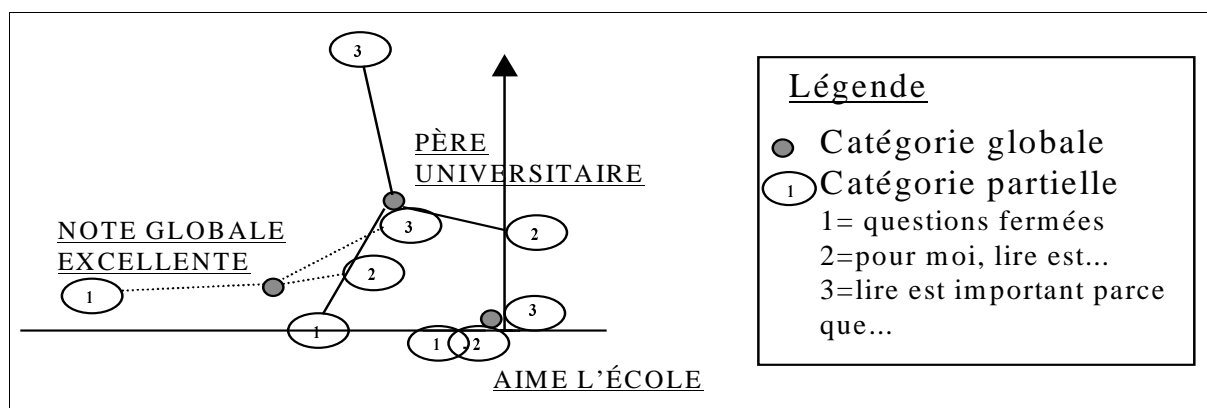


Figure 4. Extrait de la superposition de la représentation globale des catégories et de leurs représentations partielles, induites par chacun des groupes

Il est clair que, dans cet exemple, il existe une tendance très marquée à donner la "réponse correcte", à se déclarer "bon lecteur", peut-être à cause du cadre scolaire dans lequel s'est

réalisé l'enquête. Le mode de questionnement ouvert permet moins de se cacher derrière des réponses conventionnelles.

La méthodologie présentée offre un nouveau point de vue sur le traitement des questions ouvertes en prenant comme actives simultanément des questions fermées et des questions ouvertes. La comparaison des structures partielles, induites par chaque groupe de questions, permet d'appréhender les aspects que chaque type de question est capable ou non de capter.

Cette démarche permet une analyse des réponses non regroupées en fonction de catégories choisies *a priori*. L'introduction simultanée des questions fermées et des questions ouvertes permet d'éviter les écueils rencontrés lors de l'analyse directe des réponses ouvertes seules, en particulier la construction d'axes très instables, influencés par la présence de petits groupes aux réponses très spécifiques.

Abdessemed L. and Escofier B. (1996). Analyse factorielle multiple de tableaux de fréquences; comparaison avec l'analyse canonique des correspondances. *Journal de la Société de Statistique de Paris*, vol.(137-2):3-18.

Bécue M. and Pagès J. (1999). Intra-sets multiple factor analysis. Application to textual data. In J. Jansen *et al.*, editors, *Proc. of ASMDA'99 (9th International Symposium on Applied Stochastic Models and Data Analysis)*, pages 51-60.

Benzécri J.-P. (1981). *Pratique de l'analyse des données*, T.3, Linguistique & Lexicologie. Dunod.

Cazes P. and Chessel D. and Doledec S. (1988). L'analyse des correspondances internes d'un tableau partitionné : son usage en hydrobiologie. *Rev. de Statistique Appliquée*, vol. (XXXVI-1):39-54.

Escofier B. and Drouet D. (1983). Analyse des différences entre plusieurs tableaux de fréquence. *Les cahiers de l'analyse des données*, vol (8-4):491-499.

Escofier B. and Pagès J. (1988-1998). *Analyses factorielles simples et multiples; objectifs, méthodes et interprétation*. Dunod.

Lebart L. (1982). L'analyse statistique des réponses libres dans les enquêtes socio-économiques. *Consommation* vol. (1):39-62.

Lebart L. and Salem A. (1994). *Statistique textuelle*, Dunod.

Rajadell Puiggros N. (1990). *Les actitudes envers la lectura, un model d'anàlisi a l'educació primària*. Thèse Universidad de Barcelona.

