
Des méthodes statistiques originales pour analyser les conceptions d'enseignants de plusieurs pays à partir d'un questionnaire sur des questions vives

François Munoz*, Pierre Clément*

** LIRDHIST (Laboratoire Interdisciplinaire de Recherche en Didactique et en Histoire des Sciences et des Techniques), Université Claude Bernard Lyon 1 (UCBL), 43, Boulevard du 11 novembre 1918
69622 Villeurbanne Cedex
fmunoz@netcourrier.com
Pierre.Clement@univ-lyon1.fr*

RÉSUMÉ. Le questionnaire Biohead-Citizen inclut des questions vives relatives à l'éducation à l'environnement, à la santé, à la sexualité, aux déterminismes génétiques et épigénétiques, et aux théories de l'évolution. Il a été proposé à plus de 6000 enseignants et futurs enseignants de 16 pays. Nous présentons ici nos principaux choix méthodologiques pour l'analyse de ces données, avec le souci de répondre clairement à des hypothèses théoriques précises. L'analyse statistique multivariée est le socle de notre approche, et permet d'exprimer les conceptions des individus en terme d'interaction plus ou moins forte entre valeurs et connaissances scientifiques. L'analyse inter-classe étudie les différences de conceptions entre groupes d'individus, et l'analyse sur variable instrumentale explore plus généralement la relation entre les conceptions et le contexte socioculturel. Des résultats issus de l'analyse générale du questionnaire Biohead-Citizen illustrent notre propos méthodologique.

MOTS-CLÉS : Analyse multivariée, ACP, Analyse inter-classe, Variable instrumentale, Didactique des sciences, Biohead-Citizen.

1. Introduction

Le schéma méthodologique proposé ici se situe volontairement dans la perspective d'un raisonnement et d'un cheminement scientifique propres à notre recherche dans le cadre du projet européen Biohead-Citizen (Carvalho et al 2005), afin de montrer qu'il est possible de choisir des outils adéquats pour répondre aussi clairement et objectivement que possible aux questions posées.

L'objectif du questionnaire Biohead-Citizen est l'étude des conceptions d'enseignants et futurs enseignants en tant qu'interaction entre connaissances scientifiques (K), valeurs (V) et pratiques sociales (P) (modèle KVP ; Clément, 1998, 2004). Nous traitons des thèmes relatifs à des questions vives sur l'éducation à l'environnement, à la santé et à la biologie. Nous étudions également le lien entre ces conceptions et des paramètres individuels contrôlés (âge, sexe, niveau de formation, matière enseignée, croyance religieuse, opinions politiques, économiques et sur la défense de l'environnement...).

L'analyse statistique multivariée est le socle de notre approche, car elle est spécialement adaptée à l'analyse quantitative fine des réponses à un important questionnaire. Après avoir présenté le contenu du questionnaire et le contexte d'échantillonnage, nous introduirons la notion d'*orientation conceptionnelle* synthétique, issue de l'analyse multivariée.

L'analyse inter-classe permet spécifiquement d'élucider les différences de conceptions entre groupes d'individus, et l'analyse sur variables instrumentales étudie plus généralement la relation entre orientations conceptionnelles et différents paramètres personnels des individus. Sans entrer dans des développements techniques détaillés, nous illustrons notre approche par des résultats originaux issus de l'analyse générale du questionnaire Biohead-Citizen.

2. Le questionnaire Biohead-Citizen

Le questionnaire traite différents thèmes, principalement l'éducation à l'environnement, l'éducation à la santé, l'éducation à la sexualité, les déterminismes génétiques et épigénétiques de traits relatifs à la personnalité humaine, ainsi que les théories de l'évolution. Les questions sont organisées en deux sections principales, notées A (l'ensemble des thèmes + des compléments sur l'Environnement) et B (compléments sur la santé et les questions biologiques), et sont proposées aux enseignants et futurs enseignants des pays participant au projet (actuellement 16 pays ont traité la section A, 12 d'entre eux ayant aussi traité la section B).

L'échantillonnage a été conçu de manière à représenter de manière équilibrée six catégories d'enseignants, suivant deux classes de statut (enseignant en exercice, *In*, vs. futur enseignant, *Pre*) et trois classes de discipline (primaire, *P*, secondaire lettres, *L*, et secondaire biologie, *B*). Le Tableau 1 indique les effectifs des individus échantillonnés par pays et par groupe d'enseignement.

Pays	InB	InL	InP	PreB	PreL	PreP	Total
PT-Portugal	49	62	67	53	61	59	351
FI-Finlande	66	28	32	55	69	56	306
HU-Hongrie	52	68	52	56	57	49	334
LI-Lituanie	48	58	59	50	48	53	316
MT-Malte	27	51	53	14	27	26	198
PL-Pologne	50	57	53	50	48	53	311
RO-Roumanie	44	42	46	49	45	47	273
MA-Maroc	66	42	49	62	50	61	330
FR-France	100	110	114	149	101	158	732
SN-Sénégal	52	54	54	62	51	51	324
GE-Allemagne	55	80	60	54	51	65	365
IT-Italie	53	107	97	66	54	182	559
CY-Chypre	38	97	27	25	60	75	322
ES-Estonie	31	31	31	30	30	30	183
LB-Liban	153	111	246	59	56	97	722
TN-Tunisie	125	116	106	193	110	103	753

Tableau 1. Effectifs des échantillons dans le cadre de la recherche Biohead-Citizen, par pays (lignes) et par groupes d'enseignement (colonnes). InB = enseignants en exercice dans le secondaire en biologie, PreB = futurs enseignants dans le secondaire en biologie, InL = enseignants en exercice dans le secondaire en lettres, PreL = futurs enseignants dans le secondaire en lettres, InP = enseignants en exercice dans le primaire, PreL = futurs enseignants dans le primaire.

Les réponses aux questions sont recueillies et codées suivant une échelle quantitative, en général de 1 à 4 sur une gradation entre « d'accord » et « pas d'accord » (échelle de Likert), ou de 1 à 5 dans le cas d'un différenciateur sémantique d'Osgood entre deux adjectifs antonymes.

Le tableau des réponses récapitule, pour chaque individu en ligne, les valeurs codées de ses réponses aux questions, en colonnes. L'objectif est l'analyse du tableau afin d'identifier les principales orientations conceptionnelles qui émergent de l'ensemble des réponses, c'est-à-dire des ensembles cohérents de réponses traduisant des conceptions particulières. Nous présentons ci-après l'intérêt de l'analyse statistique multivariée pour une telle recherche. Nous soulignerons en particulier la nécessité d'utiliser des méthodes adéquates pour tester des hypothèses précises sur la nature des conceptions.

3. Orientations conceptionnelles

L'analyse exploratoire multidimensionnelle (Lebart, Morineau & Piron, 1995) permet d'étudier la structure des réponses à un grand nombre de questions. La synthèse ainsi réalisée met en exergue les redondances ou les corrélations possibles

entre questions, pour lesquelles on obtient des réponses globalement similaires (corrélation positive) ou dissimilaires (corrélation négative). On identifie les orientations conceptionnelles les plus importantes, c'est-à-dire celles pour lesquelles il y a une forte variation des réponses entre les individus.

Les outils statistiques les plus utilisés sont l'*Analyse en Composantes Principales* (ACP), pour les données quantitatives, et l'*Analyse des Correspondances Multiples* (ACM) pour les données qualitatives factorielles (Tenenhaus & Young, 1985 ; Lebart *et al.*, 1995). De nombreuses variantes méthodologiques existent, mais elles reposent toutes sur un même support algébrique, le schéma de dualité (Escoufier, 1987) qui effectue la transformation du tableau des réponses aux questions en un nouveau tableau définissant les orientations conceptionnelles prépondérantes :

- Ces composantes sont des variables nouvelles, V_i , qui expriment chacune une proportion p_i de la variance globale du tableau initial. Plus une composante exprime une proportion élevée de variance, plus elle traduit une divergence conceptionnelle importante au sein des individus. Le support graphique de cette analyse de variance est l'*histogramme des valeurs propres*.

- On sélectionne les composantes présentant les valeurs p_i les plus élevées, et on identifie leur signification en étudiant leur relation avec les questions Q_i . Dans le cas de l'ACP, la relation entre orientations conceptionnelles et questions est en général exprimée graphiquement au moyen du *cercle des corrélations*.

Nous présentons ici l'ACP réalisée sur la section A du questionnaire, proposée dans 16 pays dans le cadre du projet Biohead-Citizen (Figure 1). Une pondération des individus en fonction de l'effectif relatif des pays est réalisée, afin de ne pas donner plus de poids aux pays pour lesquels davantage d'individus ont été échantillonnés. Les deux premières composantes V_i expriment nettement plus de variance que les autres (Figure 1a), et sont seules prises en compte dans la suite de l'analyse.

Le cercle des corrélations dans le plan (V_1, V_2) permet de caractériser deux orientations conceptionnelles (Figure 1b) :

- Une composante « systèmes de valeurs », caractérisée notamment par l'importance des conceptions sur l'Evolution (créationnisme vs. évolutionnisme), l'inquiétude vis-à-vis de la pollution et la pénurie des ressources naturelles, l'âge souhaité pour l'enseignement de la sexualité humaine, ainsi que par la question du déterminisme génétique des différences ethniques et homme-femme.

- Une composante « perception de la nature et de l'environnement », exprimant également des valeurs en terme de préservation des espèces et d'aménagement du territoire, avec une connotation affective marquée mais pas exclusive.

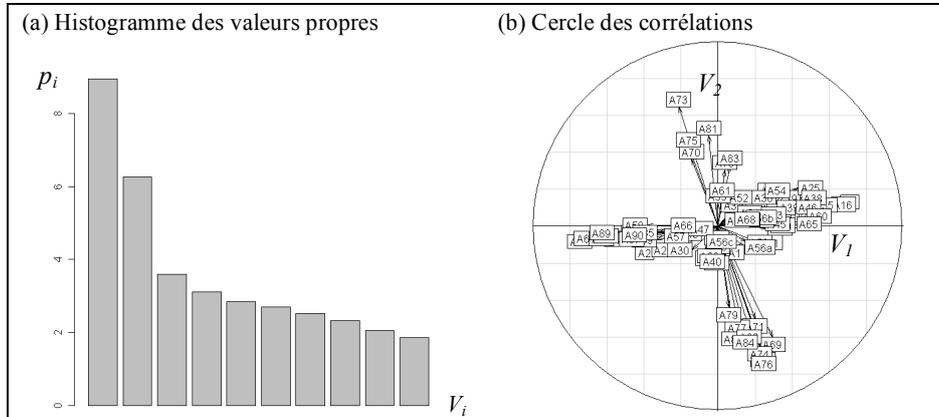


Figure 1. Analyse en Composantes Principales (ACP) réalisée sur l'ensemble des questions de la section A, pour 16 pays. L'histogramme des valeurs propres (a) représente la variance des individus exprimée par les composantes successives V_i issues de l'analyse. Le cercle des corrélations (b) représente la relation entre les deux premières composantes V_1 et V_2 et les questions de la section A, afin d'éclairer la signification de V_1 et V_2 .

Cette analyse réalise la synthèse d'un grand nombre de questions, de manière à identifier un ensemble réduit d'orientations conceptionnelles remarquables (ici V_1 et V_2). Celles-ci sont caractérisées par un ensemble cohérent de questions, dont l'importance est quantifiée au moyen des coordonnées des questions sur V_1 et V_2 . Cette analyse est le support des tests d'hypothèses que nous présentons ci-après.

4. Hypothèses de causalité

La construction du questionnaire et la mise en place de l'enquête ont été motivées par le test d'hypothèses théoriques. D'une part les questions ont été choisies et formulées afin de tester certaines interactions entre valeurs et connaissances. D'autre part les individus sont échantillonnés suivant un plan précis, six groupes d'enseignement et 16 pays, afin d'évaluer en priorité un effet de la nationalité et du « statut des enseignants » (échantillons équilibrés PreP, InP, PreL, InL, PreB, InB). A cette information d'échantillonnage s'ajoutent des données personnelles enrichissant nos informations sur les individus et sur leurs opinions. Elles sont essentiellement recueillies dans la section P du questionnaire, mais certaines questions à caractère politique ont également été disséminées dans la section A

Une question importante est l'identification des facteurs potentiellement à l'origine des orientations conceptionnelles des individus (§3). Nous examinons ici deux solutions apparentées :

- Etudier les conceptions des individus en fonction de leur appartenance à des catégories différentes, par exemple statut d'enseignement, nationalité, religion... (§4.1).

- Etudier la relation entre les orientations conceptionnelles et un ensemble de variables personnelles, catégorielles ou non, susceptibles d'être associées (§4.2).

4.0. Groupes d'individus

Le questionnaire a été proposé à un ensemble d'individus de nationalités, de religions ou de langues très variées. Nous souhaitons évaluer dans quelle mesure les conceptions des individus diffèrent suivant les groupes échantillonnés, en traitant par exemple ici l'appartenance aux groupes nationaux. L'hypothèse nulle est alors que les réponses des individus ne diffèrent pas entre les pays. Ce type d'analyse peut être mené sur n'importe quelle variable catégorielle pour laquelle les individus peuvent être répartis dans différents groupes ou classes.

4.0.0. Analyse inter-classe (between)

Une analyse inter-classe permet de rechercher et de représenter de manière synthétique ce qui différencie le plus, en moyenne, les réponses des groupes nationaux. Elle opère une combinaison des axes de l'ACP (ou de l'ACM suivant le choix initial), de manière à maximiser la variance entre les groupes nationaux (Dolédec et Chessel, 1987). On en déduit des axes discriminants, exprimant une variance décroissante entre groupes nationaux, que nous pouvons étudier de la même manière que les composantes issues de l'analyse exploratoire.

Nous représentons (Figure 2) la structure des deux premiers axes (D_1 , D_2) prépondérants dans l'analyse inter-classe du questionnaire Biohead-Citizen, relativement aux groupes nationaux. Nous indiquons la structure des composantes V_i par rapport aux axes D_i (Figure 2a), ainsi que la distribution des groupes nationaux dans cet espace (Figure 2b). L'analyse met en évidence une opposition entre pays extra-européens et pays européens sur le premier axe discriminant (52.6% de la différenciation totale entre pays).

Le cercle des corrélations souligne la prédominance dans cette différenciation du premier axe de l'ACP initiale (Figure 2a, V_1 en relation avec D_1), et exprime donc des différences conceptuelles en termes d'évolutionnisme / créationnisme, de préoccupations environnementales, d'enseignement de la sexualité et d'égalité homme-femme.

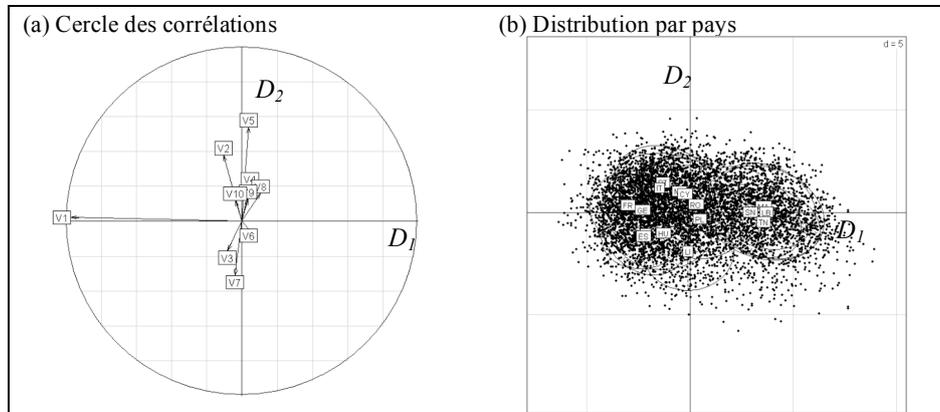


Figure 2. Analyse inter-classe sur les 16 groupes nationaux, pour la section A du questionnaire. Le cercle des corrélations (a) identifie la relation entre les axes discriminants prépondérants (D_1 , abscisses, et D_2 , ordonnées) et les composantes V_i issues de l'analyse en composantes principales (Figure 1). La distribution des individus (b) est représentée dans le plan (D_1 , D_2), en fonction de l'appartenance nationale : le label des pays (voir Tableau 1) est situé au centre des groupes nationaux. Une ellipse enveloppe la majorité des individus pour chaque pays.

Notre objectif ici étant de présenter avant tout nos méthodes, nous ne développerons pas l'analyse de l'axe D_2 , qui dépend de plusieurs composantes de l'ACP initiale, tout en discriminant moins les pays entre eux que l'axe D_1 (Figure 2a).

Nous testons l'hypothèse nulle d'absence de différenciation entre pays au moyen d'une procédure de type Monte-Carlo (test non paramétrique ; Romesburg, 1985). Nous rejetons au seuil de 0.1% l'hypothèse nulle, et concluons à l'existence de différences de conceptions significatives entre pays.

4.0.1. Analyse intra-classe (within)

La recherche de ce qui différencie le plus les individus au sein des pays est complémentaire de la recherche de ce qui discrimine le plus les pays. Alors que l'analyse inter-classe maximise la variance entre les groupes, il est possible également de maximiser la variance au sein des groupes. Cette approche intra-classe permet de rechercher des conceptions communes sur lesquelles il y a une forte variation des individus au sein de tous les pays. Nous menons cette analyse sur les 16 pays et considérons les composantes prédominantes résultantes (Figure 3).

De manière remarquable, une orientation conceptionnelle « consensuelle » émerge sur le premier axe, W_1 , de cette analyse : la Figure 3b montre une forte variation des réponses des individus sur W_1 , qui est essentiellement indépendante de la variation entre les pays (les centres des groupes sont proches de l'abscisse 0). En

revanche l'axe W_2 indique une source de variation importante à la fois au sein des pays et entre les pays, car il y a une forte dispersion des centres des groupes nationaux (Figure 3b).

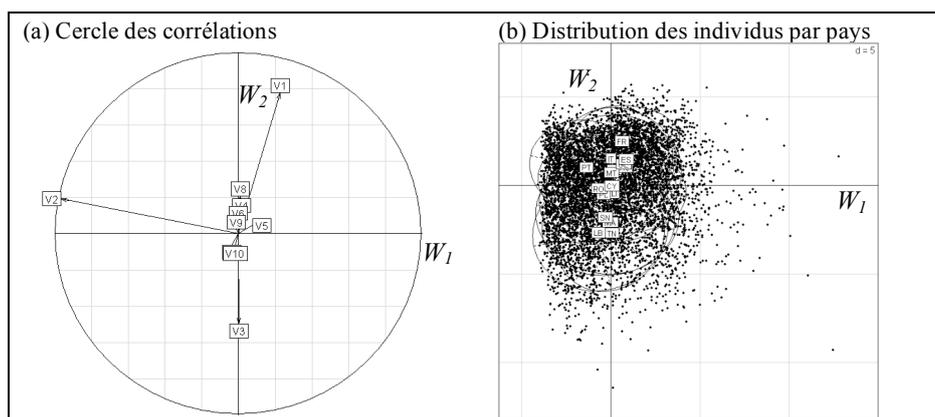


Figure 3. Analyse intra-classe réalisée entre les pays sur le questionnaire A. Le cercle des corrélations (a) identifie la relation entre les axes prépondérants (W_1 , abscisses, et W_2 , ordonnées) et les composantes V_i issues de l'Analyse en Composantes Principales (voir Figure 1). La distribution des individus (b) est représentée dans le plan (W_1 , W_2), en fonction de l'appartenance nationale : le label des pays est situé au centre des groupes nationaux. Une ellipse enveloppe la majorité des individus pour chaque pays.

L'orientation conceptionnelle « consensuelle » est clairement associée à la composante V_2 de l'ACP initiale (Figure 3a). Cela signifie que la perception de la nature et de l'environnement est très largement indépendante de la nationalité des individus, mais varie beaucoup au sein des pays. *A contrario*, la composante V_1 de l'ACP est associée à une variation à la fois entre pays et au sein des pays, à travers W_2 (Figure 3a). Nous montrons ainsi que les deux orientations conceptionnelles issues de l'ACP initiale, V_1 et V_2 , représentent essentiellement la différenciation des individus entre les pays et au sein des pays.

4.1. Variables instrumentales

De manière plus générale, nous pouvons rechercher, dans la section A du questionnaire, les orientations conceptionnelles spécifiquement reliées à certaines variables personnelles complémentaires, dites instrumentales. Ces variables peuvent être catégorielles ou non. L'analyse inter-classe traite le cas particulier d'une seule variable instrumentale catégorielle.

Nous étudions ici les orientations conceptionnelles associées à un ensemble de variables à caractère politique et religieux. Voici les variables instrumentales politiques choisies (code des variables en gras) :

- A quelles institutions faites-vous le plus confiance (publiques ou privées) pour l'école (**PubPriv1**), les services de santé (**PubPriv2**) et les retraites (**PubPriv3**) ? Codage de 1 (public) à 5 (privé).

- Etes vous impliqué dans des activités concernant la protection de l'environnement et/ou le développement durable, à la maison, en famille, au niveau local (**EngEnv1**), dans un groupe ou association (**EngEnv2**), professionnellement (**EngEnv3**) ? Codage de 1 (souvent) à 3 (jamais).

- **Croy** : Niveau de croyance en Dieu. **Prat** : Niveau de pratique religieuse. Echelle de 1 (+) à 5 (-).

Les variables suivantes sont codées de 1 (d'accord) à 4 (pas d'accord) :

- **PolitA15** : Une priorité du gouvernement doit être de prévoir des ressources minimales pour garantir aux pauvres une couverture santé.

- **PolitA20** : Le gouvernement de mon pays devrait imposer à tous les immigrants d'apprendre à parler, lire et écrire en français.

- **PolitA26** : Il y a trop d'étrangers dans mon pays : le gouvernement devrait limiter l'immigration.

- **PolitA34** : Le gouvernement doit mettre en place des lois favorisant la création d'entreprises afin de stimuler notre économie.

- **PolitA37** : Religion et politique devraient être séparées.

- **PolitA42** : Seul un pouvoir central fort peut mettre de l'ordre dans mon pays.

- **PolitA48** : La démocratie directe (sans implication du gouvernement) est la solution idéale pour faire fonctionner notre société.

- **PolitA51** : La science et la religion devraient être séparées.

L'Analyse en Composantes Principales sur Variables Instrumentales réalise une transformation algébrique analogue à celle de l'ACP traditionnelle, mais avec la contrainte d'obtenir des orientations conceptionnelles obligatoirement liées aux variables instrumentales, en l'occurrence ici des variables politiques et religieuses. Nous réalisons cette analyse sur l'ensemble des questions de la section A, avec les questions précitées comme variables instrumentales.

On obtient ainsi de nouvelles orientations conceptionnelles, V'_i , dépendantes des variables instrumentales, que l'on peut également représenter en fonction des orientations conceptionnelles V_i non contraintes (cf. Figure 1). La composante V'_1 est très clairement analogue à V_1 (Figure 4a) et est exprimée en fonction des variables instrumentales à la Figure 4b.

Cette analyse montre la relation entre les valeurs exprimées par V_1 et les conceptions à caractère politique et religieux. Les individus les plus croyants et pratiquants, à gauche sur les axes V_1 et V'_1 , sont conservateurs sur le plan de l'enseignement de la sexualité, expriment des conceptions inégalitaires et généralement créationnistes. Ces mêmes personnes souhaitent que religion et science ne soient pas séparés (A51), sont plutôt favorables aux systèmes privés de santé et d'éducation mais pas de retraites, sont favorables à un pouvoir central fort,

plutôt libéraux sur le plan de l'économie et défavorables à l'immigration (Figure 4b). On exprime ainsi différentes dimensions de l'orientation conceptionnelle « systèmes de valeurs » (V_i , Figure 1) identifiée précédemment.

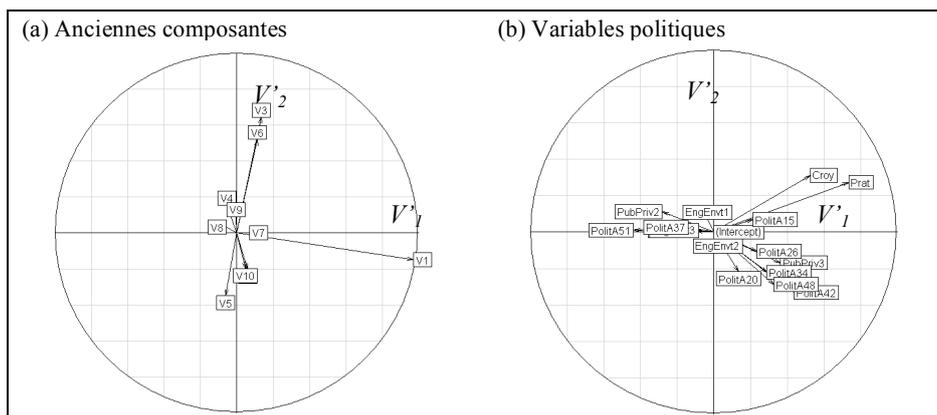


Figure 4. Analyse en composantes principales sur variables instrumentales politiques et religieuses. Le cercle des corrélations (a) identifie la relation des axes contraints V'_1 (abscisses) et V'_2 (ordonnées) avec les composantes initiales V_i issues de l'Analyse en Composantes Principales non contrainte (voir Figure 1).

5. Conclusion

Les méthodes d'analyse statistique multivariée sont utiles à la construction d'une démarche théorique rigoureuse. Au-delà d'une vocation initialement descriptive, elles permettent d'évaluer les différentes dimensions des conceptions selon des attendus théoriques précis (par exemple, sur la base du modèle KVP), ainsi que de tester les différences de conceptions entre groupes d'individus.

Des applications précises et thématiques de cette démarche sont proposées dans d'autres communications au congrès 2007 de l'AREF, relatives au projet Biohead-Citizen. D'autres résultats sont en cours d'obtention et les analyses peuvent être déclinées thème par thème, ainsi que pays par pays, en fonction des hypothèses formulées par les pays participants.

Un objectif important est l'identification des systèmes de valeurs qui ont quelque cohérence entre leurs dimensions religieuse, politique et scientifique. Jusqu'à quel point les connaissances enseignées peuvent-elles être influencées par les valeurs des enseignants (ce qu'illustrent les résultats présentés ici), ou en sont-elles parfois indépendantes (ce que suggèrent également certains de nos premiers résultats) ?

Remerciements

Ce travail s'inscrit dans le cadre du projet de recherche européen Biohead-Citizen (STREP CIT2-CT-2004-506015, E. C., Brussels, FP6, Priority 7), coordonné par Graça Carvalho (Portugal), Pierre Clément (France) et Franz Bogner (Allemagne). Nous remercions très vivement Graça Carvalho et Franz Bogner. Nous sommes également très redevables de la qualité du travail fourni par l'ensemble des équipes nationales.

Bibliographie

- Carvalho, G., Clément, P., & Bogner, F. (2005). *Biology, Health and Environmental Education for better Citizenship*. Rapport de recherche 1^{ère} année pour la Communauté Européenne (FP6, Priority 7, <http://www.biohead-citizen.net/>).
- Clément, P. (1998). La biologie et sa didactique. Dix ans de recherches. *Aster*, 27, 57-93.
- Clément, P. (2004). Science et idéologie : exemples en didactique et en épistémologie de la biologie. In *Actes du colloque Science - Médias – Société*. Lyon : ENS-LSH (<http://sciences-medias.ens-lsh.fr>).
- Dolédec, S., & Chessel, D. (1987). Rythmes saisonniers et composantes stationnelles en milieu aquatique I- Description d'un plan d'observations complet par projection de variables. *Acta Oecologica, Oecologia Generalis*, 8, 3, 403-426.
- Escoufier, Y. (1987). The duality diagram: a means of better practical applications. In Legendre, P. & Legendre, L. (Eds.), *Development in numerical ecology*. Berlin : Springer Verlag, 139-156.
- Lebart, L., Morineau, A., & Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Paris : Dunod.
- Romesburg, H.C. (1985). Exploring, confirming and randomization tests. *Computers and Geosciences*, 11, 19-37.
- Tenenhaus, M., & Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, 50, 1, 91-119.