

# La Sémiométrie

LUDOVIC LEBART  
MARIE PIRON  
JEAN-FRANCOIS STEINER

DUNOD

Facebook : *La culture ne s'hérite pas elle se conquiert*

Ludovic Lebart

Marie Piron

Jean-François Steiner

# **La Sémiométrie**

**Essai de statistique structurale**



# **La Sémiométrie**

## **Essai de statistique structurale**



## Avant-propos

Cet ouvrage s'adresse aussi bien à des statisticiens qu'à des psychosociologues, des spécialistes du marketing, de la communication, des linguistes, et plus généralement au grand public cultivé. La sémiométrie est en effet à la fois une curiosité philosophique, un outil du marketing et un champ d'expérimentation exceptionnel pour les statisticiens.

Dans ces diverses disciplines, l'ouvrage devrait intéresser un public très large allant des praticiens des bureaux d'études ou des instituts de sondage, aux étudiants, professeurs et chercheurs des universités ou des grandes écoles.

Plusieurs lectures sont possibles selon les connaissances du lecteur en mathématique et statistique, car des annexes et un glossaire, en fin d'ouvrage, permettent d'isoler les parties très techniques du texte, tout en en donnant une présentation assez simple, renvoyant sur des références accessibles et plus complètes.

L'ensemble doit beaucoup à des collaborations et des cadres de travail divers : au sein du département Economie Gestion et Sciences Humaines de l'Ecole Nationale Supérieure des Télécommunications, dirigé par Laurent Gille, et de l'URA 820 du Centre National de la Recherche Scientifique (Laboratoire de Traitement et Communication de l'Information, dirigé par Henri Maître) ; au sein de l'Institut de Recherche pour le Développement (IRD).

Nous sommes redevables de très nombreuses corrections, remarques, et suggestions à cinq personnes dont le travail de relecture et de critique fut en fait un véritable travail de collaboration. Il s'agit de Zysla Belliat et Hélène Haering, toutes deux membres du Comité Scientifique du Centre d'Etude des Supports de Publicité ; de Belaïd Ghermani de l'Université de Paris-Créteil ; de Françoise Potier de l'Institut National de Recherche sur les Transports et leur Sécurité ; de Jacqueline Sultan de la Société EMSIS.

Nous remercions également pour leurs remarques et observations pertinentes Christine Bellavoine, Etienne Brunet, Michel Grojnowski, Marie-Claude Lebart, Gérard Miltelberg, Armand Morgenzstern, Christine Tobin.

Nous restons évidemment seuls responsables des erreurs ou omissions qui pourraient subsister dans le texte.

Nous voudrions aussi dire toute notre gratitude à Jeannine Bouvet qui s'est livrée, avec une parfaite intelligence et le plus total désintéressement aux multiples travaux de saisie qui ont été à la base des différentes expériences décrites aux chapitres 3 et 4.

Nous sommes heureux d'adresser ici nos remerciements à Bruno Colin, Directeur des Opérations de la Société Taylor Nelson Sofres, notamment pour l'accès aux corpus de données sémiométriques.

Nos remerciements vont également à Jean Henriet, des éditions Dunod, ainsi qu'à Virginie Catoni, pour l'accueil qu'ils ont réservé à cet ouvrage.

L.L., M. P, J.-F. S.

Paris, 2003

# Sommaire

<b>Préambule</b>	1
<b>Introduction : Qu'est-ce que la sémiométrie ?</b>	5
<b>Chapitre 1</b>	
<b>Des mots, des individus... une structure</b>	9
<b>1.1 Des mots, un questionnaire</b>	10
<b>1.2 L'enquête représentative</b>	14
<b>1.3 L'analyse en composantes principales</b>	17
<b>1.4 Les axes principaux stables : « axes sémiométriques »</b>	19
1.4.1 Axe 1 : Un axe méthodologique de participation	19
1.4.2 Axe 2 : « Devoir / Plaisir »	21
1.4.3 Axe 3 : « Attachement / Détachement »	22
1.4.4 Le plan sémiométrique des axes 2 et 3	24
1.4.5 Axe 4 : « Esprit / Matière »	24
1.4.6 Axe 5 : « Cœur / Raison »	25
1.4.7 Axe 6 : « Humilité / Souveraineté »	26
1.4.8 Les axes suivants	27
<b>1.5 Les plans sémiométriques</b>	29
1.5.1 Les multiples facettes du « Plaisir »	29
1.5.2 Les multiples facettes du « Devoir »	35
1.5.3 Les dimensions sémantiques des mots eux-mêmes	36
<b>1.6 Conclusion</b>	37
<b>Chapitre 2</b>	
<b>Stabilité de la structure sémiométrique</b>	39
<b>2.1 Stabilité empirique</b>	40
2.1.1 Stabilité dans le temps	40
2.1.2 Stabilité par pays	41
2.1.3 Stabilité selon le sexe	51
2.1.4 Stabilité selon l'âge	53
<b>2.2 Intervalles de confiance des valeurs propres</b>	55
<b>2.3 Stabilité par rééchantillonnage</b>	58
2.3.1 Bootstrap partiel	59
2.3.2 Bootstrap total	59
<b>2.4 Stabilité structurelle</b>	64
<b>2.5 Stabilité vis-à-vis du codage</b>	67
<b>2.6 Analyse en facteurs communs et spécifiques</b>	69
<b>2.7 Conclusion</b>	71

<b>Chapitre 3</b>	
<b>Structures sémantiques et sémiométrie</b>	73
<b>3.1 Les « champs sémantiques internes »</b>	75
3.1.1 Représentation hiérarchique des proximités sémantiques	76
3.1.2 Représentation plane des proximités sémantiques	77
<b>3.2 Voisinage sémantique du questionnaire sémiométrique</b>	81
<b>3.3 Réseau sémantique des verbes français</b>	86
<b>Chapitre 4</b>	
<b>Choix spontané des mots</b>	93
<b>4.1 Mise en place de l'expérience</b>	94
<b>4.2 Premières explorations des réponses</b>	99
<b>4.3 Etude des axes principaux des réponses spontanées</b>	102
<b>4.4 Choix spontané et caractéristiques des répondants</b>	105
<b>4.5 Rapprochement entre sémiométrie et questions ouvertes</b>	110
<b>4.6 Conclusion</b>	116
<b>Chapitre 5</b>	
<b>Notations, participation, attitudes</b>	119
<b>5.1 Notation, participation et facteur de taille</b>	120
5.1.1 Les effets <i>notation</i>	120
5.1.2 Les effets <i>participation</i>	120
5.1.3 Les facteurs de taille	121
<b>5.2 Création et positionnement de variables de contrôle</b>	121
<b>5.3 Les notations dans quatre pays européens</b>	124
<b>5.4 Typologie à partir des seules variables de contrôle</b>	127
<b>5.5 Quels rapports entre les notes données aux mots, la façon de noter, et les caractéristiques des individus ?</b>	130
<b>5.6 Les plans de notation en Espagne, Grande Bretagne et Allemagne</b>	135
5.6.1 Effets <i>notation</i> en Espagne	135
5.6.2 Effets <i>notation</i> en Grande Bretagne	136
5.6.3 Effets <i>notation</i> en Allemagne	137
<b>5.7 Correction des notes par double centrage</b>	138
<b>5.8 Evolution des notations, en France, de 1990 à 2002</b>	142
<b>5.9 Conclusion</b>	145
<b>Chapitre 6</b>	
<b>Essais d'interprétation</b>	147
<b>6.1 Une première approche</b>	148
<b>6.2 Une interprétation psychanalytique</b>	152

<b>Chapitre 7</b>	
<b>Quelques applications de la sémiométrie</b>	159
7.1 Les mots caractéristiques des hommes et des femmes	161
7.2 Les mots caractéristiques de deux classes d'âge	162
7.3 Les grands amateurs de télévision	163
7.4 Les mots caractéristiques de deux classes de jeunes	165
7.5 Les mots caractéristiques de deux classes d'âge au delà de 60 ans	166
7.6 Hommes et femmes ayant attribué la note 7 au mot Dieu	168
7.7 Evolution des valeurs des Français entre 1996 et 1998	170
7.8 Conclusion	173
<b>Conclusion</b>	175
<b>Annexe 1</b>	
<b>Quelques éléments de statistique multidimensionnelle</b>	183
<b>A1.1 Rappel des principes des méthodes exploratoires         multidimensionnelles</b>	184
A1.1.1 Représentation géométrique et nuages de points	184
A1.1.2 Principe et méthodes d'analyse	185
<b>A1.2 Les méthodes factorielles : aspects techniques</b>	187
<b>A1.3 L'Analyse en Composantes Principales :         aspects techniques</b>	190
A1.3.1 Interprétations géométriques	190
A1.3.2 Problème d'échelle de mesure et transformation des données	190
A1.3.3 Analyse du nuage des répondants	191
A1.3.4 Analyse du nuage des variables (mots)	192
<b>A1.4 L'Analyse des correspondances</b>	194
<b>A1.5 L'analyse logarithmique</b>	195
<b>A1.6 L'analyse factorielle en facteurs communs et spécifiques</b>	196
<b>A1.7 Méthodes de classification hiérarchique</b>	197
<b>A1.8 Les cartes auto-organisées de Kohonen</b>	199
<b>A1.9 Outils de validation</b>	200
A1.9.1 Qu'est-ce qu'une valeur-test ?	201
A1.9.2 Le problème des comparaisons multiples	203
A1.9.3 Utilité des éléments supplémentaires	204
A1.9.4 Intervalles de confiance d'Anderson	205
A1.9.5. Les techniques de <i>bootstrap</i>	205
<b>Annexe 2</b>	
<b>Quelques éléments et résultats d'analyse</b>	211
<b>A2.1 Dictionnaire multilingue des mots</b>	211

<b>A2.2 Moyenne des notes</b>	216
<b>A2.3 Corrélations des mots par axe</b>	218
<b>Glossaire</b>	221
<b>Bibliographie</b>	225

# Préambule

## *A l'origine de la Sémiométrie...*

La Sémiométrie est née d'une expérience que tout homme a faite au moins une fois dans sa vie, d'une manière consciente et lucide ou si fugace que l'esprit n'en a gardé aucune trace mémorable.

Je veux parler de cet instant où, soudain, pour quelque raison mystérieuse, nous cessons de communiquer avec les êtres et les objets qui nous entourent, où ils cessent de nous parler, où ils ne nous disent plus rien, où nous nous retrouvons tout seuls au milieu d'un monde minéral peuplé d'objets morts, et dans lequel, les sens glacés, l'esprit figé, nous ne savons plus nous orienter. Quelques instants ou quelques jours auparavant – le phénomène se déclenche d'une manière plus ou moins brutale et sa durée est extrêmement variable – le monde était plein de vie, de sons, de couleurs, de saveurs, d'odeurs, de sensations agréables ou désagréables, d'objets et d'êtres qui nous attiraient, d'autres qui nous repoussaient, que nous aimions ou que nous haïssions, mais que nous pouvions nommer et dont le nom signifiait quelque chose pour nous, était doué d'un pouvoir évocateur qui faisait naître, à sa seule énonciation, un cortège de sensations qui en étaient comme la signature : un ami était quelque chose de chaud et de précieux, de rare aussi, de solide, très solide, et pourtant, en même temps, de très fragile ; une maison était aussi quelque chose de chaud, mais d'une autre chaleur ; alors que la chaleur de l'ami irradiait à partir d'un point situé au milieu de la poitrine, celle de la maison semblait plutôt descendre de quelque

lieu inconnu et extérieur à nous et nous pénétrer lentement de bonheur ; la chaleur de l'ami était aussi plus intense, plus violente que celle de la maison ; c'était une chaleur rouge alors que celle de la maison tirait plutôt sur le bleu. Le mot *guerre* aussi signifiait quelque chose ; c'était un tout autre monde de sensations qui semblaient appartenir à une zone qui se situait aux antipodes de celle à laquelle appartenait le mot *maison*, mais qui semblait avoir un territoire commun avec le mot *ami*. Le mot *arbre* aussi avait une signification et le mot *rivière* et le mot *chemin* et le mot *école* et le mot *été* et le mot *forêt* et le mot *tristesse* et même le mot *ennui*, et jusqu'au mot *désespoir*.

Bref, grâce aux signes qui permettent de le décrire, le monde avait un sens ; il nous était intelligible ; nous savions nous y orienter ; il était notre milieu naturel, si naturel même que la pensée ne nous effleurait pas que nous pourrions en être exclus autrement que par la mort. Nous étions si accoutumés à le retrouver chaque matin à notre réveil, toujours identique à lui-même – à l'exception de quelques objets qui de temps en temps, migraient d'une zone dans une autre –, que nous n'avions pas réfléchi à cette évidence – si forte, pourtant, dès lors qu'on la conçoit – que le sens n'est pas consubstantiel au monde qui nous entoure, qu'il n'en est qu'une propriété que nous lui attribuons, qu'une autre réalité que nous projetons sur lui. Il avait fallu cette expérience de perte du sens – dans la triple acception du mot : sensation, signification et direction – pour nous faire appréhender ce fait bien banal que le sens est un phénomène autonome et indépendant des objets à travers lesquels il s'incarne, qu'il est une réalité, au moins psychique sinon physique, et donc, qu'il doit pouvoir être étudié indépendamment des objets auquel il sert d'attribut et par la médiation desquels on le perçoit généralement.

Mais cette première expérience en contenait une autre, qui en était comme la suite logique : non seulement il y avait DU sens, mais il y avait aussi DES sens. Ainsi le mot *château*, non seulement pouvait être doué (ou dénué) de sens mais, dès lors qu'il possédait cette propriété, il en découlait nécessairement qu'il possédait aussi UN (ou plusieurs) sens.

Et cette deuxième expérience, à son tour, en contenait implicitement une troisième : ces différents sens révélés par les mots étaient plus ou moins proches les uns des autres. Ainsi le sens du mot *château* était intuitivement plus proche de celui du mot *maison* que de celui du mot *mathématique*, mais plus proche encore de celui du mot *palais*. Le sens du mot *maison* semblait équidistant de celui des mots *château* et *palais*. Le sens du mot *mathématique* était si éloigné de celui de ces trois

derniers qu'il était difficile de dire s'il était plus proche de l'un d'entre eux, de la même manière que trois points très proches semblent pareillement distants d'un quatrième, si celui-ci est situé à une grande distance ; quoique, à y regarder de plus près, le sens du mot *château* paraissait très légèrement moins éloigné de celui du mot *mathématique* que ne l'était celui du mot *maison*. Comme pour le sens des autres mots, ce n'était qu'une intuition, mais si ténue, si subtile, si insaisissable qu'elle s'évanouissait dès que nous essayions de la raisonner, et qui pourtant renaissait presque instantanément, comme quelque tache indélébile, dès que nous détournions d'elle la lumière trop crue de notre attention. Ces intuitions qui nous faisaient attribuer une distance de sens entre les mots possédaient donc un fondement ; elles étaient l'expression d'une réalité qui, pour apparemment inaccessible à un instrument d'observation objectif, n'en était pas moins stable et, donc, théoriquement mesurable. On pouvait donc imaginer un espace qui contiendrait l'ensemble des mots d'une langue, disposés les uns par rapport aux autres en fonction de la distance séparant leur sens. Ce serait une sorte de dictionnaire géographique dans lequel les mots ne seraient pas rangés dans l'ordre alphabétique – ce qui, pour pratique qu'il soit, n'en est pas moins un classement relativement arbitraire – mais disposés, les uns par rapport aux autres, en fonction de leur sens, un dictionnaire organique dans lequel chaque mot serait défini (mathématiquement) par sa position dans l'ensemble de tous les autres. Par référence à Ferdinand de Saussure qui, imaginant une science des signes, avait proposé de l'appeler sémiologie, je baptisai mon entreprise, sémiométrie<sup>1</sup>.

\*

\* \*

*J'éprouve une très grande gratitude à l'égard des personnes qui m'ont aidé dans cette entreprise. Je voudrais d'abord citer le nom de Philippe Court dont l'amitié indéfectible et une foi inébranlable m'ont soutenu dès le premier instant. Je voudrais aussi citer celui d'Eric Stemmelen, à l'époque ingénieur au Commissariat à l'Energie Atomique, qui m'a appris les premiers rudiments de l'analyse de données, avant de diriger le département que la Sofrès avait créé pour développer commercialement les*

---

1. L'idée qu'il existe un espace du sens pouvant être modélisé revient à Charles E. Osgood. C'est lui, en effet, qui, dans les années 50, imagina le premier un dispositif destiné à reconstruire cet espace. Ce dispositif diffère radicalement de celui mis en oeuvre par la Sémiométrie et j'ignorais tout des travaux d'Osgood lorsque je conçus celui-ci. Mais il convenait de citer le nom de celui qui avait introduit ce concept si riche d'espace de sens, cf. Osgood (1965), Osgood *et al* (1957).

*applications de la sémiométrie. Dans l'ordre chronologique, il me faut ensuite citer ceux de Georges Levitte et de Léon Szary qui a écrit le premier programme grâce auquel j'ai pu procéder aux premières expérimentations.*

*S'il ne fut pas mon premier mentor, Emeric Deutsch, à l'époque l'un des dirigeants de la Sofrès, fut certainement le plus décisif, grâce à la profondeur de ses connaissances tant en psychanalyse et en linguistique que dans le domaine des études de marché.*

*Je dois aussi remercier la Sofrès, son Président, Pierre Weill qui a mis à ma disposition des moyens dont peu de chercheurs disposent, son Directeur-Général, Henri Wallard, dont l'amitié me fut bien précieuse, Jean-Michel Portier, Jean-Paul Aimetti, Joël Boillot, Mireille Gettler-Summa, Jean-Pierre Dreyfus, Valérie Morrison, et tant d'autres.*

Jean-François Steiner

## Introduction :

### Qu'est-ce que la sémiométrie ?

A la question « Qu'est-ce que la sémiométrie ? », il est facile de répondre naïvement : « C'est une longue liste de mots auxquels on fait attribuer une note (selon leur caractère plus ou moins agréable ou désagréable) par des milliers de gens ».

Tentons quand même des caractérisations plus académiques :

- la sémiométrie est une technique de description de certains types de liens sémantiques entre mots ;
- la sémiométrie est un instrument de recherche plus fondamentale permettant d'approcher (avec prudence et circonspection) des notions telles que l'« inconscient collectif » ou individuel, permettant aussi de mettre en évidence et de valider des traits structuraux (dont la finesse laisse parfois pantois...) ;
- la sémiométrie est un outil destiné à décrire des systèmes de valeurs et des styles de vie, en vue d'études psycho-sociologiques ou d'applications en marketing.

La sémiométrie est en fait un peu tout cela, et, c'est ce que cet ouvrage propose de montrer.

La première définition « naïve » est l'évocation d'un protocole expérimental élaboré et rigoureux, auquel sera d'ailleurs dévolu le premier chapitre de cet ouvrage. Cette phase expérimentale qui comporte de nombreuses enquêtes répétées dans le temps et dans

l'espace (notamment dans plusieurs pays occidentaux) constitue la base de toute la méthodologie sémiométrique, base qui sera validée par des traitements statistiques puissants, mais aveugles, c'est-à-dire indépendants de tout cadre interprétatif. Notons que, au niveau de la présentation, dans ce premier chapitre comme dans ceux qui vont suivre, les développements techniques seront reportés en annexe de façon à permettre une lecture plus directe du contenu, des résultats, et de la méthodologie générale, sans heurt ni digression.

Le thème de la stabilité de structures (stabilité dans le temps et l'espace) qui fait l'objet du deuxième chapitre, est central dans la problématique de la sémiométrie. Cette stabilité est établie à partir de la répétition d'enquêtes indépendantes, mais aussi à partir d'épreuves de validité plus techniques (méthodes statistiques dites de rééchantillonnage). C'est cette stabilité qui permet d'affirmer la réalité de certains traits structuraux observés et qui donne aux cartes sémiométriques, largement utilisées lors des applications, leur valeur de grilles de repérage.

La stabilité des structures pour un questionnaire donné (comportant le choix d'une liste de mots) étant établie dans différents pays et à différents moments, il reste le problème du caractère intrinsèque de cette structure, c'est-à-dire de son indépendance vis-à-vis du questionnaire lui-même, et non pas seulement vis-à-vis des populations auxquelles ce questionnaire est soumis.

Les trois chapitres suivants évoquent nos recherches méthodologiques autour de ce thème. Pour mieux comprendre la nature des faits statistiques observés, diverses expériences ont été réalisées : étude de certains réseaux sémantiques attachés aux mots choisis, puis étude de réseaux sémantiques plus généraux (chapitre 3), questionnaire « ouvert » laissant la personne interrogée choisir spontanément, sans liste préalable, les mots qu'elle trouve agréables ou désagréables (chapitre 4). Ces expériences sont d'ailleurs l'occasion d'aborder d'autres directions de recherche suscitées par l'outil sémiométrique et la structure multidimensionnelle qu'il révèle. Le cinquième chapitre est dévolu au problème technique, souvent éludé, du lien entre l'attitude de la personne interrogée par rapport au questionnaire ou à l'enquête et le contenu même du questionnaire ou la finalité de l'enquête.

Légèrement en retrait par rapport aux considérations empiriques et techniques précédentes, un essai d'interprétation de la partie stable des

structures sémiométriques est proposé au sixième chapitre. Quel que soit le degré d'adhésion du lecteur aux thèses présentées, il ne sera probablement pas insensible aux étonnantes convergences mentionnées.

Enfin, au cours du septième chapitre, dévolu aux applications de l'outil sémiométrique, nous serons amenés à présenter des applications pratiques ou pragmatiques (en général commerciales : média-planning, positionnement ou caractérisation de marques), des applications à des problèmes plus généraux (évolution des valeurs, caractérisation de groupes socio-démographiques particuliers).

Le chapitre 8 est en fait une conclusion, qui ne peut être que provisoire dans le cadre d'un travail aussi périlleusement pluridisciplinaire ; c'est aussi une invitation à des recherches futures, dans un domaine que nous proposons d'appeler « statistique structurale ».

Viennent enfin les annexes, austères entrepôts de matériaux, de résultats, de rappels techniques, où le lecteur motivé par les chapitres précédents pourra trouver des réponses à beaucoup de questions soulevées au cours de son voyage dans cet espace multidimensionnel.



# CHAPITRE 1

## Des mots, des individus... une structure

Les études socio-économiques mettant en œuvre des enquêtes par sondage caractérisent en général les individus interrogés (consommateurs, usagers, électeurs, justiciables, patients, épargnants, lecteurs...) par ce qu'on appelle usuellement les variables signalétiques, ou encore les caractéristiques socio-démographiques de base (sexe, âge, niveau d'instruction, profession, statut matrimonial, revenu, pour citer les plus usuelles). Ces caractéristiques ont incontestablement un certain pouvoir prédictif des comportements. Dans le domaine des études marketing, il est cependant apparu nécessaire, au cours des décennies précédentes, de décrire plus finement les personnes interrogées à partir de questionnaires faisant intervenir des variables plus subjectives ou plus psycho-sociologiques. C'est en général sous la dénomination d'études de « styles de vie » ou de « courants socio-culturels » qu'ont été conduites ces tentatives de description<sup>1</sup>.

Si l'on devait trouver parmi les méthodologies existantes la démarche la plus proche de celle que nous désignons ici par « Sémiométrie », c'est probablement aux études *styles de vie* qu'il faudrait se référer, malgré de sensibles divergences de motivation, de filiation, et même de principes<sup>2</sup>.

---

1. On pourra trouver un panorama et une synthèse critique de ces études dans l'ouvrage de Pierre Valette-Florence : *Les styles de vie. Bilan critique et perspectives*. Nathan, Paris, 1994.

2. Cf. Steiner et Auliard (1992).

Ces études mettent en œuvre des questionnaires en général assez volumineux, faisant une large part aux questions d'opinion.

Prenons l'exemple de deux familles d'enquêtes apparentées aux études *styles de vie* dont les questionnaires sont ou furent transparents, c'est-à-dire accessibles dans leur intégralité aux utilisateurs : le système d'enquêtes « Agoramétrie »<sup>1</sup> et le système d'enquêtes sur les « Conditions de vie et aspirations des Français »<sup>2</sup>.

Pour la première famille citée, donnons à titre d'exemple le libellé d'une des questions permanentes du questionnaire: *Etes-vous : pas du tout d'accord, pas tellement d'accord, peut-être d'accord, bien d'accord, entièrement d'accord, avec la phrase suivante : « Il faut rétablir la peine de mort »*.

Pour la seconde famille d'enquêtes citée, et dans la même veine, à côté de questions consistant également à consigner un degré d'accord, figurent des questions du type suivant : *Parmi ces opinions, quelle est celle qui se rapproche le plus de la vôtre : Le mariage est : 1. « Une union indissoluble »; 2. « Une union qui peut être dissoute dans des cas très graves »; 3. « Une union qui peut être dissoute par simple accord des deux parties »; 4. « Ne sait pas »*.

Le questionnaire sémiométrique, lui, peut être considéré comme une forme épurée à l'extrême de questionnaire « style de vie », puisque, comme nous allons le voir, il ne s'agit plus de donner son opinion sur des propositions présentées sous forme de phrases plus ou moins lapidaires, mais d'attribuer des notes à des mots en fonction des sensations agréables ou désagréables que provoquent leurs évocations.

Nous allons présenter dans ce chapitre le questionnaire, puis le fonctionnement de l'outil sémiométrique de base : l'enquête par sondage (comment et par qui fait-on remplir les questionnaires), et les traitements statistiques avec l'émergence de la *structure sémiométrique de base*.

## 1.1 Des mots, un questionnaire

La sémiométrie repose sur une liste de 210 mots (*cf.* tableau 1.1) censés représenter, directement ou indirectement, les principales valeurs de la

---

1. Ce système d'enquête a été fondé et dirigé par Jean-Pierre Pagès et Georges Morlat (*cf.* Fabre et *al.*, 1981).

2. Sur le système d'enquête sur les « Conditions de vie et aspirations des Français » du Crédoc, mis au point en 1978 avec la collaboration de Jacques Antoine, *cf.* Lebart et Houzel (1980), Lebart (1986).

société occidentale. Ils ont été choisis en fonction de quatre critères : non-consensualité, univocité sémantique, stabilité sémantique et puissance d'évocation. Par ailleurs, et afin de lui assurer une certaine représentativité tout en évitant, du moins partiellement, une trop grossière projection de la subjectivité de l'auteur, cette liste a été établie, à l'origine, à partir des quelque trois cents racines dont sont issus 95% des mots du Pentateuque (cinq premiers livres de l'Ancien Testament)<sup>1</sup>. La liste originale qui comprenait 306 mots a été ensuite retravaillée en s'appuyant sur les résultats, d'une part d'analyses statistiques successives destinées à mettre en évidence les zones de trop forte et de trop faible densité et, d'autre part, de groupes de créativité à qui il était demandé d'imaginer les mots manquants.

Cette première liste a été ramenée, par souci d'économie et d'ergonomie, à la liste actuelle de 210 mots. Si cette réduction a inévitablement entraîné une perte d'information, elle n'a pas affecté la stabilité de la structure mise en évidence par la liste complète.

Ces mots comprennent des substantifs, des adjectifs et des verbes. Certains substantifs sont accompagnés de l'article défini et d'autres de l'indéfini. Le choix qui a été fait de la forme du mot et de l'article est intuitif et, donc, plus ou moins arbitraire. Il répondait au souci de trouver le signifiant qui, d'une part, correspondrait le mieux au concept sous-jacent qu'il s'agissait de mesurer et, d'autre part, dont l'effet de stimulation serait le plus fort, mais aussi qui serait le plus euphonique.

Si dans certains cas le choix s'imposait pour des raisons évidentes : la nudité et non une nudité, le théâtre et non un théâtre, une île et non l'île, une muraille et non la muraille, dans d'autres, il était moins immédiat : séduire et non la (ou une) séduction, critiquer et non la (ou une) critique, le désir et non un désir ou désirer, un parfum et non le parfum, un mariage et non le mariage, humble et non l'humilité, intime et non l'intimité, le désert et non un désert,....

---

1. La question de la définition d'un corpus de mots qui pourrait prétendre à une certaine « représentativité » des concepts à l'aide desquels l'homme perçoit et décrit le monde et ses relations avec lui paraissait, à Jean-François Steiner, aussi simple à poser que difficile à résoudre lorsqu'il eut, fortuitement, connaissance de cette particularité du texte biblique. La place privilégiée que ce texte avait occupé dans la culture occidentale ainsi que sa vocation à rendre compte des principales expériences spirituelles, affectives, sensibles, etc. de l'homme entre le moment de sa naissance et celui de sa mort, lui laissa penser qu'à partir du sens de ces 300 racines pourrait être établie une liste de mots qui servirait de matrice au questionnaire imaginé.

**Tableau 1.1 : Les mots de la sémiométrie**

l'absolu	un créateur	un fusil	la modération	la raison
l'acharnement	un cri	la gaieté	la modestie	une récompense
acheter	critiquer	la gloire	moelleux	le recueillement
admirer	le danger	gratuit	une montagne	réfléchir
adorer	un défi	guérir	la morale	une règle
l'ambition	le désert	la guerre	la mort	le respect
l'âme	le désir	hériter	une muraille	rêver
l'amitié	le désordre	un héros	la musique	la révolte
l'angoisse	le détachement	honnête	un mystère	la richesse
un animal	Dieu	l'honneur	nager	rigide
un arbre	différent	humble	la naissance	rire
l'argent	la discipline	l'humour	un nid	robuste
une armure	la douceur	une île	noble	rompre
l'art	le doute	immense	un nœud	rouge
astucieux	dynamique	immobile	noir	la ruse
l'attachement	l'eau	un inconnu	la nudité	sacré
attaquer	l'école	l'industrie	obéir	un sacrifice
l'audace	économiser	l'infini	l'océan	sauvage
aventurier	écrire	interdire	l'or	la science
un bâtisseur	efficace	interroger	un orage	un secret
un bijou	un effort	intime	original	séduire
bleu	l'élégance	un inventeur	la paix	sensuel
bohème	l'élite	l'ironie	le pardon	soigner
un cadeau	une émotion	un jeu	un parfum	un soldat
la campagne	l'enfance	la justice	la patience	un sommet
une caresse	enseigner	un labyrinthe	la patrie	la souplesse
une cérémonie	ensemble	la légèreté	la peau	souverain
une certitude	escalader	un livre	la perfection	sublime
le changement	éternel	la logique	la poésie	la tendresse
charitable	un étranger	la loi	la politesse	le théâtre
charnel	une évasion	la lune	précieux	la tradition
la chasse	la famille	la magie	la précision	trahir
un chercheur	une faute	une maison	un prêtre	le travail
commander	féconder	maîtriser	produire	utilitaire
le commerce	féminin	un mariage	la propriété	vert
concret	la fermeté	un masque	protéger	la victoire
la confiance	le feu	matériel	la prudence	le vide
le confort	la fidélité	maternel	la puissance	vieillir
conquérir	une fleur	la méfiance	punir	viril
consoler	un fleuve	métallique	la pureté	la vitesse
construire	la foi	la minceur	une question	volontaire
le courage	une frontière	la mode	raffine	voluptueux

La formulation généralement utilisée dans les questions d'attitude porte sur l'intérêt ou l'opinion de l'interviewé à l'égard de l'item proposé, ce qui a pour effet de placer celui-ci dans une position de juge extérieur et ainsi de l'amener à rationaliser peu ou prou sa réponse. Pressentant que l'information recueillie serait d'autant meilleure qualité qu'elle serait fournie à partir d'un niveau plus inconscient, il a semblé plus pertinent d'essayer de mesurer l'affect déclenché par le mot. Celui-ci jouant alors le rôle d'une sorte de stimulus qui contournerait la conscience et déjouerait ainsi les inévitables déformations que produisent les multiples censures que celle-ci dresse entre le sujet et son inconscient. C'est ainsi qu'il est demandé aux interviewés de noter les mots en fonction de la sensation, agréable ou désagréable, que provoque leur lecture<sup>1</sup>.

Toujours dans ce même souci d'obtenir des réponses les plus « naturelles » possible, il leur est recommandé, par ailleurs, dans un court texte introductif, de répondre en une seule fois, le plus spontanément possible en se laissant guider par leurs réactions instinctives, et conseillé, dans le cas où ils viendraient à hésiter – certains mots provoquant nécessairement des réactions ambivalentes –, de donner « n'importe quelle note » et de passer au mot suivant. Enfin, afin d'éviter les effets de halo et d'ordre, on utilise vingt rotations de questionnaires dans lesquelles l'ordre des mots est tiré aléatoirement.

L'échelle retenue comporte sept modalités variant de -3 à +3, le signe (-), qui évoque une frustration, étant assez naturellement assimilé à une sensation désagréable et le (+), qui connote, au contraire, des expériences gratifiantes, à une sensation agréable<sup>2</sup>. Mais devait-on proposer une modalité centrale, le zéro, qui correspond, en fait, à une non-réponse ? Il était tentant de ne pas inclure cette modalité dans l'échelle, afin de forcer les réponses ; il a semblé cependant que le caractère impliquant du questionnaire faisait un devoir d'ouvrir aux interviewés la possibilité de ne pas se prononcer sur certains mots.

---

1. Dans la pratique, le questionnaire sémiométrique est auto-administré. Cette procédure a certes des inconvénients connus, notamment en ce qui concerne les taux de retour, mais elle est apparue comme constituant le protocole le plus adapté à un questionnaire demandant de la part de la personne interrogée calme, concentration et discrétion ; ce mode de passation permet aussi d'éviter les biais que pourrait créer la voix d'un enquêteur interprétant les mots en fonction de ses propres sensations.

2. Du point de vue des traitements numériques qui vont suivre, cette échelle est strictement équivalente à une échelle de 1 à 7, choisie pour coder les réponses afin d'éliminer le signe moins (-) qui alourdit les fichiers de données ; et l'on parlera habituellement de note 1 (plutôt que de note -3), de note 2 (plutôt que de note -2) jusqu'à la note 7 (correspondant à la note +3). La note médiane 0 initiale devient alors la note 4.

122	La modestie	-3	-2	-1	0	+1	+2	+3
133	Mocelleux	-3	-2	-1	0	+1	+2	+3
124	La mort	-3	-2	-1	0	+1	+2	+3
100	Une muraille	-3	-2	-1	0	+1	+2	+3
085	Un mystère	-3	-2	-1	0	+1	+2	+3
105	Nager	-3	-2	-1	0	+1	+2	+3
043	Une naissance	-3	-2	-1	0	+1	+2	+3
025	Un nid	-3	-2	-1	0	+1	+2	+3
106	La nudité	-3	-2	-1	0	+1	+2	+3
071	Obéir	-3	-2	-1	0	+1	+2	+3
173	L'océan	-3	-2	-1	0	+1	+2	+3
086	Un orage	-3	-2	-1	0	+1	+2	+3

Figure 1.1 : Extrait du « questionnaire de mots »

Ce « questionnaire de mots » est ultérieurement croisé avec d'autres questionnaires comprenant des renseignements socio-démographiques et des réponses à des questions d'attitude ou de comportement recueillies auprès de la population interrogée. Ces sources d'information sont mises en relation afin de disposer simultanément pour chaque individu des deux types d'information : caractéristiques de base et variables d'intérêt, d'une part, questionnaire sémiométrique renseigné d'autre part.

## 1.2 L'enquête représentative

L'enquête par sondage reste un des outils fondamentaux de connaissance de la réalité économique et sociale. Tous les instituts de statistique nationaux et internationaux y font appel. Les sondages politiques et d'opinion, quelque peu galvaudés au cours du troisième tiers du vingtième siècle, ont peut-être terni l'image de cet outil auprès du grand public, mais ils n'ont en aucune façon altéré son utilité ni la demande ou l'intérêt des professionnels concernés.

Le calcul des probabilités dit comment et avec quelle marge d'erreur l'observation d'une petite partie d'une population peut renseigner sur la population toute entière... mais cette base mathématique rigoureuse ne doit surtout pas faire illusion. L'enquête par sondage reste un art au sens des *arts et métiers* ou des *ouvrages d'art* : beaucoup de connaissances, de savoir-faire, d'expertise, de déontologie sont nécessaires pour conférer à cet outil un statut scientifique.

Le physicien sait qu'il ne peut mesurer sans erreur de mesure ; de surcroît, il ignore l'erreur qu'il commet (sinon, il connaîtrait la vraie valeur de ce qu'il mesure) ; mais il s'efforce de connaître l'ordre de grandeur de cette erreur. Et c'est au prix de cette démarche, somme toute modeste, que se construit très progressivement la connaissance scientifique.

C'est également la démarche du statisticien qui réalise une enquête par sondage... mais dans ce cas la complexité de l'outil demande une véritable « enquête sur l'enquête » pour procéder à une évaluation des résultats.

La conception du questionnaire, le plan de sondage et le tirage de l'échantillon, la phase de recueil des données (sur le terrain en face à face, par téléphone, par courrier, par internet), la codification de l'information, les tests de cohérence de cette information et les éventuelles procédures de redressement ou de pondération sont autant d'étapes nécessitant un contrôle en temps réel. Un tel contrôle sera de préférence réalisé par une instance indépendante à la fois de l'institut qui réalise le sondage et des commanditaires de l'enquête<sup>1</sup>.

Un outil *puissant et utile, complexe et fragile*, voilà comment peut être décrite et qualifiée l'enquête représentative, et voilà pourquoi ses résultats ne peuvent être livrés sans un minimum de considérations et commentaires méthodologiques. Que le lecteur ne soit donc pas surpris si les résultats qui vont lui être présentés n'apparaissent que comme la partie émergée d'un volumineux iceberg méthodologique.

Pour savoir ce qu'évoquent les *mots* pour les Français (mais aussi les Anglais, les Italiens, les Allemands, les Espagnols, les Canadiens, et beaucoup d'autres encore), il a été ainsi procédé à de nombreuses enquêtes par sondage. Ces enquêtes ont eu lieu à des moments différents pour un même pays (c'est notamment le cas de la France), ou parfois simultanément dans des pays différents.

---

1. C'est par exemple, le rôle dévolu en France au CESP (Centre d'Etude des Supports de Publicité) pour ce qui concerne les enquêtes de mesure d'audience des médias.

**Tableau 1.2 : Les enquêtes**

<i>Pays</i>	<i>Echantillon</i>	<i>(Année)</i>
France	2764	(1990)
Grande Bretagne	1849	(1995)
France	2764	(1996)
Espagne	2983	(1996)
France	2764	(1998)
Italie	2606	(1998)
Allemagne	3065	(1999)
Grèce	1062	(2000)
France	2763	(1999)
Canada	1865	(2000)
France	2764	(2001)
France	2763	(2002)

Source : Groupe Taylor-Nelson-Sofres

Ces répétitions dans le temps et dans l'espace sont certes coûteuses, mais elles donnent de précieuses indications sur la stabilité des associations et structures observées. Elles seront complétées, au chapitre suivant, par des études plus techniques de stabilité.

Puisqu'il a été question de *structure* (et ce concept sera en fait un leitmotiv tout au long de l'ouvrage), précisons ici que les enquêtes représentatives sont des outils plus robustes pour mettre en évidence des structures que pour mesurer des moyennes. Une mesure de corrélation entre deux variables, par exemple, sera beaucoup moins sensible à la parfaite représentativité de l'échantillon qu'une mesure de la moyenne de ces variables dans l'échantillon. Cette propriété est difficile à démontrer en toute généralité, mais donne lieu à des vérifications expérimentales adaptées à chaque situation concrète à partir de simulations de perturbations d'échantillon.

Les instruments d'observation requis doivent alors être capables de structurer le corpus de mots. Pour ce faire, ils vont s'appuyer sur des techniques exploratoires qui fournissent des représentations synthétiques de vastes ensembles de données sans faire référence à des hypothèses de nature statistique ni à un modèle particulier. Ils ont, pour cela, la capacité de réduire la dimension du tableau de données, par des calculs de distances

appropriées. Ils permettent de visualiser sous forme de séries de cartes planes les associations entre les mots et entre les individus ou groupes d'individus.

Dans ce chapitre, on se référera principalement aux procédures d'analyses statistiques multidimensionnelles en considérant les méthodes factorielles (*cf.* annexe A1.2), et plus particulièrement l'analyse en composantes principales (*cf.* annexe A1.3). D'autres techniques (classification et cartes de Kohonen) interviendront dans les chapitres suivants.

Les résultats qui seront présentés sont fondés, en ce qui concerne la France, sur les six échantillons indépendants des années 1990, 1996, 1998, 1999, 2001 et 2002 regroupant les 16 582 individus interrogés, représentatifs de la population française âgée de 18 ans et plus et qui ont répondu au questionnaire de 210 mots<sup>1</sup>. D'autres résultats concernant notamment l'Allemagne, l'Espagne et le Royaume Uni seront présentés aux chapitres 2 et 5.

### 1.3 L'analyse en composantes principales

C'est une technique de description de tableaux de mesures multidimensionnelles apparentée à la technique dite d'analyse factorielle (*factor analysis*) utilisée par certains psychologues (les psycho-métriciens) depuis le début du vingtième siècle. Elle sera ici un instrument privilégié de description du corpus des notes données aux 210 mots du questionnaire par les 16 582 personnes interrogées entre 1990 et 2002.

Le principe de l'analyse est simple et intuitif<sup>2</sup> : il s'agit de représenter l'ensemble des répondants à l'enquête d'une part, et l'ensemble des mots du questionnaire d'autre part, comme deux « nuages de points », l'un étant défini par rapport à l'autre: les proximités entre mots seront définies par les notes que leur ont attribuées les répondants (deux mots notés de façon similaire seront proches, et d'autant plus proches que la similarité sera plus grande) alors que les proximités entre répondants seront définies par l'ensemble des 210 notes qu'ils ont données (deux répondants seront

---

1. Ces données sont donc traduites sous forme d'un tableau rectangulaire à 16 582 lignes (répondants) et 210 colonnes (mots).

2. Bien que reposant sur une théorie algébrique et géométrique plus élaborée qui est esquissée en annexe A1.1.

proches s'ils ont notés tous les mots d'une façon similaire : on parlera alors de personnes ayant des *profils de réponse similaires*).

Ces résultats se présentent graphiquement dans un espace réduit dont les dimensions s'appellent les *composantes principales*<sup>1</sup>. Celles-ci, associées à des axes principaux, permettent d'obtenir la meilleure visualisation approchée des distances entre les répondants d'une part et entre les mots d'autre part<sup>2</sup>. Les répondants étant anonymes, ils seront décrits par leurs caractéristiques de base (variables signalétiques). Ainsi une visualisation des proximités entre mots et catégories socio-professionnelles, sexe ou âge par exemple pourra éclairer la lecture des notations de chacune de ces catégories.

#### *Règles de lecture des graphiques*

La coordonnée d'une variable (mot) sur un axe est le coefficient de corrélation de cette variable avec la nouvelle variable synthétique (appelée *facteur*) que représente l'axe. Les axes sont qualifiés par l'ensemble des mots qui leur sont bien corrélés (coordonnées factorielles élevées). Des répondants éloignés sur une partie de cet axe auront bien noté les mots très corrélés avec cette même partie ; a contrario, de tels sujets auront sous-noté les mots anticorrélés (corrélés négativement) avec l'autre partie de l'axe. L'examen d'un plan factoriel (plan engendré par deux axes) permet de visualiser les corrélations entre les variables et d'identifier les groupes d'individus ayant donné les mêmes notes aux mêmes mots mis en évidence par ce plan.

Sur le plan factoriel, la proximité entre deux mots s'interprète en termes de corrélations de notes : deux mots seront proches s'ils sont notés de façon similaire par les répondants ; au contraire, ils seront éloignés si des notes opposées leur ont été attribuées ; deux répondants seront proches si ils ont donné des notes voisines pour chacun des mots.

En fait, les répondants (anonymes) ne nous intéressent ici que par leur signalétique (variables dites socio-démographiques telles que âge, sexe, profession, etc.) ou par leurs réponses à une question d'opinion, d'attitude ou de comportement, posée dans la même enquête. On interprète alors le nuage des répondants à partir de ces variables ou de tout autre thème abordé par l'enquête. Cette phase consiste à positionner en *éléments supplémentaires* ces variables qui n'appartiennent pas à la partie purement sémiométrique du questionnaire : ces variables n'interviennent pas dans la formation et définition de l'axe, elles l'illustrent *a posteriori*.

---

1. Celles-ci sont des nouvelles variables (que l'on peut désigner par les expressions *variables latentes*, ou *variables cachées*, ou encore *variables artificielles*) qui sont en fait des *combinaisons linéaires* des variables initiales, c'est à dire des *combinaisons linéaires* des notes données aux mots.

2. Ces nouvelles variables de synthèse sont quantitatives (ou encore : numériques), indépendantes entre elles (plus exactement : non corrélées), et d'importance décroissante (l'importance étant mesurée par la *variance* ou *dispersion*). Les premières composantes de variances plus élevées rendent compte des plus fortes structures. Elles sont alors hiérarchisées des plus dominantes aux plus marginales.

L'encadré ci-dessus fournit les premières règles d'interprétation des composantes principales ; pour chaque axe, seront donnés les principaux mots qui lui sont corrélés, positivement et négativement, et les populations qu'il discrimine<sup>1</sup>.

## 1.4 Les axes stables : « axes sémiométriques »

Seuls les six premiers axes principaux (ou encore: « facteurs ») qui semblent interprétables seront présentés. Le chapitre 2 montrera que ces six axes ont des propriétés de stabilité satisfaisantes (stabilité mesurée selon des critères expérimentaux - stabilité dans le temps, dans l'espace - et selon des critères statistiques plus techniques). Les axes suivants sont moins stables sur des échantillons de taille usuelle (de 1 000 à 2 000 personnes), mais cela ne signifie pas pour autant qu'ils ne représentent que du bruit<sup>2</sup>. On se limite en fait ici à la partie la plus invariante et la plus robuste de la structure observée.

Une première tentative d'interprétation sera alors proposée, et une qualification de chaque axe sera suggérée<sup>3</sup>. A chaque pôle d'un axe, émerge ainsi un *concept*. Chaque axe est alors défini par deux concepts opposés.

Nous reviendrons au chapitre 6 d'une manière beaucoup plus détaillée sur les problèmes posés par l'interprétation et nous proposerons alors une grille de lecture personnelle du sens des axes, et, au-delà, une hypothèse sur la nature de « l'objet » que la sémiométrie observe.

### 1.4.1 Axe 1 : Un axe méthodologique de participation

Le premier axe auquel sera donné le nom d'« *axe de participation* » n'a pas le même statut que les autres. Il s'agit d'un axe lié à la méthodologie

---

1. Ces populations sont identifiées par la technique dite des « variables nominales supplémentaires »: les caractéristiques socio-démographiques des sujets enquêtés sur l'axe apparaissent comme les point moyens des individus concernés (*cf.* annexe A1.9.3).

2. La stabilité des axes plus éloignés dépend de la taille de l'échantillon. Ainsi seuls les quatre premiers facteurs sont explicites sur un échantillon de 500 sujets ; les six premiers axes le sont à partir d'un échantillon de 1 000 enquêtés mais les deux autres axes (7 et 8) ne commencent à se clarifier qu'à partir de 5 000 individus.

3. Si les calculs apparaissent comme « objectifs », l'interprétation est de l'ordre du commentaire plus libre suggérant plusieurs dénominations possibles des axes. Mais, rappelons qu'il ne s'agit que d'étiquettes qui résument et, donc, déforment la réalité sémantique et psychologique des axes, laquelle dépend de l'ensemble des mots qui leur sont corrélés.

même de l'enquête qui traduit l'attitude des personnes interrogées à l'égard du questionnaire, alors que les autres axes concernent le contenu même du questionnaire, c'est-à-dire la *structure sémiométrique* proprement dite<sup>1</sup>.

D'un côté de l'axe, on trouve les mots suivants<sup>2</sup>:

(+) *Courage, Efficace, Dynamique, Récompense, Protéger, Honnête, Robuste, Confiance, Élégance, Respect, Pureté, Précision, Douceur, Admirer, Politesse, Construire, Charitable, Confort, Gaieté, Honneur, Produire, Raison, Guérir, Précieux, Maternel.*

De l'autre côté :

(-) *Trahir, Désordre, Guerre, Angoisse, Faute, Danger, Rompre, Vide, Mort, Révolte, Orage, Noir, Doute, Attaquer, Ironie, Labyrinthe, Critiquer, Fusil, Cri, Vieillir, Détachement, Muraille, Feu, Sauvage, Désert.*

Les catégories d'individus plus spécialement concernées par ces pôles sont respectivement :

(+) Femmes ; personnes âgées ; personnes ayant un niveau d'instruction peu élevé ; retraités ; personnes habitant les régions Est ou Nord et/ou des petites communes.

(-) Jeunes ; hommes ; actifs ; personnes ayant fait des études supérieures ; personnes habitant la région parisienne.

Cet axe oppose des mots qui exaltent des valeurs assez consensuelles et expriment un certain conformisme social (*Courage<sup>3</sup>, Efficace, Dynamisme, Précision, Récompense...*) à des mots qui évoquent plutôt des valeurs de transgression de ce qui est socialement valorisé (*Trahir, Guerre, Désordre, Angoisse...*).

Au « conformisme » correspond une population plutôt âgée, féminine et de niveau d'instruction peu élevé opposée à une population représentée par les hommes, les jeunes, les cadres et les diplômés de l'enseignement supérieur.

1. Ce premier axe représente 12 % de la variance totale. Rappelons que, contrairement à un préjugé technique assez répandu, ce pourcentage ne peut en aucune façon être interprété comme un *pourcentage d'information*. La variance totale est lestée par un *bruit statistique* considérable, et ne sert de mesure de référence que pour comparer des analyses entre elles (*cf.* annexe A1.9 consacrée aux outils de validation).

2. L'orientation des axes est arbitraire. Le fait que le signe (+) correspond à des valeurs communément considérées comme « positives » est purement fortuit. Les catégories sont ordonnées des plus significatives (statistiquement) aux moins significatives.

3. Par convention, dans le corps du texte, les mots du questionnaire sémiométrique sont écrits en italique avec une majuscule.

Toutefois, le fait que, généralement, le premier axe d'une analyse en composantes principales effectuée sur des notes centrées réduites rend compte d'un effet de taille doit nous inciter à la plus grande prudence dans l'exercice toujours périlleux de l'interprétation<sup>1</sup>. En effet, les mots *Efficace*, *Dynamisme*, etc., sont parmi les mots qui recueillent les meilleurs scores tandis que *Trahir*, *Guerre*, etc., ont des scores parmi les plus faibles. Nous reviendrons de façon beaucoup plus détaillée et plus technique sur l'interprétation de cet axe au chapitre 5 consacré aux problèmes méthodologiques liés à la notation.

#### 1.4.2 Axe 2 : « Devoir / Plaisir »

D'un côté de l'axe, se trouvent les mots suivants :

(-) *Discipline, Obéir, Patrie, Morale, Soldat, Economiser, Industrie, Prêtre, Règle, Loi, Fermeté, Foi, Travail, Sacrifice, Honneur, Dieu, Elite, Interdire, Rigide, Punir, Frontière, Chasse, Tradition, Recueillement, Métallique.*

De l'autre côté :

(+) *Sensuel, Rêver, Aventurier, Original, Ile, Nudité, Sauvage, Volupté, Légèreté, Océan, Bohème, Désir, Charnel, Emotion, Séduire, Evasion, Lune, Humour, Moelleux, Orage, Révolte, Magie, Mystère, Musique, Peau.*

Les catégories d'individus plus spécialement concernées par ces deux pôles de l'axe sont respectivement :

(-) Individus de plus de 55 ans ; retraités ou agriculteurs-exploitant.

(+) Individus de moins de 30 ans ; diplômés entre 30 et 55 ans; cadres ou employés.

Les mots *Discipline, Morale, Patrie, Obéir, Soldat, Prêtre...*, évoquent des valeurs d'ordre et de tradition qui renvoient, de façon peut-être caricaturale, au « mariage du sabre et du goupillon » évoqué à certaines époques de l'histoire de France, tandis que les mots *Sensuel, Rêver, Légèreté, Sauvage, Aventurier, Voluptueux...*, semblent correspondre à un autre volet, celui du mouvement et du plaisir qui connut son heure de gloire

---

1. L'effet de taille est en général dû au fait que, dans beaucoup de recueils de mesures, les corrélations sont, en grande majorité, positives : dans le cas de mensurations sur des individus ou des animaux, la taille est en effet la principale source de dispersion: brièvement, toutes les mensurations sont grandes chez les individus grands, et petites chez les petits. Dans la plupart des applications, l'axe suivant le facteur de taille, *facteur de forme*, est plus structurel : il renseigne sur le réseau plus subtil et caché de corrélations qui subsistent « à taille égale ».

à l'occasion des événements de mai 68. Par convention, nous appellerons cet axe: « Devoir / Plaisir »<sup>1</sup>.

Cet axe est fortement structuré par l'âge ; le « Devoir » est une valeur qui caractérise de façon assez nette les personnes âgées et le « Plaisir » est une valeur plus marquée chez les plus jeunes. Mais cet axe subsiste si on analyse seulement l'ensemble des individus ayant entre 35 et 55 ans. C'est un axe certainement lié à l'âge, mais non caractérisé uniquement par l'âge. Son interprétation est plus riche et plus complexe<sup>2</sup>.

#### 1.4.3 Axe 3 : « Attachement / Détachement »

D'un côté de l'axe, se trouvent les mots suivants :

(+) *Bijou, Cadeau, Confort, Récompense, Or, Maison, Argent, Mariage, Richesse, Gaieté, Famille, Parfum, Fidélité, Hériter, Fleur, Gratuit, Mode, Tendresse, Douceur, Élégance, Guérir, Paix, Maternel, Caresse, Rire.*

De l'autre côté :

(-) *Danger, Mort, Rompre, Orage, Angoisse, Vide, Punir, Labyrinthe, Cri, Critiquer, Doute, Révolte, Attaquer, Désert, Détachement, Faute, Feu, Ironie, Vieillir, Muraille, Effort, Méfiance, Question, Interroger, Inconnu.*

Les catégories d'individus plus spécialement concernées par les pôles de cet axe sont respectivement :

(+) Femmes ; personnes non titulaires du baccalauréat ; les ouvriers ou employés.

(-) Hommes ; diplômés ; cadres ; les artisans-commerçants ou les professions intermédiaires.

Les mots *Bijou, Cadeau, Confort, Mariage, Famille, Maison, Gaieté...* semblent parler d'un certain bonheur de vivre ensemble. Ce sont aussi des mots surnotés par les femmes par rapport aux hommes, tandis qu'à l'opposé, les mots *Danger, Mort, Punir, Vide, Rompre, Attaquer...* qui sont plutôt des mots surnotés par les hommes, apparaissent comme des mots de rupture et de déréliction. Par convention, nous appellerons cet axe : « Attachement / Détachement ». Cet axe<sup>3</sup> est marqué par une opposition Homme / Femme. Il ne s'agit cependant pas d'un axe qui se limite à une

---

1. Cet axe représente 5 % de la variance totale. Ce pourcentage est hautement significatif statistiquement, comme les pourcentages relatifs aux axes dont la description va suivre.

2. Une interprétation plus complète est proposée au chapitre 6.

3. Cet axe explique 4,2 % de la variance.

opposition entre les sexes. En effet, cet axe se retrouve si on procède à des analyses séparées des répondants hommes et des répondants femmes<sup>1</sup>.

#### 1.4.4 Le plan sémiométrique des axes 2 et 3

Ce *plan sémiométrique* (figure 1.2) sera appelé le *plan sémiométrique principal* (il est en effet engendré par les deux axes les plus importants, si l'on omet le premier axe qui est un axe de notation). Ce plan est très utilisé dans l'iconographie et la communication des analyses sémiométriques<sup>2</sup>. Il décrit simultanément les axes « Devoir / Plaisir » (axe horizontal) et « Attachement / Détachement » (axe vertical). Cette représentation plane permet en quelque sorte de compléter l'interprétation de chacun des pôles en les déclinant selon une dimension supplémentaire. On peut ainsi parler, à propos des quatre quadrants de ce plan, de « Plaisir-Attachement », de « Plaisir-Détachement », de « Devoir-Attachement », de « Devoir-Détachement ». L'interprétation est plus riche et plus nuancée. Elle est détaillée dans la cinquième section de ce chapitre consacrée à l'analyse de quelques croisements d'axes.

#### 1.4.5 Axe 4 : « Esprit / Matière »

D'un côté de l'axe, on trouve les mots suivants :

(+) *Livre, Art, Poésie, Théâtre, Arbre, Ecole, Enseigner, Pardon, Ecrire, Recueillement, Charitable, Etranger, Réfléchir, Nid, Paix, Fleur, Âme, Modération, Patience, Chercheur, Modestie, Consoler, Fleuve.*

Et de l'autre côté :

(-) *Richesse, Puissance, Or, Argent, Gloire, Conquérir, Vitesse, Ambition, Commander, Ruse, Fusil, Hériter, Séduire, Propriété, Attaquer, Victoire, Bijou, Mode, Désir, Précieux, Sensuel, Armure, Rigide, Elite, Gratuit.*

Les catégories d'individus plus spécialement concernées sont respectivement :

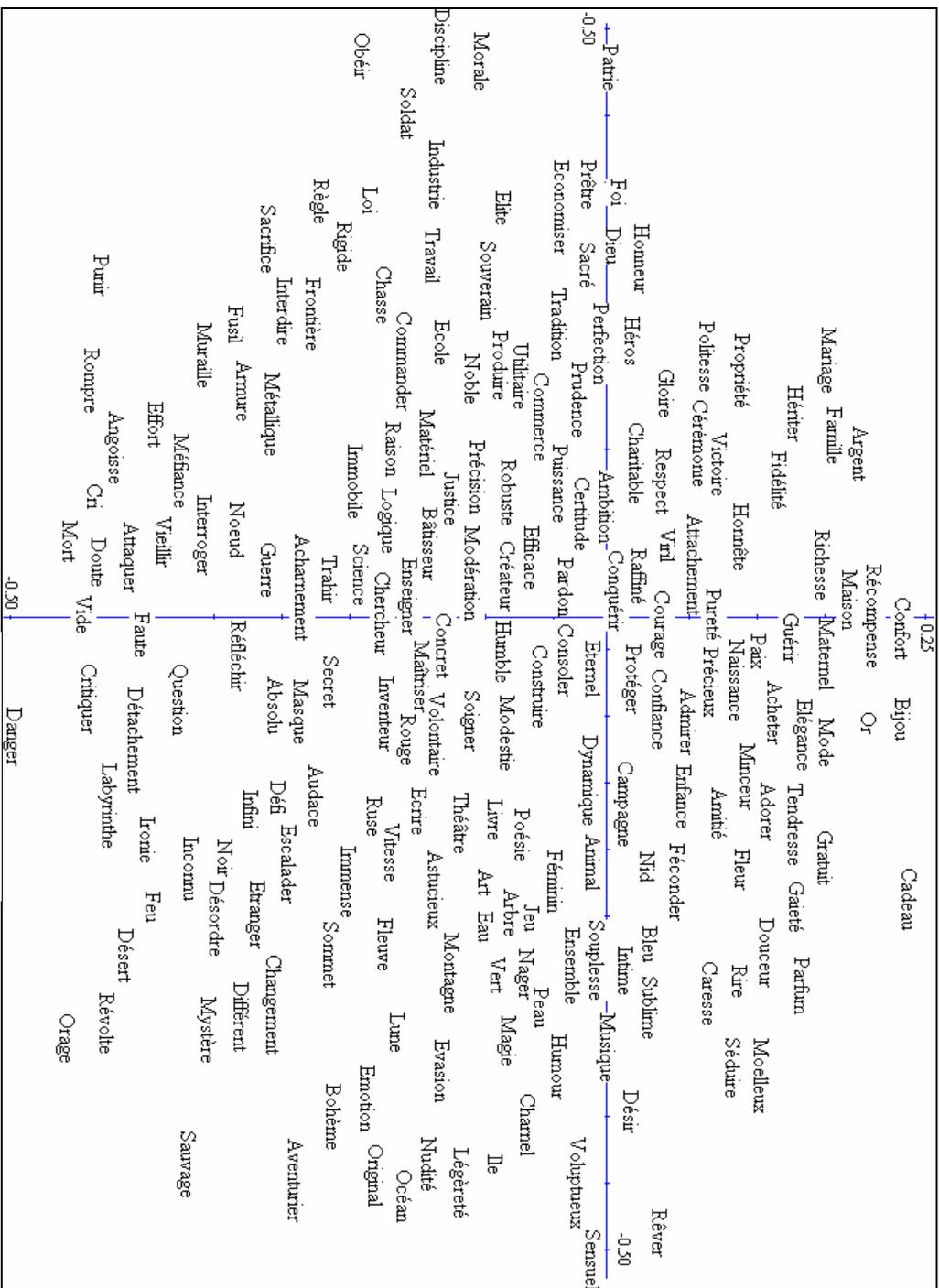
(+) personnes de plus de 55 ans ; diplômés ; femmes ; cadres.

---

1. Ce point sera développé dans la section 2.1.3 du second chapitre consacré à la stabilité des structures observées.

2. Le chapitre 7 qui présente un certain nombre d'applications de la sémiométrie fera souvent appel à une représentation des mots dans ce plan engendré par les axes 2 et 3.

ATTACHEMENT



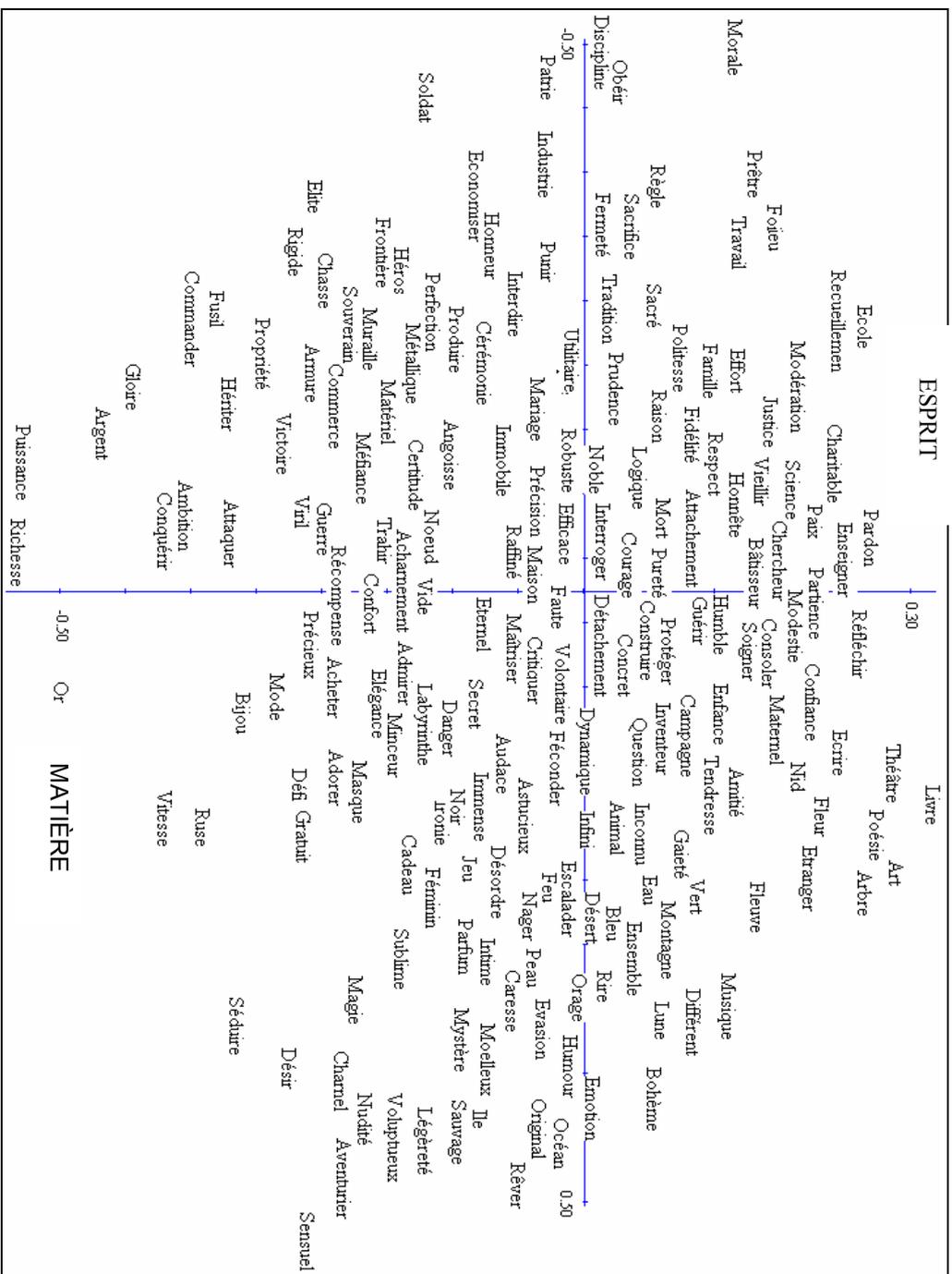


Figure 1.3: Plan sémiométrique (2,4) - Croisement de « Devoir / Plaisir » avec « Esprit / Matière »

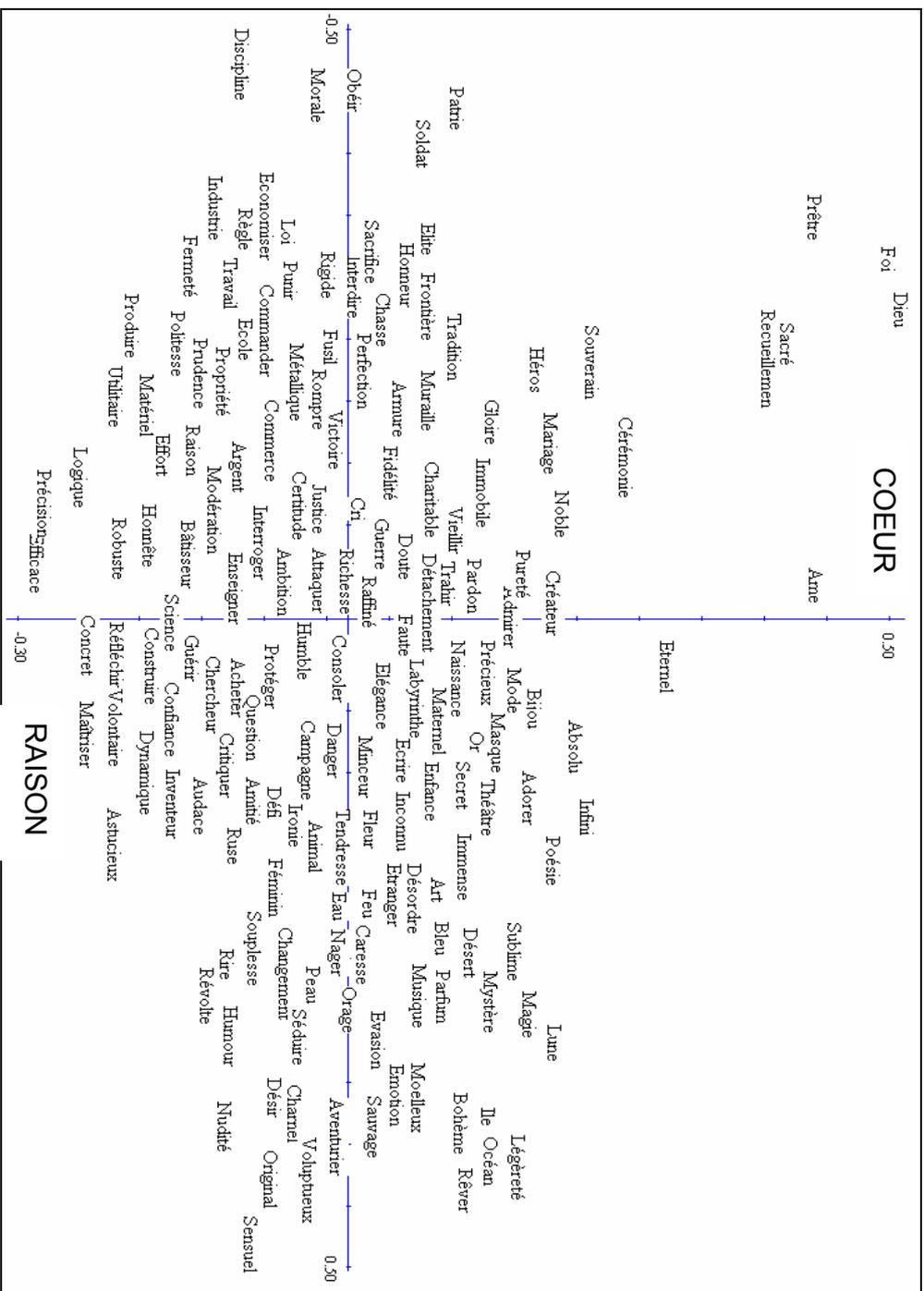


Figure 1.4 : Plan sémiotique (2,5) - Croisement de « Devoir / Plaisir » avec « Cœur / Raison »

HUMILITÉ

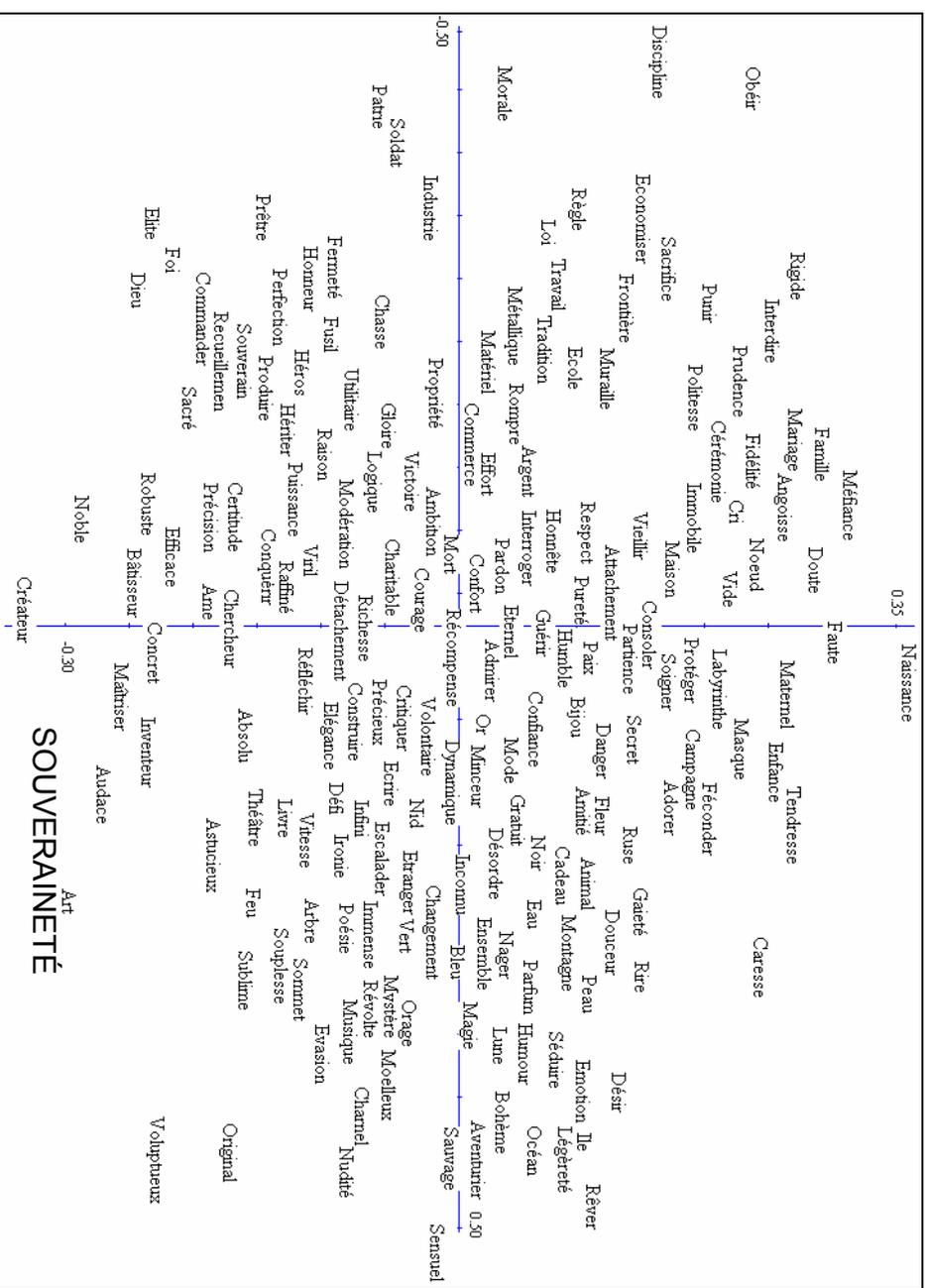


Figure 1.5 : Plan sémiotique (2,6) - Croisement de « Devoir / Plaisir » avec « Humilité / Souveraineté »

(-) personnes de moins de 30 ans, non titulaires du baccalauréat ; ouvriers, employés, artisans-commerçants, agriculteurs-exploitant ; hommes.

Les mots *Richesse, Or, Argent, Gloire, Conquérir, Vitesse, Commander...* évoquent l'attachement aux biens de ce monde, la foi dans des valeurs matérialistes, en somme le culte de Mammon, tandis que les mots qui s'y opposent, à l'autre extrémité de l'axe, *Livre, Art, Théâtre, Poésie, Âme, Ecole, Etranger, Réfléchir, ...* nous parlent d'un autre système de valeurs, plus spirituelles, plus élaborées, plus sublimées qui seraient celles d'un royaume céleste sécularisé. Par convention, nous appellerons cet axe : « Esprit / Matière »<sup>1</sup>.

Cet axe (3.3 % de la variance) oppose deux profils très distincts de répondants : les jeunes hommes non titulaires du baccalauréat aux femmes plus âgées et diplômées. Comme les axes précédents, il ne se limite pas à une opposition entre ces catégories, car on retrouve un axe similaire si l'on procède à une analyse à l'intérieur même de ces catégories.

#### 1.4.6 Axe 5 : « Cœur / Raison »

D'un côté de l'axe, se trouvent les mots suivants :

(+) *Dieu, Foi, Âme, Prêtre, Sacre, Recueillement, Eternel, Cérémonie, Souverain, Infini, Noble, Bijou, Poésie, Absolu, Mariage, Adorer, Héros, Lune, Magie, Légèreté, Mode, Masque, Gloire, Pureté, Mystère.*

De l'autre côté :

(-) *Efficace, Précision, Logique, Concret, Maîtriser, Astucieux, Utilitaire, Robuste, Produire, Réfléchir, Volontaire, Construire, Dynamique, Matériel, Effort, Audace, Honnête, Bâisseur, Inventeur, Confiance, Fermeté, Raison, Nudité, Science, Industrie.*

Les catégories d'individus plus spécialement concernées par les deux pôles de cet axe sont respectivement :

(+) femmes ; jeunes ; personnes habitant les régions Nord ou Est.

(-) hommes ; artisans-commerçant, professions intermédiaires, cadres, ouvriers ; personnes entre 40 et 50 ans ; personnes habitant la région parisienne.

---

1. Cet axe sera ultérieurement nommé « Sublimation / Matérialisme » pour des raisons qui seront explicitées au chapitre 6 (Essais d'interprétation).

Les mots *Dieu, Foi, Âme, Sacré, Eternel, Infini, etc.*, expriment une aspiration à une transcendance dont la raison ne saurait rendre compte, un besoin de rechercher, au-delà d'une logique trop rigoureuse, des formes d'expression plus oniriques ou irrationnelles ; tandis que les mots *Efficace, Précision, Dynamique, Construire, Utilitaire, Volontaire*, connotent un souci de rationalisme, de pragmatisme et le désir de vivre dans un monde parfaitement logique et tout dédié à l'action, qui n'est pas sans faire penser à l'idéal des Lumières. Par convention, cet axe sera appelé : « Cœur / Raison »<sup>1</sup>.

Ce cinquième axe<sup>2</sup> oppose les hommes et les actifs aux femmes et aux personnes jeunes.

#### 1.4.7 Axe 6 : « Humilité / Souveraineté »

D'un côté de l'axe, on trouve les mots suivants :

(+) *Naissance, Méfiance, Faute, Doute, Famille, Rigide, Tendresse, Mariage, Interdire, Caresse, Masque, Obéir, Enfance, Prudence, Maternel, Cri, Angoisse, Immobile, Nœud, Fidélité, Punir, Vide, Politesse, Cérémonie.*

De l'autre côté :

(-) *Créateur, Noble, Art, Dieu, Audace, Elite, Foi, Maîtriser, Robuste, Bâtitteur, Efficace, Concret, Sacre, Volupté, Inventeur, Âme, Commander, Recueillement, Astucieux, Absolu, Prêtre, Souverain, Sublime, Théâtre, Certitude.*

Les catégories d'individus plus spécialement concernées par ces deux pôles sont respectivement :

(+) ouvriers, employés ; personnes de moins de 40 ans ; personnes non titulaires du baccalauréat ; femmes ; personnes vivant en région rurale.

(-) cadres ; personnes âgées ; retraités ; personnes ayant un niveau d'instruction élevé ; personne habitant une agglomération importante dans les régions parisienne ou méditerranéenne. L'interprétation de cet axe<sup>3</sup> est moins aisée. En effet, si l'on trouve à l'une de ses extrémités des mots

---

1. Il s'agit ici du cœur et de la raison « pascaliens ». En tout état de cause, si les concepts correspondant aux demi-axes peuvent être considérés comme des réalités sémantiques relativement objectives puisque issus d'une analyse statistique, les noms qui leur sont attribués ne sont que des étiquettes provisoires qui seront justifiées ou modifiées ultérieurement dans le chapitre 6 (Essais d'interprétation). Nous verrons qu'il peut être également nommé « Idéalisation / Pragmatisme ».

2. Cet axe correspond à 2 % de la variance totale, pourcentage qui reste hautement significatif statistiquement compte tenu notamment du nombre de variables.

3. Ce sixième axe correspond à 1,8 % de la variance totale.

comme *Noble, Créateur, Art, Dieu, Foi, Elite, Audace...* qui évoquent une sorte d'élitisme, une aspiration à une certaine souveraineté ou distinction, pour reprendre la terminologie de Pierre Bourdieu, les mots de l'autre extrémité se décomposent en deux sous-ensembles sémantiques, *Naissance, Famille, Tendresse, Mariage, Enfance, Maternel*, d'une part et, *Méfiance, Faute, Doute, Rigide, Obéir, Interdire*, d'autre part, dont la synthèse semble, à première vue, difficile à réaliser. Pourtant, le fait que cet axe oppose des individus plutôt âgés et d'un niveau culturel élevé à des individus dont le statut social est plus modeste permet de penser qu'il rend compte de la confrontation entre deux systèmes de valeurs bien réels. Par convention, et sans entrer plus avant dans son interprétation, cet axe sera appelé : « Humilité / Souveraineté ».

#### 1.4.8 Les axes suivants

Les axes suivants sont plus difficilement interprétables. Lorsque les analyses statistiques portent sur des échantillons inférieurs à mille individus, les axes 7 et 8 n'apparaissent pas systématiquement. Ils n'acquièrent une bonne stabilité qu'à partir d'un échantillon de cinq mille individus<sup>1</sup>. On tente ici seulement une description de ces axes sans se hasarder à les interpréter et sans chercher à leur donner un nom<sup>2</sup>.

##### - Axe 7

D'un côté de l'axe, on trouve les mots suivants :

(-) *Audace, Foi, Secret, Sublime, Ambition, Maîtriser, Pureté, Charnel, Désir, Séduire, Consoler, Respect, Défi, Ruse, Acharnement, Sacré, Adorer, Eternel, Sensuel, Volontaire, Pardon, Âme, Intime.*

---

1. Cette analyse, effectuée ici sur un échantillon de plus 11 000 individus totalisant l'ensemble des quatre premières enquêtes françaises, a aussi été réalisée sur un échantillon de 1 000, puis 5 000 individus. La taille de l'échantillon améliore la qualité des axes de rang élevé. Les axes 7 et 8 représentent respectivement 1,3 et 1,2 % de la variance totale (pourcentages hautement significatifs statistiquement pour 210 variables et 5 000 individus).

2. Il faut bien faire la distinction entre : a) la mise en évidence de la stabilité d'un axe qui constitue un « fait statistique » établi de façon objective, répétitive, automatique à partir du recueil de données ; b) la description de cet axe, qui comporte une part de subjectivité comme toute description en langage naturel impliquant de choisir des mots et des concepts ; c) le fait de nommer un axe, qui implique une simplification et un choix drastique, et donc une part d'interprétation plus importante ; d) l'interprétation proprement dite de l'axe, qui n'est pas toujours possible, qui fait appel à des catégories, des théories ou des modèles, explicites ou implicites. Dans ce chapitre, les six premiers axes ont été décrits et nommés. Une interprétation sera proposée plus loin au chapitre 6. Les axes suivants seront simplement décrits, avec des commentaires interprétatifs qui n'ont d'autre ambition que de rendre la description moins aride.

(+) *Fleuve, Arbre, Montagne, Théâtre, Livre, Océan, Campagne, Fleur, Nager, Eau, Métallique, Maison, Animal, Sommet, Art, Science, Industrie, Ile, Lune, Hériter, Commerce, Rouge, Bijou, Poésie.*

Les catégories d'individus plus spécialement concernées sont respectivement :

(-) Personnes de moins de 30 ans et plus particulièrement de la tranche d'âge 15-19 ans; appartenant à une famille de cadres ; personne ayant un niveau d'instruction supérieur.

(+) Personnes de plus de 55 ans ; retraité(e)s et conjoint(e)s de retraité(e)s ; niveau d'instruction bas ; Bassin Parisien Ouest et région méditerranéenne ;

Si l'on distingue bien les deux groupes qui s'opposent sur l'axe 7, d'une part des personnes âgées n'ayant plus d'activité professionnelle, sans instruction et, d'autre part, des jeunes issus de milieux intellectuellement évolués, étudiants ou lycéens et, donc, non encore entrés dans la vie active, il semble plus difficile d'identifier les concepts et, derrière eux, les valeurs sur lesquelles ils s'opposent.

Il s'agit plus d'un clivage entre mentalités, les plus jeunes et les plus instruits affectionnant les concepts abstraits, *Audace, Foi, Secret, Sublime, Ambition, Maîtriser, Pureté*, etc., s'opposant aux plus âgés, sans instruction, s'étant déjà confrontés aux réalités de la vie, privilégiant les objets concrets (et plus particulièrement ceux évoquant la nature), *Fleuve, Arbre, Montagne, Livre, Océan, Fleur, Maison, Lune, Bijou*, etc..

#### - Axe 8

D'un côté de l'axe on trouve les mots suivants :

(+) *Théâtre, Livre, Ecrire, Raffiné, Poésie, Art, Mode, Parfum, Bijou, Élégance, Critiquer, Ecole, Noble, Enseigner, Minceur, Interroger, Doute, Légèreté, Justice, Ironie, Question, Précieux, Rompre, Réfléchir.*

De l'autre côté :

(-) *Montagne, Campagne, Sommet, Fusil, Ile, Eau, Aventurier, Escalader, Animal, Océan, Fleuve, Foi, Armure, Nager, Dieu, Arbre, Famille, Prêtre, Chasse, Sacré, Vitesse, Sauvage, Orage, Soldat.*

Les catégories d'individus plus spécialement concernées sont respectivement :

(+) Femmes âgées, d'un niveau d'instruction élevé ; cadres et conjointes de cadres.

(-) Hommes de moins de trente ans ; ouvriers et agriculteurs ; niveau d'instruction bas.

C'est un peu l'opposition de la Belle et de la Bête ou des Précieuses et des ruffians, avec d'un côté l'élégance et le raffinement féminin, *Théâtre, Livre, Ecrire, Raffiné, Poésie, Art, Mode, Parfum...*, et, de l'autre, *Montagne, Campagne, Sommet, Fusil, Aventurier, Escalader, Animal*<sup>1</sup>.

Comme le précédent, cet axe n'apparaît que lorsqu'on travaille sur des échantillons de plus cinq mille individus.

## 1.5 Les plans sémiométriques

Nous venons de présenter les six premières dimensions stables, et de décrire brièvement les deux dimensions suivantes. La prise en compte simultanée des axes deux à deux (il s'agira donc de plans principaux, et non plus d'axes principaux) apporte un nouvel éclairage sur la structure sémiométrique. Elle nuance et affine la description des axes et leurs interprétations. L'exemple du plan des axes 2 et 3 a déjà été évoqué (figure 1.2). Nous allons commenter de façon plus approfondie ce plan [noté plan (2,3)], puis examiner quelques autres plans.

### 1.5.1 Les multiples facettes du « Plaisir »

En croisant le deuxième axe « Devoir / Plaisir » avec le troisième axe « Attachement / Détachement », on obtient deux composantes opposées du Plaisir, que nous appellerons « Plaisir-Attachement » et « Plaisir-Détachement ».

Le « Plaisir-Attachement » (en haut et à droite de la figure 1.2)<sup>2</sup> est caractérisé par les mots : *Douceur, Caresse, Parfum, Gaieté, Cadeau, Rire, Tendresse, Moelleux, Séduire*, le Plaisir-Détachement (en bas et à droite de la figure 1.2) l'est, lui, par les mots : *Aventurier, Sauvage, Orage, Révolte, Mystère, Feu, Bohème, Original, Emotion*.

*Commentaires / Suggestions d'interprétation*

---

1. Opposition qui, d'ailleurs, peut être source, en certaines circonstances, d'une troublante attraction. On pense au couple formé par Lady Chatterley et son garde chasse. Cette ambivalence nous laisse penser que cet axe n'est pas seulement sociologique (on veut dire ici : opposant des catégories sociales) mais, dans une certaine mesure, psychologique.

<sup>2</sup> Cette figure se trouve ci-dessus page 24.

Si le « Plaisir-Attachement » évoque la joie de vivre à travers une douce exaltation des sens, un univers de béatitude dans lequel chaque instant semble une éternité de suave jouissance, le « Plaisir-Détachement » a un goût plus âcre, il fait entendre une note plus âpre, plus rude, plus sombre, plus violente, plus tragique. Il s'agit bien de deux modalités du plaisir de vivre, mais celui-ci du guerrier et celui-là .... de son repos.

On voit alors quelle subtile contribution peut être apportée à l'étude sémantique des mots par... les résultats d'une enquête par sondage auprès de la population générale ! Cette contribution est probablement assez inattendue, voire incongrue, pour les linguistes et les sémiologues. Elle est d'autant plus remarquable qu'elle ne relève absolument pas d'une enquête d'opinion sur le thème abordé ici « le plaisir » ou les « proximités sémantiques » (dont le protocole serait probablement très lourd) mais d'une simple notation de mots un par un<sup>1</sup>.

Continuons notre déclinaison du Plaisir en croisant maintenant l'axe 2 « Devoir / Plaisir » avec l'axe 4 « Esprit / Matière ». La figure 1.3 représente le plan [2,4] ainsi obtenu.

On peut alors distinguer, dans la figure 1.3, le « Plaisir-Esprit » (en haut et à droite) qui est caractérisé par les mots : *Art, Fleuve, Etranger, Musique, Bohème, Arbre, Poésie*, et le « Plaisir-Matière » (en bas et à droite), par les mots : *Sensuel, Désir, Séduire, Charnel, Magie, Aventurier, Nudité, Voluptueux, Légèreté*.

#### *Commentaires / Suggestions d'interprétation*

Le « Plaisir-Esprit » est celui, subtil, des sens apaisés par la contemplation des beautés de l'art et de la nature, tandis que le « Plaisir-Matière » recherche des bénéfices plus frustes, plus immédiats, plus proches de cette autre nature que représentent nos instincts. On notera que cette diagonale oppose l'art à la sexualité ou, plus exactement, une sexualité domestiquée, canalisée, élaborée à une sexualité naturelle et spontanée.

En croisant encore l'axe 2 « Devoir / Plaisir » avec l'axe 5 « Cœur / Raison », on obtient le « Plaisir-Cœur » (en haut et à droite de la figure 1.4) qui est caractérisé par les mots : *Ile, Rêver, Légèreté, Océan, Bohème, Lune, Magie, Mystère, Sublime, Parfum, Désert, Infini*, et le

---

1. Les 16 582 personnes interrogées ne peuvent d'ailleurs soupçonner en donnant ces notes que des faits statistiques stables de cette nature seront extraits automatiquement de la matrice des corrélations des notes (matrice symétrique à 210 lignes et 210 colonnes contenant les 21 945 coefficients de corrélation).

Plaisir-Raison (en bas et à droite de la figure 1.4), caractérisé par les mots : *Nudité, Audace, Astucieux, Sensuel, Révolte, Humour.*

*Commentaires / Suggestions d'interprétation*

Le « Plaisir-Cœur » est le plaisir lié au rêve et à la contemplation de phénomènes dont le territoire se situe aux marches évanescentes de la réalité, dans ce *no man's land* où les objets semblent ne pas avoir pris encore une forme claire et définitive.

Le « Plaisir-Raison », au contraire, ne s'en laisse pas conter et aime à poser sur les objets un regard cru et sans illusion.

Enfin, en croisant toujours l'axe 2 « Devoir / Plaisir » avec l'axe 6 « Humilité / Souveraineté », on obtient le « Plaisir-Humilité » (en haut et à droite de la figure 1.5 représentant le plan [2,6]) qui est caractérisé par les mots : *Rire, Désir, Caresse, Douceur, Gaieté, Tendresse, Ruse*, et le « Plaisir-Souveraineté » (en bas et à droite de la figure 1.5), caractérisé par les mots : *Original, Voluptueux, Sublime, Feu, Sommet, Art.*

*Commentaires / Suggestions d'interprétation*

Le « Plaisir-Humilité » n'est pas sans rappeler le « Plaisir-Attachement », quoiqu'en plus innocent, en plus enfantin, qui semble comme une réminiscence de l'état paradisiaque d'avant le « Pêché ». Le Plaisir-Souveraineté se complairait, lui, plutôt, dans la recherche de sensations sublimes et capiteuses.

Bien que ces déclinaisons du Plaisir s'estompent progressivement lorsque l'ordre des axes augmente, rendant ainsi l'interprétation et les commentaires de moins en moins assurés, on peut s'étonner de la finesse et de la cohérence de cette « décomposition sémantique »<sup>1</sup>.

### 1.5.2 Les multiples facettes du « Devoir »

L'examen des mêmes plans factoriels (figures 1.2 à 1.5), en focalisant cette fois sur la partie gauche de l'axe horizontal (axe 2, « Devoir / Plaisir ») fournit de la même manière les différentes facettes du « Devoir ». Nous invitons cette fois le lecteur à écrire son propre commentaire à partir de la consultation des figures précitées.

---

1. Nos commentaires et suggestions sont illustratifs, et peut-on dire, facultatifs. Leur présence ne doit pas obérer le statut scientifique des figures 1.2 à 1.5, qui sont dérivées du tableau de données de façon automatique et reproductible.

En croisant l'axe 2 avec l'axe 3, (« Attachement / Détachement »), on obtient le « Devoir-Attachement » (en haut et à gauche de la figure 1.2), caractérisé par les mots : *Fidélité, Cérémonie, Propriété, Politesse, Mariage, Famille*, et le « Devoir-Détachement », (en bas et à gauche de la figure 1.2), par les mots : *Sacrifice, Interdire, Fermeté, Loi, Règle, Obéir, Punir, Rigide, Frontière, Métallique, Muraille, Mort*.

En croisant l'axe 2 avec l'axe 4 (« Esprit / Matière »), on obtient le « Devoir-Esprit » (en haut et à gauche de la figure 1.3), caractérisé par les mots : *Modestie, Recueillement, Charitable, Justice, Modération, Respect, Dieu, Âme*, et le « Devoir-Matière » (en bas et à gauche de la figure 1.3), par les mots : *Propriété, Gloire, Argent, Puissance, Elite, Souverain, Commerce*.

En croisant l'axe 2 avec l'axe 5 (« Cœur / Raison »), on obtient le « Devoir-Cœur » (en haut et à gauche de la figure 1.4), caractérisé par les mots : *Eternel, Immobile, Âme, Sacré, Cérémonie, Souverain, Tradition*, et le « Devoir-Raison » (en bas et à gauche de la figure 1.4), caractérisé par les mots : *Propriété, Prudence, Matériel, Raison, Honnête, Produire, Utilitaire*.

En croisant l'axe 2 avec l'axe 6, « Humilité / Souveraineté », on obtient le « Devoir-Humilité » (en haut et à gauche de la figure 1.5), caractérisé par les mots : *Méfiance, Rigide, Fidélité, Prudence, Sacrifice, Mariage, Honnête, Obéir*, et le « Devoir-Souveraineté » (en bas et à gauche de la figure 1.5), par les mots : *Sacré, Dieu, Souverain, Elite, Commander, Héros, Perfection, Gloire*.

Cette simple évocation des facettes ou dimensions du « Devoir » confirme le constat fait à propos des dimensions du « Plaisir » : constat de cohérence, de finesse des résultats. Ajoutons que cette procédure de déclinaison systématique d'un concept suivant différents axes permet de percevoir le caractère profondément multidimensionnel de la topographie des mots, irréductible à un axe ou un plan.

### 1.5.3 Les dimensions sémantiques des mots eux-mêmes

Après avoir décliné les concepts que représentent les demi-axes principaux : « Devoir », « Plaisir », « Attachement », « Détachement », « Esprit », « Matière », « Humilité » et « Souveraineté », nous allons nous intéresser aux mots mêmes de la liste.

Un même mot peut, bien sûr, corrélérer avec plusieurs axes, chacune de ses « épiphanies » révélant une de ses dimensions sémantiques.

### Le mot *Mariage*

Ainsi le mot *Mariage* du « Devoir »<sup>1</sup>, que connotent les mots *Fidélité, Honnête, Respect, Politesse...*, n'est pas le même que celui de l'« Attachement », que connotent les mots *Confort, Maternel, Naissance, Confiance, Amitié, Tendresse et Douceur*, celui-là évoquant plutôt le mariage vu comme une institution, celui-ci, comme le cadre d'une relation d'amour entre deux êtres (cf. figure 1.2).

### Le mot *Révolte*

De même, la *Révolte* du « Plaisir », que connotent les mots *Aventurier, Original, Bohème, Evasion*, n'est pas la même que celle du « Détachement », que connotent les mots, *Rompre, Critiquer, Doute, Danger, Mort*. Dans un cas, il s'agit de la manifestation d'un désir d'échapper aux limites imposées par les contraintes sociales pour accéder à une vie plus large, dans l'autre, de briser ces contraintes quitte à se détruire soi-même (figure 1.2).

### Le mot *Frontière*

De même, la *Frontière* du « Détachement », que connotent les mots *Rigide, Fermer, Rompre, Muraille, Masque*, n'est pas la même que la *Frontière* du « Devoir », que connotent les mots *Mariage, Propriété, Règle, Morale*. La première est la frontière qui isole le sujet en rejetant l'autre, la seconde est celle qui distingue, organise, et garantit l'ordre, autant social que mental (figure 1.2).

### Le mot *Théâtre*

De même le *Théâtre* du « Cœur » ou encore de l'« Idéalisation » (axe 5), que connotent les mots *Cérémonie, Bijou, Masque, Magie*, n'est pas le même que le *Théâtre* de l'« Esprit » (axe 4), que connotent les mots *Art, Livre, Musique, Poésie, Ecrire*. L'un est le spectacle auquel on assiste, l'autre, l'œuvre que l'on crée.

### Le mot *Dieu*

Quant au mot *Dieu*, il corrèle avec quatre des axes de la sémiométrie, nous découvrant ainsi quelques unes de ses composantes sémantiques : le *Dieu* du « Devoir » (sur l'axe 2) que connotent les mots *Vertu, Sacrifice, Règle, Morale, Discipline* ; le *Dieu* de l'« Esprit » (sur l'axe 4) que connotent les mots *Pardon, Consoler, Charitable* ; le *Dieu* du « Cœur » (sur

---

1. On désigne par l'expression *Mariage* du « Devoir » la composante (ou encore : coordonnée) du mot *Mariage* sur l'axe du « Devoir ». De même, *Révolte* du « Plaisir » veut dire : composante du mot *Révolte* sur l'axe du « Plaisir ».

l'axe 5) que connotent les mots *Cérémonie, Théâtre, Magie, Mystère* ; et le Dieu de la « Souveraineté » (sur l'axe 6) que connotent les mots *Absolu, Créateur, Unique*. Le premier évoque le Dieu de la Loi de l'Ancien Testament, le deuxième, celui d'amour, du Nouveau Testament, le troisième, celui de la religion spectacle, et pour finir, le quatrième, le Dieu ineffable de la transcendance.

## 1.6 Conclusion

Dressons un premier bilan des résultats acquis, et aussi un inventaire des questions qui se posent à ce stade, inventaire qui annoncera, de façon motivée cette fois, la teneur des chapitres à venir.

La démarche suivie est très empirique et exploratoire : quelques idées simples, une coûteuse campagne de mesures, puis un traitement statistique assez classique et polyvalent destiné à extraire d'éventuels traits structuraux.

Il restera bien sûr à s'assurer de la réalité de ce qui a été observé, puis à comprendre la nature et la portée des nouvelles informations mises en évidence.

Il a beaucoup été question de mots, et peu d'individus ou de répondants. On a bien mentionné quelles catégories d'individus caractérisaient certains axes (les personnes âgées plutôt du côté du « Devoir », les jeunes plutôt de celui du « Plaisir », par exemple). Mais les individus interrogés sont anonymes, et ne pouvaient intervenir à cette étape de travail que par l'intermédiaire des catégories auxquelles ils appartenaient.

Or c'est une des propriétés de la méthode statistique utilisée que de positionner les individus sur les mêmes axes... le point moyen des individus d'une catégorie étant celui qui nous a servi à positionner la catégorie en question. On aura en général beaucoup d'autres informations sur ces individus en dehors des variables signalétiques : par exemple, la marque de leur véhicule, les titres de magazines qu'ils lisent habituellement, le temps et les modes de transport domicile-travail, autant d'informations que l'on peut rapprocher des coordonnées sémiométriques des individus sur les axes principaux<sup>1</sup>.

---

1. Ceci ouvre la voie aux applications marketing de la sémiométrie qui justifient le rapprochement opéré en début de chapitre avec les enquêtes styles de vie.

Les résultats des analyses ont mis en évidence des traits structuraux, sous la forme d'associations privilégiées de mots. L'existence de ces traits structuraux, c'est-à-dire en fait leur stabilité, leur invariance, leur éventuelle généralisation (ou inférence statistique) à l'ensemble des populations concernées, sera étudiée de façon détaillée sous trois angles : dans le temps, par pays, et à partir de procédures statistiques de validation dites de *rééchantillonnage*. Ce travail de validation des structures observées sera la matière du chapitre 2.

La richesse et la finesse des structures observées en termes sémantiques tout au long des sections précédentes posent le problème des relations entre sémantique et sémiométrie. La démarche est difficile car la sémantique n'est pas un champ disciplinaire balisé, plutôt un thème de recherche écartelé par son interdisciplinarité même. Le chapitre 3 apportera à ce chantier quelques matériaux, fondés sur de nouvelles expériences.

Les résultats obtenus en soumettant une liste de mots à quelques milliers de personnes étant éprouvés, consolidés et mieux compris, il reste à étudier leur degré de dépendance vis-à-vis de cette liste. Le caractère opératoire de la liste est clairement établi, mais qu'en est-il de son éventuelle universalité ? Le chapitre 4, fondé sur un nouveau recueil empirique, aborde ce questionnement ambitieux.

Le lecteur intrigué tout au long du présent chapitre par le traitement particulier réservé au premier axe trouvera les explications requises au chapitre 5 qui est une sorte de « zoom technique » sur les spécificités de cet axe, avec le renfort de données internationales et des retombées qui concernent en fait la méthodologie de toutes les enquêtes par sondage.

Enfin, le chapitre 6 approfondira les interprétations esquissées précédemment, en toute conscience des risques que représente un tel travail, alors que le chapitre 7 mettra le maçon au pied du mur en montrant un petit échantillon d'applications de la méthodologie.

## CHAPITRE 2

# Stabilité de la structure sémiométrique

Nous allons montrer dans ce chapitre que la structure mise en évidence au chapitre premier est stable, c'est-à-dire n'est ni un accident statistique, ni un artefact des méthodes utilisées. Cette structure est un ensemble de six axes principaux bien identifiables, auxquels nous avons donné des noms conventionnels, mais que tout un chacun peut reconnaître, indépendamment de cette dénomination, par les mots qui leurs sont associés.

Dans un premier temps, la stabilité sera éprouvée empiriquement sur différents échantillons recueillis à des moments différents pour un même pays (la France); il s'agit donc de *stabilité dans le temps*. Puis des échantillons relatifs à des pays différents seront analysés; on parlera alors de *stabilité dans l'espace*. Enfin, les axes sémiométriques seront recalculés dans des sous-populations particulières à l'intérieur d'un même pays, étudiant séparément les hommes et les femmes, puis différentes classes d'âge. On parlera alors de *stabilité interne*. Quelques exceptions et singularités seront néanmoins observées. Elles seront limitées, mais intéressantes à interpréter, comme peuvent l'être de petites déviations par rapport à un modèle ou à une norme. On constatera ainsi quelques différences subtiles entre Europe du nord et du sud. Les axes sont stables, mais leur ordre d'apparition peut varier. On constatera aussi, au chapitre 5 cette fois, la lente dérive de la population française au sein de cette structure au cours des douze dernières années.

Dans un second temps, cette stabilité sera testée de façon plus technique, utilisant les outils de la validation statistique et de la simulation. Les

échantillons seront perturbés par plusieurs variantes d'une méthode de *rééchantillonnage* connue sous le nom de *bootstrap*<sup>1</sup> pour mettre à l'épreuve la structure sémiométrique. Celle-ci sera, disons-le d'emblée, largement insensible à ces perturbations.

Evidemment, ce chapitre va se livrer à un exercice peu captivant pour un lecteur en quête d'aventures inédites... puisqu'il s'agit en somme de montrer que l'on trouve toujours la même chose, dans le temps, dans l'espace, voire à l'intérieur même de certaines sous-catégories d'individus. Le lecteur qui nous ferait confiance sur la surprenante stabilité de la structure sémiométrique peut donc passer au chapitre suivant, et simplement revenir consulter celui-ci si un doute le saisit lors de ses cheminements à venir dans la richesse et la complexité de l'espace des mots.

## 2.1 Stabilité empirique par sous-population

### 2.1.1 Stabilité dans le temps

Le questionnaire sémiométrique a été passé en France durant six années réparties sur une période de dix ans (1990, 1996, 1998, 1999, 2001 et 2002) en vue d'applications marketing<sup>2</sup> et pour suivre certains aspects de l'évolution des valeurs des Français. La comparaison de ces vagues successives d'enquêtes obtenues à partir d'échantillons indépendants met en évidence une grande stabilité des structures (*cf.* tableau 2.1).

Mais, si les structures sémiométriques ne changent pas en elles-mêmes, on observe cependant de légères différences au niveau du positionnement de certaines catégories. Des modifications de faible amplitude peuvent être relevées selon le sexe, l'âge mais surtout l'activité. Ainsi, les exploitants-agricoles expriment de plus en plus depuis ces dix dernières années des valeurs proches du pôle « Détachement » (un des pôles de l'axe 3 : « Attachement / Détachement ») et semblent moins caractérisés par le pôle « Devoir » de l'axe 2 : « Devoir / Plaisir ». Les artisans-commerçants et les ouvriers se rapprochent en 1998 du pôle « Plaisir » de ce même axe, caractérisé avant cette date surtout par les cadres et les professions intermédiaires. Ces déplacements de catégories dans l'espace sémiométrique, faibles mais significatifs statistiquement, demandent pour être interprétés un travail pluridisciplinaire impliquant aussi bien des statisticiens que des sociologues et des économistes. Un tel travail sort du

---

1. Le principe de la méthode « *bootstrap* » est esquissé en annexe A1.9.5.

2. Des exemples d'application font l'objet du chapitre 7.

cadre de la présentation plus spécifiquement méthodologique de cet ouvrage.

Nous verrons au chapitre 5 que la manière de noter rend compte d'une autre évolution significative qu'il n'est pas possible de déceler ici sur la simple comparaison d'analyses effectuées indépendamment sur chaque échantillon. Nous renvoyons pour une analyse de l'évolution des valeurs des Français au paragraphe 7.7 du chapitre 7 consacré aux comparaisons diachroniques entre des échantillons d'une même population interrogés à des moments différents.

Bien que l'on observe des mouvements significatifs de catégories dans les plans sémiométriques, les positions des mots dans ces plans restent stables, comme l'attestent les tableaux 2.1 à 2.6 qui décrivent systématiquement les évolutions des axes 1 à 6 pour les quatre années 2002, 1999, 1996, 1990. Dans cette série de tableaux, chaque axe, pour chaque année, est décrit par ses 20 mots les plus extrêmes, c'est-à-dire par les dix mots qui lui sont le plus corrélés positivement et les dix mots qui lui sont le plus corrélés négativement.

### **2.1.2 Stabilité par pays**

Le questionnaire sémiométrique a été soumis à des échantillons appartenant à plusieurs pays européens<sup>1</sup>, toujours en vue d'applications de type marketing. Une des inconnues, dans cette opération d'envergure, était l'effet de la traduction du questionnaire sur les résultats. Le tableau A2.1 présenté en annexe A2 donne les correspondances entre mots dans les cinq langues concernées. Il montre à la fois l'étendue et les limites des difficultés rencontrées. Beaucoup de mots ne posent pas de gros problèmes et d'autres conduisent à faire des choix parfois difficiles<sup>2</sup>.

---

1. France, Grande Bretagne, Allemagne, Italie, Espagne, Finlande, Norvège, ....

2. Ainsi, *Dieu*, *L'eau*, *Un arbre*, *Métallique*, *Un nid*, ne posent pas de gros problèmes, en première approximation, et pour les pays concernés. En revanche, *Argent*, *Réfléchir*, *un Prêtre*, sont des mots dont la traduction n'est pas immédiate (*Silver* ou *Money* pour *Argent* en Anglais ?) et nécessitent un choix raisonné, voire une convention arbitraire. Mais comme on le verra plus loin, les résultats seront également stables par *perturbation* de la liste de mots, et la traduction peut être considérée comme une opération entraînant une perturbation dans des limites tolérables.

**Tableau 2.1 :**  
**Mots caractérisant les « axes 1 » de 1990 à 2002**  
 « Axe de participation »

axe 1 : 2002	axe 1 : 1999	axe 1 : 1996	axe 1 : 1990
courage	courage	courage	protéger
dynamique	efficace	protéger	pureté
douceur	dynamique	efficace	courage
efficace	précision	dynamique	honnête
gaieté	récompense	robuste	récompense
respect	élégance	honnête	efficace
récompense	robuste	politesse	admirer
honnête	charitable	pureté	dynamique
confiance	élégance	élégance	respect
élégance	protéger	récompense	politesse
zone centrale			
faute	noir	orage	doute
rompre	mort	doute	mort
attaquer	danger	rompre	vide
mort	rompre	guerre	faute
danger	faute	vide	danger
vide	vide	faute	guerre
angoisse	angoisse	danger	angoisse
guerre	désordre	angoisse	rompre
trahir	guerre	désordre	désordre
désordre	trahir	trahir	trahir

**Tableau 2.2 :**  
**Mots caractérisant les « axes 2 » de 1990 à 2002**  
 « Devoir / Plaisir »

axe 2 : 2002	axe 2 : 1999	axe 2 : 1996	axe 2 : 1990
sensuel	sensuel	sensuel	sensuel
rêver	rêver	rêver	aventurier
voluptueux	légèreté	aventurier	rêver
île	sauvage	nudité	sauvage
légèreté	aventurier	original	original
moelleux	île	voluptueux	nudité
original	original	désir	bohème
océan	voluptueux	bohème	légèreté
sauvage	nudité	île	voluptueux
nudité	charnel	émotion	orage
zone centrale			
règle	économiser	interdire	prêtre
rigide	règle	punir	loi
prêtre	loi	règle	honneur
élite	industrie	industrie	interdire
soldat	prêtre	économiser	économiser
industrie	soldat	morale	soldat
patrie	obeir	soldat	patrie
obeir	patrie	patrie	discipline
morale	morale	discipline	morale
discipline	discipline	obeir	obeir

Exemple de lecture du tableau 2.1 : Pour l'axe 1, en 2002, le mot le plus corrélé positivement à l'axe est le mot courage, suivi par dynamique, douceur, efficace, etc. Le mot le plus corrélé négativement à ce même axe est le mot désordre, suivi de trahir, guerre, angoisse, etc.

**Tableau 2.3 :**  
Mots caractérisant les « axes 3 » de 1990 à 2002  
« Attachement / Détachement »

axe 3 : 2002	axe 3 : 1999	axe 3 : 1996	axe 3 : 1990
danger orage révolte critiquer mort attaquer cri feu détachement vide	danger mort punir vide rompre attaquer angoisse labyrinthe critiquer révolte	danger orage mort désert labyrinthe révolte rompre critiquer cri angoisse	danger doute angoisse rompre cri vide labyrinthe mort orage méfiance
zone centrale			
fidélité richesse mariage récompense cadeau maison argent confort famille bijou	fleur or gaîté maison récompense famille mariage confort cadeau bijou	mariage gratuit hériter récompense or argent richesse cadeau bijou confort	argent richesse récompense caresse gaîté maison or confort cadeau bijou

**Tableau 2.4 :**  
Mots caractérisant les « axes 4 » de 1990 à 2002  
« Esprit / Matière »

axe 4 : 2002	axe 4 : 1999	axe 4 : 1996	axe 4 : 1990
richesse puissance or gloire vitesse argent ambition conquérir ruse séduire	richesse puissance or argent gloire conquérir vitesse commander hériter ambition	puissance argent richesse or gloire ruse vitesse ambition conquérir fusil	richesse puissance argent or gloire conquérir vitesse propriété ambition hériter
zone centrale			
paix écrire enseigner réfléchir école arbre théâtre poésie art livre	pardon enseigner écrire arbre étranger école poésie théâtre art livre	école enseigner art pardon charitable poésie recueillement nid arbre livre	école recueillement pardon âme fleur charitable art arbre poésie livre

**Tableau 2.5 :**  
**Mots caractérisant les « axes 5 » de 1990 à 2002**  
 « Cœur / Raison »

axe 5 : 2002	axe 5 : 1999	axe 5 : 1996	axe 5 : 1990
dieu foi prêtre sacré âme recueillement créateur éternel infini souverain	dieu foi âme recueillement sacré prêtre créateur éternel noble souverain	dieu foi âme prêtre sacré éternel cérémonie recueillement souverain poésie	dieu foi prêtre âme éternel sacre recueillement bijou cérémonie noble
zone centrale			
confort réfléchir matériel maîtriser volontaire utilitaire concret logique précision efficace	volontaire produire maîtriser utilitaire concret astucieux construire dynamique précision efficace	construire utilitaire astucieux produire réfléchir maîtriser efficace précision concret logique	industrie volontaire matériel robuste maîtriser efficace précision produire concret logique

**Tableau 2.6 :**  
**Mots caractérisant les « axes 6 » de 1990 à 2002**  
 « Humilité / Souveraineté »

axe 6 : 2002	axe 6 : 1999	axe 6 : 1996	axe 6 : 1990
créateur art audace robuste inventeur efficace bâtisseur précision voluptueux noble	bâtisseur créateur maîtriser art robuste concret inventeur audace noble efficace	noble créateur art élite dieu audace souverain foi âme robuste	noble élite commander audace dieu créateur art absolu bâtisseur maîtriser
zone centrale			
obéir féconder rigide faute vide mariage méfiance angoisse famille naissance	cri interdire obéir mariage famille tendresse méfiance doute faute naissance	enfance rigide obéir doute mariage interdire famille faute méfiance naissance	enfance famille tendresse angoisse obéir rigide faute doute méfiance naissance

Les résultats présentés concernent six de ces pays : la France, la Grande Bretagne, l'Allemagne, l'Italie, l'Espagne, le Canada (Québec). La stabilité des structures peut être vérifiée à la lecture des figures 2.1 à 2.5.

On relève cependant certaines singularités qui, sans remettre en cause la stabilité d'ensemble, la nuancent, mettant en évidence des particularités culturelles sur lesquelles il est intéressant de s'arrêter.

La première de ces singularités concerne l'interversion des axes 2 et 3, « Devoir / Plaisir » et « Attachement / Détachement », entre les pays du sud de l'Europe et ceux du nord : pour les Espagnols, les Français et les Italiens, l'axe « Devoir / Plaisir » apparaît avant l'axe « Attachement / Détachement ». A l'inverse, pour les Allemands, les Anglais et les Canadiens, c'est la dimension « Attachement / Détachement » qui apparaît en seconde position, et qui précède donc l'axe « Devoir / Plaisir » occupant maintenant la troisième position<sup>1</sup>. Ce ne sont pas les axes qui sont en cause, mais leur hiérarchie<sup>2</sup>.

Le chapitre 5 montrera que cette interversion d'axes peut être imputable, du moins en partie, aux effets de notation qui caractérisent les premier et troisième axes et qui traduisent des différences culturelles d'attitude par rapport à l'enquête et/ou son questionnaire ainsi que des différences de comportement vis-à-vis d'une échelle de notes.

Cette singularité se retrouve également dans les analyses que l'on peut faire à l'intérieur de certaines classes d'âges. Ce point sera repris lorsque seront présentées ces analyses (section 2.1.4).

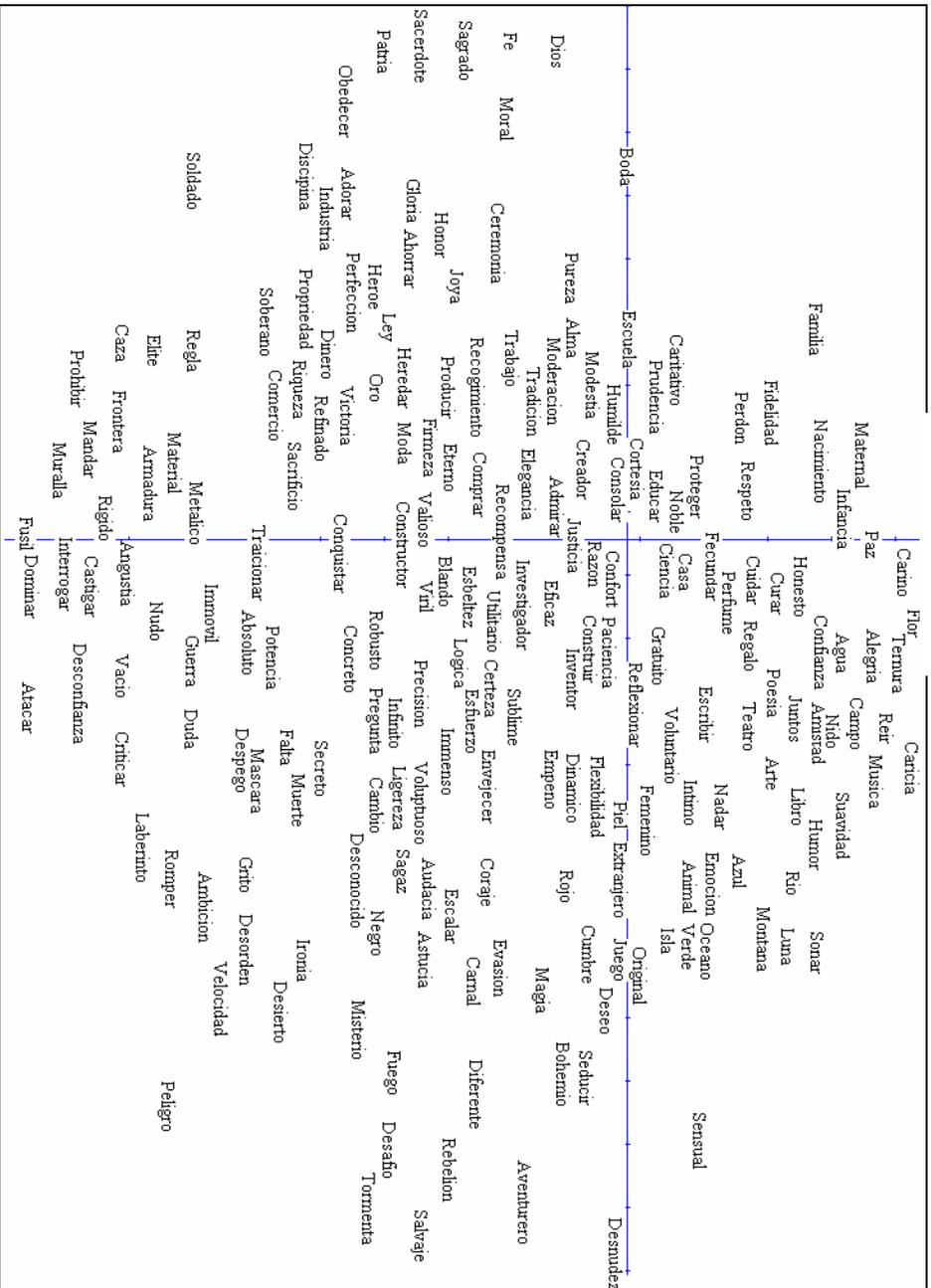
La deuxième singularité concerne l'axe 4 appelé conventionnellement « Esprit / Matière ». Si, du côté du pôle « Matière », on retrouve, pour les cinq pays considérés, un grand nombre de mots communs tels que *Or*, *Argent*, *Bijou* (sauf Canada), *Richesse*, *Conquérir*, *Victoire*, *Séduire*, etc., qui expriment très clairement la recherche d'une jouissance à travers la conquête et la possession de bien matériels, les mots qui s'y opposent à l'autre extrémité de l'axe diffèrent assez nettement selon les pays.

---

1. Ce résultat est confirmé par les axes de la Norvège, de la Finlande et de la République Tchèque, dont il n'est pas rendu compte ici.

2. On peut donc dire que, pour décrire l'ensemble des corrélations entre mots, l'opposition « Devoir / Plaisir » est plus importante que l'opposition « Attachement / Détachement » dans le sud de l'Europe que dans le nord de l'Europe.

ATTACHEMENT



DÉTACHEMENT

Figure 2.1 : Plan sémiotique (2,3) - Espagne

ATTACHEMENT

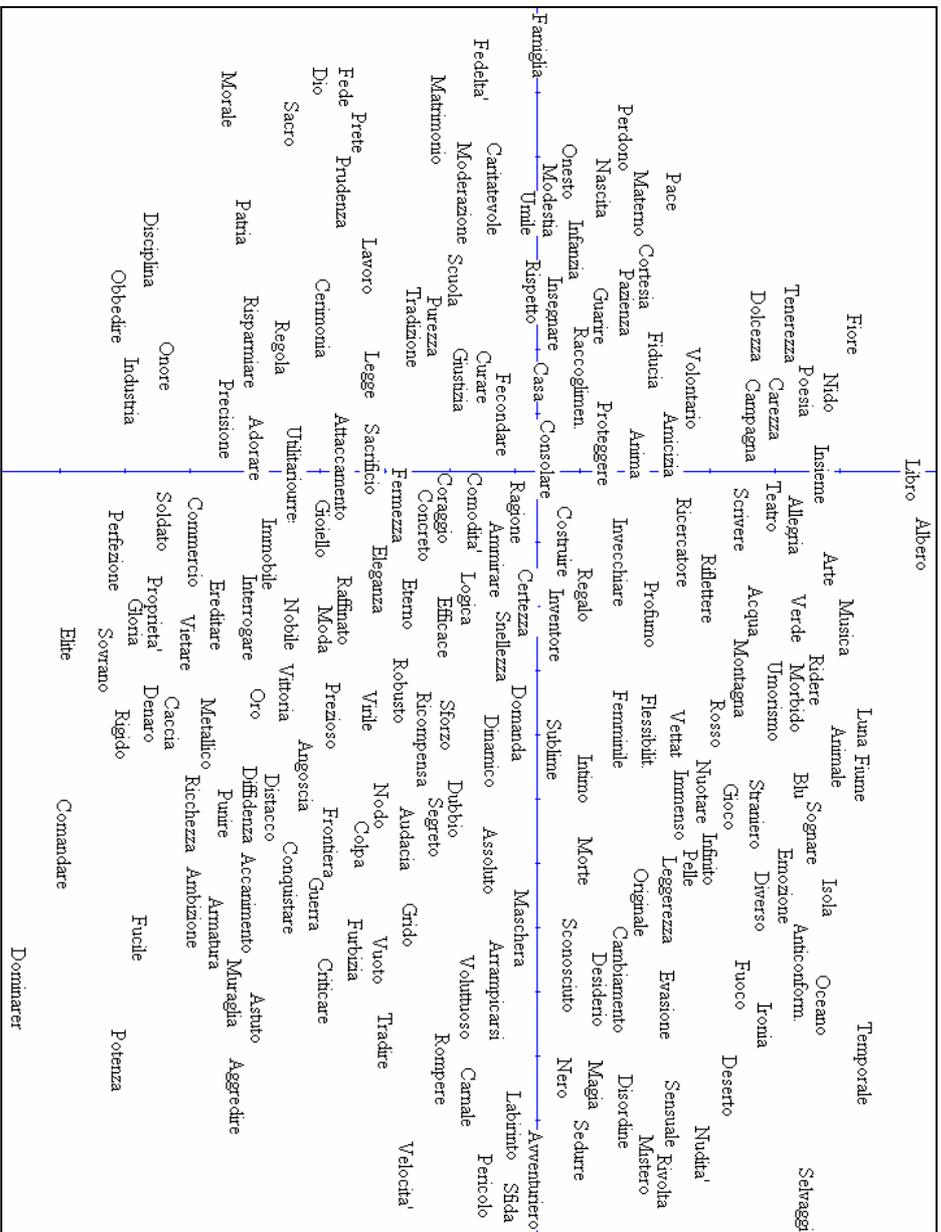


Figure 2.2 : Plan s miotique (2,3) - Italie







On peut diviser ceux-ci en trois groupes : la France où la notion d'« Esprit » ou encore de « Sublimation » est avant tout – mais non exclusivement – exprimée par l'ensemble des mots du questionnaire évoquant les activités artistiques (*Livre, Art, Théâtre et Poésie*), le Canada mais surtout la Grande-Bretagne où cette notion est principalement rendue par des mots liés à la nature (*Arbre, Montagne, Fleuve, Océan, Lune et Vert*), et l'Espagne et l'Italie où apparaissent des mots comme *Mort, Sacrifice, Humble, Faute* (sauf Espagne), *Doute*, qui sont complètement absents dans les autres pays ; l'Allemagne présentant une synthèse de ces trois groupes<sup>1</sup>.

### 2.1.3 Stabilité selon le sexe

Nous avons vu au cours du chapitre précédent que les axes 3 (intitulé conventionnel : « Attachement / Détachement ») et 5 (intitulé conventionnel : « Cœur / Raison »), obtenus à partir d'analyses de la population globale, opposaient des populations féminines et masculines. Si la variable « sexe » est vraiment la variable explicative cachée responsable des grandes oppositions observées, les analyses réalisées uniquement auprès d'un échantillon de femmes ou uniquement auprès d'un échantillon d'hommes ne doivent plus faire apparaître d'axes de ce type.

Or, l'ensemble des structures sémiométriques restent, à quelques nuances près, conservées lors d'analyses ne portant que sur des hommes ou sur des femmes. Il y a là une stabilité et une permanence assez inattendues. La structure est la même pour les hommes et pour les femmes, mais certains des traits de cette structure peuvent séparer de façon significative les hommes des femmes<sup>2</sup>.

---

1. On pourrait, à propos de tels résultats, faire l'hypothèse que des particularités culturelles peuvent être à l'origine des différences observées, les Français opposant l'art à l'argent, les Anglais, l'or de la City à la campagne anglaise, et les habitants des pays du sud opposant alors l'or à la mort. Mais, conscients de nos limites dans un domaine aussi subtil, nous laisserons ce soin à nos lecteurs et aux spécialistes concernés.

2. Ce phénomène d'apparition d'une même structure à des échelles différentes évoque irrésistiblement les structures dites *fractales*. Ce phénomène, tout comme l'analogie qui y est attachée, fut déjà remarqué et mentionné par Jean-Pierre Pagès, dans le cadre des enquêtes Agoramétrie. L'analogie avec le monde fractal est d'autant plus prégnante qu'elle se prolonge, de subdivision en subdivision de l'échantillon, jusqu'à l'individu, qui porterait en lui une structure du même type. Nous reviendrons à diverses reprises sur ce thème, et ce jusqu'au moment où il s'agira pour nous de conclure et d'évoquer problèmes ouverts et recherches futures.

Tableau 2.7 : Confrontation, axe par axe, des analyses réalisées séparément sur les hommes et sur les femmes

axe 2 : « Devoir / Plaisir »		axe 3 : « Attachement / Détachement »		axe 4 : « Esprit / Matière »		axe 5 : « Cœur / Raison »		axe 6 : « Humilité / Souveraineté »	
homme	femme	homme	femme	homme	femme	homme	femme	homme	femme
sauvage	sensuel	punir	danger	puissance	puissance	dieu	dieu	naissance	mariage
aventurier	rêver	danger	rompre	richesse	richesse	foi	foi	méfiante	naissance
sensuel	voluptueux	mort	mort	or	or	âme	âme	faute	famille
rêver	nudité	rompre	cri	argent	argent	prêtre	sacre	doute	cérémonie
orage	aventurier	angoisse	orage	gloire	gloire	sacre	prêtre	rigide	adorer
nudité	original	labyrinthe	désert	conquérir	conquérir	recueillem.	recueillem.	caresse	doute
original	légèreté	doute	critiquer	vitesse	ruse	cérémonie	noble	masque	enfance
île	bohème	cri	révolte	fusil	ambition	éternel	créateur	tendresse	éternel
révolte	désir	vide	punir	commander	séduire	bijou	souverain	angoisse	dieu
océan	île	critiquer	vide	attaquer	bijou	adorer	absolu	cri	foi
-----									
loi	sacrifice	caresse	acheter	charitable	pardon	utilitaire	confiance	volupté	inventeur
dieu	règle	richesse	hérier	arbre	charitable	réfléchir	ruse	art	bâtisseur
foi	prêtre	argent	maison	pardon	paix	produire	gaieté	prêtre	produire
prêtre	économiser	gratuit	récompense	écrire	poésie	robuste	volontaire	recueille	astucieux
économiser	industrie	or	argent	enseigner	théâtre	astucieux	honnête	âme	précision
soldat	soldat	gaieté	richesse	théâtre	nid	logique	punir	noble	audace
morale	patrie	bijou	cadeau	école	recueillem	concret	politesse	sacre	maîtriser
obéir	morale	récompense	or	art	art	maîtriser	méfiante	foi	robuste
discipline	obéir	confort	confort	poésie	livre	précision	effort	dieu	concret
patrie	discipline	cadeau	bijou	livre	arbre	efficace	puissance	créateur	efficace

Tableau 2.8 : Confrontation, axe par axe, des analyses réalisées séparément sur deux classes d'âges

axe 2 : « Devoir/ Plaisir »	axe 3 : « Attach. Détach. »	axe 3 : « Attach. Détach. »	axe 2 : « Devoir/ Plaisir »	axe 4 : « Esprit / Matière »		axe 5 : « Cœur / Raison »		axe 6 : « Humilité / Souveraineté »	
				- 45 ans	+ 45 ans	- 45 ans	+ 45 ans	- 45 ans	+ 45 ans
orage	révolte	fleur	cadeau	richesse	richesse	dieu	dieu	faute	naissance
sauvage	orage	gaieté	parfum	or	puissance	foi	foi	naissance	doute
étranger	sauvage	douceur	moelleux	vitesse	or	âme	âme	rigide	masque
lune	bohème	tendresse	voluptueux	puissance	argent	prêtre	prêtre	doute	caresse
révolte	désert	paix	bijou	séduire	conquérir	saacre	saacre	méfiante	tendresse
feu	nudité	cadeau	rire	argent	fusil	recueillem.	recueillem.	caresse	cérémonie
bohème	original	maternel	rêver	sensuel	gloire	cérémonie	éternel	noeud	faute
mystère	feu	caresse	gaieté	désir	commander	éternel	absolu	interdire	mariage
désert	aventurier	parfum	sensuel	conquérir	hériter	bijou	infini	tendresse	immobile
différent	ironie	rêver	caresse	gloire	vitesse	souverain	noble	masque	méfiante
économiser	discipline	méfiante	rigide	travail	modestie	construire	sensuel	audace	sublime
hériter	soldat	vide	danger	vieillir	amitié	réfléchir	dynamique	art	astucieux
obéir	mariage	armure	interdire	enseigner	enseigner	robuste	utilitaire	noble	noble
mariage	honnneur	rompre	cri	dieu	livre	astucieux	astucieux	âme	robuste
patrie	richesse	muraille	discipline	obéir	paix	utilitaire	produire	sacre	élite
propriété	économiser	angoisse	rompre	prêtre	consoler	logique	matériel	prêtre	audace
gloire	hériter	fusil	angoisse	morale	charitable	maîtriser	volontaire	recueillem.	créateur
richesse	patrie	danger	sacrifice	foi	école	précision	efficace	créateur	efficace
discipline	propriété	punir	obéir	recueillem.	confiance	concret	logique	foi	concret
argent	argent	attaquer	punir	école	pardon	efficace	précision	dieu	maîtriser

Les nuances relevées proviennent essentiellement des classements des mots dans l'explication des axes concernés, classements qui peuvent varier légèrement sans perturber leur interprétation initiale. Le tableau 2.7 présente, pour chaque axe<sup>1</sup>, les dix premiers mots qui leur sont le plus fortement corrélés.

Les trois premiers axes de la structure sémiométrique (axes 2-3-4) sont stables. Les valeurs révélées par ces axes sont similaires chez les hommes et chez les femmes et ce malgré des différences notables observées dans l'analyse commune, notamment pour le troisième facteur. Aussi la notion d'« Attachement » de l'axe 3 est une valeur plus féminine qui existe cependant chez les hommes et celle de « Détachement » est une valeur plus masculine que l'on retrouve aussi chez les femmes. On note pour le pôle « Attachement », chez les femmes, la présence des mots *Acheter*, *Hériter*, *Fidélité* et l'absence de *Gaieté*, *Caresse*, *Rire*, *Douceur* dans la liste des dix premiers mots. Il existe donc des nuances dans la caractérisation de l'axe, nuances qui s'estompent ou évoluent si l'on retient plus de mots pour caractériser cette extrémité de l'axe.

C'est sans doute pour les axes 5 et 6 que se produisent quelques irrégularités dans la mesure où les quatre concepts émergeant de ces axes dans l'analyse globale ne correspondent pas exactement aux axes de même rang dans l'analyse des hommes et dans celle des femmes. Mais, hormis ce décalage sur les deux derniers axes qui s'estomperait sans doute avec un échantillon plus grand, la structure sémiométrique observée sur ces deux sous-populations est stable, stabilité que l'on retrouve d'ailleurs dans l'analyse suivante, celle par âge.

#### 2.1.4 Stabilité selon l'âge

Nous venons de voir que les axes 3 « Attachement / Détachement » et 5 « Raison / Cœur » étaient liés, dans l'analyse globale, au sexe des répondants, mais qu'ils subsistaient lorsqu'on élimine l'effet *sexe*, c'est-à-dire en travaillant séparément sur deux sous-populations de même sexe. Or l'axe 2 (« Devoir / Plaisir ») et, dans une moindre mesure, l'axe 4 (« Esprit / Matière ») sont, eux, liés à la variable « âge ». Que se passe-t-il si l'on scinde la population en deux catégories d'âge : les « moins de 45 ans » et les « plus de 45 ans », et si l'on procède aux analyses à l'intérieur de chacune de ces deux sous-populations ?

---

1. Le premier facteur dont le statut est particulier n'est pas pris en compte ici. Toutefois, il se révèle être également stable.

Là encore, on observe une permanence des structures à l'exception d'une simple interversion des axes 2 et 3 pour les personnes de plus de 45 ans, mettant alors en second plan le « Plaisir / Devoir » par rapport à l'« Attachement / Détachement » (cf. tableau 2.8)<sup>1</sup>.

Ainsi, la notion de « Plaisir », chez les « moins de 45 ans » se rapproche, du point de vue des mots qui la caractérise, de la notion d'« Esprit » (*Lune, Mystère, Art, Inconnu, Fleuve, Océan*). De la même façon, la notion de « Devoir » pour cette même sous-population, fait plus référence à la notion de discipline et à une certaine réussite matérielle, proche du pôle « Matière » de l'axe 4.

L'« Attachement » de l'axe 3 est moins matériel chez les « moins de 45 ans » (absence, parmi les 15 premiers mots, des mots : *Or, Récompense, Argent*) et plus lié aux mots du « Plaisir » chez les « plus de 45 ans » (*Voluptueux, Sensuel, Séduire*). Le « Détachement » est plus agressif et guerrier pour les plus jeunes (*Attaquer, Fusil, Armure, Guerre*) et plus oblatif pour les plus âgés (*Obéir, Sacrifice, Interdire, Effort*).

La notion d'« Esprit » de l'axe 4 est moins bien définie pour les moins de 45 ans dont les valeurs se retrouvent en partie dans le « Plaisir » de l'axe 2.

Précisons que pour ces études de stabilité selon l'âge, comme pour celles relatives au sexe dans la section précédente, la division en deux de l'échantillon déséquilibre fortement celui-ci, de façon évidente pour les critères incriminés, mais aussi pour d'autres critères (profession, niveau d'éducation, région...) qui n'étaient pas forcément équilibrés à l'intérieur de chaque tranche d'âge ou de chaque sexe. Les sous-échantillons n'étant plus représentatifs, la permanence globale des structures n'en est que plus surprenante, au contraire des petites différences observées qui paraissent finalement modestes.

## 2.2 Intervalles de confiance des valeurs propres

Dans cette section, plus technique, ainsi que dans les sections suivantes, la stabilité et la cohérence de la structure sémiométrique sera éprouvée en utilisant les outils de la validation statistique et de la simulation. On est naturellement conduit à poser un certain nombre de questions sur la qualité des représentations : Observe-t-on vraiment quelque chose ? Les données ont-elle une structure ? Ou, au contraire, de simples fluctuations

---

1. On ne s'étonnera pas, compte tenu de l'intervention d'axes signalée, de voir sur le tableau 2.8 l'axe 2 des « moins de 45 ans » comparé à l'axe 3 des « plus de 45 ans » (deux premières colonnes) et réciproquement (deux colonnes suivantes).

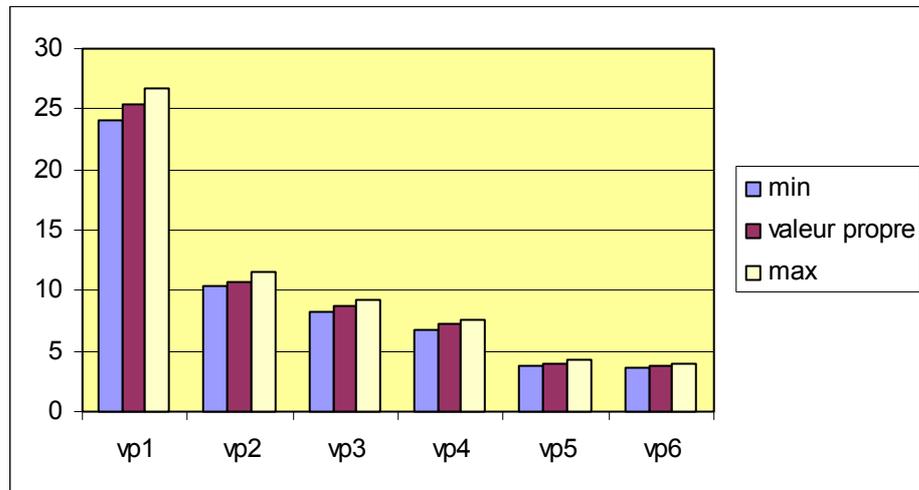
d'échantillonnage suffiraient-elles à expliquer les pourcentages de variance obtenus (pourcentages qui, rappelons-le, mesurent l'importance des axes) ?

Cette question sera tout d'abord traitée dans un cadre statistique classique : il existe en effet une procédure permettant d'évaluer la confiance que l'on peut accorder aux premières *valeurs propres*, c'est-à-dire finalement aux premiers axes. L'ampleur de l'intervalle de confiance qui va être calculé donne une indication sur la stabilité de la valeur propre vis-à-vis des fluctuations dues à l'échantillonnage. L'intervalle de confiance d'Anderson (*cf.* annexe A1.9.4) pour chaque valeur propre décrit en quelque sorte l'incertitude sur celle-ci. Le tableau 2.9, puis la figure 2.6 dans le cas des deux premières valeurs propres, présentent les intervalles de confiances pour les 6 premières valeurs propres dans le cas de trois populations de tailles différentes.

**Tableau 2.9 : Intervalles de confiance des 6 premières valeurs propres (cas de trois tailles d'échantillon différentes)**

		borne inférieure	valeur propre	borne supérieure
Echantillon 2 000	vp1	24.00	25.35	26.77
	vp2	10.40	10.98	11.60
	vp3	8.24	8.70	9.19
	vp4	6.80	7.18	7.58
	vp5	3.80	4.01	4.23
	vp6	3.59	3.79	4.00
Echantillon 10 000	vp1	25.49	26.19	26.91
	vp2	10.07	10.35	10.63
	vp3	8.58	8.82	9.06
	vp4	6.82	7.01	7.20
	vp5	4.04	4.15	4.26
	vp6	3.58	3.68	3.78
Echantillon 15 000	vp1	26.40	26.99	27.59
	vp2	10.13	10.36	10.59
	vp3	8.64	8.84	9.03
	vp4	6.78	6.93	7.09
	vp5	4.07	4.16	4.26
	vp6	3.61	3.69	3.77

Les calculs portent sur trois échantillons sémiométriques de 2 000, 10 000 et 15 000 individus. Les résultats sont très satisfaisants, en ce sens que, pour une taille d'échantillon donnée, les intervalles de confiance des premières valeurs propres n'empiètent pas, ce qui traduit bien que celles-ci sont significativement distinctes.



**Figure 2.6 : Variations des intervalles de confiance des six premières valeurs propres (vp1,...vp6) pour une taille de l'échantillon de 2000.**

On lit, par exemple, sur les deux premières lignes du tableau 2.9, comme sur la figure 2.6, que la borne supérieure de l'intervalle de la seconde valeur propre (11.60) est plus petit que la borne inférieure de celui de la première (24.00).

La figure 2.6 montre aussi, dans le cas le plus défavorable (échantillon de taille 2000), que les intervalles de confiance restent petits par rapport aux valeurs propres. On note cependant que pour cette taille d'échantillon, les valeurs propres 5 et 6 sont voisines et que leurs intervalles de confiance empiètent.

Tout ceci prouve que les premiers axes principaux occupent des directions significatives, et qu'ils sont bien individualisés.

Ces méthodes s'appliquent en fait à des observations distribuées selon une loi normale, mais leur robustesse leur permet d'être utilisées en dehors de ce cadre. Pour être plus rigoureux et s'affranchir de ces hypothèses, nous allons utiliser des validations par la méthode de *bootstrap non paramétrique*.

Nous verrons donc au paragraphe suivant comment les méthodes de validation par rééchantillonnage s'appliquent à l'étude de la stabilité des formes sans hypothèse au préalable sur la distribution des données.

## 2.3 Stabilité par rééchantillonnage

La permanence des structures observée sur des échantillons indépendants devrait nous satisfaire et nous convaincre que, lorsqu'un ensemble de mots a été choisi, une série d'axes ou de dimensions apparaissent assez systématiquement, pas toujours exactement dans le même ordre, mais avec des caractérisations par les mots très voisines.

On ne sait que peu de choses, cependant, sur la stabilité des positions des mots les uns par rapport aux autres. Leur classement sur les axes a bien montré qu'ils devaient être proches d'une population à une autre. Mais on peut se poser légitimement la question de savoir quelle est la précision de la position d'un mot sur les plans engendrés par deux axes principaux, compte tenu des fluctuations d'échantillonnage de nos recueils de réponses.

Les statisticiens ne pouvaient pas répondre à cette question de façon satisfaisante jusqu'à une période récente au cours de laquelle la puissance des nouveaux moyens de calcul a permis de systématiser les méthodes dites de *rééchantillonnage*.

Parmi ces méthodes, une méthode statistique de validation, désignée sous le nom de *bootstrap*<sup>1</sup> permet de vérifier à la fois la stabilité des axes (ce qui ne fera que confirmer les résultats des sections précédentes de ce chapitre), et d'évaluer la confiance que l'on peut accorder à la position de chaque point dans les plans principaux.

La méthode de *bootstrap* classique utilisée ici consiste à perturber l'échantillon initial en simulant un grand nombre d'échantillons de même taille<sup>2</sup>. Dans ce contexte, les simulations sont appelées *réplications*.

Deux variantes de cette méthode seront utilisées, le *bootstrap* partiel<sup>3</sup>, et le *bootstrap* total<sup>4</sup>.

Elles donneront des résultats convergents, attestant la stabilité des structures observées jusqu'au sixième axe.

---

1. Cette terminologie imagée due à Efron (1979) se réfère à l'expression « *to pull oneself up by one's own bootstraps* » qui signifie littéralement « s'élever en tirant ses propres lanières de bottes », ou encore « progresser sans demander d'aide » (cf. annexe A1.9.5).

2. Ces échantillons sont obtenus par  $n$  tirages au hasard *avec remise* des individus dans l'échantillon initial de  $n$  individus.

3. Le *bootstrap* partiel consiste à projeter les colonnes des tableaux répliqués (mots) comme des éléments supplémentaires sur les axes de l'analyse de référence, c'est-à-dire l'analyse de l'échantillon initial non perturbé. Avec 30 réplications, on obtient 30 positions (permettant de calculer des ellipses de densité) pour chaque mot (cf. annexe A1.9.5).

4. Le *bootstrap* total consiste à refaire des analyses en composantes principales complètes sur chaque échantillon répliqué. Comme dans le cas partiel, avec 30 réplications, on obtient 30 positions (permettant de calculer des ellipses de densité) pour chaque mot (cf. annexe A1.9.5).

### 2.3.1 Bootstrap partiel

Les figures 2.6 et 2.7 montrent quelques ellipses de confiance obtenues par la méthode de validation selon les principes du *bootstrap* partiel respectivement pour les plans (2, 3) et (5, 6). Le plan (2, 3) est le premier plan sémiométrique<sup>1</sup>, le plus utilisé dans les visualisations, et le plan (5, 6) est le dernier plan sémiométrique.

Il s'agit d'un échantillon « France », regroupant les quatre échantillons indépendants des années 1990, 1996, 1998 et 1999, et totalisant 11055 individus, représentatifs de la population française âgée de 18 ans et plus. Les ellipses sont de tailles très réduites, et donc les fluctuations de la position des points sont très faibles, malgré la sévérité du principe des réplifications.

En effectuant un tirage avec remise dans l'échantillon, on montre qu'environ 30% des individus sont exclus de chaque réplification ; d'autres individus en revanche apparaissent deux, trois fois, jusqu'à six fois dans certains cas exceptionnels.

On note que les ellipses relatives au plan (5, 6) sont légèrement plus grandes, tout en laissant à la position des points une précision acceptable.

### 2.3.2 Bootstrap total

Ce sont les figures 2.8 et 2.9 qui montrent les mêmes plans principaux que précédemment, avec maintenant des ellipses plus grandes<sup>2</sup>.

L'examen de la figure 2.9 montre que les zones de confiance des mots interfèrent peu. Il s'agit pourtant d'une épreuve sévère (le *bootstrap* total) et d'un plan sémiométrique *a priori* moins stable que le plan principal, le plan (5, 6) ; les effets des fluctuations d'échantillonnage sont donc faibles, confirmant ainsi la stabilité de ces facteurs et des structures obtenues.

---

1. Les axes 2 et 3 sont les deux axes les plus importants, puisque l'axe 1 ne fait pas partie de la structure sémiométrique proprement dite.

2. En fait, le *bootstrap* total donne une vue pessimiste de la précision des points, car les analyses en composantes principales sont refaites pour chaque réplification, et donc les plans de projection sont calculés sur des données perturbées : non seulement les données sont perturbées, mais l'outil de visualisation est lui-même perturbé (cf. annexe A1.9.5).



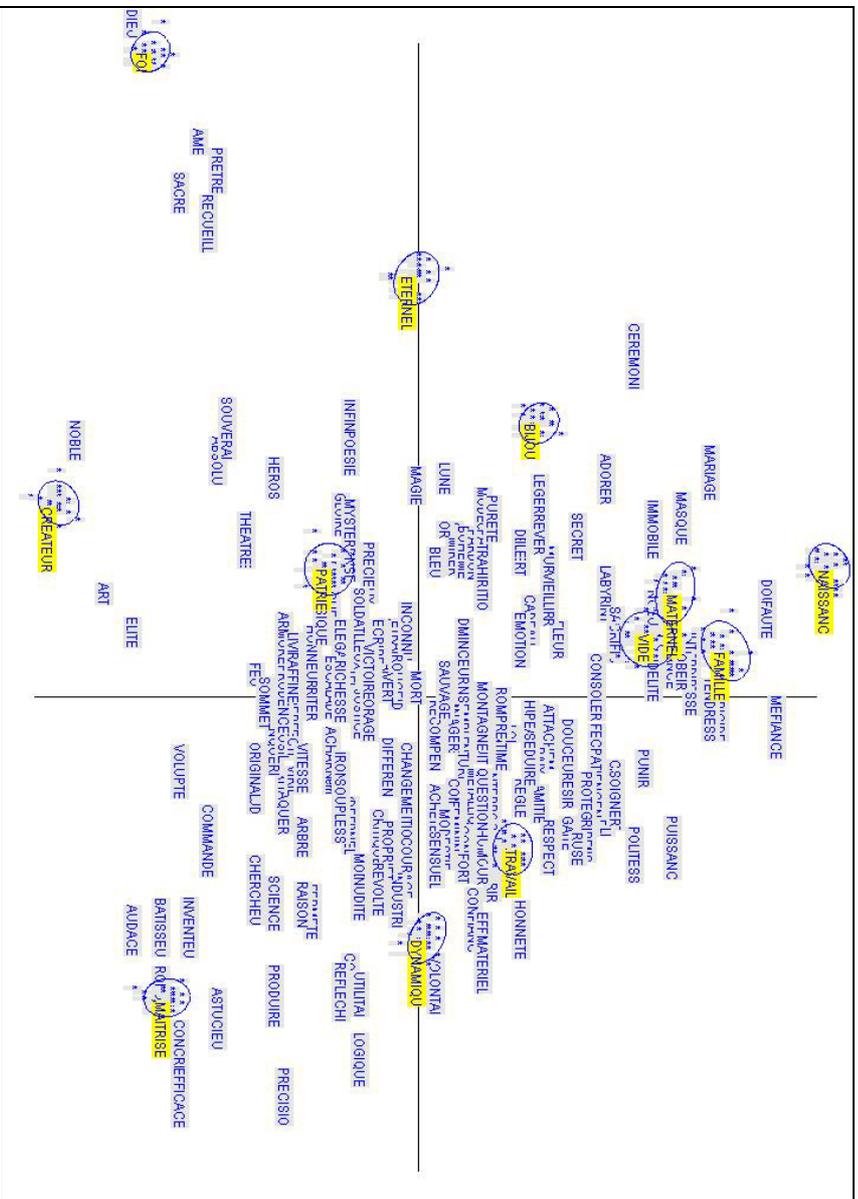


Figure 2.7 : Plan engendré par les axes 5 et 6 – Quelques zones de confiances (Bootstrap partiel)

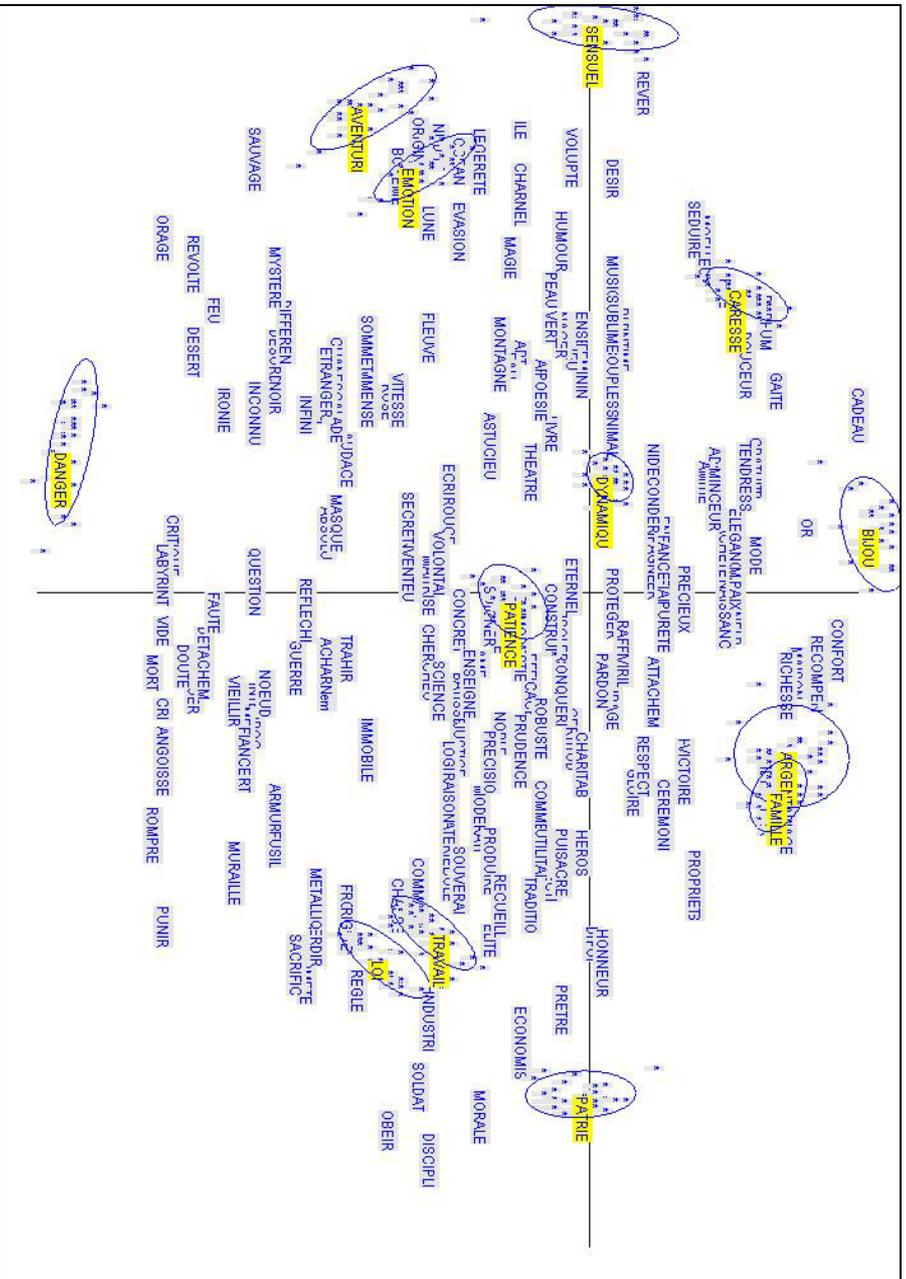


Figure 2.8 : Plan engendré par les axes 2 et 3 – Quelques zones de confiances (Bootstrap total)



## 2.4 Stabilité structurelle

Dans la section précédente de ce chapitre, c'est la stabilité des structures vis-à-vis des fluctuations d'échantillonnage qui a été éprouvée sur l'ensemble des individus selon la méthode de *bootstrap*.

Nous allons changer de registre et poser une question radicalement différente, et beaucoup plus difficile : dans quelle mesure les structures sémiométriques dépendent-elles du choix des mots ?

Cette question est effectivement plus difficile car il n'existe pas un univers des mots comparable à l'univers des individus susceptibles d'être interrogés dans une enquête, univers constitué, par exemple, par l'ensemble des individus âgés de 18 ans ou plus d'un pays donné. Les statisticiens ont l'habitude de traiter la dimension « individu » ou « observation » à laquelle on peut associer la notion de tirage dans une urne et les modèles mathématiques correspondant. Rares sont les situations où l'on est amené à parler de tirage ou d'univers de variables<sup>1</sup>.

On va néanmoins chercher à tester la stabilité des structures vis-à-vis de l'ensemble des variables. On fait alors l'hypothèse que les variables (les 210 mots) sont choisies par un tirage aléatoire sans remise dans une urne contenant l'ensemble des mots « sémiométrisables » de la langue française<sup>2</sup>. Un tirage dans cet univers suppose à chaque fois la constitution d'un nouveau questionnaire et un nouveau travail de terrain, opérations dont le coût serait prohibitif. Comme dans le cas des échantillons d'individus, le *bootstrap* fournit une solution pragmatique et économique à ce problème : on va construire des répliques de l'ensemble des mots en procédant à des tirages avec remise dans l'ensemble des 210 mots déjà sélectionnés.

Alors que le *bootstrap* classique réalisé sur les individus teste la stabilité des *patterns* de variables, c'est-à-dire de mots (*cf.* section 2.2), le *bootstrap* réalisé sur les variables permet d'éprouver les *patterns* observables au niveau des individus. Or ceux-ci sont anonymes et nombreux (plusieurs milliers). Les seules structures intéressantes sont celles qu'ils induisent sur leurs caractéristiques de base (sexe, âge, profession, niveau d'éducation, activités, types d'habitats...). La procédure décrite en annexe A1.2.4 permet

---

1. Il est des circonstances où le problème de l'échantillonnage se pose à la fois pour les variables et les individus. Ce problème ancien, posé par Hotelling (1933) [inventeur/découvreur de l'analyse en composantes principales], a été en particulier étudié par Escoufier (1970).

2. Il s'agit, rappelons-le, de mots pleins - pas de mots outils ou grammaticaux - faisant partie du vocabulaire de base, non consensuels, non ambigus sémantiquement - non polysémiques -, chargés émotionnellement ou axiologiquement - pas de mots neutres du type *table*, ni de mots techniques comme *magnétoscope*.

cependant de positionner les mots correspondant à chaque échantillon répliqué, y compris les mots qui sont absents de certaines répliques<sup>1</sup>.

L'ensemble des variables (ici 210 mots) est beaucoup plus restreint que celui des individus (plusieurs milliers). En moyenne, à chaque tirage avec remise, on démontre qu'environ deux tiers des variables initiales seront présentes, ce qui constitue une ponction très sévère dans l'ensemble initial, et donc une perturbation que l'on peut juger excessive du tableau de données.

La figure 2.10 montre les résultats de cette procédure sur le plan des axes [2,3] : les ellipses de confiance des mots sont bien sûr très dilatées par rapport à celles des graphiques précédents, mais les résultats n'ont rien de catastrophique. Le long du second axe (horizontal), la zone de confiance elliptique du mot *Sensuel* reste opposée à celle du mot *Patrie*, de même, le long du troisième axe (vertical) la zone de confiance du mot *Bijou* reste opposée à celle du mot *Danger*. Les interprétations de ces axes résistent aux très fortes perturbations imposées au recueil de mots.

Les axes suivants montrent une stabilité moins nette, mais on peut penser qu'une liste de mots plus étendue (par exemple une liste de 306 mots, ce qui était le cas lors des premières versions de la sémiométrie) permettrait d'aller plus loin dans l'application de cette épreuve de validité par ailleurs extrêmement sévère.

Il faut retenir de cette section que cette procédure (novatrice) de *bootstrap* sur variables fournit une présomption - et non une preuve - de stabilité de la structure vis-à-vis du choix des mots. Le chapitre 4, consacré à la sémiométrie « ouverte » apportera également des contributions à ce problème de la stabilité vis-à-vis de la composition de la liste, problème intéressant et important, mais pas crucial si la sémiométrie n'est utilisée que pour comparer et différencier, ce qui est le cas dans le cadre des applications actuelles.

---

1. En bref : une réplique *bootstrap* consiste en un tirage avec remise au sein des 210 mots de départ. Certains mots seront absents (ils auront un poids nul), d'autres apparaîtront une fois (poids 1), deux fois (poids 2), etc. En remplaçant les poids nuls par des poids infinitésimaux, ce qui nécessite quelques aménagements techniques supplémentaires, on verra se positionner automatiquement les mots absents comme mots supplémentaires (cf. annexe A1.9.3).

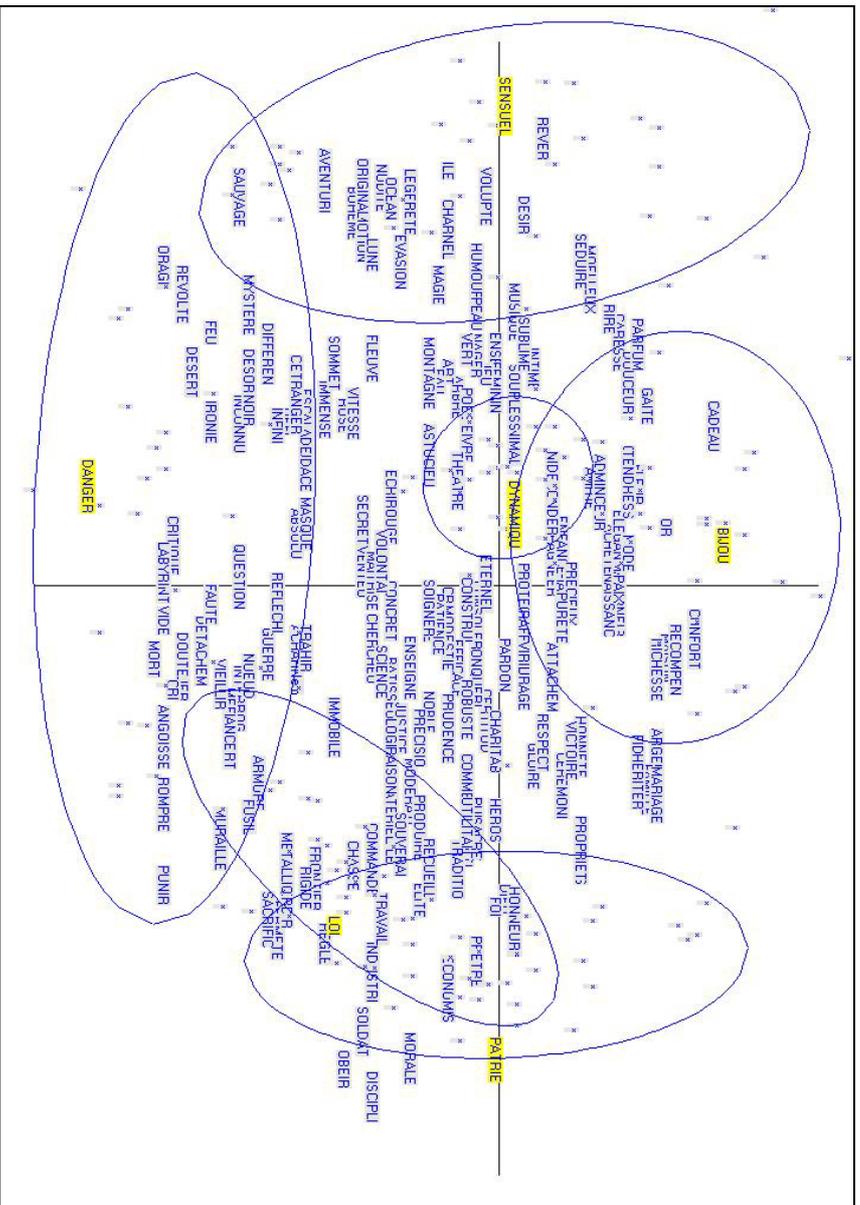


Figure 2.10 : Plan engendré par les axes 2 et 3 – Zones de confiances (Bootstrap sur variables), ellipses de confiance des mots : *Bijou, Danger, Dynamique, Loi, Patrie, Sensuel*

## 2.5 Stabilité vis-à-vis du codage

La structure sémiométrique (existence et permanence de six axes principaux caractérisés par les mêmes mots) se révèle stable dans l'espace, dans le temps, et vis-à-vis d'éventuelles fluctuations d'échantillonnage.

On pourrait aussi objecter, assez légitimement, que le codage de base relève d'une convention assez arbitraire : pour le répondant, c'est surtout l'ordre des notes qui importe, et non la matérialisation de celles-ci sous forme d'échelle de 1 à 7, avec des notes en progression arithmétique.

Est-ce que la structure sémiométrique mise en évidence serait la même si le codage proposé était différent, en respectant toutefois l'ordre des notes ?

Notons déjà que toute transformation  $y$  du codage initial  $x$  de la forme  $y = ax + b$  laisse invariante les corrélations, et donc laisse identique toute la structure sémiométrique. Ainsi, on ne change pas les résultats en remplaçant (1, 2, 3, 4, 5, 6, 7) par (2, 4, 6, 8, 10, 12, 14) ou encore par (11, 12, 13, 14, 15, 16, 17).

Pour déformer encore plus le codage initial, il sera donc procédé à des transformations *non-linéaires*<sup>1</sup> (voir figure 2.11). On va voir que la structure est invariante pour une large famille de telles transformations.

Le résultat est le suivant : pour toutes ces transformations, les six premiers axes sémiométriques sont conservés. Ceci signifie que les axes apparaissent dans le même ordre et sont caractérisés par les mêmes mots.

---

1. Les transformations non-linéaires suivantes du codage arithmétique de départ ont été réalisées (voir figure 2.11) : a)  $y = \sqrt{x}$  ; b)  $y = x^2$  ; c)  $y = \log_{10}(1+x)$  ; d) codage « ternaire » selon lequel  $y = 1$  pour  $x = 1$  ou  $2$  ;  $y = 2$  pour  $x = 3$  ou  $4$  ou  $5$  ;  $y = 3$  pour  $x = 6$  ou  $7$ .

Sur la figure 2.11, le carré a été divisé par 5 et le logarithme multiplié par 10 pour ramener les différents codages à une échelle comparable à celle du codage initial (ou codage de base, en pointillé).

Partant un échantillon de 5 527 individus (France), et donc d'un tableau de notes (5 527 x 210), quatre nouveaux tableaux de mêmes dimensions ont été calculés en appliquant aux notes les transformations précédentes.

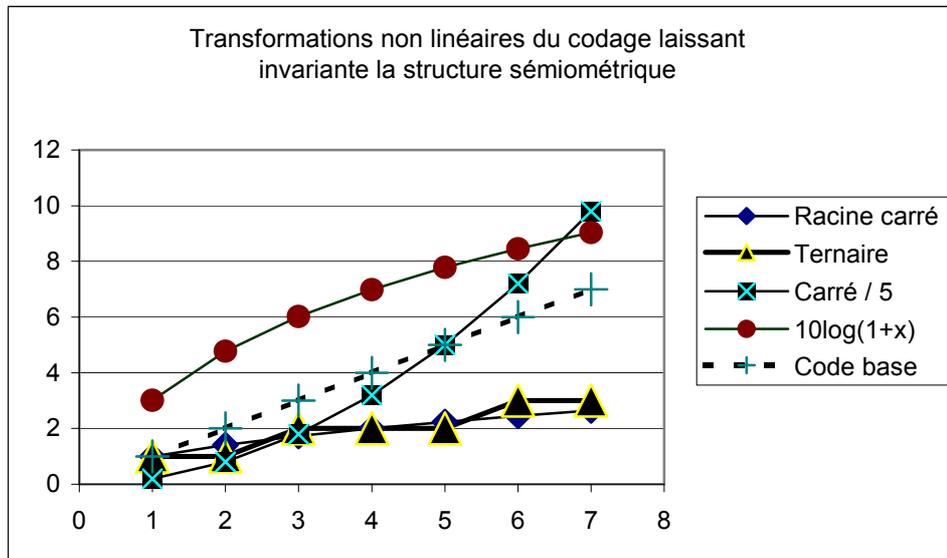


Figure 2.11 : Diverses transformations de l'échelle des notes

Tableau 2.10 : Valeurs propres des différents tableaux recodés

Numéro	Base	Carré	Log	Ternaire	Racine
1	26.6	28.7	23.5	22.1	25.1
2	10.2	10.3	9.6	8.2	9.9
3	8.7	8.4	8.7	7.1	8.8
4	7.0	7.0	6.4	5.9	6.8
5	4.1	4.0	4.0	3.6	4.1
6	3.6	3.7	3.3	3.2	3.5

*On constate que la première valeur propre est maximale pour les carrés des notes (codes fortement dilatés, de 1 à 49), est minimale pour le codage ternaire, qui contracte les 7 valeurs en trois valeurs.*

Ce sont les directions propres (ou axes principaux) qui sont stables, c'est-à-dire caractérisées par les mêmes groupes de mots. Les valeurs propres, qui décrivent les allongements le long de ces axes, sont, elles, plus sensibles au codage, sans pour autant modifier l'ordre des axes.

Il est intéressant de noter que le codage ternaire n'altère pas la structure : cela indique que l'échelle en 7 positions est probablement assez redondante. Mais cela ne prouve pas obligatoirement (il faudrait le vérifier, mais cela impliquerait de retourner sur le terrain) qu'une échelle en trois positions, proposée directement aux personnes interrogées, restituerait la structure initiale.

En revanche, un codage binaire, (par exemple :  $y = 1$  pour  $x = 1,2,3,4$  ;  $y = 2$  pour  $x = 5,6,7$ ) trop destructeur d'information, altère sensiblement la structure. Le premier axe est à peu près conservé (valeur propre 19.3,

inférieure à toutes les premières valeurs propres du tableau 2.4.1) mais on observe une rotation à l'intérieur du plan (2,3). De façon surprenante, les axes 4, 5, 6, sont cependant conservés. Avec un tel codage, certains mots comme *guerre*, *mort*, *trahir*, n'obtiennent que des notes 1 sur tous les individus, d'autres, comme *paix* ou *tendresse* n'obtiennent que des notes 2.

De tels mots, dont les variances sont nulles, n'interviennent plus dans les corrélations et sont exclus de la structure.

## 2.6 Analyse en facteurs communs et spécifiques

La méthode d'analyse en facteurs communs et spécifiques, qui remonte au tout début du vingtième siècle (*cf.* annexe A1.6), est apparentée à l'analyse en composantes principales, mais intervient en général dans un contexte fort différent : il ne s'agit plus d'un simple instrument d'observation, mais d'un modèle.

Ce modèle stipule ici que chacune des 210 notes (correspondant donc à chacun des mots) dépend d'un petit nombre de facteurs communs et d'un facteur spécifique (voir la formulation exacte en annexe A1.6).

Les facteurs communs correspondraient, par exemple, aux six premières dimensions<sup>1</sup>, et les facteurs spécifiques résumeraient en quelque sorte les 204 axes suivants, qui ne sont pas pris en compte.

Ce modèle analytique, qui exige d'ailleurs quelques contraintes supplémentaires, a bien peu de chance d'être validé par un test statistique rigoureux, même en prenant une trentaine ou une cinquantaine de facteurs communs, ne serait-ce que parce que les notes de départ sont des variables discrètes et non des variables continues. Et pourtant, l'ajustement d'un tel modèle (en utilisant la procédure évoquée dans l'annexe A1.6) avec, par exemple, 12 facteurs communs, livre pour les six premiers facteurs (en fait les coefficients des mots sur ces facteurs, appelés  *saturations* ) les six premiers axes de la structure sémiométrique.

L'application de cette méthode figure dans ce chapitre dévolu à la stabilité des résultats car la quasi-totalité des conclusions précédentes sur l'interprétation des axes s'applique aussi à l'analyse factorielle en facteurs communs et spécifiques. Les résultats ont donc résisté à des perturbations importantes des données de base, à des changements de codage, et également à des modifications notables de la méthode de réduction des données.

---

1. L'analyse en composantes principales est dans ce cas un cas particulier d'analyse en facteurs communs et spécifiques avec 210 facteurs communs (on garde tous les axes) et aucun facteur spécifique. Le facteur spécifique est présent pour modéliser les axes abandonnés.

Un des sous-produits intéressants de ce dernier type d'analyse est le calcul, pour chaque mot, de sa variance spécifique (variance non expliquée par les facteurs communs) dont le complément à 1 est la variance commune (appelée parfois *communauté*).

**Tableau 2.11 : Mots ayant les variances communes les plus fortes... et les plus faibles**

Communautés les plus fortes			Communautés les plus faibles		
Mot	Spécificité	Commun.	Mot	Spécificité	Commun.
foi	0.332	0.668	justice	0.823	0.177
dieu	0.362	0.638	chasse	0.825	0.175
prêtre	0.447	0.553	frontière	0.825	0.175
recueillement	0.480	0.520	détachement	0.828	0.172
sacré	0.491	0.509	nager	0.829	0.171
puissance	0.525	0.475	interdire	0.832	0.168
discipline	0.538	0.462	métallique	0.836	0.164
efficace	0.538	0.462	changement	0.839	0.161
sensuel	0.559	0.441	jeu	0.849	0.151
richesse	0.562	0.438	inconnu	0.853	0.147
argent	0.568	0.432	nœud	0.854	0.146
gloire	0.571	0.429	évasion	0.856	0.144
or	0.585	0.415	noir	0.866	0.134
poésie	0.587	0.413	acharnement	0.870	0.130
séduire	0.588	0.412	guerre	0.891	0.109
arbre	0.592	0.408	trahir	0.893	0.107
âme	0.593	0.407	immobile	0.914	0.086
élégance	0.595	0.405	rouge	0.925	0.075

Le tableau 2.11 donne la liste des 18 mots dont les notes sont les mieux expliquées par les facteurs communs (mots ayant les *communautés* les plus fortes) et des 18 mots dont les notes sont les moins bien expliquées par les facteurs communs (mots ayant les spécificités les plus fortes)<sup>1</sup>.

1. Un résultat assez similaire pourrait être trouvé en sommant, pour les six premiers axes, les carrés des coordonnées des mots sur les axes, qui sont d'ailleurs les corrélations des mots avec ces axes (tableau A2.3 de l'annexe 2). Ces sommes représentent, très approximativement, les communautés, alors que leur complément à 1 représente les spécificités. Il ne s'agit ici que d'une approximation car les vraies communautés de l'analyse factorielle en facteurs communs et spécifiques s'obtiennent par un processus itératif (annexe A1.6).

On retrouve dans la première colonne les mots bien expliqués par l'ensemble des premiers axes de la sémiométrie, et dans la seconde des mots qui sont peu ou mal expliqués par ces premiers axes.

## 2.7 Conclusion

Une fois définie la liste de mots et le protocole d'enquête (le questionnaire et le libellé de la question de base), la structure extraite du tableau de données recueillies apparaît stable. Cette structure, formée de six directions principales ordonnées, ne change pas d'un échantillon à un autre, d'une année à une autre, ne subit pas de modifications (ou subit des modifications minimales) d'un pays à un autre – et ce malgré les risques pris lors des traductions du questionnaire.

La structure sémiométrique se retrouve et se répète à l'intérieur de sous-populations particulières (sexe, classes d'âge) suggérant un phénomène fractal, mais il ne s'agit à ce stade que d'une analogie.

Cette stabilité est également confirmée par les épreuves statistiques standards de rééchantillonnage (*bootstrap*) qui confirment les expériences plus empiriques précédentes. Une épreuve non standard (*bootstrap* sur variables) fut une première tentative d'évaluation du degré de dépendance de la structure obtenue vis-à-vis de la composition de la liste de mots.

Le codage numérique choisi pour les réponses peut-être décliné suivant plusieurs fonctions sans altération de la structure, ce qui lève l'hypothèque d'une sensibilité des corrélations au codage somme toute conventionnel de départ.

Enfin, l'analyse factorielle classique, ou analyse en facteurs communs et spécifiques, reproduit les mêmes premières dimensions, comme d'ailleurs l'analyse logarithmique évoquée plus loin au chapitre 5. Il ne s'agit plus exactement de stabilité, mais de présomption de robustesse de la structure sémiométrique.

Les chapitres 3 et 4 qui vont suivre vont relater plusieurs expériences apportant d'autres éléments de réponse à la question beaucoup plus complexe et ambitieuse de l'éventuel caractère intrinsèque de la structure.

## CHAPITRE 3

# Structures sémantiques et sémiométrie

La stabilité mise en évidence au chapitre précédent suscite une série de questions, relatives aussi bien à la nature qu'à l'origine même des traits structuraux.

Une des questions, fondamentale, concerne le rôle et l'importance des relations de type purement sémantique dans la structure sémiométrique. Plusieurs travaux ont été réalisés pour tenter de répondre à cette question. Les relations sémantiques ne sont certes pas faciles à définir ni à mesurer, et dans la discipline connue sous le nom de Traitement du Langage Naturel<sup>1</sup>, les questions qui touchent la sémantique restent parmi les plus ardues.

En admettant que l'on puisse vaincre ces difficultés, il subsiste l'interrogation suivante : *les proximités sémantiques entre mots ne sont-elles pas responsables de l'essentiel de la structure observée ?* Autrement dit, de façon un peu sommaire, *la structure que l'on observe est-elle une structure linguistique plutôt que psychologique ou psycho-sociologique ?*

Deux mots ayant des sens voisins seraient notés de façon similaire et donc corrélés, et le *pattern* observé ne serait autre que le réseau de ces liens sémantiques.

La stabilité de la structure en découlerait, puisque la langue est relativement stable dans le temps. Elle est aussi pratiquement la même pour

---

1. Discipline d'où sont dérivés les produits usuels que sont les correcteurs d'orthographe, les logiciels de traduction, les moteurs de recherche documentaire.

les différentes classes d'âge, le sexe, etc.. Un tel réseau sémantique doit aussi résister - plus ou moins - à une traduction du questionnaire, d'où la relative stabilité observée d'un pays à un autre<sup>1</sup>.

Pour apporter des éléments de réponse à ces interrogations, le présent chapitre relate trois séries de travaux, formant autant de parties.

La première partie, intitulée « Les champs sémantiques internes » montre comment les réponses au questionnaire sémiométrique et les corrélations entre mots calculées à partir de ces réponses décrivent effectivement, et assez finement, des liens de nature sémantique. Ce résultat n'était pas évident *a priori*, car, ne l'oublions pas, les questions posées ne portent que sur le caractère *agréable* ou *désagréable* des mots de la liste.

La deuxième partie intitulée « Voisinage sémantique du questionnaire sémiométrique », est dévolue à l'expérience suivante : les 210 mots de la sémiométrie vont être décrits par leurs *synonymes*, ou au moins par leurs *voisins sémantiques*, ceux-ci étant fournis par un dictionnaire de synonymes. Autrement dit, plus de questionnaire à faire remplir par un échantillon représentatif... c'est la langue elle-même qui répond, à travers certains des outils qui ont été forgés pour la décrire et la comprendre. Bien sûr, les dictionnaires de synonymes ne sont pas des instruments parfaits, et ils diffèrent notablement entre eux. L'expérience sera d'ailleurs l'occasion de voir l'étendue des difficultés qui nous attendent.

La troisième partie intitulée « Réseau sémantique des verbes français », relate une expérience beaucoup plus générale qui a consisté à prendre tous les verbes de la langue française (en fait les 829 verbes les plus usuels figurant dans le manuel classique de grammaire Bescherelle, parmi lesquels se trouvent déjà de nombreux verbes très peu usités, voire quasi-inconnus) et à étudier la structure sémantique de ces verbes à partir de leurs voisins.

Cette expérience confirmera qu'il existe des proximités locales sémantiques, mais qu'il n'existe pas d'axes principaux dominants et stables.

---

1. Disons d'emblée, en anticipant sur les résultats présentés dans ce chapitre, que la structure sémiométrique ne se réduit que très partiellement à une structure sémantique. Au niveau local, autour d'un point représentant un mot, on pourra trouver des voisins sémantiques, mais les grandes oppositions responsables d'axes stables ne sont pas observées spontanément à partir d'analyses de simples proximités sémantiques.

### 3.1 Les « champs sémantiques internes »

On appellera champ sémantique interne pour un mot donné l'ensemble des mots (pris dans une liste de mots fixée *a priori*, d'où le qualificatif d'*interne*) qui lui sont corrélés. La distance utilisée est d'autant plus petite que le coefficient de corrélation entre les deux mots est élevé<sup>1</sup>. Cette dénomination est justifiée par l'interprétation *a posteriori* des proximités observées. Il n'était pas évident que des notes fondées seulement sur l'agrément ou le désagrément engendrent des proximités sémantiques. Cette section montre la cohérence et la finesse des proximités observées.

Certains mots possèdent un champ sémantique interne riche et dense défini par des corrélations élevées avec de nombreux autres mots de la liste. C'est le cas pour les mots *Efficace* et *Courage*. Des mots comme *Montagne*, *Voluptueux*, *Mystère*, proches d'un petit nombre de mots, ont chacun un champ plus restreint.

Un même mot peut bien évidemment, dans certaines limites, être corrélé avec des mots non corrélés entre eux. C'est ainsi que les champs sémantiques de *Courage* et *Efficace* ne sont pas les mêmes malgré leur forte corrélation et l'appartenance de certains mots comme *Robuste* et *Dynamique* aux deux champs.

#### Champ sémantique du mot *Efficace*

*Précision* (.51)<sup>2</sup>, *Concret* (.40), *Robuste* (.40), *Courage* (.39), *Astucieux* (.38), *Maîtriser* (.38), *Dynamique* (.37), *Raison* (.37), *Confiance* (.35), *Produire* (.35), *Utilitaire* (.35), *Certitude* (.34), *Honnête* (.34), *Volontaire* (.34), *Construire* (.33), *Logique* (.32), *Récompense* (.32), *Réfléchir* (.32), *Respect* (.32), *Bâtitteur* (.31), *Fermeté* (.31), *Honneur* (.31), *Douceur* (.30).

#### Champ sémantique du mot *Courage*

*Honnête* (.41), *Dynamique* (.40), *Respect* (.40), *Efficace* (.39), *Confiance* (.38), *Honneur* (.38), *Elégance* (.36), *Politesse* (.36), *Amitié* (.35), *Guérir* (.35), *Robuste* (.35), *Charitable* (.34), *Précision* (.33), *Volontaire* (.33), *Douceur* (.32), *Protéger* (.32), *Pureté* (.32), *Raison* (.32), *Gaieté* (.31), *Récompense* (.31), *Certitude* (.30).

---

1. Le carré de la distance entre deux mots *a* et *b* dont le coefficient de corrélation vaut  $r(a,b)$  s'écrit  $d^2(a,b) = (1 - r(a,b))$ .

2. Les quantités entre parenthèses sont les coefficients de corrélation avec le mot en caractères gras qui précède. Notons que tous les coefficients de corrélation cités sont hautement significatifs statistiquement compte tenu de la taille de l'échantillon (11 055).

### Champ sémantique du mot *Voluptueux*

*Sensuel (.47), Charnel (.45), Désir (.34), Moelleux (.34), Nudité (.32), Séduire (.31), Souplesse (.31), Sublime (.30), Féminin (.25), Rêver (.24), Astucieux (.23), Audace (.23), Douceur (.23), Intime (.23), Peau (.23), Caresse (.22), Légèreté (.22), Original (.22), Art (.20), Élégance (.20), Viril (.20).*

### Champ sémantique du mot *Montagne*

*Sommet (.47), Escalader (.34), Campagne (.33), Fleuve (.32), Arbre (.31), Océan (.31), Ile (.26), Lune (.26), Fleur (.24), Musique (.23), Eau (.22), Animal (.21), Bleu (.21), Construire (.21), Nid (.21), Douceur (.20).*

### Champ sémantique du mot *Mystère*

*Orage (.26), Magie (.25), Aventurier (.24), Emotion (.24), Original (.24), Secret (.24), Sauvage (.23), Infini (.22), Feu (.21), Océan (.20).*

### Champ sémantique du mot *Dieu*

*Foi (.74), Prêtre (.65), Recueillement (.60), Sacré (.55), Ame (.48), Créateur (.33), Patrie (.32), Pardon (.29), Charitable (.27), Morale (.27), Héros (.26), Honneur (.26), Noble (.26), Cérémonie (.25), Eternel (.25), Tradition (.25), Mariage (.23), Souverain (.22), Famille (.21), Pureté (.21), Discipline (.20), Elite (.20).*

Ce premier cadrage d'un espace des mots, par la constitution de quelques champs sémantiques internes, peut se généraliser à l'ensemble des mots du questionnaire. On utilise pour cela les méthodes de classification qui consistent à effectuer une partition de l'ensemble des éléments à classer c'est-à-dire les mots. Cependant, dans ce cas, un mot ne peut appartenir à plusieurs champs.

### 3.1.1 Représentation hiérarchique des proximités sémantique

Les mots qui se ressemblent le plus quant à leur manière d'être notés et donc ressentis vont être regroupés maintenant de façon automatique. La méthode d'analyse en composantes principales présentée au premier chapitre réalise une telle représentation globale des corrélations, et donc des proximités. Mais les deux méthodes de classification appliquées ici, *classification hiérarchique*<sup>1</sup> et *cartes auto-organisées* de Kohonen<sup>2</sup>, décrivent peut-être avec plus de finesse les proximités locales qui sont, nous le verrons, les seules à être interprétables en termes sémantiques.

Pour ce faire, on réalise dans un premier temps une *classification hiérarchique* sur les 210 mots munis des distances définies précédemment.

---

1. Le principe des méthodes de classification hiérarchique est brièvement présenté en annexe A1.7.

2. Le principe des cartes auto-organisées est brièvement présenté en annexe A1.8.

Le principe de ce traitement (l'algorithme, en fait) est de regrouper les mots deux par deux par agglomération progressive fournissant ainsi une hiérarchie de *partitions* des mots. Nous retenons ici les partitions emboîtées en 12, 24 et 36 classes représentées par le tableau 3.1 qui schématise le *dendrogramme* de la hiérarchie.

Les résultats parlent d'eux-mêmes. Les classes obtenues regroupent des mots qui ne sont pas de vrais synonymes<sup>1</sup>, mais appartiennent au même *halo sémantique*.

Examinons, par exemple, la partition en 12 classes (la moins fine des trois partitions) dont les classes sont repérées dans la première colonne du tableau 3.1 par les symboles 1/12, 2/12, ..., 12/12.

Sont regroupés au niveau de cette partition, les mots ayant trait au concept de « Sublimation » (classe 1/12, *Absolu, Immense, Infini, Admirer, Adorer, ..., Secret, Sublime*). D'autres classes correspondent à des thèmes comme l'art de l'écriture (classe 3/12 : *Art, Ecrire, livre, Poésie, Théâtre*), la nature (classe 4/12 ; *Escalader, Île, .., Sommet, Animal, ..., Vert*), l'émotion (classe 5/12 ; *Emotion, ..., Rêver, Humour, ..., Rouge, Bohème, .., Révolte*).

Notons, en passant, que l'on retrouve dans la classe 28/36 le champ sémantique du mot *Efficace* ; le mot *Courage* n'en fait pas partie et se regroupe avec les mots *Respect, Politesse, Honnête*.

### 3.1.2 Représentation plane des proximités sémantiques

La représentation des classes emboîtées du tableau 3.1 ne permet pas d'apprécier les distances relatives entre les groupements de mots obtenus par la classification hiérarchique. La carte de Kohonen (figure 3.1) est un mode de visualisation de données qui regroupe les mots en classes en conservant autant que faire se peut la topologie initiale de l'espace des mots. Elle rend compte ainsi des proximités entre classes qui se ressemblent (*cf.* annexe A1.8).

Ainsi, les mots *Théâtre* et *Poésie* sont proches à la fois des regroupements « *Nid, Livre, Fleuve, Ecrire, Art* », « *Vert, Sublime, Nager, Musique, Montagne, Eau, Bleu* », « *Sommet, Secret Lune, Immense* », « *Justice* », « *Recueillement, Créateur, Ame* » (*cf.* figure 3.1).

---

1. La liste de mots excluait *a priori* la présence de synonymes qui auraient introduit des redondances inutiles.

**Tableau 3.1 : Classification hiérarchique en 12, 24 et 36 classes emboîtées (cf. annexe 1.7)**

1/12	1/24	1/36	absolu, immense, infini
		2/36	admirer, adorer, éternel, précieux, secret, sublime
2/12	2/24	3/36	douceur, féminin, intime, moelleux, peau, souplesse
		4/36	caresse, désir, séduire
		5/36	charnel, nudité, sensuel, voluptueux
3/12	4/24	6/36	art, écrire, livre, poésie, théâtre
	5/24	7/36	escalader, île, lune, montagne, océan, sommet
4/12	6/24	8/36	animal, arbre, bleu, campagne, eau, fleur, fleuve, nager, nid, vert
		9/36	émotion, évasion, légèreté, magie, original, rêver
5/12	7/24	10/36	humour, jeu, musique, rouge
		11/36	bohème, changement, différent, étranger, inconnu
		12/36	acharnement, aventurier, défi, ruse, vitesse
		13/36	désert, feu, mystère, noir, orage, sauvage
6/12	9/24	14/36	critiquer, désordre, ironie, révolte
		15/36	guerre, trahir
		16/36	angoisse, cri, danger, détachement, doute, faute, labyrinthe, méfiance, mort, rompre, vide, vieillir
		17/36	armure, attaquer, chasse, fusil
6/12	11/24	18/36	immobile, masque, muraille, nœud
		19/36	frontière, interdire, métallique, punir, rigide, sacrifice
		12/24	

**Tableau 3.1 (suite) : Classification hiérarchique en 12, 24 et 36 classes emboîtées**

7/12	13/24	20/36	bijou, élégance, minceur, mode, parfum
		21/36	acheter, cadeau, confort, gratuit, maison, récompense
	14/24	22/36	ambition, conquérir, gloire, puissance, victoire
		23/36	argent, hériter, or, propriété, richesse
8/12	15/24	24/36	noble, perfection, raffiné
		25/36	commander, élite, héros, honneur, patrie, soldat, souverain, tradition, viril
	16/24	26/36	commerce, industrie, matériel, produire, utilitaire
	17/24	27/36	discipline, économiser, fermeté, loi, morale, obéir, règle
9/12	18/24	28/36	astucieux, audace, certitude, concret, dynamique, efficace, logique, maîtriser, précision, raison, robuste, volontaire
		19/24	effort, interroger, question, réfléchir
	20/24	30/36	bâtisseur, chercheur, construire, école, enseigner, inventeur, science, travail
10/12	21/24:	31/36	âme, créateur, dieu, foi, prêtre, recueillement, sacré
11/12	22/24	32/36	enfance, ensemble, féconder, maternel
		33/36	cérémonie, famille, mariage, naissance
	23/24	34/36	courage, fidélité, honnête, politesse, respect
		35/36	amitié, confiance, gaieté, guérir, paix, rire, tendresse
12/12	24/24	36/36	attachement, charitable, consoler, humble, justice, modération, modestie, pardon, patience, protéger, prudence, pureté, soigner

RICHESSE OR HERITER GLOIRE ARGENT	RECOMPENSE PRECIEUX MODE MINCEUR GRATUIT ELEGANCE CONFORT CADEAU BIJOU ADORER ACHETER	MAISON FECONDER DOUCEUR	TENDRESSE PROTEGER PAIX NAISSANCE MATERNEL GUERIR GAITE FLEUR FIDELITE FAMILLE ENFANCE CONFIANCE CAMPAGNE ATTACHEMENT AMITIE	SOIGNER RESPECT PUISSANCE POLITESSE PATIENCE HUMBLE HONNETE CONSOLER	TRAVAIL REGLE MORALE MODERATION LOI EFFORT ECOLE
VIRIL VICTOIRE PRUDENCE PROPRIETE CONQUERIR COMMERCE AMBITION	CERTITUDE	RAFFINE MARIAGE CEREMONIE ADMIRER	PURETE PARDON CHARITABLE	MODESTIE ENSEIGNER COURAGE	UTILITAIRE RAISON PRODUIRE PRECISION LOGIQUE FERMETE
SOLDAT MATERIEL INDUSTRIE ECONOMISER DISCIPLINE COMMANDER	TRADITION SOUVERAIN PERFECTION PATRIE HONNEUR HEROS ELITE	SACRE PRETRE NOBLE FOI ETERNEL DIEU	RECUEILLEMEN CREATEUR AME	JUSTICE	VOLONTAIRE SCIENCE ROBUSTE REFLECHIR MAITRISER INVENTEUR EFFICACE CONSTRUIRE CONCRET CHERCHEUR BATISSEUR
SACRIFICE RIGIDE OBEIR METALLIQUE INTERDIRE FRONTIERE CHASSE	INTERROGER	ABSOLU	THEATRE POESIE	NID LIVRE FLEUVE ECRIRE ART	DYNAMIQUE AUDACE ASTUCIEUX ARBRE
ROMPRE PUNIR NOEUD MURAILLE MEFIANCE IMMOBILE FUSIL CRI ARMURE	QUESTION MASQUE DETACHEMENT ACHARNement	ROUGE INFINI INCONNU ETRANGER ESCALADER	SOMMET SECRET LUNE IMMENSE	VERT SUBLIME NAGER MUSIQUE MONTAGNE EAU BLEU	SOUPLESSE PEAU JEU INTIME HUMOUR FEMININ ENSEMBLE ANIMAL
VIEILLIR VIDE TRAHIR MORT LABYRINTHE GUERRE FAUTE DOUTE DANGER ATTAQUER ANGOISSE	RUSE REVOLTE ORAGE NOIR IRONIE DESORDRE DESERT CRITIQUER	VITESSE SAUVAGE MYSTERE FEU DIFFERENT DEFI CHANGEMENT	ORIGINAL NUDITE MAGIE LEGERETE EVASION EMOTION BOHEME AVENTURIER	VOLUPTÉ SENSUEL OCEAN ILE CHARNEL	SEDUIRE RIRE REVER PARFUM MOELLEUX DESIR CARESSE

Figure 3.1 : Proximités entre mots décrites par une carte de Kohonen

La classe « *Recueillement, Créateur, Âme* » est elle-même proche de « *Modestie, Enseigner, Courage* » et « *Pureté, Pardon, Charitable* », « *Raffiné, Mariage, Cérémonie, Admirer* », « *Sacré, Prêtre, Foi, ...* ».

Les classes sont sans doute moins homogènes et parfois moins cohérentes que les partitions hiérarchiques mais la représentation par la carte auto-organisée de Kohonen insiste davantage sur la cohérence entre les classes.

Les cartes de Kohonen ont cependant l'inconvénient d'être plus instables que les classifications ascendantes hiérarchiques et s'adressent davantage à des ensembles de données non structurées. De fait, nous verrons au chapitre suivant qu'elles sont mieux adaptées aux questionnaires ouverts qu'à l'analyse de corpus de mots déjà fortement structuré.

### 3.2 Voisinage sémantique du questionnaire sémiométrique

Dans cette première expérience, chaque mot de la sémiométrie est décrit par ses « synonymes », collationnés à partir d'une source externe<sup>1</sup>. Le tableau 3.2 nous donne un court extrait du recueil ainsi obtenu pour les premiers mots de *Absolu* à *Angoisse*. Comme on le voit, le nombre de voisins sémantiques est très variable.

Ce nombre de voisins peut être très grand et peut atteindre 22 lignes de texte pour le mot *Changement* qui ne figure pas dans l'extrait du tableau 3.2, ou très court (voire inexistant, ainsi, l'adjectif *Maternel* – également extérieur au tableau 3.2 – n'a pas de voisin sémantique dans ce dictionnaire).

A partir de ce nouveau recueil de textes est construit un *tableau croisé* contenant en lignes les 210 mots de base, et en colonnes tous les mots rencontrés dans le recueil, c'est-à-dire tous les voisins sémantiques des 210 mots.

Le tableau, à l'intersection de la ligne *i* et de la colonne *j* contient la valeur « 1 » si le mot *j* figure parmi les voisins sémantiques du mot *i*, et contient la valeur « 0 » sinon.

Si certains mots ont des voisins sémantiques communs, un calcul de distance approprié les désignera comme proches.

---

1. Synonymes et voisins sémantiques ont été fournis par le *Dictionnaire de synonymes et contraires* de Henri Bertaud du Chazaud, publié par Robert, 1994.

**Tableau 3.2 :**  
**Exemples de voisins sémantiques des neuf premiers mots**

– **Absolu**

catégorique, complet, dirimant, discrétionnaire, dogmatique, entier, exclusif, foncier, formel, impératif, impérieux, inconditionnel, indispensable, infini, parfait, plein, radical, rédhitoire, total, unilatéral, autocratique, autoritaire, arbitraire, cassant, césarien, despotique, dictatorial, dogmatique, exclusif, impérieux, intransigeant, jupitérien, omnipotent, souverain, totalitaire, tyrannique, idéal, infini, intégrité, intransigeance, perfection, plénitude

– **Acharnement**

ardeur, effort, énergie, persévérance, ténacité, cruauté, entêtement, fureur, furie, obstination, opiniâtreté, rage, sadisme

– **Acheter**

acquérir, corrompre, soudoyer

– **Admirer**

apprécier, s'émerveiller, s'enthousiasmer, s'extasier, louer, louer

– **Adorer**

aimer, bénir, glorifier, idolâtrer, admirer, honorer, respecter, révéler, vénérer, aduler, courtoiser, encenser, flagorner, flatter

– **Ambition**

appétit, ardeur, aspiration, brigue, convoitise, désir, faim, fringale, idéal, passion, prétention, quête, recherche, soif, arrivisme, mégalomanie, but, dessein, fin, mobile, objet, rêve, visée, vue

– **Ame**

cœur, conscience, dedans, esprit, fond, intérieur, mystère, pensée, principe, secret, spiritualité, transcendance, vie, air, ectoplasme, émanation, essence, éther, étincelle, feu, flamme, souffle, vapeur, émotion, expression, intelligence, sensibilité, sentiment, agent, animateur, centre, cerveau, chef, instigateur, maître, moteur, nœud, organisateur, patron, responsable

– **Amitié**

accord, cordialité, entente, sympathie, amabilité, compliment, hommages, caresse, flagornerie, flatterie, grimace, affection, affinité, amour, attachement, bonté, complaisance, dévotion, dévouement, dilection, douceur, inclination, inclination, intérêt, lien, penchant, piété, respect, sollicitude, tendresse, union, vénération, goût, prédilection, vocation, bienveillance, affabilité, altruisme, compréhension, convivialité, cordialité, gentillesse, indulgence, mansuétude, obligeance, prévenance, bonté, charité, clémence, compassion, générosité, indulgence, mansuétude, miséricorde, indulgence, pitié, serviabilité, socialité,

– **Angoisse**

inquiétude, anxiété, appréhension, crainte, émotion, ennui, malaise, peine, préoccupation, scrupule, souci, stress, supplice, tension, trac, transe, trouble, alarme, alerte, émoi, affolement, agitation, désarroi, détresse, effarement, effroi, épouvante, panique, peur, terreur

Selon une méthode d'application courante pour les analyses statistiques de texte, on peut donc construire un tableau des distances entre les mots, et représenter graphiquement ces distances par des proximités entre points.

Le premier résultat obtenu, décevant, mais prévisible, est une séparation entre les noms et les verbes... qui ne peuvent pas être synonymes, et assez rarement voisins sémantiques dans les dictionnaires usuels, alors que noms et verbes figurent parmi les mots de la sémiométrie.

On procédera donc à l'analyse des seuls noms et adjectifs, qui, au nombre de 177, sont donc très majoritaires dans la liste originale des 210 mots du questionnaire.

On n'observe plus des nuages de points réguliers et équilibrés comme ceux des graphiques sémiométriques présentés dans les sections précédentes, mais des grappes de mots qui s'opposent à tous les autres.

Ce phénomène semble dû à la *non-transitivité* des similitudes sémantiques, et à l'absence de pertinence de la notion d'éloignement sémantique. La non-transitivité peut s'exprimer de la façon suivante : si le mot A a un sens voisin de celui du mot B, et si le mot B a un sens voisin de celui du mot C, alors, A n'a pas forcément un sens voisin de celui du mot C. On peut même assez rapidement aboutir à un antonyme de A<sup>1</sup>.

L'observation des liens sémantiques issus d'un dictionnaire avec ceux que l'on peut dériver des notes sémiométriques aide à comprendre la nature du fait statistique et de la structure que l'on observe.

Tout d'abord, il existe, parmi les 210 mots de la sémiométrie, des mots qui n'ont pas de voisins sémantiques communs avec les autres (en fait avec plus de deux autres) comme par exemple les mots : *Lune, Peau, Arbre, Fusil, Ile, Théâtre, Soldat*. On comprend bien que pour *Lune*, par exemple, ses synonymes (*Artémis, Diane, Hécate, Phoebé, Séléné*<sup>2</sup>), trop particuliers, ont peu de chance de se retrouver parmi les voisins, même éloignés, des autres mots. De même pour *Fusil* (*carabine, lebel, mousqueton, arquebuse, mousquet, tromblon*).

---

1. Ainsi, *abandonner* est un voisin sémantique de *donner*, voisin de *partager*, qui est lui-même un voisin de *participer*, à son tour voisin de *se joindre*. On rencontre en général des chaînes plus longues, qui conduisent plus simplement à une indifférence sémantique. Comme la séquence *littérature – poésie – musique – vibration*. Le premier et le dernier terme de la chaîne n'ont pas – ou peu – de rapports, sans pour autant avoir des sens contraires.

2. Bien que n'appartenant pas au questionnaire sémiométrique, ces mots, étant des noms propres, figurent avec une majuscule.

Les corrélations observées entre ces mots et les autres ne sont donc pas des artefacts sémantiques, elles relèvent d'un contexte perceptif plus général.

Il existe, en revanche, des mots ou groupes de mots qui ont entre eux des liens sémantiques incontestables, lesquels vont se traduire simultanément par des proximités sur les cartes sémiométriques et sur les graphes sémantiques<sup>1</sup>.

Ainsi, les mots *Charnel*, *Sensuel*, *Voluptueux* qui figurent dans le questionnaire sémiométrique sont des voisins sémantiques les uns des autres, et ont plusieurs voisins sémantiques en commun dans le dictionnaire de synonymes que nous avons utilisé (comme par exemple : *animal*, *bestial*, *érotique*, *lascif*, *libidineux*, *luxurieux*). Il s'agit cependant d'une exception dans le questionnaire. Il n'existe pas d'autres groupes de trois mots aussi proches sémantiquement.

On pourrait alors objecter que cette redondance partielle (partielle car ces mots ont malgré tout des sens clairement distincts) donne un poids excessif aux notions sous-jacentes. Ce groupe de trois mots serait alors responsable de certains des traits de structure observés et décrits au chapitre précédent. Mais ce n'est pas le cas : comme on le vérifie aisément, les axes de base dont nous avons décrit la stabilité ne sont pas modifiés si l'on supprime, dans le fichier des notes, deux sur trois des mots incriminés.

Fait encore plus marquant, les six premiers axes sont conservés si l'on supprime de la liste l'ensemble des trois mots *Charnel*, *Sensuel* et *Voluptueux*.

Le second axe, le plus corrélé aux trois mots supprimés, et donc le plus susceptible d'être affecté par cette suppression, est étonnamment stable, comme l'atteste le tableau 3.3, qui nous montre les mots les plus extrêmes sur l'axe. Dans les deux cas, ces listes s'opposent à *Patrie*, *Argent*, *Propriété*, *Soldat*, *Rigide*, *Economiser*, *Obéir*, *Discipline*.

Ainsi, les proximités sémantiques, particulièrement importantes dans le cas de ces trois mots, semblent jouer un rôle secondaire dans l'établissement de la structure stable sous-jacente.

Il existe aussi des couples de mots, comme *Ironie* et *Humour*, qui ont un grand nombre de voisins sémantiques communs (*esprit*, *causticité*, *malice*, *verve*, *spirituel*, *comique*, *drôle*, *facétieux*, *malicieux*, *piquant*, *satirique*, *moquerie*, *saillie*, *satire*) et qui sont donc très proches sémantiquement, que

---

1. L'outil retenu pour donner une appréciation visuelle agréable des graphes sémantiques est de nouveau la carte auto-organisée de Kohonen. Cf. annexe A1.8 pour plus de détails sur cette méthode neuronale et la justification de son emploi dans ce contexte.

la sémiométrie sépare de façon significative. Ainsi, sur le troisième axe « Attachement / Détachement », *Ironie* se situe du côté du « Détachement » et « Humour », de celui de l' « Attachement » .

**Tableau 3.3 :**

<b>Liste des mots les plus corrélés positivement avec le deuxième axe</b>	
Avec 210 mots	Avec 207 mots
<i>Art</i>	<i>Art</i>
<i>Bohème</i>	<i>Bohème</i>
<i>Livre</i>	<i>Livre</i>
<i>Original</i>	<i>Original</i>
<i>Sensuel</i>	<i>Poésie</i>
<i>Poésie</i>	<i>Orage</i>
<i>Révolte</i>	<i>Théâtre</i>
<i>Voluptueux</i>	<i>Musique</i>
<i>Orage</i>	<i>Révolte</i>
<i>Sauvage</i>	<i>Océan</i>
<i>Océan</i>	<i>Sauvage</i>

Autre exemple de proximité sémantique non traduite par des proximités sémiométriques : le mot *Animal*, dont on a vu qu'il figurait parmi les voisins sémantiques des mots *Charnel*<sup>1</sup>, *Sensuel*, *Voluptueux* ne leur est pas voisin dans les plans sémiométriques. Il s'en éloigne notamment le long du second axe (« Devoir /Plaisir »). Il faut cependant préciser qu'il y a ici ambiguïté entre le substantif *animal* et l'adjectif *animal* qui peuvent incontestablement avoir des voisins sémantiques différents. Le questionnaire sémiométrique spécifie bien « Un animal ».

En conclusion de ce paragraphe, on retiendra que :

- la structure sémantique de la liste sémiométrique telle qu'elle est décrite (grossièrement, il est vrai) par un dictionnaire de synonymes ne donne pas lieu à des axes (ou directions principales) stables.
- localement, on retrouve des associations sémantiques sur les cartes sémiométriques, mais il existe aussi des exceptions notables.
- les grandes oppositions stables observées dans le champ sémiométrique ont un caractère psycho-sociologique, voire socio-démographique marqué (liens de certains axes avec le sexe, l'âge, la catégorie socio-

<sup>1</sup> Le mot *Charnel* n'apparaît pas dans le tableau 3.3 car il n'occupe que le treizième rang sur le deuxième axe.

professionnelle), et ne relèvent pas ou peu du registre synonymie-antonymie.

### 3.3 Un réseau sémantique des verbes français

Cette seconde expérience, contrairement à la précédente, est indépendante du corpus de 210 mots. C'est l'ensemble des verbes français usuels (les 829 verbes les plus fréquents figurant dans le manuel classique de grammaire « Bescherelle »<sup>1</sup>) qui est maintenant pris en considération.

La question posée est aussi beaucoup plus générale et ambitieuse : se dégage-t-il une structure analogue à celle observée lors des analyses sémiométriques lorsque l'on décrit l'ensemble des distances sémantiques entre les verbes de la langue française ?

Chaque verbe sera encore décrit par l'ensemble de ses synonymes, ce qui va permettre de calculer des distances entre verbes, et donc d'observer la structure induite par ces distances.

Pourquoi avoir choisi les verbes ? Parce que ceux-ci constituent un ensemble beaucoup plus restreint et homogène que celui des noms ou des adjectifs, et qu'il est donc possible d'être (quasi) exhaustif.

Les résultats confirmeront ceux observés au paragraphe précédent : la structure sémantique est essentiellement locale, et les « grandes distances sémantiques » ne donnent pas lieu à des axes stables, ni à des oppositions interprétables.

Le corpus formé par les verbes et leurs synonymes (analogue au texte du tableau 3.2 ci-dessus, mais relatif cette fois à 829 verbes) comporte 17 446 occurrences de mots (qui sont évidemment tous des verbes).

Ce même texte comporte 3 839 verbes distincts. Ce nombre est beaucoup plus grand que les 829 verbes de départ, car des verbes moins usités peuvent figurer parmi les synonymes des verbes du *Bescherelle*. Le tableau 3.4 donne, à titre d'exemple, les synonymes retenus du verbe *couper*.

Deux verbes seront d'autant plus proches qu'ils auront beaucoup de synonymes communs.

Evidemment, beaucoup de paires de verbes n'auront aucun synonyme commun, et les distances entre ces paires seront équivalentes.

---

1. Bescherelle, *La conjugaison, Dictionnaire de 12 000 verbes*, Hatier, Paris, 1990.

### Tableau 3.4 Synonymes du verbe *couper*

*amputer, champlever, cisailer, coupailier, débiller, découper, diviser, ébarber, élaguer, entrecouper, hacher, inciser, massicoter, réséquer, scarifier, sectionner, taillader, tailler, trancher, tronçonner, essoriller, émonder, séparer, supprimer, soustraire, censurer, exclure, châtrer, castrer, chaponner, émasculer, hongrer, traverser, franchir, mélanger, abattre, scier, faucher.*

Comme cela est habituel lors des analyses statistiques de textes, il est nécessaire de faire intervenir un seuil de fréquence minimale des mots.

Ne seront retenus ici dans un premier temps que les verbes apparaissant au moins quatre fois dans le corpus formé de l'ensemble des verbes et de leurs synonymes. Ceci laisse 1 619 mots distincts, et réduit, dans une proportion moindre, la taille du corpus global qui comporte maintenant 13 660 occurrences.

Cette fréquence minimale de quatre laisse quand même quelques verbes relativement peu courants dans le corpus, tels que : *contrister, lancer, transmuier, gourmander, congutiner, épisser, déconfire, carguer, excaver, ravauder, désencombrer.*

Le tableau 3.5 donne les verbes les plus fréquents dans le corpus, en se limitant maintenant, pour des questions d'encombrement, aux verbes apparaissant plus de 16 fois<sup>1</sup>.

Afin d'obtenir rapidement une vue d'ensemble du graphe sémantique des verbes de la langue française, on est finalement conduit à procéder à une analyse des correspondances (cf. annexe A1.4) d'un tableau binaire ayant 1 619 lignes et 829 colonnes<sup>2</sup>.

---

1. Le tableau dans son intégralité est disponible auprès des auteurs.

2. Le terme situé à l'intersection de la ligne *i* et de la colonne *j* de ce tableau vaut 1 si le verbe *i* est un des synonymes du verbe *j*, et vaut 0 sinon.

**Tableau 3.5 : Extrait de la liste des verbes (classés ici par fréquences décroissantes (de 50 à 17) du nombre de synonymes)**

<i>verbes</i>	<i>frequence</i>	<i>verbes</i>	<i>frequence</i>	<i>verbes</i>	<i>frequence</i>
enlever	50	régler	22	organiser	18
fixer	39	comprendre	22	incorporer	18
prendre	36	conduire	22	aider	18
examiner	33	encourager	22	entamer	18
joindre	33	assurer	22	accabler	18
arrêter	33	lier	22	apprendre	18
entraîner	32	écarter	22	étendre	18
assembler	32	céder	22	tirer	18
pousser	31	supprimer	22	condamner	18
soutenir	31	diriger	22	juger	18
abandonner	30	trancher	21	signaler	18
augmenter	30	garantir	21	remplir	18
atteindre	29	déterminer	21	transporter	17
montrer	28	attirer	21	communiquer	17
protéger	28	approuver	21	rassembler	17
maintenir	28	marquer	21	emporter	17
admettre	28	élever	21	percer	17
disposer	28	déclarer	20	attraper	17
couper	28	attaquer	20	considérer	17
couvrir	28	retourner	20	garder	17
exposer	27	former	20	ravir	17
arranger	27	développer	20	éloigner	17
toucher	26	renverser	20	renvoyer	17
relever	26	unir	20	discerner	17
gagner	26	imaginer	20	assujettir	17
appuyer	26	apprécier	20	dire	17
découvrir	26	conclure	20	émouvoir	17
agiter	25	estimer	20	diminuer	17
détruire	25	tenir	20	extraire	17
dégager	25	combiner	20	introduire	17
attacher	24	accorder	20	bouleverser	17
abattre	24	instruire	20	corriger	17
indiquer	24	frapper	19	contraindre	17
saisir	24	représenter	19	exalter	17
observer	24	concevoir	19	évaluer	17
retenir	24	reprendre	19	ordonner	17
soustraire	24	avancer	19	étudier	17
écraser	24	recueillir	19	repousser	17
exciter	24	animer	19	rapprocher	17
percevoir	23	anéantir	19	placer	17
établir	23	exprimer	19	remuer	17
associer	22	donner	18	scruter	17
distinguer	22	mener	18	désigner	17
réunir	22	rejeter	18	tourmenter	17
troubler	22	rapporter	18		

Avec le seuil de fréquence choisi de 4, le même groupe de mots (*trépider, s'agiter, trembler, grelotter, trembloter, frissonner, frémir...*) s'oppose au groupe (*crier, gémir, clamer, mander, confirmer, hurler,*

*promettre, notifier, corroborer, certifier, témoigner, narrer, relater, proclamer, réciter...*) sur le premier axe, et au groupe (*démoraliser, exténué, ternir, lasser, massacrer, assombrir, abîmer, désespérer, contrister, navrer, noircir, désoler, déshonorer, corrompre, mâchurer, contaminer...*) sur le second axe.

En prenant un nouveau seuil de fréquence minimale de 5 (on élimine les verbes ayant moins de cinq occurrences dans le corpus), il reste 1 265 verbes distincts dans un corpus formé maintenant de 12 244 occurrences.

Avec ce nouveau seuil de cinq, c'est le groupe de mots (*désirer, rêver, vouloir, espérer, ambitionner, souhaiter, songer, prier, se représenter, supplier, implorer, demander...*) qui s'oppose à deux nouveaux groupes sur les deux premiers axes : le groupe – moins homogène – (*manger, désespérer, mâchurer, trembler, refroidir, décourager, lasser, déprimer, démanteler, abîmer, raser...*) sur le premier axe.

Le même groupe de mots s'oppose au groupe (*embrancher, agglutiner, corrélér, agréger, conjuguer, conjoindre, souder, jumeler, juxtaposer, fusionner, raccorder, connecter, moiser, annexer, marier, braser...*) sur le deuxième axe.

On pourrait encore incrémenter le seuil de fréquence pour montrer que l'on continue à obtenir des structures triangulaires (un groupe s'oppose à deux autres dans le premier plan factoriel engendré par les deux premiers axes), et, de surcroît, des structures instables, puisque les groupes caractérisant les axes principaux changent lorsque le seuil de fréquence minimale change.

En fait, l'analyse géométrique du nuage multidimensionnel de points-verbes montre que ce nuage est presque sphérique<sup>1</sup>. Cette quasi-sphère comporte à la périphérie des « grumeaux » qui sont des amas de verbes sémantiquement voisins. Ces « grumeaux » créent les axes principaux au gré de leur taille, qui dépend d'ailleurs du seuil de fréquence minimale choisi au départ. On est donc très loin de la structure elliptique du nuage de mots de la sémiométrie<sup>2</sup>, avec ses axes principaux stables et significativement distincts.

Pour des seuils de l'ordre de 12 à 18, c'est à dire pour des verbes apparaissant de douze à dix huit fois comme synonymes d'autres verbes de

---

1. Cette sphéricité se vérifie en particulier par la similitude des première *valeurs propres* dont les pourcentages de variance valent, pour les cinq premières : 0.77, 0.73, 0.69, 0.68, 0.66 (cas de l'analyse avec le seuil 4).

2. Rappelons que la signification statistique des axes sémiométriques a été éprouvée par *les intervalles de confiance d'Anderson*, des répétitions d'échantillons indépendants dans le temps et dans l'espace, et les méthodes de *bootstrap*, cf. chapitre 2.

la langue française<sup>1</sup>, on voit bien apparaître les groupements précités, mais les grandes oppositions observées sur les axes principaux sont décevantes, inattendues.

Ainsi, pour plusieurs de ces seuils, on voit apparaître, à l'extrémité d'un des axes, le groupe :

*[conjoindre, annexer, marier, accoupler, allier, relier, rattacher, rapprocher, rassembler, réunir, associer, assembler, rallier, lier, mélanger, joindre, enchaîner, mêler].*

Mais ce groupe de verbes relativement homogène et assez lié au pôle « Attachement » du troisième axe de la sémiométrie, s'oppose souvent au regroupement :

*[réclamer, demander, souhaiter, requérir, aspirer, solliciter, explorer, inspecter, sonder, scruter, fouiller, appeler, penser, estimer, considérer].*

Il ne se retrouve jamais opposé au regroupement fréquent pourtant plus antonymique que le précédent :

*[dégager, séparer, distinguer, discerner, abandonner, écarter, renvoyer, rejeter, éloigner, repousser].*

En fait, la notion de synonymie telle qu'elle est traitée dans les dictionnaires ou les thésaurus nous éloigne de l'esprit du questionnaire sémiométrique où il est demandé si les mots présentés évoquent des sensations agréables ou désagréables, et où le contexte même du questionnaire (liste de mots volontairement hétéroclite) et le temps imparti à sa complétion impliquent de prendre les mots suggérés dans un de leurs sens les plus courants. Au contraire, un dictionnaire de synonymes s'efforce d'être exhaustif et de ne pas omettre les emplois secondaires, métaphoriques, familiers ou argotiques<sup>2</sup>.

Illustrons, en marge de ces analyses statistiques, ces difficultés de la notion d'éloignement sémantique par une des réflexions de Gaston Bachelard, reprise par Roland Barthes dans ses *Mythologies*, pour qui, d'une certaine façon, « le vin est le contraire de l'eau... »<sup>3</sup>. La distance est ici plus axiologique que sémantique, et l'eau pourra être le contraire de la

1. Avec des seuils aussi élevés, il n'est pas question de voir apparaître les verbes rares absents de la liste des 829 verbes de base de notre corpus.

2. Le lecteur aura peut-être été surpris, lors de l'analyse précédente avec le seuil 4, de voir le verbe *manger* en compagnie de *démanteler*, de *raser*... c'est que *manger* comporte aussi comme synonyme *dilapider*, *dissiper*, *gaspiller*.

3. « Bachelard a déjà donné la psychanalyse substantielle de ce liquide (*le vin*) à la fin de son essai sur les rêveries de la volonté, montrant que le vin est suc de soleil et de terre, que son état de base est, non pas l'humide, mais le sec, et qu'à ce titre, la substance mythique qui lui est le plus contraire, c'est l'eau. » (Roland Barthes, *Mythologies*, Seuil, 1957).

sécheresse pour un agriculteur, du froid pour un esquimau, du solide pour un physicien, de la terre pour un marin, ou du vin pour le Français moyen de 1950 dont Barthes décrypte avec humour les valeurs et les mythes qui leurs sont associés.

Dans la même veine, un simple proverbe comme « le mieux est l'ennemi du bien » doit sa force lapidaire à la remise en question – légèrement paradoxale – de certaines proximités sémantiques avérées<sup>1</sup>.

A propos de l'expérience sur l'ensemble des verbes français et de sa grande ambition : découvrir dans le réseau sémantique de ces verbes une structure stable qui aurait un rapport avec celle observée en sémiométrie, on doit aussi constater, comme cela pouvait être pressenti, que les verbes porteurs de valeurs sont très peu fréquents dans l'ensemble des 829 verbes retenus.

Par nature, le verbe est lié à l'action, au mouvement, au temps, au changement. Les verbes techniques ou descriptifs abondent et en quelque sorte étouffent, dans le cas de notre corpus de synonymes, les verbes plus abstraits, affectifs ou relationnels (*aimer, juger, connaître, soigner, admirer, interdire...*).

Enfin, nous concluons par une remarque qui concerne à la fois les deux expériences précédentes : qu'il s'agisse de noms, d'adjectifs ou de verbes, le réseau sémantique n'est pas homogène<sup>2</sup>.

Certains mots ont beaucoup de synonymes, d'autres en ont peu, ou pas du tout. Un mot qui n'a pas de synonyme n'est pas moins important qu'un autre. C'est le cas par exemple des verbes *menacer* ou *voter*, mais c'est le cas aussi du mot *maternel*. Il faudra, dans une phase de recherche ultérieure, étendre les expériences précédentes en utilisant des corpus faisant appel, au delà de la synonymie, à des analogies, voire à des définitions de dictionnaire et des recueils de citations.

---

1. Beaucoup de clauses de style exploitent cette complexité topologique de l'espace sémantique, qui, pour les statisticiens, plus prosaïquement, est un espace que l'on ne peut pas munir facilement d'une distance euclidienne.

2. En termes plus techniques, mais plus précis, le graphe sémantique est loin d'être un graphe régulier, ses sommets (les verbes) ont des degrés très variables, certains sommets sont même isolés.

## CHAPITRE 4

### Le choix spontané des mots

Ce chapitre tente de répondre à la question suivante : peut-on réaliser une sorte de *sémiométrie ouverte*, ou de *sémiométrie spontanée* en demandant directement aux personnes interrogées de citer spontanément les mots qui leur paraissent agréables ou désagréables ?

Il ne s'agit donc plus maintenant de compiler des dictionnaires de synonymes comme au chapitre 3, mais de réaliser une enquête spécifique sur le terrain.

L'expérience a consisté à incorporer les deux *questions ouvertes* suivantes dans une enquête par sondage auprès de la population générale :

*« Sans qu'on puisse expliquer pourquoi, peut-être à cause de ce qu'ils évoquent, certains mots nous sont agréables, d'autres désagréables.*

- 1. En ce qui vous concerne personnellement, quels sont les mots qui vous sont le plus agréable ? (Citez le plus de mots possible).*
- 2. Quels sont maintenant les mots qui vous sont le plus désagréable ? (Citez le plus de mots possible) ».*

Le principe de ce type de questionnement n'est pas nouveau. Un précédent célèbre concerne Albert Camus, à qui on avait demandé quels

étaient ses dix mots préférés<sup>1</sup>. Les vocables *agréable* et *désagréable* ont été utilisés à dessein ici dans le libellé des questions pour rester le plus proche possible de l'esprit et de la formulation du questionnaire sémiométrique (mais une enquête similaire portant sur des préférences serait certainement tout aussi intéressante). Si l'exercice n'est pas nouveau dans son principe, sa réalisation dans le cas d'une enquête par sondage auprès d'un échantillon censé représenter la population générale est, à notre connaissance, une expérience nouvelle.

#### 4.1 Mise en place de l'expérience

On dispose des questionnaires de 1 191 répondants<sup>2</sup>. Les réponses correspondantes ont produit un « texte artificiel » long de 41 547 mots (*occurrences* de mots, en fait) à partir de 7 170 mots distincts cités.

Beaucoup de mots n'ont été cités qu'une fois, et donc ne jouent aucun rôle dans le calcul des distances entre répondants.

Parmi les 7 170 mots distincts, il y en a 1 466, soit 20.4 %, qui sont cités quatre fois ou plus. Si l'on se restreint au texte artificiel formé par ces 1 466 mots apparaissant au moins quatre fois, on obtient un texte de 33 950 occurrences, ce qui constitue 82 % du corpus initial (comme c'est le cas pour un texte courant, la distribution dissymétrique des fréquences d'utilisation des mots permet de reconstituer une part notable du texte avec un nombre restreint de mots).

Le tableau 4.1 donne, à titre d'exemple, les réponses des trois premiers répondants de l'enquête (les mots précédés d'un dièse (#) sont les mots cités comme désagréables<sup>3</sup>). On retrouve, certes, parmi les mots cités, des mots figurant dans le questionnaire sémiométrique, mais on trouve évidemment aussi beaucoup de mots «consensuels», qui ont été éliminés, par construction, de la liste des 210 mots.

---

1. Albert Camus, Carnets. Sa réponse fut : *le monde, la douleur, la terre, la mère, les hommes, le désert, l'honneur, la misère, l'été, la mer.*

2. L'enquête a été réalisée par la Sofres en 1995.

3. Certains mots, comme *feu, argent, cigarette, pluie, ordre*, peuvent figurer (pour des individus différents) avec ou sans dièse, c'est-à-dire être cités comme désagréables par les uns, et comme agréables par les autres. Il était donc nécessaire de donner une codification particulière à l'une des famille de mots. On a choisi d'accoler un symbole particulier (#) aux mots considérés comme désagréables.

### Tableau 4.1 : Exemples de réponses libres

#### Répondant 1:

doux mignon merveilleux charme plénitude enfant philosophie jardin plage animaux chaud finesse soleil soie nature joie liberté amie voyage gentillesse tendresse paisible caresse sérénité naturel sentiment amour musique vacances bonheur

#haine #méchanceté #assoiffé #vilain #débecter #désinvolture #serpent #nerf #squelette #graisse #arnaque #trahison #ballonner #chérubin #vomir #solitude #hypocrisie #hémoglobine #grosse #arrogance #ennemi #chagrin #déblatérer #évier #entartre #séquestrer #saleté #vicieux

#### Répondant 2:

printemps soleil santé joie bonheur enfant voyage vacances argent maison cérémonie restaurant promenade ami fleur jardin beauté compagnie bricoler tricot loisirs magasin fête visite découverte lecture mer montagne destinée

#hiver #froid #neige #verglas #maladie #pollution #secte #drogue #solitude #inactivité #laideur #mort #guerre #souffrance #pauvreté #misère #corruption #attente

#### Répondant 3:

amour fleur joie gentillesse tendre gaie aimable jolie souriant nature forêt montagne caresse beauté aide chocolat cerise lit sommeil pain vin femme enfant vie voyage famille ami

#méchanceté #tuerie #drame #hypocrite #égoïste #viol

Le tableau 4.2 donne la liste des mots apparaissant plus de 63 fois<sup>1</sup>. Les quatre mots les plus fréquemment cités sont absents du questionnaire sémiométrique. Il s'agit de trois mots cités comme agréables (*amour*, *soleil*, *vacances*), et d'un mot cité comme désagréable (*#maladie*).

Puis apparaissent les mots *#guerre*, *#mort*, *amitié*, *enfant*, *famille*, *fleur*, figurant tous les six (*Enfance*<sup>2</sup> pour *enfant*) dans le questionnaire sémiométrique. Les trois mots suivants : *bonheur*, *voyage*, *mer*, en sont absents (il y a bien *Eau*, *Océan*, et *Fleuve*, dans le questionnaire sémiométrique, mais pas : *mer*). Il y a évidemment beaucoup de redondances ou de relations d'implication dans les mots cités spontanément comme (*santé*, *#maladie*, *#cancer*, *#sida*) ou encore (*amitié*, *ami*), (*#mort*, *#décès*), (*#vol*, *#voleur*), (*#crime*, *#meurtre*), (*amour*, *aimer*)<sup>3</sup>.

1. Ce seuil très élevé est encore ici choisi pour des raisons d'encombrement. Le tableau est disponible dans son intégralité auprès des auteurs.

2. Rappelons que les mots du questionnaire sémiométrique sont écrits, par convention, en italique avec majuscule.

3. Les mots préférés d'Albert Camus apparaissent dans cet ordre de fréquence : *mer* (286), *#douleur* (104), *#misère* (99), *été* (91), *mère* (22), [*#mère* (1)], *honneur* (16), *terre* (14), *désert* (11) [*#désert* (6)], *homme* (et non *les hommes*) (6) [*#homme* (2)], *monde* (1) [*#monde* (1)]. Ce rapprochement avec les mots de Camus est certes anecdotique, mais il a cependant le mérite de

**Tableau 4.2 : Mots les plus fréquents cités spontanément**

<i>mots</i>	<i>freq</i>	<i>mots</i>	<i>freq</i>	<i>mots</i>	<i>freq</i>
amour	731	#pluie	135	forêt	88
soleil	628	fleurs	135	#bruit	87
vacances	498	#drogue	133	#prison	86
#maladie	475	#viol	131	animaux	85
#guerre	439	printemps	129	manger	83
#mort	421	gentillesse	129	#peur	83
amitié	379	jardin	124	livre	82
enfant	366	vie	123	#politique	82
famille	358	ami	122	#jalousie	82
fleur	328	promenade	121	lecture	80
bonheur	305	#malheur	119	eau	80
voyage	292	chaleur	116	bonjour	80
mer	286	naissance	115	tolérance	78
musique	277	#sida	113	#tristesse	78
#chômage	265	#solitude	107	#secte	77
#accident	263	#noir	106	ciel	76
joie	258	calme	104	#pédophilie	75
nature	209	#douleur	104	#décès	75
#violence	191	enfants	102	maman	72
santé	189	plaisir	101	lumière	71
tendresse	184	#pauvreté	100	plage	71
beauté	180	chocolat	100	#vol	70
#méchanceté	173	mariage	99	loisirs	69
argent	171	#misère	99	beau	69
#mensonge	167	#pollution	96	gentil	68
liberté	167	repos	96	parfum	68
#racisme	165	campagne	96	arbre	67
#haine	165	#impôt	95	cinéma	67
paix	156	bleu	94	restaurant	67
montagne	156	sport	93	chat	67
#cancer	156	#hôpital	93	honnêteté	65
rire	153	#divorce	91	merci	64
douceur	151	#égoïsme	91	#faim	64
#froid	148	été	91	#saleté	64
maison	147	#hypocrisie	90	#méchant	64
sourire	145	#injustice	90	dormir	64
fête	143	#famine	88	partage	63
bébé	141	#souffrance	88	#voleur	63
travail	137	#intolérance	88	aimer	63

On constate aussi que les substantifs constituent une écrasante majorité des mots cités ; les verbes et les adjectifs apparaissent dans cette liste

---

bien faire ressortir la vraie nature de notre questionnaire : il ne s'agit pas de thèmes de préoccupation, ni de concepts, ni de mots-clés renvoyant à des problèmes fondamentaux, mais simplement de sensations, agréables ou désagréables.

d'abord sous des formes ambiguës ; *rire* (153) et *sourire* (145), sont probablement cités la plupart du temps comme des substantifs et non comme des verbes, compte tenu de leur contexte – nous verrons que les catégories grammaticales apparaissent souvent par séquences dans une même réponse – de même que *ami* (122), plus souvent substantif qu'adjectif. Les trois premiers adjectifs cités seraient alors *#noir* (106), *bleu* (94), *beau* (69) (ce sont d'ailleurs aussi des substantifs), et les premiers verbes cités *manger* (83), *dormir* (64), *aimer* (63).

On note également des thèmes plus conjoncturels ou liés à l'actualité immédiate (*#pédophilie*, *#impôts* (95), *#drogue* (133), *#attentat* (59), *#secte* (77)) qui avaient été exclus *a priori* du questionnaire sémiométrique.

Une des constatations que l'on peut faire en considérant l'ensemble du corpus des réponses originales, et non seulement les mots les plus fréquents, est l'absence totale de citation spontanée de certains mots du questionnaire sémiométrique fermé.

Pour cet échantillon de 1 191 répondants, 173 mots (sur les 210 que compte le questionnaire sémiométrique) sont cités spontanément comme agréables ou désagréables. Si l'on ne retient que les 600 premiers répondants, seulement la moitié des mots du questionnaire sémiométrique sont cités spontanément par les personnes interrogées (certains d'entre eux peuvent cependant apparaître plus de 200 fois – ce qui est le cas de : *fleur*, *guerre*, *mort*, *amitié*).<sup>1</sup>

**Tableau 4.3 :**  
**Les 37 mots du questionnaire sémiométrique non cités spontanément**

<i>absolu</i>	<i>conquérir</i>	<i>humble</i>	<i>modération</i>	<i>réfléchir</i>
<i>acharnement</i>	<i>défi</i>	<i>interroger</i>	<i>muraille</i>	<i>règle</i>
<i>armure</i>	<i>détachement</i>	<i>inventeur</i>	<i>noeud</i>	<i>rigide</i>
<i>astucieux</i>	<i>élite</i>	<i>labyrinthe</i>	<i>or</i>	<i>robuste</i>
<i>attachement</i>	<i>enseigner</i>	<i>logique</i>	<i>produire</i>	<i>sacré</i>
<i>bâtisseur</i>	<i>escalader</i>	<i>magie</i>	<i>question</i>	<i>utilitaire</i>
<i>cérémonie</i>	<i>féconder</i>	<i>masque</i>	<i>recueillement</i>	<i>viril</i>
<i>concret</i>	<i>fermeté</i>			

Cette expérience de citation spontanée de mots dans une enquête représentative est probablement assez originale, et l'« aptitude à être cité

1. Bien entendu, si l'on accroissait la taille de l'échantillon, on recueillerait de plus en plus de mots distincts, et l'on arriverait à obtenir la liste exhaustive des mots du questionnaire. Toutefois, comme c'est le cas pour tout corpus de texte, le nombre de mots distincts du recueil de réponses augmenterait beaucoup moins vite que le nombre total de leurs occurrences.

spontanément » de certains mots n'a pas été étudiée, à notre connaissance, de façon systématique. Il s'agit, au moins en partie, de mots très fréquents dans la langue courante, encore qu'il n'existe pas de corpus écrits permettant d'estimer ces fréquences. Ce critère de fréquence extrême n'a pas été pris en considération pour composer la liste sémiométrique.

Il y a dans les réponses une concentration des fréquences de mots consensuels, et aussi une extrême dispersion sur les mots relativement rares ou idiosyncratiques (par exemple : *coquecigrue*, *chèvrefeuille*, *abdomen*, *Popocatepelt*, *clafoutis*, *candélabre*) tenant parfois au caractère ludique de l'exercice de complétion d'un tel questionnaire.

La probabilité de trouver des mots assez neutres (comme les mots du questionnaire sémiométrique : *enseigner*, *interroger*, *question*, *utilitaire*) est alors très faible. Le nombre de mots qui n'apparaissent qu'une fois (ou : *hapax*) dans ces réponses libres est considérable (4 266, pour 7 170 mots distincts)<sup>1</sup>.

Le caractère ouvert des questions favorise la citation de synonymes, de voisins sémantiques (n'appartenant pas forcément à la liste sémiométrique) ou de flexions d'un même mot (singulier – pluriel, présent – infinitif, masculin – féminin, etc.). Il y a dispersion de la fréquence sur les synonymes et les flexions. Le synonyme le plus fréquent cache alors les autres. La fréquence du premier est exagérément augmentée, au détriment de celle des derniers<sup>2</sup>.

Les sections 4.2 et 4.3 qui suivent sont dévolues à une première exploration du corpus de réponses. Puis la section 4.5 utilisera le fait qu'une petite partie des répondants avait répondu au questionnaire sémiométrique au cours de l'année précédente pour tenter un rapprochement entre les deux types de questionnaires.

---

1. Sur les problèmes de distributions lexicales, cf. Muller (1977, 1979) ; il existe par ailleurs une modélisation statistique des événements rares appliquée au domaine lexicométrique, cf. Efron and Thisted (1976), Baayen (2000).

2. Le mot *Armure* présent dans le questionnaire sémiométrique n'est jamais cité spontanément dans cet échantillon ; mais sont citées les formes graphiques suivantes de même racine : *#arme* (49), *#armement* (1), *#armes* (12), *#armé* (1), *#armée* (11). De même ne sont jamais cités spontanément les mots (absents aussi du questionnaire sémiométrique) *cuirasse*, *bouclier*, *épée*, *sabre*, *glaive*, qui, comme *Armure* ou *Muraille*, évoquent plus l'histoire ou la fiction qu'un danger ou un débat actuel. En revanche, on relève : *#fusil* (14), *#bombe* (34), *#bombes* (2), *#canon* (6), *#bombardement* (3), *#mitrailleuse* (1), *#mitrailleuse* (1), *#missile* (1).

## 4.2 Premières explorations des réponses

Les analyses qui suivent permettent de mieux comprendre la nature des réponses à ces questions ouvertes.

Comme dans les sections précédentes, les tableaux croisant les mots et les réponses sont décrits en utilisant la technique d'analyse des correspondances, brièvement présentées en annexe A1.4. L'analyse est effectuée dans un premier temps sur les mots cités comme agréables. Le premier des tableaux analysés correspond au seuil de fréquence minimale quatre, c'est-à-dire aux mots cités plus de quatre fois. Il a 1 191 lignes (répondants) et 592 colonnes (mots).

Le tableau 4.4 montre les mots caractéristiques du premier axe principal (ou factoriel) de ce tableau à 1 191 lignes et 592 colonnes.

**Tableau 4.4 : Mots extrêmes sur le premier axe  
(Analyse des réponses libres sur les mots agréables)**

Partie gauche		Partie droite	
<i>dessin</i>	-.48	<i>polie</i>	9.23
<i>copains</i>	-.46	<i>souriante</i>	8.60
<i>orage</i>	-.44	<i>dynamique</i>	7.81
<i>horizon</i>	-.44	<i>courageux</i>	7.03
<i>concert</i>	-.43	<i>serviable</i>	6.58
<i>comète</i>	-.41	<i>adorable</i>	5.35
<i>lait</i>	-.40	<i>gentille</i>	5.23
<i>radio</i>	-.40	<i>propre</i>	4.78
<i>plante</i>	-.39	<i>sincère</i>	4.58
<i>feu-de-bois</i>	-.39	<i>poli</i>	4.55
<i>fille</i>	-.39	<i>juste</i>	4.54
<i>train</i>	-.39	<i>bonsoir</i>	4.41
<i>farniente</i>	-.39	<i>sociable</i>	4.34

Fait remarquable, les mots de la colonne de droite (coordonnées positives) sont beaucoup plus éloignés de l'origine (près de vingt fois pour les premiers) que ceux de la colonne de gauche (coordonnées négatives). D'autre part, les mots de la colonne de droite sont assez homogènes sémantiquement, alors que ceux de la colonne de gauche auraient pu inspirer Jacques Prévert.

L'explication du phénomène est simple, et rejoint les constatations faites au cours des analyses du chapitre précédent : la grosse masse des points-mots est concentrée autour de l'origine, et un petit sous-groupe (une vingtaine de mots sur 592), parce qu'il est très éloigné et formé de mots corrélés, est seul responsable d'un axe.

On note, que, sur la droite de l'axe, il n'y a pratiquement que des *adjectifs*, s'opposant, à gauche, à la masse des substantifs, renvoyant souvent à des objets, des substances, des concepts matériels.

Les axes suivants mettent en exergue des petits groupes de mots, sans qu'aucun axe ait une interprétation bipolaire. Il est clair, ici encore, qu'il existe des regroupements privilégiés, au niveau local, mais pas de structuration globale de ces regroupements.

Pour éprouver alors l'hypothèse selon laquelle la complexité structurelle observée serait produite par les mots peu fréquents, les analyses ont été répétées en augmentant le seuil de fréquence minimale.

En ne retenant que les 158 mots apparaissant plus de 25 fois dans le corpus des réponses à la première question ouverte (mots agréables), on améliore sensiblement la structure globale : les deux premiers axes restent définis par très peu de mots, mais ceux-ci ont le mérite d'apparaître au moins 25 fois, et donc les résultats obtenus sont plus significatifs du point de vue statistique.

Devant l'insuffisance des méthodes factorielles pour décrire les structures de ce type<sup>1</sup>, on a de nouveau eu recours aux *cartes de Kohonen*, qui, rappelons-le, fournissent une tentative de synthèse entre une procédure de classification et un plan factoriel (cf. annexe A1.8).

La figure 4.1 présente une telle carte, dite « carte auto-organisée de Kohonen ». Il s'agit donc d'une description graphique des associations entre mots apparaissant plus de 25 fois dans l'ensemble des réponses. Deux mots appartenant à une même case sont souvent cités simultanément par la même personne. Ceci reste encore vrai, mais dans une moindre mesure, s'ils sont situés dans des cases voisines. Si les cases sont très éloignées, les mots correspondants sont, au contraire, rarement cités simultanément.

La première constatation qui peut être faite au vu de ce graphique a trait à une faiblesse du questionnaire spontané : le rôle des associations d'idées au moment du choix des mots. On observe de telles associations à la simple lecture des réponses brutes, mais la figure 4.1 confirme qu'il ne s'agit pas d'un phénomène accidentel<sup>2</sup>.

---

1. Il s'agit dans ce cas du manque de lisibilité imputables à la surcharge des graphiques (nombreux points superposés). Les analyses sont néanmoins intéressantes.

2. Ce phénomène est encore plus visible avec un seuil plus bas, donc des mots moins fréquents, mais les cartes de Kohonen correspondantes sont alors trop encombrantes pour une publication de format normal.

égalité justice fraternité franchise	respect intelligence honnêteté bonté	politesse gentillesse entente affection	harmonie		joli beau		pardon merci bonjour
solidarité sincérité	générosité fidélité compréhensio	tendresse	sérénité douceur calme		bon		
tolérance humour courage	partage paix liberté confiance	nature espoir confort	vivre tranquillité plaisir	île sieste manger femme dormir			heureux aimer
jeunesse gaieté convivialité	joie beauté amitié	vie rire détente amour	repos chaleur	rêve lit	étoile oiseau eau couleur	vert caresse bleu	doux câlin
bien-être	réussite passion chance bonheur	fête cadeau	sourire loisir enfant chanson		sable lumière livre fleur ciel	parfum baiser	rose papa maman bébé
	propreté	naissance mariage jeux argent anniversaire	été vacances	voyage sport soleil repas rencontre printemps plage musique	mer campagne	chat ami	gâteau chocolat Noël
gentil aimable		travail santé cadeaux	week-end loisirs famille	voiture promenade maison champagne animaux	danse	neige chien	théâtre peinture
agréable		voyages petits-enfan enfants amis	fleurs	télévision restaurant	vélo randonnée montagne lecture cinéma	pluie oiseaux jardin cuisine bateau	vin verdure fruit forêt chant arbre

**Figure 4.1 : Carte de Kohonen représentant les associations entre mots dans les réponses libres**

Des regroupements comme (*été, vacances*), (*papa, maman*), (*gâteau, chocolat*), (*pardon, merci, bonjour*) peuvent en effet être l'indice d'un remplissage un peu automatique du questionnaire.

Voici trois exemples de réponses réelles pour lesquelles ce type d'association est assez clair :

- « *aimable gentillesse agréable utile merci bonjour pardon excusez-moi obéissant* »
- « *bonjour bonsoir merci pardon excusez-moi* »
- « *bonjour merci s'il vous plaît pardon* »

On trouvera également ci-dessous une réponse longue qui montre comment les associations<sup>1</sup> se sont progressivement formées au fil de la réponse :

- « *paysage campagne amitié bonheur intelligence sensibilité mesure égard vieillard bébé femme compagne mère père enfant grand-mère grand-père oncle tante cousin herbe arbre rivière ruisseau océan lune étoile aube crépuscule innocence* ».

Il est clair que le questionnaire fermé a le mérite d'éviter ces dérives et les surpondérations accidentelles de certains thèmes. Cette tendance à l'association des mots énoncés consécutivement concerne aussi la catégorie grammaticale des mots : on a vu que des adjectifs s'opposaient à des noms sur le premier axe (tableau 4.4). Une liste qui commence par un adjectif (ou un nom, ou un verbe) aura tendance à se poursuivre avec des adjectifs (ou des noms, ou des verbes).

### 4.3 Les axes principaux des réponses spontanées

En ne retenant que les mots apparaissant plus de 25 fois, on commence à apercevoir des dimensions stables qui se rapprochent de certains axes sémiométriques.

Le plan principal – plan (1, 2) – de l'analyse réalisée avec ce seuil de fréquence (analyse non représentée ici) oppose la quasi-totalité des mots à deux petits conglomerats « *merci, pardon, bonjour* » d'une part, et « *agréable, gentil, aimable* » d'autre part (structure triangulaire déjà évoquée au début de ce chapitre). Ce phénomène est bien visible sur la carte de Kohonen (figure 4.1) qui concerne d'ailleurs le même tableau (individus – mots). En effet, on trouve bien isolées dans le coin supérieur droit de la figure 4.1 les trois formules de politesse précédentes, entourées de cases vides, et les autres mots évoqués (« *agréable, gentil, aimable* ») dans le coin inférieur gauche de la figure<sup>2</sup>.

---

1. Les mécanismes d'associations de mots ont été étudiés, dans d'autres contextes, par les psychologues ; cf. notamment : Ferrand et Alario (1998).

2. Ceci permet de montrer en passant que les cartes de Kohonen ont un grand pouvoir de compression, puisqu'on y lit des informations relatives à plusieurs axes simultanément.



La figure 4.2 montre la disposition des mots dans le plan (3, 4), plan engendré par les troisième et quatrième axes. On note dans ce plan beaucoup de mots évoquant des loisirs et des plaisirs sur la partie gauche de l'axe horizontal, (on verrait d'ailleurs le mot « #travail », c'est-à-dire le rejet du travail, dans cette zone). Sur la partie droite, on peut lire des mots tels que « *respect, politesse, égalité, justice* »... Voilà une opposition « Devoir / Plaisir » qui n'est pas sans rappeler l'axe 2 de la structure sémiométrique... obtenu ici sans faire appel à une liste de mots. On verrait de même que l'axe 4 (axe vertical de la figure 4.2) n'est pas sans rapport avec l'axe 3 de la structure sémiométrique.

En conclusion de cette première exploration des réponses aux deux questions ouvertes, on doit relever le caractère extrêmement *bruité* des données recueillies de cette façon, par comparaison à un recueil de notes attribuées à une liste de mots identiques pour toutes les personnes interrogées. La distance entre individus va dépendre en fait du petit nombre de mots qu'ils peuvent avoir en commun, et les individus n'ayant aucun mot en commun auront des distances indifférenciées. L'existence de ce bruit considérable n'exclut pas de trouver des traits structuraux, mais la taille actuelle de l'échantillon limite la portée du travail sur données individuelles.

On observe cependant des convergences qui laissent penser qu'une structure apparentée à la structure sémiométrique peut apparaître spontanément et indépendamment de tout questionnaire fermé.

#### 4.4 Choix spontané et caractéristiques des répondants

Une technique utilisée fréquemment, dans le cadre du traitement statistique des réponses aux questions ouvertes, consiste à analyser non pas les réponses elles-mêmes, mais des tableaux de réponses regroupées, qui sont moins sensibles aux importantes fluctuations individuelles. Les réponses sont en effet très « bruitées », mais on peut espérer trouver par cette technique des régularités indécélables directement à partir des réponses non agrégées.

Des regroupements *a priori* des répondants sont réalisés à partir de quelques unes de leurs caractéristiques disponibles. Ce sera l'occasion de voir que le sexe et l'âge, considérés isolément, ou mieux encore simultanément, ne sont pas indépendants des mots cités comme agréables ou désagréables.

La première variable de base servant au regroupement des réponses sera l'âge, pour lequel sont retenues deux catégories extrêmes : les personnes âgées de moins de 30 ans et celles âgées de plus de 55 ans (tableau 4.5). Seuls sont pris en compte les mots agréables apparaissant au moins 16 fois.

Les mots caractéristiques<sup>1</sup> des jeunes (*plaisir, manger, dormir, ...*) et des personnes plus âgées (*politesse, courage, fraternité, ...*) ne sont pas sans rappeler les deux extrémités de l'axe horizontal de la figure 4.2, axe assez similaire à l'axe deux de la sémiométrie.

**Tableau 4.5 : Mots caractéristiques des âges extrêmes**

Mots spontanés caractéristiques	Valeurs-test	Probabilités
<i>Moins de 30 ans</i>		
1 plaisir	4.00	.000
2 manger	3.78	.000
3 dormir	3.48	.000
4 câlin	2.63	.004
5 bébé	2.37	.009
6 repos	2.29	.011
7 chocolat	2.22	.013
<i>Plus de 55 ans</i>		
1 politesse	3.38	.000
2 courage	3.00	.001
3 fraternité	2.77	.003
4 voyage	2.74	.003
5 merci	2.42	.008
6 lecture	2.38	.009
7 affection	2.37	.009
8 propreté	2.37	.009
9 pardon	2.22	.013

La seconde variable de base sera le sexe. Les mots les plus caractéristiques des hommes et des femmes figurent dans le tableau 4.6. Les

1. Ce sont les mots anormalement fréquents dans la catégorie par rapport à leur fréquence moyenne dans l'ensemble de l'échantillon. La différence entre la fréquence interne à la catégorie et la fréquence globale est convertie en valeur-test, c'est à dire en variable normale centrée réduite dans l'hypothèse d'indépendance des fréquences (cf. annexe A1.9.1). La dernière colonne du tableau 4.5 donne une information équivalente en termes de probabilités de dépassement (information moins précise mais peut-être plus suggestive).

valeurs-test, moins élevées, montrent que le sexe est moins discriminant que l'âge, lequel introduit une grande hétérogénéité à l'intérieur même de chacun des deux groupes hommes et femmes.

**Tableau 4.6 : Mots caractéristiques des hommes et des femmes**

Mots spontanés caractéristiques	Valeurs-test	Probabilités
<i>Hommes</i>		
1 promenade	2.82	.002
2 joie	2.41	.008
3 courage	2.30	.011
4 sport	2.25	.012
<i>Femmes</i>		
1 maman	3.32	.000
2 chocolat	2.53	.006
3 agréable	2.20	.014
4 livre	2.09	.018

Quatre nouvelles catégories de répondants sont maintenant obtenues en croisant le sexe avec deux classes d'âge : hommes de moins de 30 ans, hommes de plus de 55 ans, femmes de moins de 30 ans, femmes de plus de 55 ans (tableau 4.7).

Les tableaux 4.6 et 4.7 ont alors peu de mots en commun : le croisement âge-sexe est une variable plus pertinente que le sexe seul. Parmi les mots communs aux deux tableaux, on note : *courage* pour les hommes, *maman* et *chocolat* pour les femmes.

En revanche, les mots caractéristiques des jeunes et des personnes plus âgées dans le tableau 4.5 caractérisaient en fait, soit les hommes, soit les femmes, mais pas les deux. Ainsi, parmi les mots caractéristiques des jeunes, les mots *plaisir*, *dormir*, *manger*, caractérisent surtout les hommes, et les mots *câlin*, *bébé*, *chocolat*, les femmes. Il était donc important de croiser les deux variables sexe et âge.

La taille de l'échantillon ne permet cependant pas de prendre en compte des croisements plus élaborés.

La figure 4.3 représente, sous forme de proximités graphiques, une synthèse des liens existant entre les six catégories (deux catégories proches

ont des profils lexicaux communs) et entre les mots (deux mots proches ont des profils socio-démographiques similaires)<sup>1</sup>.

**Tableau 4.7 : Mots caractéristiques de quatre catégories sexe-âge**

<b>Mots spontanés caractéristiques</b>	<b>Valeurs-test</b>	<b>Probabilités</b>
<i>Homme moins de 30 ans</i>		
1 dormir	3.24	.001
2 plaisir	3.03	.001
3 manger	2.84	.002
4 loisir	2.29	.011
<i>Homme plus de 55 ans</i>		
1 courage	3.59	.000
2 fraternité	2.80	.003
3 propreté	2.68	.004
4 santé	2.24	.012
<i>Femme moins de 30 ans</i>		
1 chocolat	3.06	.001
2 bébé	3.02	.001
3 animaux	2.46	.007
4 maman	2.26	.012
5 câlin	2.13	.016
6 été	2.12	.017
<i>Femme plus de 50 ans</i>		
1 merci	3.00	.001
2 affection	2.83	.002
3 politesse	2.53	.006
4 bonjour	2.14	.016

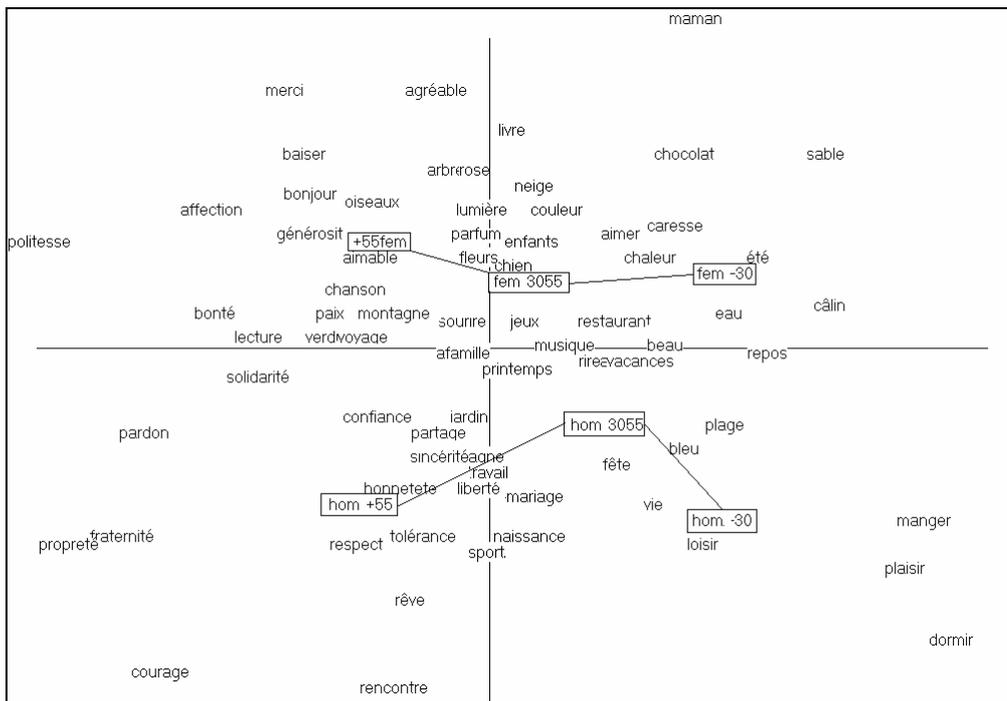
La représentation obtenue confirme la complémentarité, on pourrait presque dire l'additivité des effets de ces deux variables de base.

En effet, ces deux variables se déploient selon des directions orthogonales, l'âge, horizontalement, opposant les catégories les plus âgées à gauche, aux plus jeunes à droite ; le sexe, selon une direction verticale,

---

1. Il s'agit du premier plan principal d'une analyse des correspondances de la table de contingence (dite table lexicale) dont les lignes sont les mots apparaissant au moins 16 fois, et les colonnes les six catégories sexe-âge, incluant, cette fois, la classe d'âge intermédiaire. L'outil de description est le même que celui de la figure 4.2, mais il s'applique maintenant à des données agrégées, et non plus aux données individuelles.

opposant les hommes, en bas, aux femmes, en haut. Fait remarquable, et qui a pu échapper aux lecteurs peu familiers du fonctionnement de ces méthodes de description : rien, dans les six colonnes du tableau d'entrée soumis à l'analyse, n'indique qu'elles proviennent du croisement de deux variables. La seule information qui décide de la position des points-colonnes sur le graphique est le profil lexical de ces colonnes, donc l'ensemble des mots cités spontanément comme agréables par les catégories correspondant à ces colonnes.



**Figure 4.3 : Citation spontanée des mots selon le sexe et l'âge (plan des deux premiers axes).**

Finalement, cette figure est une synthèse des résultats lisibles sur les tableaux 4.5, 4.6 et 4.7. Elle décrit bien, dans le cadre d'un continuum nuancé, les oppositions entre sexes pour une classe d'âge donnée, et l'étendue des variations lexicales internes à chaque sexe, en fonction de l'âge<sup>1</sup>.

Il reste à apporter les éléments de validation que nous fournissaient les valeurs-test dans le cas des tableaux. Ce sont les *ellipses de confiance*

1. Rappelons que les axes 2 et 3 de la sémiométrie opposent respectivement, de la même manière, les hommes aux femmes et les jeunes aux personnes plus âgées.



répondants, on prendra note de la richesse et de la cohérence du matériau recueilli en laissant une liberté totale de réponse aux personnes interrogées. Les réponses ont beau être bruitées, bigarrées pourrait-on dire pour certaines, leur regroupement fait apparaître des régularités qui sont un indice des potentialités de ce type de recueil.

#### 4.5 Rapprochement entre sémiométrie et questions ouvertes

Nous avons pu disposer des notes sémiométriques pour 335 personnes parmi les 1 191 répondants aux deux questions ouvertes invitant à citer spontanément des mots agréables ou désagréables<sup>1</sup>.

La structure sémiométrique de ce groupe<sup>2</sup>, pourtant restreint, de 335 individus est bien conforme, au moins en ce qui concerne les quatre premiers axes, à celle présentée au chapitre I. Cela ne surprendrait pas outre mesure si ce groupe avait été pris au hasard dans l'échantillon du panel. Or il s'agit de personnes ayant accepté de répondre au questionnaire ouvert, et l'on aurait pu craindre qu'ils ne constituent un échantillon biaisé. Apparemment, il n'en est rien.

Cela nous autorise à répondre à la question suivante : quels sont les mots cités spontanément qui caractérisent les premiers axes sémiométriques. Les mots cités spontanément vont être considérés comme des variables nominales supplémentaires (au même titre que le sexe ou l'âge) et seront donc projetés sur les axes principaux (*cf.* annexe A1.9.3).

On a vu que, sur la totalité du corpus des réponses aux questions ouvertes, la moitié environ des mots du questionnaire sémiométrique apparaissent au moins une fois. Mais une étude statistique de rapprochement entre les deux questionnaires demande de ne retenir que les mots cités avec une certaine fréquence. Le rapprochement sera alors parfois surprenant car peu de mots cités spontanément appartiendront au questionnaire sémiométrique.

Alors que la position des mots de la sémiométrie sur les axes est caractérisée par leurs coefficients de corrélation avec les axes, la position des mots cités spontanément projetés comme des catégories supplémentaires sur les mêmes axes est caractérisée par les valeurs-test qui

---

1. Ces 335 personnes appartiennent au panel d'où a été tiré l'échantillon de 1 191 personnes qui avaient répondu au questionnaire ouvert.

2. Structure issue de l'analyse en composantes principales du tableau de note (335 x 210).

prennent en compte l'effectif concerné (nombre de personnes ayant cité le mot) en convertissant la coordonnée sur l'axe en une variable normale centrée réduite<sup>1</sup>.

– *Le premier axe*

Le premier axe qui est un axe méthodologique de participation à l'enquête et que nous considérons comme extérieur aux structures sémiométriques fait l'objet d'un traitement particulier (chapitre 5).

La position des mots sur ce premier axe sera fort intéressante d'un point de vue méthodologique. Cet axe oppose, en effet, les individus utilisant pleinement l'échelle proposée pour les notes à des individus qui n'utilisent que la partie centrale de l'échelle.

Les mots les plus caractéristiques des individus qui utilisent pleinement l'échelle des notes sont, pour l'analyse sémiométrique des 335 individus : *Courage, Politesse, Héros, Honneur, Protéger, Robuste, Tradition, Dynamique, Raffiné, Élégance, Honnête...*, avec des coefficients de corrélation avec l'axe variant de  $-0.59$  à  $-0.49$ .

Les mots caractérisant ceux qui n'utilisent que la partie centrale de l'échelle sont : *Trahir, Angoisse, Révolte, Faute, Danger, Désordre, Mort...*, dont les coefficients de corrélation avec l'axe, nettement plus faibles, varient de  $0.29$  à  $0.17$ .

Les mots *cités spontanément* caractérisant les individus qui utilisent pleinement l'échelle des notes sont, suivis de leurs valeurs-test entre parenthèses, *confiance* ( $-2.9$ ), *aimer* ( $-2.8$ ), *bonjour* ( $-2.7$ ), *merci* ( $-2.4$ ), *courtoisie* ( $-2.1$ ), *honnête* ( $-2.1$ ). On retrouve bien les notions de politesse et d'honnêteté. Notons également, pour les mêmes individus, les mots considérés comme les plus désagréables : *#inceste* ( $-2.2$ ), *#assassin* ( $-2.1$ ), *#pédophile* ( $-2.1$ ). On ne pouvait évidemment voir apparaître sur cet axe les mots du questionnaire sémiométrique : *Héros, Protéger, Robuste, Tradition, Raffiné, Élégance*, qui ne sont jamais cités spontanément comme étant des mots agréables.

On a vu qu'il n'existait pas, à partir du questionnaire sémiométrique, de mots caractérisant fortement les répondants occupant la partie positive de l'axe, c'est-à-dire les répondants utilisant la partie centrale de l'échelle. L'hypothèse a été faite qu'il s'agit de répondants peu enclins à se livrer, ou encore « ne jouant pas pleinement le jeu du questionnaire ». Ces répondants

---

1. Rappelons que la valeur de celle-ci sera approximativement comprise entre  $-2$  et  $+2$  si les personnes ayant cité le mot sont réparties aléatoirement sur l'axe concerné. (Voir annexe A1.9.1.)

posent des problèmes dans toutes les procédures d'enquêtes par sondage. Comme les abstentionnistes, dont ils constituent la frange encore observable, ils sont le point faible ou la zone d'ombre de cet instrument d'observation<sup>1</sup>.

On ne peut s'attendre, non plus, à voir citer comme agréables les mots [*Trahir, Angoisse, Révolte, Faute, Danger, Désordre, Mort...*]. De façon d'ailleurs très parlante, le mot le plus caractéristique de ces répondants est *#contrainte* (considéré donc ici comme un mot particulièrement désagréable, avec une valeur-test de 3.7, qui est la plus élevée de l'ensemble des mots sur l'axe). Puis viennent les mots *vie* (2.7), *pluie* (2.5), *océan* (2.5), *intelligence* (2.3), *femme* (2.2), et les mots désagréables (outre *#contrainte* déjà cité) *#tristesse* (2.4), *#araignée* (2.3), *#raciste* (2.3), *#travail* (2.2).

Il s'agit par ailleurs de personnes plus jeunes, plus instruites, et ce sont le plus souvent des hommes. On apprend maintenant qu'il s'agit avant tout de personnes qui rejettent les contraintes (traduction du fait statistique : qui sont caractérisées de façon très significatives par la citation spontanée du mot *contrainte* comme mot désagréable). Le rejet du mot *travail* (et probablement des contraintes qu'il représente) va dans le même sens. Les mots *vie*, *pluie*, *océan*, *intelligence* occupent également des positions significatives (à plus de 2.3 écart-types de ce que seraient leurs positions dans l'hypothèse d'une répartition aléatoire des réponses) mais l'interprétation est plus délicate. Le mot *pluie* est intéressant car il figure beaucoup plus comme mot désagréable (37 citations dans le sous-échantillon utilisé ici) que comme mot agréable (seulement 5 citations).

Mais la position des personnes qui ont choisi *pluie* comme mot agréable est suffisamment typée sur ce premier axe pour que leur point moyen soit significatif statistiquement. On peut regretter que la taille de l'échantillon de cette expérience ne permette pas de mieux caractériser ces répondants, qui semblent être des personnes moins conventionnelles, plus individualistes, mais qui acceptent pourtant de répondre !

---

1. On désigne parfois, non sans malaise, ces différences d'attitudes sous le nom d'effet *notation*. On ne peut exclure un tel effet, mais le nommer ne résout pas tous les problèmes méthodologiques posés par l'attitude de la personne interrogée par rapport au questionnaire et à l'enquête (cf. chapitre 5).

– *Les axes sémiométriques*

Le tableau 4.8 donne les mots cités spontanément les plus caractéristiques des axes 2 à 4. Rappelons qu'il s'agit des axes issus de l'analyse en composantes principales des 210 mots de la sémiométrie, les mots cités spontanément y étant projetés *a posteriori*.

Les axes 5 et 6 n'étant pas bien reconstitués sur notre petit échantillon de 335 répondants, ils ne seront pas analysés ici. Les zones ombragées correspondent à des valeurs-test inférieures à 2. Elles ne seront pas interprétées, bien qu'elles fournissent un vivier de mots qui seraient peut-être candidats à être beaucoup plus significatifs statistiquement si l'échantillon était plus grand.

Les mots en gras ont des valeurs-test supérieures ou égales à 2.3, et occupent donc des positions qui méritent d'être interprétées.

– *L'axe 2*

La première colonne de ce tableau concerne l'axe 2, conventionnellement désigné par axe « Devoir / Plaisir » dont on rappelle qu'il oppose les mots : *Discipline, Obéir, Patrie, Morale, Soldat, Economiser, Industrie, Prêtre, Règle...* aux mots : *Sensuel, Rêver, Aventurier, Original, Ile, Nudité, Sauvage...*].

Les mots caractérisant la partie « Devoir » imposés dans le questionnaire sémiométrique, ne sont pas cités spontanément comme mots agréables ; il ne faut donc pas s'attendre à les retrouver dans les réponses à notre question ouverte.

On trouve en premier lieu *éducation* (-2.7), *heureux* (-2.4), *généreux* (-2.3), *union* (-2.3), qui sont, pourrait-on dire, les composantes agréables du devoir (assez subtilement, *heureux* s'opposera à *plaisir* sur cet axe). En revanche, la première colonne contient beaucoup plus de mots significatifs (ayant des valeurs-test supérieures ou égales à 2.2) dans sa partie basse qui correspond au demi-axe « Plaisir ».

Parmi les plus significatifs, on trouve *#argent* (donc citation du mot comme désagréable, avec une valeur test de 4.9), puis le mot *plaisir* (3.9), qui n'est pas un mot du questionnaire sémiométrique, mais qui a été choisi pour caractériser l'axe depuis pratiquement l'origine de la méthode, donc bien avant cette expérience de questionnement ouvert. Puis on trouve les mots *musique, livre, sommeil, campagne*, etc.. En tout plus de 28 mots cités spontanément ont une valeur-test supérieure ou égale à 2.2.

**Tableau 4.8 :**  
**Projections sur les axes sémiométriques des mots cités spontanément**

AXE 2	vtest	AXE 3	vtest	AXE 4	vtest
<b>éducation</b>	<b>-2,7</b>	<b>fille</b>	<b>-3,4</b>	<b>grand-père</b>	<b>-3,7</b>
<b>heureux</b>	<b>-2,4</b>	<b>randonnée</b>	<b>-2,9</b>	<b>#haine</b>	<b>-3,1</b>
<b>généreux</b>	<b>-2,3</b>	<b>rivière</b>	<b>-2,8</b>	<b>grand-mère</b>	<b>-3</b>
<b>union</b>	<b>-2,3</b>	<b>#nazisme</b>	<b>-2,7</b>	<b>paix</b>	<b>-3</b>
propre	-2,2	<b>#mépris</b>	<b>-2,4</b>	<b>#alcoolisme</b>	<b>-2,8</b>
serviable	-2,2	<b>vent</b>	<b>-2,4</b>	<b>fraternité</b>	<b>-2,8</b>
visite	-2,2	<b>#fatigue</b>	<b>-2,3</b>	<b>travail</b>	<b>-2,8</b>
merci	-2,1	<b>#haine</b>	<b>-2,3</b>	<b>petits-enfants</b>	<b>-2,7</b>
bonne-humeur	-2	manger	-2,2	<b>#guerre</b>	<b>-2,6</b>
sagesse	-2	vitesse	-2,2	<b>respect</b>	<b>-2,6</b>
sentiment	-2	affection	-2,1	<b>rencontre</b>	<b>-2,5</b>
travail	-2	boire	-2,1	<b>arbre</b>	<b>-2,4</b>
avion	-1,9	farniente	-2,1	<b>#drogue</b>	<b>-2,3</b>
entraide	-1,9	pain	-2,1	<b>vérité</b>	<b>-2,3</b>
-----	----	-----	----	-----	----
<b>#arme</b>	2,2	champagne	1,8	dormir	1,5
<b>#racisme</b>	2,2	fête	1,8	manger	1,5
bisou	2,2	réunion	1,8	océan	1,5
délice	2,2	télévision	1,8	plaisir	1,5
verdure	2,2	vacances	1,8	sport	1,5
<b>chaleur</b>	<b>2,3</b>	amabilité	1,9	week-end	1,5
<b>océan</b>	<b>2,3</b>	belle	1,9	argent	1,6
<b>#pollution</b>	<b>2,4</b>	<b>#désagréable</b>	2	belle	1,6
<b>couleur</b>	<b>2,4</b>	beau	2	cheval	1,6
<b>câlin</b>	<b>2,4</b>	mari	2	restaurant	1,6
<b>désir</b>	<b>2,4</b>	mignon	2	soirée	1,6
<b>mari</b>	<b>2,4</b>	<b>#divorce</b>	2,1	<b>#angoisse</b>	1,7
<b>étoile</b>	<b>2,4</b>	jolie	2,1	mignon	1,7
<b>rire</b>	<b>2,5</b>	réussite	2,1	nourriture	1,7
<b>vent</b>	<b>2,5</b>	<b>#tempête</b>	2,2	<b>#cauchemar</b>	1,8
<b>#obligation</b>	<b>2,6</b>	agréable	2,2	<b>#chagrin</b>	1,8
<b>bébé</b>	<b>2,7</b>	bijou	2,2	<b>#payer</b>	1,8
<b>île</b>	<b>2,7</b>	été	2,2	mariage	1,8
<b>#intolérance</b>	<b>2,8</b>	<b>#feu</b>	<b>2,3</b>	filles	1,9
<b>#violence</b>	<b>2,8</b>	<b>anniversaire</b>	<b>2,3</b>	joyeux	1,9
<b>maman</b>	<b>2,8</b>	<b>cadeaux</b>	<b>2,3</b>	vacances	1,9
<b>#pédophile</b>	<b>2,9</b>	<b>chéri</b>	<b>2,4</b>	<b>#obligation</b>	2
<b>campagne</b>	<b>2,9</b>	<b>famille</b>	<b>2,4</b>	gentillesse	2
<b>sommeil</b>	<b>2,9</b>	<b>mariage</b>	<b>2,4</b>	<b>chance</b>	<b>2,3</b>
<b>livre</b>	<b>3,1</b>	<b>naissance</b>	<b>2,8</b>	<b>prairie</b>	<b>2,5</b>
<b>musique</b>	<b>3,1</b>	<b>aimable</b>	<b>3</b>	<b>#facture</b>	<b>3</b>
<b>plaisir</b>	<b>3,9</b>	<b>gentillesse</b>	<b>3,1</b>	<b>désir</b>	<b>3,1</b>
<b>#argent</b>	<b>4,9</b>	<b>#pluie</b>	<b>3,2</b>	<b>vitesse</b>	<b>5,3</b>

En bref, les mots cités spontanément éclairent l'interprétation du premier axe sémiométrique, et confirment celle du second.

– *L'axe 3*

La seconde colonne du tableau 4.8 représente le positionnement des mots le long du troisième axe sémiométrique, nommé conventionnellement axe « Attachement / Détachement ». Ici encore, les mots du « Détachement » du questionnaire sémiométrique (*Danger, Mort, Rompre, Orage, Angoisse, Vide, Punir...*) ont peu de chance d'être cités spontanément comme faisant partie des mots les plus agréables.

On trouve, pour les individus situés du côté du « Détachement » (qui sont majoritairement des hommes, souvent jeunes), les mots cités spontanément: *filles* (-3.4), *randonnée* (-2.9), *rivière* (-2.8), *vent* (-2.4).

C'est un détachement beaucoup moins inhumain et moins caricatural que celui suggéré par la sémiométrie, surtout si l'on y ajoute *manger, vitesse, affection, boire, farniente, pain...* pour les mots agréables, et *#nazisme, #fatigue, #haine* pour les mots cités comme désagréables<sup>1</sup>.

Du côté de l'« Attachement » (partie basse de la seconde colonne), en revanche, on trouve des mots présents dans le questionnaire sémiométrique comme *naissance* (2.8), *mariage* (2.4), *cadeau(x)* (2.3), *bijou* (2.2). Les mots les plus caractéristiques sont *#pluie* (3.2) considéré cette fois comme mot désagréable, et *gentillesse* (3.1).

Le questionnaire ouvert confirme bien, et enrichit, les interprétations de la sémiométrie pour cet axe.

– *L'axe 4*

Cet axe caractérisé par l'opposition « Esprit/Matière », dans sa partie négative (« Esprit ») contient parmi les mots cités spontanément *grand-père, grand-mère, paix, fraternité, travail, petits-enfants, respect, rencontre, arbre, vérité...*

On retrouve bien les mots *arbre* et *paix* qui sont communs avec ceux de la sémiométrie. Le mot cité spontanément *livre*, on l'a vu, caractérisait nettement la partie positive (« Plaisir ») de l'axe 2, à côté des mots *plaisir* et *musique*. Lorsque le mot *livre* est cité spontanément comme un mot agréable, il s'agit bien du livre de loisir, et non du livre d'école ou

---

1. Rappelons encore une fois que les mots surnotés en sémiométrie ne sont pas forcément des mots bien notés, ce sont des mots qui obtiennent une note supérieure à la moyenne (moyenne qui peut être très basse, et qu'ignore le répondant, puisqu'elle est calculée *a posteriori* sur l'ensemble des réponses).

d'enseignement (ces deux derniers mots accompagnant *livre* sur l'axe 4 de la sémiométrie) . Alors que donner une note au mot *Livre* de la liste sémiométrique, implique, peut-être, une prise en compte de plus de facettes sémantiques du mot.

*Grand-père* et *grand-mère* (comme *petits-enfants*) sont absents du questionnaire sémiométrique, mais sont associés à la paix et la sérénité qui caractérisent cet axe.

Du côté positif de l'axe (« Matière »), le mot cité spontanément qui caractérise le plus le demi axe : *vitesse* (avec la valeur-test exceptionnelle de 5.3) est également un des jalons sémiométriques de cette dimension, comme d'ailleurs le second mot le plus caractéristique : *désir* (3.1)<sup>1</sup>.

Malgré la différence de nature des questionnements et la taille modeste de cet échantillon, on observe donc une incontestable cohérence entre les réponses aux deux types de questions fermées et ouvertes.

## 4.6 Conclusion

Ces expériences de questionnement ouvert sont riches d'enseignements. Mentionnons seulement les trois principaux résultats obtenus à l'issue de ce recueil de données originales et du traitement statistique opéré sur ce recueil.

Le premier résultat est qu'on ne peut obtenir, sur un échantillon de taille modeste, une *sémiométrie spontanée* par un questionnement ouvert du type proposé dans cette section, c'est à dire sans aucune contrainte. On obtient bien des associations locales, schématisées par la carte de Kohonen de la figure 4.1, des plans factoriels présentant une parenté assez marquée avec des plans de la sémiométrie (figure 4.2) mais pas toutes les dimensions latentes stables et interprétables que révèle le questionnaire fermé.

Ce résultat partiellement négatif aide à mieux comprendre quel peut être le rôle d'une liste de mots imposée, identique pour chaque répondant (ce qui n'exclut d'ailleurs pas de faire varier l'ordre des mots à l'intérieur de cette liste de façon à éliminer un éventuel effet d'ordre).

La citation spontanée induit une dispersion sans limite du vocabulaire, réduisant de façon corrélative la signification des distances entre individus. De plus, beaucoup de mots riches de sens et de valeurs ne sont, *a priori*, ni

---

1. Cette analyse de l'axe 4 sera reçue avec prudence compte tenu de la taille de l'échantillon.

agréables, ni désagréables, et donc ont peu de chance d'apparaître dans les réponses spontanées. Notons aussi que les mots consensuels (*amour*, *vacances*, etc.) lestent le recueil sans apporter une information décisive.

Les résultats suivants concernent l'étude des répondants au questionnaire ouvert qui avaient aussi répondu, antérieurement, au questionnaire sémiométrique.

Le deuxième résultat, qui est ici plus confirmé que révélé, concerne le rôle des notes par opposition à une simple mention de présence ou d'absence. La note permet de se référer à une note moyenne pour chaque mot, note moyenne qui n'est pas connue des répondants. On peut mal noter un mot, et pourtant le noter au dessus d'une note moyenne que l'on ignore au moment de l'interview.

Ce traitement pourtant élémentaire permet de travailler exclusivement sur des différences entre individus, et d'obtenir des axes bipolaires ayant des propriétés de stabilité acceptables<sup>1</sup>. Ceci explique le caractère parfois caricatural des axes sémiométriques.

Le troisième résultat concerne la richesse du spontané comme complément à l'illustration et à l'interprétation des axes sémiométriques. Incapable de produire seule des axes bipolaires, le corpus des réponses aux questions ouvertes permet d'étoffer l'interprétation des axes calculés à partir du questionnaire sémiométrique. Car, si la citation spontanée ne peut munir toutes les paires de répondants d'une distance opératoire (ainsi, par exemple, toutes les paires d'individus n'ayant cité aucun mot commun sont à des distances comparables), elle permet de caractériser presque sans limite ces individus<sup>2</sup>.

Il suffit, en effet, que quelques individus aient cité un mot pour que la position de celui-ci puisse être testée statistiquement sur les axes sémiométriques.

---

1. Ainsi, pour l'axe 3, *Mort* (note moyenne : 1.76), *Angoisse* (note moyenne : 1.84) caractérisent les individus sur le demi-axe « Détachement », alors que *Famille* (note moyenne : 6.48) et *Tendresse* (note moyenne : 6.67) font partie des mots qui caractérisent le demi-axe opposé, « Attachement ». Nous avons calculé ces notes moyennes sur l'ensemble des individus à partir de notes de 1 à 7. Il est clair dans ces conditions qu'un individu ayant mis la note 2 à *mort* et *angoisse* (donc au dessus de la moyenne des notes de ces mots) et 6 à *famille* et *tendresse* (donc au dessous des notes moyennes) pourra se trouver du côté « Détachement » tout en ayant légitimement le sentiment d'avoir mal noté les deux premiers mots et bien noté les deux derniers.

2. En termes plus techniques, l'ouvert fournit de mauvaises variables actives, mais d'inépuisables variables supplémentaires ou illustratives (cf. annexe A1.9.3).

Ceci nous aura permis, par exemple, d'habiller de termes nuancés les structures sémiométriques quelque peu austères du « Devoir » et du « Détachement ». Et aussi de mieux comprendre la nature du premier axe qui fait l'objet du chapitre suivant.

## CHAPITRE 5

### Notations, participation, attitudes

Nous revenons, dans ce chapitre méthodologique, sur le premier axe qualifié d'*axe de participation et de notation* (chapitre 1), et exclu, au cours des chapitres précédents, de la structure sémiométrique proprement dite. Il nous incombe en effet de rendre compte au lecteur des résultats et arguments ayant conduit à donner un statut particulier à cet axe. L'étude des styles de notation et du degré de participation à l'enquête débouche naturellement sur le problème mal connu et souvent éludé de la relation entre, d'une part l'attitude par rapport à une enquête et, d'autre part, le contenu et la qualité des réponses de cette enquête. Ce premier axe des analyses sémiométriques est, on l'a vu, toujours stable et dominant<sup>1</sup>.

Ce chapitre confirmera l'hypothèse selon laquelle ce premier axe est un axe que l'on peut qualifier de *méthodologique*, c'est-à-dire, en bref, un axe qui ne concerne pas le contenu du questionnaire et qui est beaucoup moins lié que les axes suivants aux caractéristiques socio-démographiques des personnes interrogées.

Les spécialistes de traitements multidimensionnels de données d'enquêtes rencontrent souvent de tels axes, qui sont la plupart du temps dominants en termes de variance expliquée.

---

1. Il représente toujours de l'ordre de 12% de la variance, alors que le second axe ne dépasse que rarement 5%.

On parle parfois *d'axe de notation*, *d'axe d'attitude par rapport à l'enquête*, et, dans un contexte sensiblement différent (surtout en biométrie) de *facteur de taille*. Le premier axe des analyses sémiométriques est tout cela à la fois.

La richesse et l'étendue de la base de données disponible a permis de faire une étude approfondie de ce phénomène, et même, ce faisant, d'apporter quelques contributions au problème général, souvent épineux et pourtant crucial, de la participation des personnes interrogées à une enquête.

## 5.1 Notation, participation et facteur de taille

### 5.1.1 Les effets notation

Certains effets *notation* sont connus et identifiés lors de la réalisation des tests psycho-sociologiques ou psychométriques et des enquêtes d'opinion et/ou de marketing. Le premier effet généralement décrit est celui du *niveau* de notation. Il existerait - il existe - des bons noteurs et des mauvais noteurs<sup>1</sup>. Le second effet concerne l'*intervalle* de notation : certains individus utilisent tout l'intervalle de notation proposé (ici : toutes les notes de 1 à 7)<sup>2</sup>, d'autres n'utilisent que la partie centrale de l'intervalle (par exemple les notes de 2 à 6, ou seulement les notes 3, 4, 5). La combinaison de ces deux effets produit des styles de notation idiosyncrasiques conduisant à une spécialisation sur des zones particulières de l'intervalle (par exemple les notes 1, 6, 7).

### 5.1.2 Les effets participation

Il s'agit là d'une notion fondamentale et très dérangeante pour les statisticiens : c'est une sorte de prolongement du « refus d'interview » qui constitue probablement la plus grande menace vis-à-vis de la qualité et de la pérennité des systèmes d'information statistique en général, et des enquêtes par sondage en particulier. Est-on sûr que le fait d'accepter de remplir un questionnaire (cette acceptation est en elle-même à l'origine d'un biais, mais sur lequel on ne sait que peu de choses)

---

1. Ou notateurs : les deux formes dérivées *notateur* et *noteur* existent ; nous avons choisi d'utiliser la plus courte.

2. Rappelons que les notes: 1, 2, 3, 4, 5, 6, 7, correspondent respectivement, au niveau du questionnaire, aux notes: -3, -2, -1, 0, +1, +2, +3. Les deux codages sont équivalents pour les calculs de corrélation et pour tous les traitements statistiques réalisés.

implique de la part de tous les répondants une même adhésion à la règle du jeu, à l'esprit de l'enquête ?<sup>1</sup> Cette adhésion ne sera pas indépendante des effets *notation* : un questionnaire truffé de non-réponses ou de réponses systématiquement moyennes est finalement bien proche d'un refus d'interview. C'est même un refus qui ne s'avoue pas, presque une non-réponse sur le refus.

### 5.1.3 Les facteurs de taille

Il s'agit là d'un phénomène tout à fait différent, parce que lié à la composition même du questionnaire, mais dans le cas de la sémiométrie, il est indissolublement lié aux effets *notation*. Les premières études sur l'analyse factorielle au début du vingtième siècle ont concerné des notes à des tests d'intelligence ou à des matières scolaires. Prenons ce dernier exemple : le premier axe sera souvent le « facteur général d'aptitude » baptisé parfois, mais ce fut un sujet de débat, « intelligence » (Spearman, 1904). À de rares exceptions près, toutes les notes sont corrélées positivement entre elles. Il y a, en première approximation, de bons élèves et de mauvais élèves. Les premiers ont des notes supérieures à la moyenne dans pratiquement toutes les matières, les seconds des notes inférieures.

De façon beaucoup plus simple, si l'on mesure la longueur et le diamètre d'œufs de poules, on trouvera que ces deux mesures sont corrélées, car la principale source de variabilité est la taille de l'œuf : le premier axe principal va opposer les gros œufs aux petits œufs, d'où le nom de facteur de taille. Le second axe principal, parfois dénommé « facteur de forme » va au contraire montrer que, à taille égale, les œufs sont plus ou moins sphériques. Nous allons voir qu'un tel phénomène vient se superposer aux effets *notation* dans le cas de la sémiométrie.

## 5.2 Création et positionnement de variables de contrôle

Chaque répondant est, au départ, caractérisé par ses 210 notes au questionnaire sémiométrique et par ses réponses aux questions socio-démographiques (sexe, âge, etc.). À partir des 210 notes, nous avons créé de nouvelles variables techniques appelées variables de contrôle.

---

1. En fait, on est sûr du contraire, mais on a rarement à notre disposition, dans les bases de données recueillies, des éléments pour le prouver ou l'analyser. De ce point de vue, les différents fichiers sémiométriques disponibles et la nature *généraliste* du questionnaire offrent des circonstances favorables à cette étude méthodologique.

Chaque individu est donc caractérisé également par :

- La somme de ses notes calculées sur l'ensemble des 210 mots (ou de façon équivalente, sa note moyenne, qui sera notée  $Note\_moy$ )
- Sept nouvelles variables qui sont : le nombre (noté :  $Note\_1$ ) de notes « 1 », le nombre (noté :  $Note\_2$ ) de notes « 2 », ..., le nombre (noté :  $Note\_7$ ) de notes « 7 » (la somme de ces 7 variables vaut 210, nombre total de notes).
- Deux variables dérivées des précédentes qui sont les notes médianes regroupées, notées :  $Note\_345 = Note\_3 + Note\_4 + Note\_5$ , et les notes fortes :  $Note\_67 = Note\_6 + Note\_7$ .

Les statistiques de base de ces variables de contrôle ont été calculées sur un échantillon représentatif de 5 527 répondants en France (cf. tableau 5.1 pour les variables numériques).

On lit, par exemple, sur la première ligne du tableau 5.1, que le nombre de notes 1 utilisées par un répondant est en moyenne de 11.43 (ce qui représente 5.4 % des 210 notes). La colonne Min (minimum) nous montre qu'il existe au moins un individu qui n'a jamais utilisé la note 1, puisque le minimum vaut 0. La colonne Max (maximum) nous montre qu'il existe au moins un répondant ayant utilisé 89 fois la note 1, et que personne n'a utilisé cette note 1 plus souvent.

**Tableau 5.1 : Statistique de base des variables de contrôle**

	Moyenne	(%)	Minimum	Maximum
$Note\_1$	11.43	( 5.4%)	0	89
$Note\_2$	9.89	( 4.7%)	0	69
$Note\_3$	15.51	( 7.4%)	0	78
$Note\_4$	40.05	(19.1%)	0	146
$Note\_5$	41.67	(19.8%)	0	124
$Note\_6$	46.83	(22.3%)	0	139
$Note\_7$	44.62	(21.2%)	0	148
(Total) (210.00)		(100.0%)		
$Note\_345$	97.23	(46.3%)	0	188
$Note\_67$	91.45	(43.5%)	0	156
$Note\_Moy$	4.95		2.39	6.21

Les trois premières notes sont moins employées que les notes 4 et 5, elles mêmes moins employées que les notes 6 (la plus fréquente) et 7.

La dernière ligne du tableau 5.1 est de nature très différente, puisqu'il ne s'agit plus de comptages de notes par individu, mais de moyennes de l'ensemble des notes données par un répondant. On voit que la note moyenne décernée par un individu est de 4.95 (supérieure à 4, moyenne arithmétique des sept nombres 1, 2, 3, 4, 5, 6, 7) mais qu'elle peut varier de 2.39, pour certains questionnaires, jusqu'à 6.21<sup>1</sup>

Le tableau 5.2 montre, cette fois, les corrélations de ces variables de contrôle avec les six premiers axes sémiométriques tels que ceux-ci sont calculés au cours des chapitres 1 et 2.

Au premier coup d'œil, on voit que les corrélations élevées de toutes ces variables (qui ne sont pas liées au contenu du questionnaire), concernent principalement le premier axe (et, de façon bien moindre, le troisième axe avec lequel la variable *Note\_1* a une corrélation de 0.48).

La variable de contrôle *note moyenne* (rappelons qu'il s'agit de la moyenne des 210 notes attribuées par chaque individu) est fortement corrélée (négativement : -0.92) à l'axe 1, de même que la fréquence des notes 6 et 7 (-0.91).

**Tableau 5.2 :**  
**Corrélations des variables de contrôles avec les axes sémiométriques**

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
<i>Note_1</i>	-.08	-.19	.48	-.13	-.16	.06
<i>Note_2</i>	.31	-.07	.31	-.10	-.06	.20
<i>Note_3</i>	.46	-.06	.12	-.03	.07	.14
<i>Note_4</i>	.55	.17	-.01	.04	-.01	-.22
<i>Note_5</i>	.41	.11	-.20	-.00	.07	.02
<i>Note_6</i>	-.23	.11	-.15	.02	.02	.13
<i>Note_7</i>	-.77	-.20	-.01	.02	-.01	-.02
<i>Note_345</i>	.79	.18	-.09	.02	.06	-.11
<i>Note_67</i>	-.91	-.12	-.11	.03	.00	.06
<i>Note_moy</i>	-.92	-.05	-.36	.10	.07	-.03

Toutefois, ces notes 6 et 7 ne s'opposent pas sur cet axe aux notes 1 ou 2, comme on aurait pu le croire, mais à l'agrégat des notes 3, 4, et 5

---

1. Il ne faut pas confondre cette moyenne des notes pour un individu (calculées sur l'ensemble des 210 notes attribuées aux mots par cet individu) avec les moyennes des notes données à chacun des 210 mots (calculées sur l'ensemble des individus d'un échantillon), qui figurent en annexe A2.2. On apprend ainsi par exemple dans cette annexe que *Guerre* est le mot le plus mal noté en moyenne (1.26) suivi par *Trahir* (1.42), alors que *Paix* est le mot le mieux noté en moyenne (6.72), suivi de *Tendresse* (6.67).

(dont la corrélation de 0.79 avec le premier axe est la plus forte corrélation positive).

Ainsi le premier axe, très lié aux fréquences des notes les plus fortes (6 et 7), et opposant celles-ci aux fréquences des notes médianes (3, 4, 5), est à la fois un facteur de taille et un facteur de notation.

Le fait que des individus choisissent de noter en n'utilisant que la partie centrale de l'échelle (notes 3, 4, 5) alors que d'autres n'hésitent pas à utiliser des notes extrêmes (7 ou 1) constitue le facteur *notation*. Le fait que l'axe 1 (et dans une moindre mesure l'axe 3) soit fortement corrélé (en valeur absolue) avec la moyenne des notes constitue le *facteur de taille*. Mais les notes extrêmes n'interviennent pas de façon équilibrée, car les notes fortes sont beaucoup plus fréquentes que les notes faibles (les notes 6 et 7 représentent 43.5 % des notes attribuées, alors que les notes 1 et 2 ne représentent que 10.1 % de ces notes). Les corrélations des variables de contrôle avec l'axe 1 (et l'axe 3) montrent que le facteur *notation* et le *facteur de taille* sont liés. Nous reviendrons de façon détaillée sur ces liens complexes lors des sections 5.5 et 5.6.

### 5.3 Les notations dans quatre pays européens

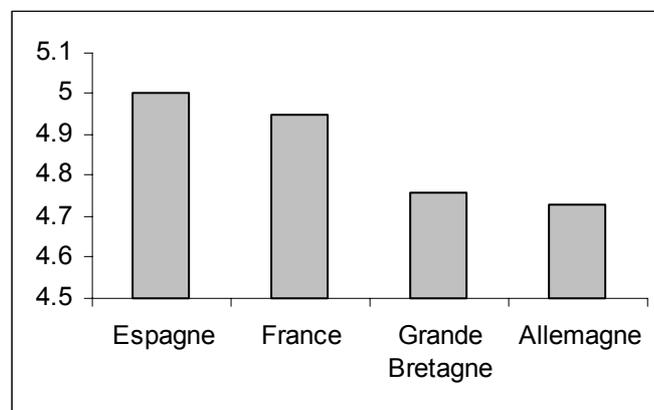
Les variables de contrôle qui caractérisent globalement la façon de noter ont été calculées pour quatre pays (France, Espagne, Grande Bretagne, Allemagne). Ce paragraphe montre que les façons de noter sont éminemment variables.

**Tableau 5.3 : Styles de notation dans quatre pays européens.**

Pays :	Espagne	France	Grande Bretagne.	Allemagne
note_1	14.71	11.43	11.11	10.07
note_2	9.67	9.89	11.87	11.21
note_3	14.89	15.51	19.93	22.49
note_4	33.31	40.05	44.36	43.01
note_5	39.72	41.67	46.46	50.4
note_6	44.17	46.83	38.27	40.15
note_7	53.32	44.62	37.7	32.4
note_345	87.88	97.23	111.09	116.19
note_67	97.61	91.45	75.88	72.43
Taille d'échantillon	2984	5527	924	3066
Note moy.	5.0	4.95	4.76	4.73

L'effet de ces styles de notation sur les structures sémiométriques est cependant limité. Le tableau 5.3 donne un bilan généralisant la première colonne du tableau 5.1.

On lit sur ce tableau 5.3, par exemple, que 14.71 est le nombre moyen de notes « 1 » attribuées par un répondant espagnol, que 32.4 est le nombre moyen de notes « 7 » attribuées par un répondant allemand. L'ordre des colonnes de ce tableau n'est pas neutre : on a volontairement classé les pays par ordre de *moyennes des notes* décroissantes (voir aussi la figure 5.1, qui schématise la dernière ligne du tableau 5.1) : l'Espagne est le pays qui donne en moyenne les notes les plus élevées aux mots (5.0), alors que l'Allemagne donne les notes les plus faibles (4.73).



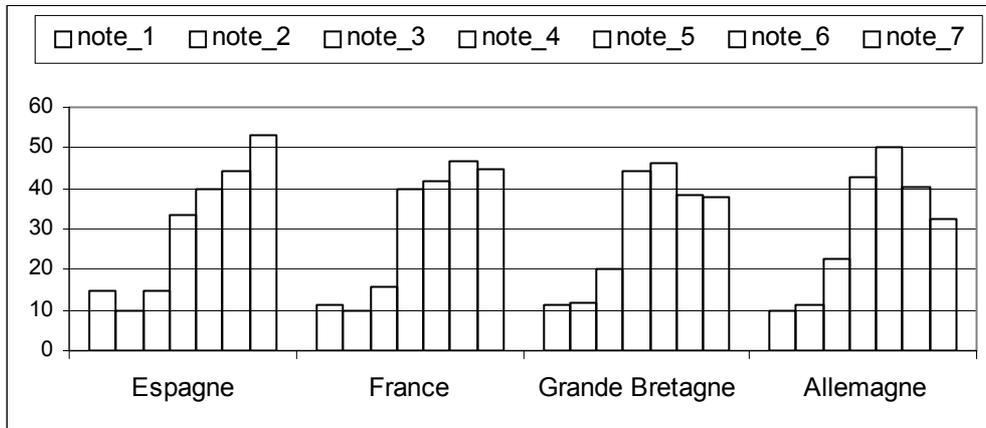
**Figure 5.1 : Moyennes des notes données aux mots dans les quatre pays**

On constate une opposition entre le Sud et le Nord de l'Europe, opposition souvent observée dans les enquêtes internationales<sup>1</sup>.

#### - Noteurs frileux ou résolus ?

La figure 5.2 illustre la variété de l'éventail des fréquences de notes. Il est clair par exemple que la barre de droite de l'histogramme, pour chaque pays (cette barre correspond à la note 7), décroît régulièrement de l'Espagne à l'Allemagne, de même que décroissent les notes moyennes.

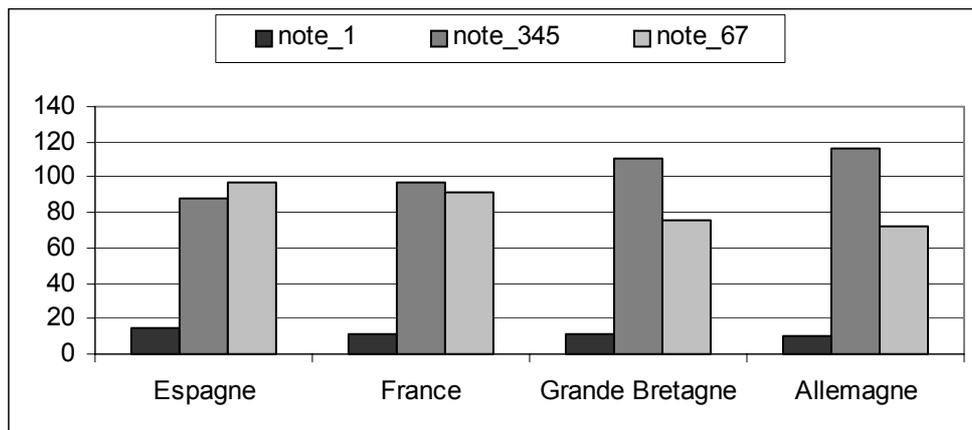
1. Ces différences entre notes moyennes sont statistiquement très significatives pour tous les couples de pays sauf pour le couple Grande Bretagne – Allemagne (la taille de l'échantillon britannique ne permettant pas de conclure avec netteté). Pour la France, par exemple, l'écart-type des notes moyennes (il s'agit, rappelons-le, des moyennes calculées pour chaque individu sur ses 210 notes) est de 0.38, la moyenne de ces *notes moyennes* étant de 4.95. Sur un échantillon de 5527 répondants, l'intervalle de confiance au seuil de 5% de cette moyenne est [4.94, 4.96]



**Figure 5.2 : Bilan des fréquences des sept notes par pays.**

La figure 5.3, enfin, par les regroupements des notes 3, 4 et 5 d'une part, des notes 6 et 7 d'autre part, met mieux en évidence les variations transnationales de l'effet *notation* : les pays du Nord de l'Europe utilisant beaucoup plus volontiers les notes intermédiaires, sans que l'on puisse vraiment décider s'il s'agit de nuance dans la façon de noter, de prudence, de méfiance, voire de désintérêt.

Dans tous les cas, nous avons affaire ici à des différences culturelles importantes, qui hypothèquent la pertinence des comparaisons internationales de *niveaux* ou de *moyennes* à l'intérieur même de l'Europe, sous-continent qui représente pourtant un ensemble relativement homogène à l'échelle mondiale (structure par âge, niveau d'éducation, niveau de vie, etc.).



**Figure 5.3 : Notes extrêmes ( 1 et 6/7) et notes médianes (3/4/5)**

Mais, curieusement peut-être, ces effets *notation* ne parviennent pas à ébranler fortement la stabilité structurelle observée, qui dépend, elle, des *corrélations* entre notes, et non du *niveau* des notes.

## 5.4 Typologie à partir des seules variables de contrôle

Pour approfondir cette étude sur le volet français de l'enquête, une « analyse de contrôle » va être effectuée. Il s'agit d'une nouvelle analyse en axes principaux des individus. Ceux-ci ne seront plus caractérisés par leurs 210 notes, mais par les 7 variables de contrôle, c'est-à-dire par les nombres de notes 1, 2, ...7 qu'ils ont données, sans se préoccuper des mots auxquels ils ont attribués ces notes.

Un des premiers sous-produits de cette analyse est la matrice des corrélations des fréquences des notes (tableau 5.4). Ces corrélations, calculées sur 5 527 individus, sont toutes très significatives statistiquement (à deux exceptions près : les valeurs 0.01 et -0.02)<sup>1</sup>.

**Tableau 5.4 : Corrélacion entre les fréquences de chacune des sept notes chez les répondants**

	Note_1	Note_2	Note_3	Note_4	Note_5	Note_6	Note_7
Note_1	1.00						
Note_2	0.07	1.00					
Note_3	-0.22	0.31	1.00				
Note_4	-0.15	-0.21	-0.24	1.00			
Note_5	-0.49	-0.02	0.46	-0.11	1.00		
Note_6	-0.37	0.17	0.01	-0.41	0.10	1.00	
Note_7	0.45	-0.26	-0.45	-0.34	-0.65	-0.31	1.00

Les plus fortes corrélations positives concernent, d'une part le nombre de notes 5 et le nombre de notes 3 (0.46 : les personnes qui utilisent beaucoup de notes 5 utilisent aussi beaucoup de notes 3, ce sont des *noteurs modérés*), et, d'autre part le nombre de notes 1 et le nombre de notes 7 (0.45 : ce lien est d'autant plus remarquable que la note 1 est, selon le tableau 5.1, quatre fois moins fréquente que la note 7 ; il s'agit ici de *noteurs extrêmes*).

La plus forte corrélacion négative (-0.65) en valeur absolue concerne la note 7 et la note 5, elle traduit une liaison négative entre une note forte

---

1. Pour 5 527 individus, toute corrélacion supérieure en valeur absolue à la valeur 0.03 peut être considérée comme significative au seuil 0.05.

et une note moyenne, c'est-à-dire une opposition entre *noteurs extrêmes* et *noteurs modérés*.

Le premier axe principal de cette « analyse de contrôle », déduit de la matrice de corrélation précédente, est décrit par le tableau 5.5, qui contient les variables les plus caractéristiques de l'axe : des variables actives (variables : *Note\_1*, ...*Note\_7* définies plus haut) et des variables supplémentaires ou illustratives<sup>1</sup> qui sont les notes composites (*Note\_345*, *Note\_67*) et la moyenne individuelle des notes. Ce tableau contient de plus les mots notés par les mêmes 5 527 individus.

Le résultat est assez spectaculaire si on garde en mémoire le fait que les individus ont été caractérisés par un simple bilan numérique des notes données (les sept variables de contrôle), sans se soucier des mots auxquels ces notes ont été attribuées.

**Tableau 5.5 : Analyse des fréquences des sept notes :**  
**Description du premier axe de l'« analyse de contrôle »**  
*Attention : les mots sont ici des variables « supplémentaires »*

<i>Variables les plus corrélées négativement au premier axe</i>		<i>Variables les plus corrélées positivement au premier axe</i>	
<i>Note_5</i>	- .80	<i>Note_7</i>	.85
<i>Note_3</i>	- .66	<i>Note_1</i>	.68
<i>Note_345</i>	- .64	<i>Note_67</i>	.56
<i>Note_6</i>	- .44	Moyenne_des_notes	.42
<i>Note_2</i>	- .29	DOUCEUR	.39
TRAHIR	- .17	HONNETE	.36
ROMPRE	- .16	COURAGE	.35
FAUTE	- .15	CONFIANCE	.35
DESORDRE	- .14	DYNAMIQUE	.34
PUNIR	- .13	CONFORT	.34
ATTAQUER	- .13	GAIETE	.34
ANGOISSE	- .13	POLITESSE	.34
CRITIQUER	- .13	RESPECT	.34
DANGER	- .12	PROTEGER	.34
MURAILLE	- .11	RECOMPENSE	.33
DOUTE	- .10	TENDRESSE	.31
VIDE	- .09	AMITIE	.31

1. Voir annexe A1.9.3 pour quelques notions sur les éléments supplémentaires : dans le cas présent de l'analyse en composantes principales, la coordonnée d'une variable supplémentaire sur un axe est son coefficient de corrélation avec l'axe, calculé sur l'ensemble des coordonnées des individus sur l'axe.

Au niveau des variables actives, ce premier axe de contrôle oppose bien les nombres de notes extrêmes (7 et 1) aux nombres de notes moyennes (3, 4, 5).

Au niveau des variables supplémentaires, la moyenne des notes se situe du côté des notes extrêmes, et l'agrégat des notes (3, 4, 5) du côté des notes moyennes. Mais surtout, et c'est là le fait le plus marquant, on retrouve à peu près les mêmes mots que ceux qui caractérisaient le premier axe sémiométrique.

Si l'on se réfère à la description de ce premier axe (chapitre 1), on trouve en commun, du côté positif, les onze premiers mots de la colonne de droite du tableau 5.5. Du côté négatif, on trouve sept mots communs avec la colonne de gauche du même tableau.

En ce sens, le premier axe sémiométrique mérite son nom d'*axe méthodologique*. Il n'est pas nécessaire de disposer des 210 notes (c'est-à-dire du tableau 5 527 x 210) pour le construire. Les sept variables de contrôle (tableau 5 527 x 7) suffisent.

Il reste maintenant à comprendre pourquoi ces mots particuliers caractérisent l'axe.

Les noteurs dits « modérés » (les répondants qui n'utilisent que la partie centrale de l'échelle de notes) sont caractérisés par des mots dont la note moyenne est basse (colonne de gauche du tableau 5.5 ; voir le tableau des notes moyennes des mots en annexe A2.2), et « anti-caractérisés » par des mots dont la note moyenne est forte (colonne de droite du tableau 5.5). Il y a une première raison simplement mécanique à cela : les variables « mots » sont centrées par rapport à la moyenne des notes de chaque mot. Comme les *noteurs modérés* utilisent relativement peu la partie basse de l'échelle des notes, ils donnent des notes qui, une fois centrées, apparaîtront mécaniquement comme des notes élevées<sup>1</sup>. Mais si l'on effectue un centrage des notes *par individu*, autrement dit si, pour chaque répondant, on retranche à chaque note la moyenne de ses 210 notes (de façon à ce que tous les répondants aient même note moyenne), on obtient toujours la même caractérisation du premier axe de l'analyse de contrôle par les mots figurant dans le tableau 5.5. Donc

---

1. Donnons encore un exemple de ce phénomène déjà évoqué au chapitre précédent : *Trahir* a une note moyenne de 1.42 sur l'ensemble des individus. Ceux qui donnent les notes 2 ou 3 à *Trahir* se situent donc au dessus de la moyenne générale, et apparaissent donc comme ayant relativement bien noté ce mot. Le mot *Douceur* a, lui, une note moyenne de 6.37. Donc une note 5 ou 6, c'est-à-dire une bonne note pour un noteur modéré, sera une mauvaise note par rapport à la moyenne.

l'effet mécanique mentionné ne suffit pas à expliquer pourquoi ces mots caractérisent les deux styles de notation<sup>1</sup>. Nous reviendrons, plus bas, sur cette procédure de centrage des notes par individu (section 5.7).

## 5.5 Quels rapports entre les notes données aux mots, la façon de noter, et les caractéristiques des individus ?

Les sections précédentes ont montré que la moyenne des notes de chaque individu (notée *Note\_moy*) était fortement liée au premier axe sémiométrique, et, dans une moindre mesure, au troisième axe. Nous allons maintenant étudier de façon plus détaillée ces liaisons, tout d'abord à partir de la nouvelle variable nominale à 4 modalités notées Moy1, Moy2, Moy3, Moy4, qui provient d'une répartition selon 4 classes des valeurs de la note moyenne *Note\_moy*. Après consultation de l'histogramme des notes moyennes, les 5 limites de classes choisies pour cette répartition sont : 1, 4.5, 4.95, 5.4, 7.

Le tableau 5.6 donne les valeurs-test<sup>2</sup> de chacune des 4 modalités sur les 5 premiers axes sémiométriques. On voit que le premier axe est caractérisé par des valeurs-test exceptionnelles.

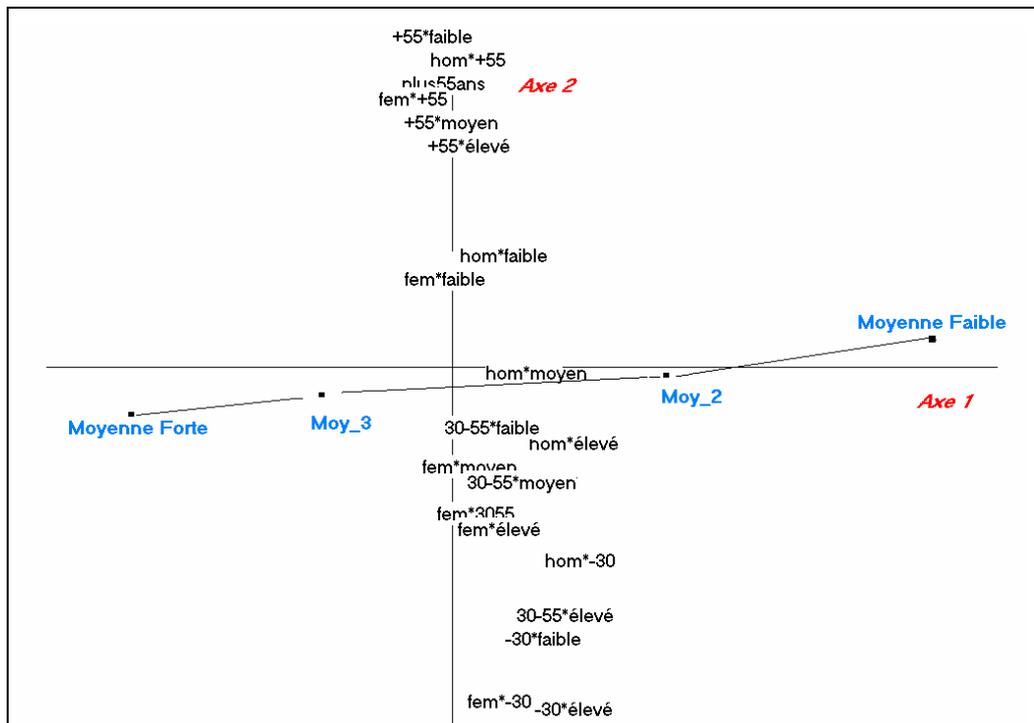
**Tableau 5.6 : Importance de la moyenne individuelle des notes**

Moyenne des notes par individu en 4 classes	Axes sémiométriques Valeurs-test				
	axe 1	axe 2	axe 3	axe 4	axe 5
Moy1 - moyenne_faible	41.9	2.5	10.8	-3.0	-2.0
Moy2 - moyenne_2	28.8	1.3	14.8	-4.1	-2.7
Moy3 - moyenne_3	-29.3	-1.7	-5.0	1.4	.3
Moy4 - moyenne_forte	-39.9	-1.9	-24.9	6.9	5.5

1. On ne peut exclure l'hypothèse selon laquelle la façon de noter serait liée à certains traits psychologiques des répondants, avec un effet sur le choix des mots, effet qui se superposerait à ce que nous avons désigné par *effet mécanique*. Le questionnaire sémiométrique lui-même ne permet pas de dissocier ces effets, mais le questionnaire ouvert complémentaire (section 5 du chapitre 4) a donné lieu à quelques résultats en faveur de cette hypothèse.

2. Rappelons que la *valeur-test* est une simple conversion de la coordonnée sur l'axe en une quantité qui, dans l'hypothèse d'indépendance (répartition aléatoire des individus de la modalité concernée sur l'axe) se comporte comme une variable normale centrée réduite. Autrement dit, si la modalité n'est pas liée à l'axe (i.e. est formée d'individus répartis de façon aléatoire sur l'axe), la valeur-test doit être (approximativement) comprise entre -2 et +2. Ce paramètre permet donc de juger très rapidement la signification statistique de la position d'une modalité sur un axe (*cf.* annexe A1.9.1).

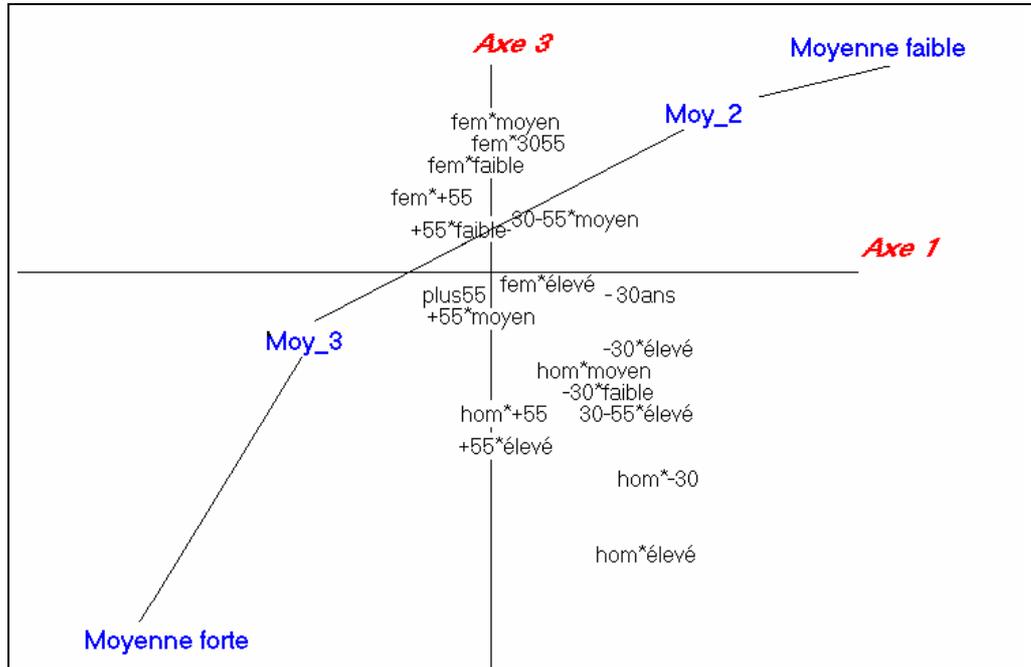
Le troisième axe se distingue également, bien que l'ordre de grandeur ne soit pas comparable. L'indépendance n'est pas totale avec les autres axes, mais la liaison est beaucoup plus faible. La figure 5.4, quant à elle, montre dans le plan des deux premiers axes sémiométriques, le comportement des quatre modalités, notées « Moy1\_faible », « Moy2 », « Moy3 », « Moy4\_forte », par rapport à celui d'autres variables nominales (croisement sexe-âge, croisement âge-éducation, croisement sexe-éducation). On voit clairement que les modalités de la note moyenne caractérisent le premier axe beaucoup plus que toute autre variable socio-démographique, et caractérisent peu le second axe, contrairement à toutes ces mêmes variables. On retrouve d'ailleurs le résultat mentionné lors de la description de l'axe 2 (chapitre 1) selon lequel le deuxième axe oppose des personnes jeunes et instruites des deux sexes à des personnes plus âgées<sup>1</sup>.



**Figure 5.4 : Position des modalités de la moyenne des notes dans le plan (1,2)**

La figure 5.5 montre que, contrairement au deuxième axe sémiométrique, le troisième axe est lié à la moyenne des notes.

1. Sur les figures 5.4 et 5.5, les symboles « -30 », « 30-55 », « +55 » caractérisent les trois classes d'âges, « hom » et « fem » les sexes, et « moyen », « élevé », « faible », les niveaux d'instruction.



**Figure 5.5 : Position des modalités de la somme des notes dans le plan (1,3)**

Bien que trois fois moins important en termes de variance expliquée que le premier axe (4.2 % contre 12 %) il est le seul axe, avec le premier, à être nettement lié à cette moyenne. Cet axe 3 (vertical sur la figure 5.5) qui oppose, on l'a vu au chapitre 1, des hommes (plutôt instruits) à des femmes (plutôt peu instruites), est quand même beaucoup plus lié aux caractéristiques socio-démographiques que l'axe 1.

Il a été montré, au chapitre 2, que cet axe subsistait sur des sous-échantillons formés uniquement d'hommes ou de femmes.

Il est clair que le facteur de taille n'est pas l'axe 1 isolé, mais une direction à peu près diagonale (un peu plus proche de l'axe 1 en fait, car l'échelle de l'axe vertical est dilatée sur cette figure) dans le plan (1, 3). Cette direction diagonale est schématisée par la trajectoire des modalités allant de *Moyenne faible* à *Moyenne forte*.

Projetons maintenant sur ce même plan (1,3) issu, rappelons-le, d'une analyse sémiométrique classique du tableau ayant 5 527 lignes et 210 colonnes, les mots (un extrait, pour plus de lisibilité) et les variables de contrôles (fréquences de chacune des 7 notes pour chaque individu). La figure 5.6 ainsi obtenue est une vue d'ensemble de la richesse et de la

complexité de la liaison entre *notes* et *façon de noter*, entre résultat d'une enquête et attitude par rapport à cette enquête<sup>1</sup>.

La position des mots dans le plan (1,3) très décentrée par rapport à l'origine, est typique de l'existence d'un facteur de taille. Les répondants, dont le point moyen est l'origine des axes ont donc des attitudes très diverses par rapport à l'ensemble des notes : ceux qui sont en haut à droite donnent des notes faibles à l'ensemble des mots. Ceux qui sont en bas à gauche des notes fortes.

Mais la trajectoire des fréquences de notes montre que la note moyenne cache des composantes très différentes. Avec un simple facteur de taille, cette trajectoire aurait dû être à peu près rectiligne, allant des notes faibles en haut à droite vers les notes fortes en bas à gauche. Il n'en est rien, car à l'effet taille se superpose un effet *attitude* ou encore *participation* : certains répondants utilisant toutes les notes proposées, d'autres se cantonnant dans les notes moyennes, d'autres enfin utilisant préférentiellement les notes extrêmes.

- 1) Les répondants situés dans le quadrant supérieur gauche (quadrant I) sont de « vrai bon noteurs » qui donnent la note 7 lorsqu'un mot leur est très agréable (*Cadeau, Guérir, Tendresse, Respect*) et la note 1 dans le cas contraire (*Danger, Vide, Faute, Punir*). Les mots de la sémiométrie permettent de rapprocher les mots de l'« Attachement » de l'acceptation de toute l'échelle de notation. **On voit donc qu'il n'y a pas indépendance entre contenu du questionnaire et attitude par rapport au questionnaire.**
- 2) Les répondants situés dans le quadrant inférieur gauche (quadrant III) sont de « bons noteurs systématiques » ; leurs notes contiennent moins de 1 et de 2, ils ne rejettent pas nettement les mots.
- 3) Les répondants situés dans le quadrant supérieur droit (vide de mots) (Quadrant II) sont les « mauvais noteurs systématiques ». Ils utilisent le bas de l'échelle de notation, mais rejettent aussi les mots (*Danger, Vide, Faute, Punir*). Le facteur de taille au sens classique oppose le quadrant II et le quadrant III.
- 4) Enfin, les répondants situés dans le quadrant inférieur droit (quadrant IV) utilisent surtout la partie centrale de l'échelle de

---

1. La figure 5.6 n'est qu'une esquisse. Les points représentant les notes 1 et 7 sont à l'extérieur du cadre dans la direction indiquée par les flèches (les coordonnées précises des points-notes sont données par les colonnes *Axe 1* et *Axe 3* du tableau 5.4).

notes. Ils utilisent relativement peu les notes 1 et 7. Les mots de la sémiométrie permettent maintenant de rapprocher les mots du « Détachement » de ce refus de toute l'échelle de notation.

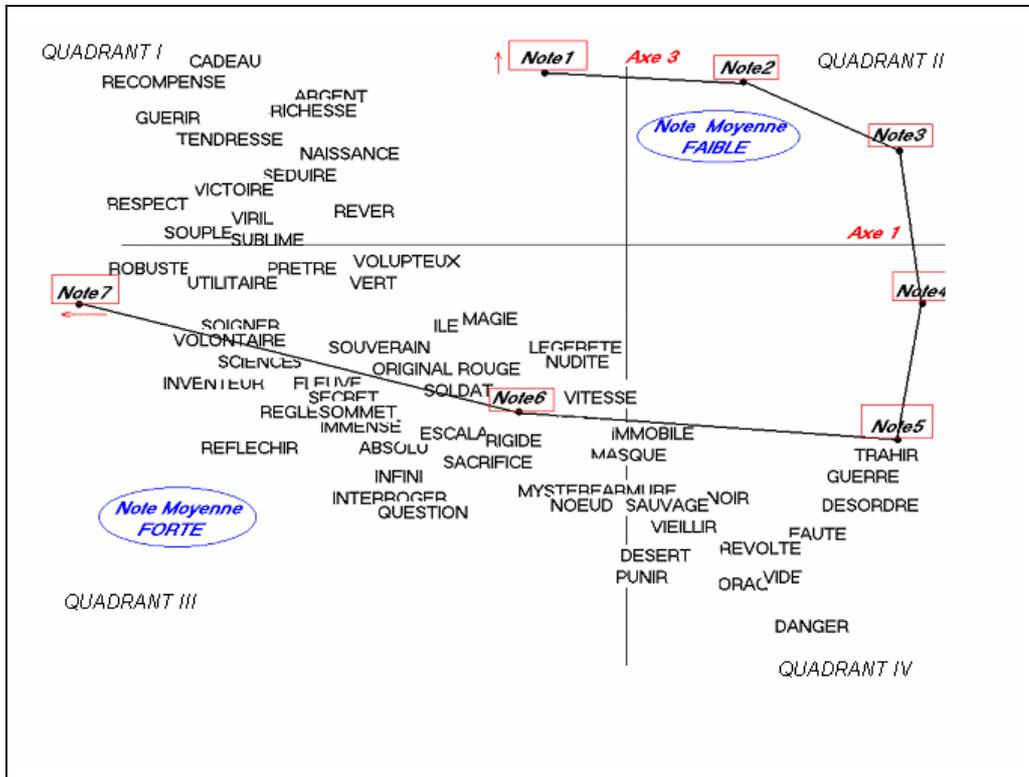


Figure 5.6 : Position des mots et des fréquences de notes dans le plan (1, 3)

Ce plan (1, 3) peut être appelé *plan de notation*. Ce sont en effet le long des deux dimensions de ce plan que l'on observe des positions significatives et excentrées des variables de contrôle (fréquences des notes, moyenne des notes).

On a noté au chapitre 2 des interversions dans l'ordre des axes pour les axes 2 et 3, l'axe qualifié conventionnellement d'axe « Attachement / Détachement » occupant la troisième position en France, en Italie et en Espagne) et la seconde position dans les pays du nord de l'Europe (Allemagne, Grande Bretagne, Finlande, Norvège). Cet axe étant le seul à être concerné par le style de notation (avec le premier axe que nous avons exclu de la structure sémiométrique proprement dite), on peut faire l'hypothèse que son changement de rang est précisément lié aux différences culturelles de notation mises en évidence au cours du présent chapitre.

## 5.6 Les plans de notations en Espagne, Grande Bretagne et Allemagne

Nous allons vérifier que les traits structuraux observés dans le cas de la France sont encore valides pour les trois pays supplémentaires que nous avons sélectionnés<sup>1</sup>.

### 5.6.1 Effets notation en Espagne

La figure 5.7 est l’analogue de la figure 5.6 pour le recueil sémiométrique espagnol, portant sur 2 984 individus. Seulement un mot sur 4 (sélection aléatoire) figure sur le graphique.

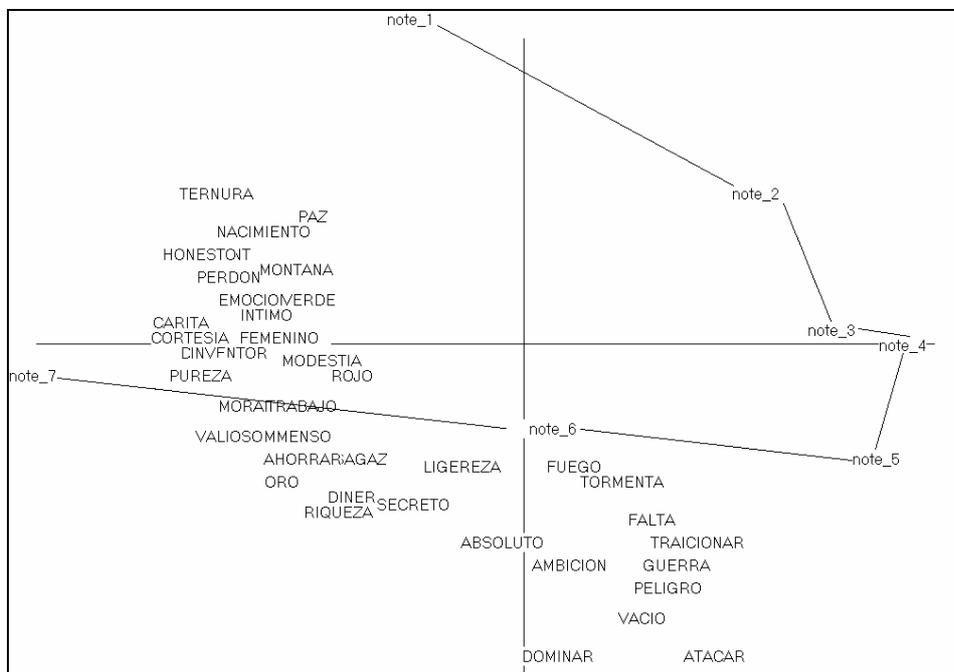


Figure 5.7 : Schématisation du plan (1, 3) (Espagne)

Comme précédemment, les fréquences de notes figurent avec le statut d’éléments supplémentaires, les variables actives de l’analyse en composantes principales étant les 210 mots.

Comme pour la France, c’est le plan engendré par les axes 1 et 3 qui rassemble les effets *notation*, avec les mêmes interprétations<sup>2</sup> : mots de

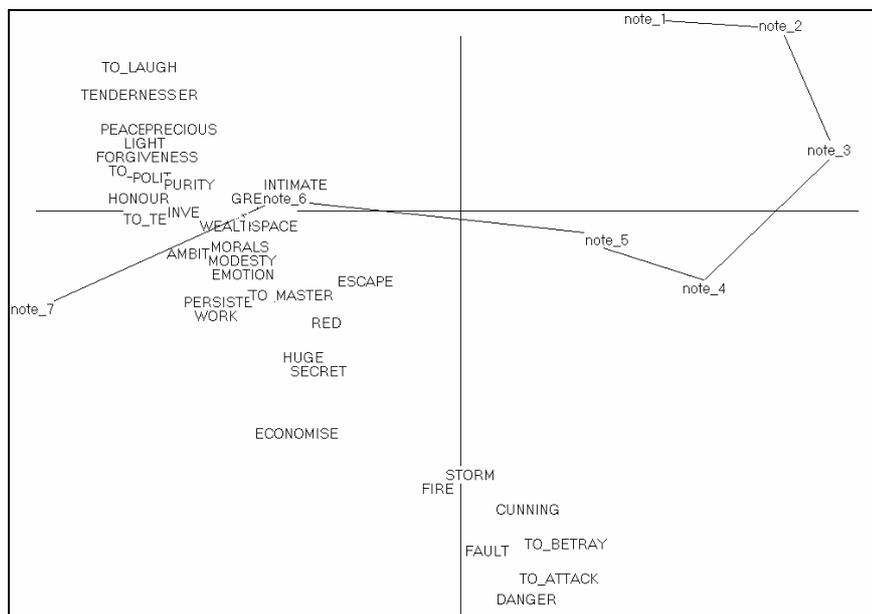
1. Des résultats similaires relatifs à l’ensemble des pays pour lesquels des données sémiométriques sont disponibles ne sont pas publiés ici, faute de place.

2. Cf. le tableau multilingue des mots en annexe A2.1 pour une traduction.

l'« Attachement » (*Tendresse, Naissance, Paix*) du côté des fréquences élevées des notes extrêmes, mots du « Détachement » (*Attaquer, Dominer, Vide, Guerre, Trahison*) (plus proche des fréquences élevées des notes moyennes (notes 4 et 5)).

### 5.6.2 Effets notation en Grande Bretagne

Le schéma de la figure 5.8 est relatif à un échantillon nettement plus restreint (924 individus) et doit donc être interprété avec plus de prudence.



**Figure 5.8 : Schématisation du plan (1, 2) (Grande Bretagne)**

En raison de l'interversion d'axes signalée à plusieurs reprises, notamment dans la section 2.1.2 du chapitre 2, nous présentons maintenant non plus le plan engendré par les axes 1 et 3, comme ce fut le cas pour la France et l'Espagne, mais le plan engendré par les axes 1 et 2. Les positions des points-mots dans ce plan sont tout à fait analogues à celles observées dans les cas de la France et de l'Espagne pour les plans définis par les axes 2 et 3. Les points représentant les fréquences de notes (positionnées, rappelons-le, en tant qu'éléments supplémentaires) ont, eux, une configuration légèrement différente. Les notes (7, 6, 5, 4, 3) s'échelonnent le long du premier axe, alors que pour la France et l'Espagne, ce n'était le cas que des seules notes (7, 6, 5).

### 5.6.3 Effets notation en Allemagne

La figure 5.9 montre comment les mots et les fréquences de notes se structurent dans le plan (1, 2) de l'Allemagne, pays dont l'enquête est effectuée sur un échantillon de 3 066 répondants.

Le coude de la courbe des fréquences de notes se fait au niveau de la note 4, alors qu'il se faisait au niveau de la note 5 pour la France et l'Espagne. L'organisation générale reste cependant la même dans les quatre pays : concentration des effets notations dans un sous-espace à deux dimensions, comportement non-linéaire des notes dont la trajectoire s'incurve de façon à opposer les notes extrêmes aux notes moyennes. Présence des mots de l'« Attachement » chez les personnes utilisant toute l'échelle des notes, des mots du « Détachement » du côté des notes intermédiaires.

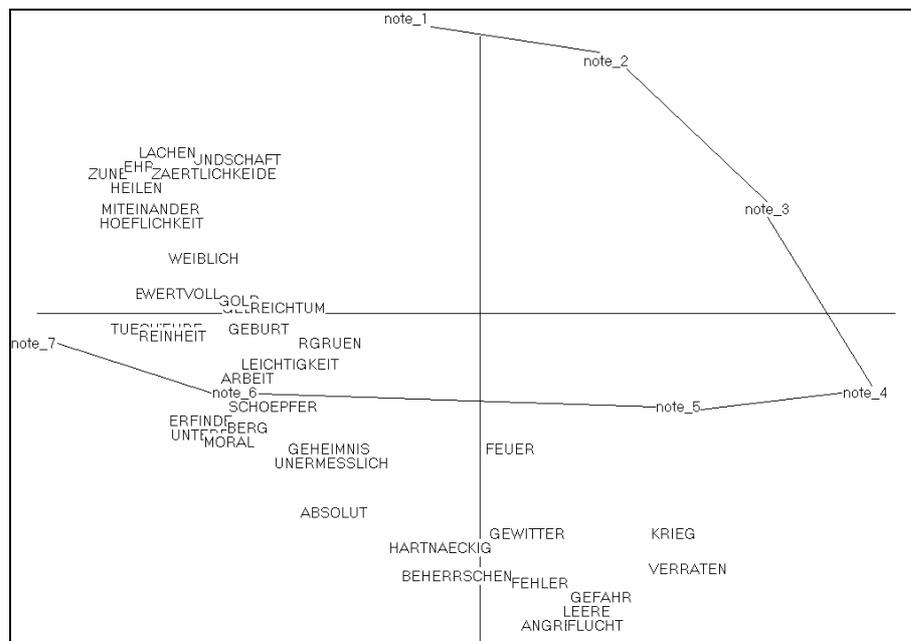


Figure 5.9 : Schématisation du plan (1, 2) (Allemagne)

## 5.7 Correction des notes par double centrage

Il arrive souvent que le *facteur taille* soit corrigé par un centrage au niveau de chaque répondant<sup>1</sup>. Afin d'éliminer les différences entre bons noteurs et mauvais noteurs, on procède de cette façon à un « centrage par individu » qui s'ajoute au centrage habituel qui est le centrage des variables : on retranche à chacune des 210 notes d'un répondant la moyenne de ses 210 notes. Les notes corrigées de tous les individus ont alors une moyenne nulle<sup>2</sup>.

On a vu plus haut (dernière ligne du tableau 5.1) que les notes moyennes des individus s'échelonnaient entre 2.39 (pour le plus *bas noteur* de l'échantillon) et 6.21 (pour le plus *haut noteur*). Après la transformation de centrage « sur individu », tous les individus auront la même note moyenne.

Cette transformation va profondément affecter le premier axe principal (en termes de variances) puisque cet axe était essentiellement lié à la note moyenne des répondants (le tableau 5.2 a en effet montré que la valeur absolue de la corrélation entre le premier axe et la note moyenne est de 0.92).

**Tableau 5.7 : Paramètres des analyses normales et centrées par individu (Echantillon France)**

	Valeur propre (Analyse normale)	Valeur propre (centrage par individu)	Pourcentage (Analyse normale)	Pourcentage (centrage par individu)
Axe 1	26.64	13.33	12.69	6.35
Axe 2	10.24	11.21	4.88	5.34
Axe 3	8.79	8.05	4.19	3.83
Axe 4	7.05	5.10	3.36	2.29

1. Des corrections analogues se font dans les enquêtes de marketing, mais aussi lors de certaines corrections d'épreuves d'examens ou de concours, lorsque de nombreuses copies sont réparties aléatoirement entre correcteurs, et que l'on désire éliminer un éventuel effet *correcteur* sur les notes, en ramenant les notes relatives à chaque correcteur à une même note moyenne commune à tous les correcteurs.

2. Cela n'est peut-être pas intuitif pour un non-mathématicien, mais le fait de centrer en colonnes (faire en sorte que tous les mots aient une même note moyenne nulle) après avoir centré en lignes (faire en sorte que tous les répondants aient une même note moyenne nulle) ne décentre pas le tableau en ligne. Autrement dit, on obtient immédiatement un tableau doublement centré.

On lit effectivement sur le tableau 5.7 que le nouveau premier axe est deux fois moins important en termes de variance. Ce n'est plus un axe fortement dominant.

En bref, cette transformation va modifier l'espace des premiers axes de la façon suivante : le nouveau premier axe reste voisin de l'ancien premier axe, tout en étant fortement corrélé à l'ancien axe 3 (« Attachement / Détachement »). Il apparaît par ailleurs que les autres axes sont à peu près conservés, mais décalés (le nouveau troisième étant l'ancien quatrième, etc.) avec cependant quelques interversions dans leur ordre.

La figure 5.10 montre le nouveau plan (1,2) qui est le lieu des principaux changements. Les fréquences d'emploi des notes sont projetées dans ce plan ; puis sont encadrés, de deux façons différentes les mots de l'« Attachement » (partie haute et gauche du graphique) et du « Détachement » (partie basse et droite du graphique).

Le nouvel axe 1 après centrage par individu (axe horizontal de la figure 5.10) est corrélé<sup>1</sup> à l'ancien axe 1 (0.92) et à l'ancien axe 3 (0.78).

Le nouvel axe 2 est fortement corrélé à l'ancien axe 2 (0.88) (axe « Devoir / Plaisir ») avec maintenant une légère liaison avec l'ancien axe 3 (0.51), liaison décelable sur la figure 5.10 à partir de la position des mots encadrés.

La trajectoire polygonale des fréquences de notes nous montre que le centrage par individu ne résout évidemment pas le problème de l'opposition entre noteurs extrêmes (du côté « Attachement » du plan) et noteurs moyens (du côté « Détachement »).

### ***Fallait-il centrer a priori les notes de chaque individu ?***

Nous avons choisi de ne pas effectuer de transformation a priori sur les notes (à l'exception de l'opération de centrage et de réduction des variables inhérente à la technique d'analyse en composantes principales) pour plusieurs raisons, que nous pouvons maintenant expliciter à la lumière des expériences présentées dans cette section.

- Dans toutes les enquêtes et tous les pays étudiés, le niveau de notation (note moyenne par individu) et la structure de notation (fréquences de chacune des sept notes) forment un *pattern* complexe non-réductible à une transformation non-linéaire préliminaire.

---

1. Il s'agit ici de corrélations calculées à partir des coordonnées des 210 mots.



- Le centrage par individu élimine le facteur *taille* (moyenne individuelle des notes) mais pas l'effet *notation* (spécialisation de la notation dans la partie centrale de l'échelle).
- L'analyse des données non-transformées montre la complexité de la relation entre attitude par rapport au questionnaire et contenu du questionnaire.
- L'option prise d'analyse les données brutes, qui consiste à considérer l'attitude par rapport à l'enquête et le style de notation comme parties intégrantes du phénomène à étudier, sans simplification mécanique préalable, nous paraît plus riche heuristiquement, plus ambitieuse scientifiquement. Elle se situe dans une optique exploratoire de « non-réduction *a priori* du champ de l'observable » qui n'hypothèque pas les recherches à venir.

### ***Une alternative intéressante : l'analyse logarithmique***

Parmi les différents essais de recodage et de transformation des données, une technique a donné des résultats intéressants : l'analyse logarithmique (*cf.* annexe A1.5).

C'est une transformation qui conduit directement aux différents axes de la sémiométrie tels qu'ils ont été étudiés aux chapitres précédents, avec cependant élimination du premier axe. Comme dans le cas du double centrage, l'axe 3 apparaît en premier, suivi des axes (2, 4, 5, 6). Mais les axes sont très purs : c'est exactement l'axe « Attachement / Détachement » qui apparaît en premier, puis l'axe « Devoir / Plaisir », etc. (les coefficients de corrélations avec les axes sémiométriques classiques sont supérieurs à 0.95).

Cela dit, les règles d'interprétation des résultats sont moins connues et moins diffusées que celles de l'analyse en composantes principales, et il n'a pas paru opportun d'utiliser cette méthode pour présenter les axes sémiométriques. Il s'agit pour l'instant d'une piste de recherche possible<sup>1</sup>.

---

1. La transformation logarithmique, d'un usage fréquent en économie, a classiquement pour effet de symétriser les distributions et de linéariser les relations. Intuitivement, il n'est donc pas trop surprenant de la voir opérer avec succès dans un contexte que nous avons largement décrit au cours des chapitres précédents comme non-linéaire et non-symétrique.

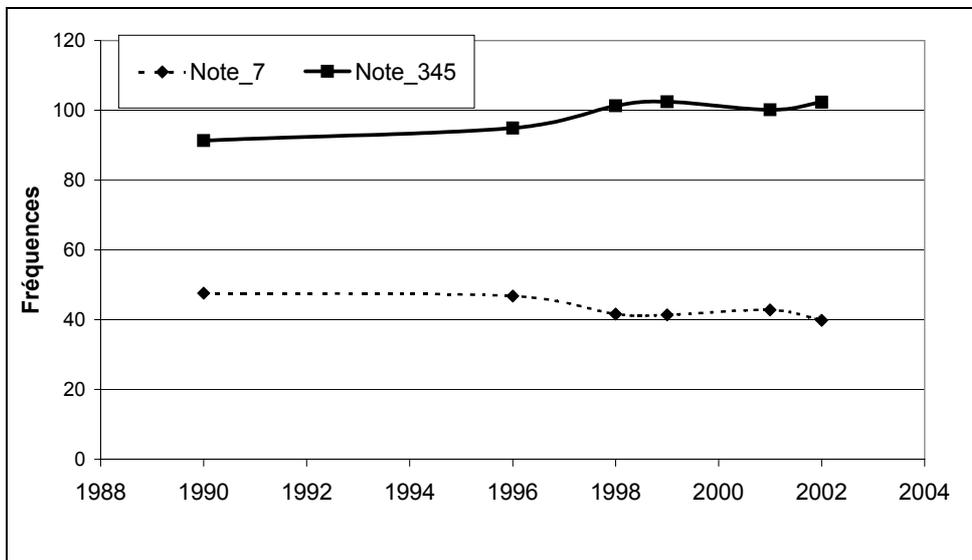
## 5.8 Evolution des notations, en France, de 1990 à 2002

Nous avons vu que la façon de noter est variable selon les pays. Nous allons maintenant voir que la façon de noter varie aussi dans le temps. L'étude portera sur le seul pays pour lequel des séries statistiques longitudinales sont disponibles : la France.

Le résultat est assez spectaculaire, mais comme nous l'avons fait jusqu'ici, nous nous attacherons à séparer les faits statistiques que nous mettons en évidence, étayés par des procédures de validation adaptées, et leur interprétation, pour laquelle nous ne faisons que des propositions, ouvertes aux apports et aux critiques d'autres équipes interdisciplinaires.

Le travail d'analyse porte sur six enquêtes d'effectifs comparables concernant les années 1990, 1996, 1998, 1999, 2001, 2002, totalisant 15 684 individus.

La structure sémiométrique (les six premiers axes, et l'ordre d'apparition de ces axes, c'est-à-dire leurs importances relatives en termes de variance) est stable au cours des six années, mais cette stabilité cache une évolution nette dans la façon de noter : les personnes interrogées utilisent de moins en moins de notes extrêmes, et donc de plus en plus de notes médianes.



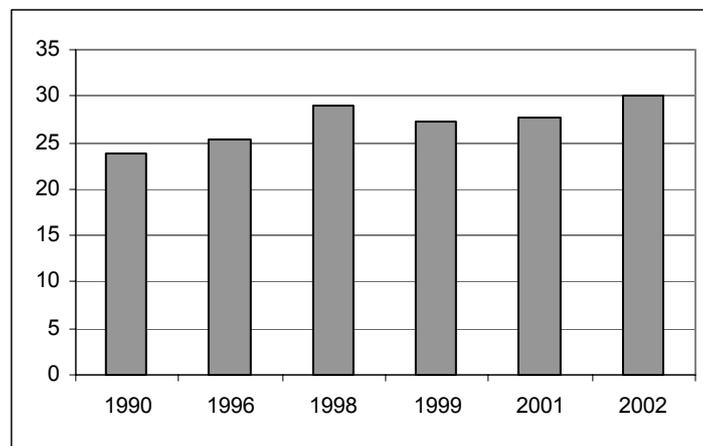
**Figure 5.11 : Evolution des fréquences des notes médianes et de note 7.  
La propension à utiliser des notes modérées augmente**

La figure 5.11 illustre cette évolution, qui oppose surtout la période 90-96 aux quatre dernières années. Un autre trait caractéristique de cette

diminution de la participation (ou de l'engagement) des répondants au questionnaire est l'importance croissante du premier axe, que nous avons qualifié d'axe méthodologique, axe qui oppose les réponses extrêmes aux réponses médianes.

La figure 5.12 retrace l'évolution de la première valeur propre (variance sur le premier axe) au cours de cette période. Mais les fréquences de notes et l'importance du premier axe ne traduisent qu'une petite partie de l'évolution observée de 1990 à 2002.

En fait, les points-années occupent des positions très typées<sup>1</sup> dans le plan des axes 1 et 3, plan qui a été étudié en détail dans les sections précédentes de ce chapitre.



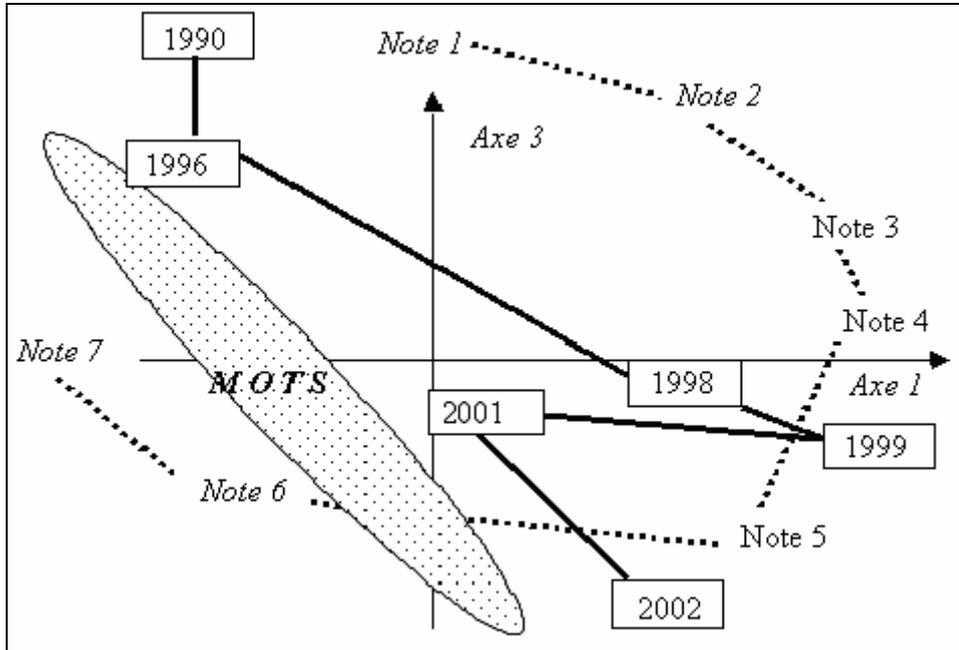
**Figure 5.12 : Evolution de la plus grande valeur propre : le premier axe prend de plus en plus d'importance au cours du temps**

Ces positions ont été reportées sur la figure 5.13 qui comporte également, en traits pointillés, la « parabole des fréquences de notes » représentée auparavant sur les figures 5.6 à 5.10, toutes dévolues à ces plans (1,3). La figure 5.13 contient également une ellipse grisée représentant approximativement la zone occupée par les mots. Dans la partie haute de cette ellipse se trouvent les mots : *Elite, Créateur, Foi, Sacré, Efficace, Robuste, Modération, Récompense, Cadeau, Bijou*. Ces mots sont caractéristiques des premières années de la période<sup>2</sup>.

1. Ces positions sont des projections des six catégories de la variable nominale « année » sur le plan sémiométrique [1,3] selon la technique des variables supplémentaires (Annexe A1.9.3).

2. La proximité de la partie haute de l'ellipse avec les points représentant les premières années sur la figure 5.13 suggère une association entre les mots correspondants et ces premières années. On vérifie ensuite, indépendamment de l'analyse en composantes

On trouve parmi ces mots ceux qui caractérisent les bons noteurs repérés par le premier axe, et les mots de l'« Attachement » du troisième axe.



**Figure 5.13 :** Esquisse de l'évolution de la position des années dans le plan des axes (1,3). L'ellipse grisée schématise la zone occupée par les mots.

Dans la partie basse de l'ellipse se trouvent les mots : *Faute, Règle, Punir, Masque, Désert, Danger, Angoisse, Désordre, Trahir*. Ces mots caractérisent les dernières années de la période considérée.

Ce sont pour les uns, les mots de la « non-participation », un des pôles du premier axe, et pour les autres, les mots du « Détachement », un des pôles du troisième axe.

Malgré la position légèrement atypique de l'année 2001, la trajectoire des années est résolument orientée vers le pôle « Détachement », avec, simultanément - et peut-être corrélativement - une tendance à noter de façon plus centrale et moins typée. En ce sens, la France se rapproche progressivement des deux pays de l'Europe du Nord étudiés (Grande Bretagne et Allemagne), dont nous avons vu la prédilection pour les notes médianes (*cf.* tableau 5.3 et figure 5.3). La fréquence des notes (3,

---

principales, que les notes moyennes de ces mots pour les années 1990 et 1996 sont très significativement supérieures à leurs notes moyennes sur toute la période (valeurs-test supérieures à 5).

4, 5) varie en France de 91 à 102, sans atteindre cependant les 116 de l'Allemagne, ni même les 111 de la Grande Bretagne. Le nombre moyen de « notes 7 » décroît de 47 à 39 entre 1990 et 2002, mais reste supérieur aux 37 de la Grande Bretagne et au 32 de l'Allemagne. En revanche, l'ordre des axes ne varie pas au cours du temps, et l'on n'observe pas l'interversion des axes 2 et 3 qui caractérise tous les pays de l'Europe du Nord pour lesquels nous disposons de données.

Il est assez délicat de faire la part de ce qui serait une évolution dans la façon de noter (dans le sens d'un désengagement croissant se manifestant par des notes de moins en moins différenciées) et une évolution socio-culturelle plus générale vers le désengagement et les valeurs du « Détachement ». Cette dernière évolution aurait de toutes façons des répercussions sur les attitudes par rapport aux enquêtes en général, et par rapport aux enquêtes sémiométriques en particulier.

L'intérêt du cadre conceptuel et technique dans lequel nous travaillons est de présenter le fait statistique dans toute sa complexité, en amont de toute modélisation, et donc de rendre le matériau de base lisible et ouvert à diverses interprétations.

Rappelons que c'est l'opposition sur le premier axe qui traduit l'opposition entre les notes extrêmes et les notes moyennes. Or, il apparaît que les années sont plus nettement (i.e. : significativement, au sens statistique) séparées par le troisième axe<sup>1</sup>. Ceci est une présomption forte en faveur d'une évolution structurelle vers les valeurs du « Détachement » qui ne serait pas un simple artefact de la façon de noter, cette façon de noter pouvant être elle-même une des composantes d'une tendance plus générale.

## 5.9 Conclusion

Comme cela a été énoncé dès le chapitre 2, les structures sémiométriques (configurations de mots) sont stables. Les matériaux statistiques et expériences de ce chapitre auront montré que cette stabilité subsiste malgré des styles de notation assez différents d'un pays à un autre.

---

1. Les valeurs- test des positions des années 1990, 1996, 2002 valent respectivement 9.1, 7.7 et -10.6 sur le troisième axe (axe vertical de la figure 5.13) alors qu'elles valent -5.1, -5.1, et 2.5 sur le premier axe (axe horizontal de la figure 5.13)..

En montrant que le premier axe des analyses pouvait être obtenu simplement à partir de la fréquence d'emploi de chacune des sept notes (sans se soucier des mots auxquels sont attribuées ces notes), le caractère « méthodologique » de cet axe que nous avons jusqu'ici tenu à l'écart de la structure sémiométrique, a été établi : il concerne essentiellement la forme et non le contenu des questionnaires et des enquêtes.

Les *notes moyennes* sont beaucoup plus affectées que les *corrélations entre notes* par les styles de notation. L'ordre des axes est sensible aux codages de l'information, alors que les axes eux-mêmes sont beaucoup plus robustes.

Les interversions d'axes observées entre les pays de l'Europe du Sud (France et Espagne dans ce chapitre) d'une part, et les pays de l'Europe du Nord (Allemagne et Grande Bretagne dans ce chapitre) d'autre part, peuvent être dues à ces différences de styles de notation, ou à des différences structurelles plus profondes<sup>1</sup>.

Un des principaux enseignements de ce chapitre est que l'attitude par rapport à un questionnaire fait partie intégrante de l'information à analyser et à interpréter ; que cette attitude conduit à des *patterns* stables, que l'on peut difficilement écarter sans mutiler l'information de base.

En général, les questionnaires d'enquêtes ne comportent pas d'éléments permettant de caractériser les attitudes par rapport aux conditions mêmes de recueil de l'information et l'adhésion de la personne interrogée à la problématique du questionnaire.

C'est un peu le mérite du questionnaire généraliste que constitue le questionnaire sémiométrique de permettre cette caractérisation. On peut étudier avec cet outil ce que disent les réponses, mais aussi, dans une certaine mesure, ce que répondre veut dire.

---

1. Le fait statistique est le suivant : l'axe « Devoir / Plaisir » est moins important, en terme de variance expliquée, en Europe du Nord qu'en Europe du Sud. Une des pistes d'interprétation possible est, par exemple, la suivante: cet axe opposant des personnes jeunes à des personnes plus âgées, on peut imaginer que, pour des raisons historiques, culturelles, (voire climatiques, ethniques, etc.) l'opposition entre les générations est moins marquée dans l'Europe du Nord que dans celle du Sud.

## CHAPITRE 6

### Essais d'interprétation

Les chapitres précédents ont présenté les structures sémiométriques révélées par six dimensions stables. Ces structures ont été validées à partir d'épreuves empiriques et à partir de procédures techniques de rééchantillonnage. Toujours à partir de nouvelles expériences de terrain ou d'analyses statistiques, la parenté de cette structure avec certains réseaux sémantiques a été étudiée, puis le recueil actuel de données a été confronté à ce que serait un recueil par questions ouvertes. Enfin, les problèmes de notation et d'attitudes vis-à-vis des questionnaires et des enquêtes ont été étudiés, en insistant sur l'interaction inévitable entre, d'une part, ces effets *notation* et ces attitudes et, d'autre part, le contenu même des résultats.

Nous avons en quelque sorte élaboré un support « objectif », un ensemble de *faits statistiques*, en restant avares de commentaires substantifs. Les éventuelles remarques interprétatives ont eu jusqu'à présent pour seul rôle, pédagogique, de couvrir d'un peu de chair un squelette technique et méthodologique. C'est le caractère pluridisciplinaire du champ étudié, et donc la multiplicité des interprétations possibles, qui a conduit à préserver autant que cela a été possible le socle factuel, reproductible, à partir duquel plusieurs interprétations « subjectives » peuvent maintenant être formulées. Ce chapitre, rédigé plus particulièrement par Jean-François Steiner, change résolument de ton et prend le risque de proposer et d'argumenter en

faveur de quelques interprétations possibles des structures observées situées à plusieurs niveaux, historique, littéraire et psychologique.

## 6.1 Une première approche

Quand on la contemple un peu attentivement, la structure sémiométrique, telle qu'elle apparaît sur les cartes des différents plans factoriels, provoque en nous cette curieuse impression – quasi proustienne – faite d'un mélange de sentiment d'étrangeté et de familiarité, que l'on éprouve face à un objet, ou un être, bien connu mais que l'on découvre dans une situation où on ne s'attend pas à le rencontrer, comme, par exemple, lorsqu'on se trouve assis dans un restaurant à une table voisine d'une personnalité célèbre. On reconnaît parfaitement ses traits mais on n'ose croire que c'est elle, car elle appartenait pour nous jusque là à un autre monde avec lequel le nôtre ne communiquait pas.

Ainsi, le deuxième axe « Devoir / Plaisir », par exemple, avec d'un côté les mots *Sensuel, Rêver, Aventurier, Original, Nudité, Bohème, Evasion, Révolte* et de l'autre, *Discipline, Obéir, Morale, Soldat, Règle, Loi, Travail, Dieu, Famille, Tradition*, etc., évoque une opposition qui nous semble traverser aussi bien notre propre vie que la vie politique, être au cœur de nos conflits les plus intimes comme de ceux de la société.

D'un côté, c'est le fameux « Law and Order » (*Loi, Règle, Discipline*) des Anglo-saxons mais aussi l'illustre trilogie vichyssoise (*Travail, Famille, Patrie*) ou encore la célèbre, quoique quelque peu surannée, alliance du sabre et du goupillon (*Soldat, Prêtre*). De l'autre côté, on retrouve tous les thèmes de mai 68, le rêve d'un monde tout dédié à la jouissance, sans contrainte et sans loi, où il est interdit d'interdire et où le rêve doit être pris pour la réalité (*Sensuel, Rêver, Nudité, Bohème, Evasion, Révolte*).

Mais cette opposition, et c'est là où cette structure mise en évidence par la sémiométrie ne cesse d'être troublante, n'est pas seulement un phénomène politique contemporain. Il y a deux siècles, Goethe l'avait déjà décrite, mais en la situant au cœur de son être : « Deux âmes, hélas !, fait-il dire à Faust, habitent ma poitrine et qui se repoussent l'une l'autre. L'une assoiffée de plaisir, est attachée au monde de toute la force

de ma chair ; l'autre, mue par une force inconnue, cherche à m'entraîner vers les hautes demeures de mes ancêtres »<sup>1</sup>.

Le troisième axe « Attachement/Détachement »<sup>2</sup>, avec d'un côté les mots *Bijou, Cadeau, Mariage, Gaieté, Famille, Parfum, Fleur, Gratuit*, etc., et, de l'autre, *Danger, Mort, Rompre, Angoisse, Vide, Punir, Cri, Doute, Révolte, Question, Inconnu*, etc., évoque l'ultime injonction que Dieu fait, par la bouche de Moïse, au peuple hébreu, juste avant qu'il ne pénètre dans le pays de Canaan : « Vois, j'ai mis devant toi la vie et le bien, la mort et le mal »<sup>3</sup>, injonction dont le célèbre slogan de *beatniks*, « *Make love not war* », peut être considéré, par certain côté, comme un très lointain avatar. Cette opposition entre l'état de dérégulation et celui de plénitude de l'être est au cœur de la problématique biblique et d'une grande partie de la mystique juive avec les deux concepts antagonistes fondamentaux de *Shechina* (présence divine) et de *Galouth* (exil de l'âme), « Du fond de l'angoisse, j'ai crié vers toi et tu m'as répondu, tu as délivré mon âme »<sup>4</sup>, dit le psalmiste. Toute la dramaturgie de la descente en Egypte – *Mitzraïm*, en hébreu, mot qui signifie littéralement les deux angoisses – et de la remontée vers Canaan, le pays où « coule le lait et le miel », repose sur cette même opposition entre les deux états de l'âme tantôt glacée d'angoisse et tantôt dilatée de bonheur.

Le quatrième axe « Esprit / Matière », de même, avec son opposition entre, d'un côté, les mots *Livre, Art, Poésie, Enseigner, Ecrire, Réfléchir, Chercheur, Ame, Dieu, Foi*, etc., et de l'autre, *Richesse, Or, Argent, Gloire, Conquérir, Ambition, Séduire, Désir, Sensuel*, etc., décrit une opposition qui se retrouve dans de nombreux conflits intérieurs ou de société. C'est le conflit chrétien par excellence qui oppose l'âme au corps, la richesse au salut, le monde d'ici-bas tout dédié au culte de Mamon et l'autre monde, plein de sublimes félicités, conflit que l'on

---

1. « Zwei Seelen wohnen, ach ! in meiner Brust,  
die eine will sich von der andern trennen :  
Die eine hält, ein derber Liebeslust,  
sich an die Welt mit klammernden Organen :  
Die andre hebt ewaltsam sich vom Dunst  
Zu den Gefilden hoher Ahnen. »

2. Rappelons le phénomène d'interversion des axes « Devoir / Plaisir » et « Attachement / Détachement » entre les pays « latins », France, Italie et Espagne et les pays septentrionaux, Allemagne, Grande-Bretagne, Norvège, Finlande, Hongrie et République Tchèque, l'axe « Devoir / Plaisir » ayant une valeur propre supérieure à celle de l'axe « Attachement / Détachement » dans les pays latins, et réciproquement.

3. Deutéronome 30,15.

4. Psaume 120.

retrouve laïcisé de nos jours dans l'opposition entre l'art et l'argent, la recherche et le commerce, les ONG et les multinationales, etc.. Chacun trouvera, de Bossuet à Tolstoï, mille exemples dans la littérature occidentale, d'œuvres dont le ressort dramatique est organisé autour de ce conflit. Citons, pour la beauté de la langue, cette réponse d'Henriette à Armande dans les Femmes Savantes de Molière :

« Ainsi, dans nos desseins l'une à l'autre contraire,  
 Nous saurons toutes deux imiter notre mère :  
 Vous, du côté de l'âme et des nobles désirs,  
 Moi, du côté des sens et des grossiers plaisirs ;  
 Vous, aux productions d'esprit et de lumière,  
 Moi, dans celles, ma sœur, qui sont de la matière. »<sup>1</sup>

Le cinquième axe « Cœur / Raison », avec les mots, *Dieu, Foi, Ame, Sacré, Eternel, Cérémonie, Infini, Noble, Bijou, Poésie, Absolu, Adorer, Héros, Roi<sup>2</sup>, Reine, Château, Mode, Mystère*, d'un côté et : *Efficace, Précision, Logique, Concret, Astucieux, Utilitaire, Produire, Volontaire, Inventeur, Raison, Science*, de l'autre, évoque l'opposition entre notre besoin de rêve et de dépassement, notre besoin de croire en des réalités d'un autre ordre, magique ou transcendant, et notre besoin antagoniste de comprendre et d'agir dans un monde rationnel régi par des lois simples et entièrement accessibles à notre intelligence. On peut l'interpréter comme l'axe du conflit pascalien entre le cœur, siège et source de la foi et la raison pour laquelle ses mystères demeureront à jamais impénétrables ; ou comme celui qui oppose don Quichotte, le chevalier errant qui fuit la triste réalité de sa mélancolie et de son dénuement dans un monde peuplé de chimères où les moulins à vents se confondent avec des géants, une pauvre auberge espagnole avec un puissant château et une misérable fille de ferme avec une « princesse sans pareille », qui fuit la banale et désolante quotidienneté des choses dans la sublime beauté d'un monde reconstruit par le rêve, avec le brave Sancho Pança qui tente désespérément d'opposer son bon sens terre à terre de paysan à l'imaginaire embourbé dans la glèbe au délire onirique de son maître.

Mais, cet axe peut aussi s'interpréter comme celui du conflit qui opposa, à partir de la Contre Réforme, les tenants de la grâce et ceux qui prétendaient, tels les molinistes, que le salut s'obtenait par les œuvres<sup>3</sup>,

1. *Les Femmes Savantes*, acte I scène 1.

2. Dans une version antérieure du questionnaire qui comprenait 286 mots, dont les mots Roi, Reine et Château, ceux-ci corrôlaient avec le demi axe « Cœur ».

3. Cf. l'affaire de la constitution *Unigenitus* qui empoisonna la fin du règne de Louis XIV, la Régence et une grande partie du règne de Louis XV, qui, sorte d'affaire Dreyfus avant la

ou, dans la seconde moitié du 18<sup>e</sup> siècle, le mouvement rationaliste des Lumières qui affirmait que l'homme ne devait avoir d'autre maître que sa raison, au courant religieux qui posait, à l'origine de la Création, l'existence d'un mystère indépassable, mystère qu'étaient chargées d'entretenir les somptueuses pompes de l'Eglise, la troublante et mélancolique beauté de la musique baroque et l'exubérance de son architecture.

Le sixième axe « Humilité / Souveraineté », avec les mots *Naissance, Méfiance, Doute, Famille, Rigide, Interdire, Obéir, Enfance, Maternel, Immobile, Fidélité, Vide, Politesse*, etc. auxquels il faut ajouter *Peuple, Foule et Egalité* qui appartenaient au questionnaire de 286 mots, d'un côté, et *Noble, Créateur, Art, Dieu, Audace, Elite, Maîtriser, Bâtisseur, Commander*, etc., de l'autre, évoque l'opposition entre ce qu'il est convenu d'appeler les « petites gens » ou « Les Humbles », comme les appelle Jules Romains, qui subissent leur destin et dont l'horizon s'est trouvé racorni par les pesanteurs sociologiques, par un déficit d'éducation et, peut-être, d'amour, et, pour reprendre une expression un peu désuète mais très parlante, les « grands de ce monde » ou « Les Superbes », toujours selon Jules Romains<sup>1</sup> qui, au contraire, ont été élevés dans l'idée qu'ils appartenaient à une classe supérieure qui avait vocation à dominer les autres et à maîtriser leur destin.

Dans un autre registre, cet axe est aussi un axe sociologique, puisqu'il oppose les cadres, les personnes âgées, les personnes résidant dans les agglomérations de plus de 200 000 habitants et les régions parisienne et méditerranéenne, aux ouvriers et employés, aux personnes de moins de quarante ans, aux non titulaires du baccalauréat et aux habitants des

---

lettre, déchira la société française pendant plusieurs décennies. La véhémence avec laquelle Saint Simon l'évoque dans ses Mémoires résonne comme un écho lointain du séisme qu'elle représenta pour la société française : « Ce même mois de mars (1711) vit éclore les premiers commencements de l'affaire qui produisit la constitution *Unigenitus* si fatale à l'Eglise et à l'Etat, si honteuse à Rome, si funeste à la religion, si avantageuse aux jésuites, aux sulpiciens, aux ultramontains, aux ignorants, aux gens de néant, et surtout à tout genre de fripons et de scélérats, dont les suites, dirigées autant qu'il leur a été possible sur le modèle de celle de la révocation de l'édit de Nantes, ont mis le désordre, l'ignorance, la tromperie, la confusion partout, avec une violence qui dure encore, sous l'oppression de laquelle tout le royaume tremble et gémit, et qui, après plus de trente ans de la persécution la plus effrénée, en éprouve, en tout genre et en toutes professions, un poids qui s'étend à tout, et qui s'appesantit toujours. »

1. Les Hommes de bonne volonté, livres 5 et 6.

régions rurales ; et même idéologique<sup>1</sup>, ce qui rend son maniement quelque peu scabreux.

Comme nous allons le voir, ci-dessous, dans une tentative d'interprétation de la structure sémiométrique à travers une grille de lecture freudienne, cet axe peut aussi être compris comme la projection, sinon d'un conflit intérieur, du moins d'une tension intrapsychique à laquelle l'homme est soumis du fait de la nécessité dans laquelle il se trouve, au cours de son développement, de construire son moi par un travail de colonisation de son univers pulsionnel.

Mais avant de nous lancer dans cette périlleuse aventure, notons, en passant, que, peut-être, contrairement à ce qu'affirmait Goethe, plus de deux âmes, en réalité, habitent notre poitrine, comme, d'ailleurs, Nietzsche l'avait déjà pressenti, qui lui répondait : « Si un Allemand disait : deux âmes, hélas, habitent ma poitrine, il s'en faudrait de beaucoup qu'il dise la vérité »<sup>2</sup>.

## 6.2 Une interprétation psychanalytique

A partir de 1920, Freud élabore une théorie, connue sous le nom de « deuxième topique », dans laquelle il décrit l'appareil psychique comme constitué de trois grandes « instances », « le *ça*, pôle pulsionnel de la personnalité, le *moi*, instance qui se pose en représentant des intérêts de la totalité de la personne (...), le *surmoi* enfin, instance qui juge et critique, constituée par intériorisation des exigences et des interdits parentaux. »<sup>3</sup>.

Il nous semble que la description qu'il donne du *ça* : « ...il (le *ça*) comprend tout ce qui a été constitutionnellement déterminé, donc, avant tout, les pulsions émancées de l'organisation somatique », « Il (le *ça*) tend à satisfaire les besoins innés de l'individu », « Le *ça* obéit à l'inexorable principe de plaisir », correspond assez bien au pôle « Plaisir » de l'axe 2 dont une grande partie des mots (*Sensuel, Rêver, Nudité, Sauvage, Voluptueux, Désir, Charnel, Caresse, Emotion, Moelleux, Orage, Peau, Révolte, Danger, etc.*) qui le caractérisent évoque justement l'expression

---

1. Dans le questionnaire de 286 mots, les mots *Peuple, Foule et Egalité* corrélaient avec le demi axe Humilité et s'opposaient, donc, aux mots *Noble, Elite, Roi, Reine et Château*.

2. *Par-delà le Bien et le Mal*, F. Nietzsche.

3. *Vocabulaire de la psychanalyse*, J. Laplanche et J.-B. Pontalis.

libre des pulsions, la satisfaction de besoins innés et la recherche du plaisir, ou, plus exactement, la soumission à son principe.

A l'opposé, la description du *surmoi*, « Sorte de précipité de la longue période d'enfance (...) par laquelle se prolonge l'influence parentale et (...) transmise par eux, celle des traditions familiales, raciales, nationales, ainsi que les exigences du milieu social immédiat qu'ils représentent », dont « la tâche essentielle consiste toujours à réfréner les satisfactions », qui édicte de « sévères interdictions » et qui peut même se manifester sous un jour « dur et cruel »<sup>1</sup>, correspond, de son côté, assez fidèlement au pôle « Devoir » de ce même axe avec les mots *Famille, Patrie, Loi, Tradition, Economiser, Interdire, Sacrifice, Punir, Rigide*, etc.

Réfléchissant sur la nature des pulsions, leur orientation, la manière dont elles s'expriment dans les conduites humaines, Freud proposait, dès 1920, d'en distinguer deux catégories principales : « ce qui nous conduit à distinguer deux sortes de pulsions : celles qui cherchent à conduire la vie à la mort et les autres, les pulsions sexuelles, qui indéfiniment tendent et parviennent à renouveler la vie »<sup>2</sup>. En 1938, à la veille de sa mort, il revenait sur cette dichotomie et la précisait : « Après de longues hésitations, de longues tergiversations, nous avons résolu de n'admettre l'existence que de deux pulsions fondamentales : l'Eros et la pulsion de destruction. (...) Le but de l'Eros est d'établir de toujours plus grandes unités, donc de conserver : c'est la liaison. Le but de l'autre pulsion, au contraire, est de briser les rapports, donc de détruire les choses ».

Notre propos n'est évidemment pas de nous prononcer sur la pertinence de cette métapsychologie. Elle a fait l'objet de nombreuses polémiques, qui ont, d'ailleurs, été abandonnées aujourd'hui, faute de combattants et qui ne faisaient que ranimer le vieux débat théologique inauguré par Saint Augustin, entre les dualistes, qui prétendaient que le monde avait été créé par deux forces, celles du bien et celles du mal, et les monistes qui posaient l'existence de Dieu à l'origine de toute chose. Ce qui nous semble beaucoup plus intéressant c'est de montrer comment cette dichotomie rend un compte assez fidèle de l'opposition qui s'exprime sur l'axe 3 « Attachement / Détachement » de la sémiométrie. Et, en effet, on peut considérer que pratiquement tous les mots du pôle « Attachement », *Bijou, Cadeau, Récompense, Maison, Mariage, Gaieté*,

---

1. *Abrégé de psychanalyse*, S. Freud.

2. *Au-delà du principe de plaisir*, S. Freud.

*Famille, Fidélité, Tendresse, Paix, Caresse*, mais aussi *Or, Argent, Richesse, Mode, Parfum*, dans un sens moins immédiat mais évident, parlent d'objets destinés à créer des liens entre les êtres ou entre les êtres et les choses, tandis que ceux du pôle « Détachement », *Danger, Mort, Rompre, Orage, Angoisse, Vide, Punir, Critiquer, Doute, Révolte, Attaquer, Faute, Ironie, Vieillir, Muraille, Méfiance, Inconnu*, etc., évoquent tous la destruction des relations, par la violence ou l'absence, avec soi ou avec les autres.

Si les deux axes précédents peuvent être interprétés assez aisément à partir de la « deuxième topique » freudienne, les deux suivants (« Esprit / Matière » et « Cœur / Raison ») se laissent moins aisément réduire à une interprétation psychanalytique littérale. Pourtant, le pôle « Esprit » de l'axe 4, avec les mots connotant la spiritualité religieuse, *Ame, Dieu, Prier, Recueillement*, ceux décrivant les activités intellectuelles, *Science, Chercheur, Inventeur, Réfléchir, Question*, et ceux évoquant les différentes expressions de l'art, *Poésie, Livre, Art, Musique, Théâtre*, n'est pas sans faire penser au phénomène psychique que Freud appelle sublimation : « La sublimation est un processus postulé par Freud pour rendre compte d'activités humaines apparemment sans rapport avec la sexualité, mais qui trouvent leur ressort dans la force de la pulsion sexuelle. Freud a décrit comme activités de sublimation principalement *l'activité artistique et l'investigation intellectuelle* »<sup>1</sup>, « La pulsion sexuelle met à la disposition du travail intellectuel des quantités de force extraordinairement grandes et ceci par suite de cette particularité, spécialement marquée chez elle, de pouvoir déplacer son but sans perdre, pour l'essentiel de son intensité. On nomme cette capacité d'échanger son but sexuel originaire contre un autre but, qui n'est plus sexuel mais qui lui est psychiquement apparenté, capacité de sublimation »<sup>2</sup>, « Les historiens de la civilisation semblent d'accord pour admettre que, grâce à ce détournement des forces pulsionnelles sexuelles loin des buts sexuels et cette orientation vers de nouveaux buts – processus qui mérite le nom de sublimation –, de puissantes composantes sont acquises, intervenants dans *toutes les constructions intellectuelles* »<sup>3</sup>. Certes, à aucun moment, Freud ne mentionne la spiritualité religieuse comme étant une manifestation du mécanisme de sublimation, mais on est en droit de penser qu'il ne faut voir là que

---

1. *Vocabulaire de la psychanalyse, op. cit.* C'est nous qui soulignons.

2. *Gesammelte Werke*, VII, 150, S. Freud.

3. *Trois Essais sur la théorie sexuelle*, S. Freud.

l'effet de la virulence de son athéisme. En effet, les liens entre la religion et l'art, aussi bien que la science dans ses manifestations les plus pures, sont trop connus pour qu'on ne s'autorise pas à lever cette exclusion.

Si l'on retient cette interprétation du demi axe « Esprit » comme correspondant au phénomène de sublimation, on peut comprendre son demi axe antagoniste comme celui des activités humaines dans lesquelles les pulsions sexuelles s'expriment sous une forme plus immédiate ou moins élaborée, hypothèse que le sens des mots, *Richesse, Or, Argent, Gloire, Conquérir, Ambition, Commerce, Commander, Attaquer, Séduire, Désir, Sensuel, Charnel, Voluptueux*, etc., est loin de contredire.

Dans la perspective de cette interprétation, l'axe 4 serait constitué par l'opposition entre les individus qui expriment leur sexualité sous une forme relativement basique et ceux qui, plus sensibles aux contraintes culturelles qu'ils ont subies au cours de leur enfance et à la multiplicité des interdits que toute civilisation puissante et féconde ne cesse d'édicter à l'encontre de la sexualité sont parvenus à la canaliser vers d'autres buts représentant, du moins jusqu'à une date récente, des idéaux socialement plus valorisés.

Au fur et à mesure où, déclinant les axes, on pénètre plus profondément dans la structure sémiométrique, l'interprétation de ceux-ci devient plus malaisée, ce qui ne doit pas nous surprendre puisque la quantité d'information apportée par chacun d'eux ne cesse de s'amenuiser. Cette difficulté est encore accrue par le fait que, si le pôle « Matière » de l'axe 4 et le pôle « Raison » de l'axe 5 sont définis par des ensembles de mots radicalement différents (*Or, Argent, Richesse...*, pour celui-ci et *Logique, Précision, Efficace*, pour celui-là), le pôle « Esprit » de l'axe 4 et le pôle « Cœur » de l'axe 5, présentent, eux, des ressemblances troublantes : mêmes mots de spiritualité religieuse, *Dieu, Prier, Foi, Ame...* et mêmes mots d'expression artistique, *Poésie, Théâtre, Musique, Art*. Pourtant, malgré ces similitudes, on observe deux différences importantes entre les pôles « Cœur » et « Esprit », qui montrent qu'ils ne doivent pas être confondus. Outre la présence de mots comme *Roi, Reine* et *Château* du côté du « Cœur », la grande différence entre les pôles « Esprit » et « Cœur » réside dans le fait que les mots évoquant des activités intellectuelles, *Inventeur, Chercheur, Réfléchir, Science, Raison*, qui corrôlaient positivement avec le demi axe « Esprit », corrôlent négativement avec le demi axe « Cœur » tandis que certains mots de « Matière » *Bijou, Or, Mode, Héros, Cérémonie*, qui corrôlaient

négativement avec le demi axe « Esprit », corrèlent maintenant positivement avec le demi axe « Cœur ».

Cette sorte de chassé croisé qui s'opère lorsque l'on passe de l'axe 4 à l'axe 5, s'il représenta longtemps pour nous un écueil sur lequel venait buter nos tentatives d'interprétation, fournit peut-être une clef de compréhension de l'axe 5 du point de vue de la théorie freudienne. En effet, Freud note l'existence d'un phénomène psychique qui présente un certain nombre de similitudes avec la sublimation, au point d'être parfois confondu avec elle, mais qui doit pourtant être considéré comme radicalement indépendant dans son mécanisme : l'idéalisation. Notre hypothèse est que, si le pôle « Esprit » de l'axe 4 correspond, du moins en partie, à la sublimation, le pôle « Cœur » de l'axe 5, qui présente tant de points communs avec lui, correspond au phénomène d'idéalisation.

Mais écoutons Freud : « La formation de l'idéal du moi, écrit-il<sup>1</sup>, est souvent confondue avec la sublimation des pulsions, au détriment d'une claire compréhension. Tel qui a échangé son narcissisme contre la vénération d'un idéal du moi élevé n'a pas forcément réussi pour autant à sublimer ses pulsions libidinales. L'idéal du moi requiert, il est vrai, cette sublimation mais il ne peut l'obtenir de force : la sublimation demeure un processus particulier ; l'idéal peut bien l'inciter à s'amorcer mais son accomplissement reste complètement indépendant d'une telle incitation (...) La formation d'idéal augmente (...) les exigences du moi, et c'est elle qui agit le plus fortement en faveur du refoulement ; la sublimation représente l'issue qui permet de satisfaire à ces exigences sans amener le refoulement ».

Après avoir mis en évidence les liens qui existent entre les deux phénomènes, Freud les définit, l'un par rapport à l'autre, du point de vue de sa théorie : « La sublimation est un processus qui concerne la libido d'objet et consiste en ce que la pulsion se dirige vers un autre but, éloigné de la satisfaction sexuelle ; l'accent est mis sur la déviation qui éloigne du sexuel. L'idéalisation<sup>2</sup> est un processus qui concerne l'objet et

---

1. *Pour introduire le narcissisme*, S. Freud.

2. Freud opère un glissement entre formation d'un moi idéal et idéalisation de l'objet, sans s'en expliquer plus avant, comme si, pour lui, les deux mécanismes étaient identiques. Le terme manquant de l'ellipse est, sans doute, le phénomène d'identification projective par lequel le sujet se confond avec l'objet, c'est à dire que le sujet, en fait mal débarrassé de son narcissisme, projetterait l'image idéal qu'il se fait de lui-même sur l'objet, l'investissant ainsi de libido narcissique. Freud confirme, d'ailleurs, cette hypothèse qui écrit par ailleurs : « Dans maintes formes de choix amoureux, il devient même évident que l'objet sert à remplacer un idéal du moi propre, non atteint. On l'aime à cause des perfections auxquelles on a aspiré pour

par lequel celui-ci est agrandi et exalté psychiquement sans que sa nature soit changée. (...) Ainsi, pour autant que sublimation désigne un processus qui concerne la pulsion et idéalisation un processus qui concerne l'objet, on doit maintenir les deux concepts séparés l'un de l'autre »<sup>1</sup>.

L'axe 5 opposerait, donc, d'un côté, l'idéalisation, « processus psychique par lequel les qualités et la valeur de l'objet sont portées à la perfection », (*Dieu, Prier, Reine, Roi, Foi, Prêtre, Poésie, Cérémonie, Bijou, Théâtre, Sacré, Ame, Château*) à la rationalisation, « Procédé par lequel le sujet cherche à donner une explication cohérente du point de vue logique (...) à une attitude, une action, une idée, un sentiment, dont les motifs véritables ne sont pas aperçus »<sup>2</sup>, (*Efficace, Précision, Logique, Concret, Maîtriser, Astucieux, Utilitaire, Réfléchir, Raison, Science*, etc.) On notera que l'idéalisation comme la rationalisation sont souvent utilisées dans des stratégies de défense contre l'angoisse. Le fait qu'elles ressortissent à des stratégies psychiques de nature antinomique, pourrait peut-être expliquer, du moins en partie, le côté souvent inexpiable des conflits qui se structurent autour de ces deux attitudes<sup>3</sup>.

L'interprétation du sixième axe du point de vue de la théorie freudienne est beaucoup moins évidente. Comme nous l'avons noté ci-dessus, on pourrait y voir la trace d'un conflit intrapsychique entre le *moi*, dans son effort toujours recommencé d'arrachement à la pesanteur du monde pulsionnel<sup>4</sup> (*Noble, Elite, Créateur, Fort, Audace, Ouvert, Commander*), et la tentation de céder à l'appel d'un retour impossible à l'état idyllique d'enfance et de nature (*Famille, Naissance, Enfance, Eau, Animal, Campagne*).

---

le moi propre et qu'on voudrait maintenant se procurer par ce détour pour satisfaire son narcissisme » (*Psychologie des foules et analyse du moi*).

1. Pour introduire la narcissisme, *op. cit.*

2. *Vocabulaire de la psychanalyse, op. cit.*

3. Les concepts d'« Esprit », de « Cœur » et de « Raison » relevant d'une topographie psychologique plus poétique que précise et peu en accord avec les modes de pensée d'aujourd'hui, nous proposons de renommer l'axe 4 « Esprit / Matière » par « Sublimation / Matérialisme » et l'axe 5 « Cœur / Raison » par « Idéalisation / Pragmatisme ». Mais, il est bien évident qu'il ne s'agit que d'étiquettes qui résument et, donc, déforment la réalité sémantique et psychologique des axes, laquelle est rendue par l'ensemble des mots qui corrélaient avec eux.

4. Cf. le célèbre « *Wo es war, soll ich werden* » (Où il y avait le *ça*, le *moi* doit advenir) de Freud, qui résonne comme un écho lointain du premier commandement de Dieu à l'homme après l'avoir créé à partir de la poussière : « Fructifiez et multipliez et emplissez la terre et asservissez-la », que certains commentateurs interprètent comme une injonction faite à l'homme de dominer ses instincts.

Ce pourrait donc être l'axe de l'opposition du *moi* et du *ça*, ou plus exactement l'axe de l'opposition entre les individus qui survalorisent leur *moi* comme instance souveraine, ceux, peut-être, que les marxistes qualifiaient de « petits bourgeois individualistes »<sup>1</sup> et ceux qui préfèrent renoncer à leur singularité et se fondre dans un tout qui les transcende et les absorbe.

Mais, il s'agit là d'une interprétation très libre que nous sommes incapables d'étayer sur aucun texte théorique, comme nous avons tenté de le faire, avec plus ou moins de succès, pour les axes précédents.

Notre propos, d'ailleurs, ne tendait pas à donner une interprétation exhaustive et définitive des dimensions de l'espace sémiométrique, mais seulement à en proposer quelques essais de lecture afin de faire partager au lecteur le sentiment d'émerveillement que nous éprouvons parfois devant la manière dont, à partir d'une simple note évaluant la quantité de plaisir (ou de déplaisir) qui s'attache à leurs connotations, les mots s'organisent dans un espace sémantique d'une profondeur inattendue.

---

1. Lorsque l'on projette les consommateurs des différents types de produits sur les axes, les consommateurs de produits de luxe, dont la fonction sociale de valorisation du *moi* est bien connue (marques de Champagne prestigieuses, montres de marque, etc.), apparaissent systématiquement sur l'axe 6 du côté de la Distinction.

## CHAPITRE 7

### Quelques applications de la sémiométrie

Conçue, à l'origine, comme un instrument permettant de reconstruire un hypothétique espace symbolique qui structurerait, au moins partiellement, notre perception inconsciente du monde, la Sémiométrie s'est révélée un jour, d'une manière tout à fait fortuite, être un outil que l'on pourrait utiliser avec quelque profit dans des applications psychosociologiques et de marketing. C'est ainsi qu'à partir de 1986, parallèlement à un programme de recherche plus théorique et à une réflexion sur la nature de la structure mise en évidence par la sémiométrie, une collaboration s'est instaurée entre Jean-François Steiner et le groupe international de communication Taylor Nelson Sofres pour développer cette approche dans le cadre d'études d'opinion et de marketing.

Ce n'est, bien sûr, pas ici le lieu de présenter un bilan des quelques centaines d'études ainsi réalisées, tant dans le domaine du marketing que dans ceux de la sociologie, de la psycholinguistique ou de la sémantique. Notre propos n'est que d'illustrer certaines de ces applications par quelques exemples. Les cas présentés sont loin d'être exhaustifs de l'ensemble des applications. Nous les avons choisis plus pour leur valeur didactique (la valeur exemplaire de leur pouvoir descriptif et explicatif)

que pour leur intérêt proprement opérationnel<sup>1</sup>. Les applications commerciales, qui restent la propriété des organismes ayant commandité les études, font intervenir des marques ou des noms pour lesquels une certaine confidentialité est en général souhaitée. Elles ne seront pas présentées ici. Le lecteur n'aura cependant aucune peine à imaginer qu'à la place des catégories utilisées dans ce chapitre (sexe, classes d'âge) on puisse faire intervenir les consommateurs de tel produit, les possesseurs de telle marque de voiture, les électeurs de tel candidat ou parti.

La méthode consiste à comparer la note moyenne attribuée par une première population (appelée population étudiée) aux 210 mots du questionnaire sémiométrique, à la note moyenne attribuée à ces mots par une seconde population (appelée population de référence<sup>2</sup>), puis à appliquer un test statistique aux écarts de moyennes afin de déterminer les mots significativement sur-notés et sous-notés par celle-là. Ces mots sont ensuite présentés sur le plan principal (engendré par deux axes principaux) le plus pertinent. Ce type de visualisation facilite la lecture des résultats, mais surtout permet d'apprécier la cohérence sémantique de ces résultats. On verra au fil des exemples que le regroupement des mots caractéristiques dans des zones particulières de l'espace sémiométrique permet une confirmation de la signification des résultats par une prise en compte des corrélations entre mots.

Une méthode alternative (et plus classique) pour comparer des sous-populations consiste, dans l'espace dont les points sont les répondants, à projeter les centres de gravité (ou « points moyens ») des individus appartenant à ces sous-populations sur les plans principaux (*cf.* annexe

---

1. Il existe d'autres applications qui reposent sur la faculté que la sémiométrie offre de mesurer des distances entre individus et groupes d'individus dans un espace stable dans le temps et l'espace. Elles ne seront citées ici que pour mémoire. Il s'agit des applications dans le domaine du médiaplanning qualitatif et des ressources humaines. D'autres applications, enfin, utilisent le pouvoir prédictif des notes sémiométriques pour améliorer la qualité des fusions de fichiers ou pour prédire des données manquantes (*cf.*, par exemple, Auliard et Steiner, 1992).

2. Cette seconde population est, en général, le complément de la première. Toutefois, lorsque l'on soupçonne qu'un effet de structure pourrait masquer le résultat que l'on cherche à mettre en évidence, on filtre les deux populations par la variable perturbatrice. Exemple, lorsqu'on veut étudier des produits consommés essentiellement par des individus d'un sexe donné, on filtrera les deux populations par ce sexe. C'est ainsi que pour un magazine féminin, si l'échantillon sur lequel on travaille est représentatif de la population nationale, on opposera ses lectrices à l'ensemble des femmes et non pas au reste de la population. Car dans ce cas-là, l'effet de sexe étant infiniment plus prégnant que les valeurs véhiculées par un organe de presse, les mots qui caractériseraient ce lectorat seraient essentiellement des mots surnotés par les femmes.

A1.9.3). Cette procédure est en fait complémentaire de celle qui vient d’être décrite, et, dans la pratique, doit être mise en oeuvre simultanément. Elle donne cependant des résultats moins suggestifs, et ne sera pas systématiquement présentée dans le cas des exemples de ce chapitre.

### 7.1 Les mots caractéristique des hommes et des femmes

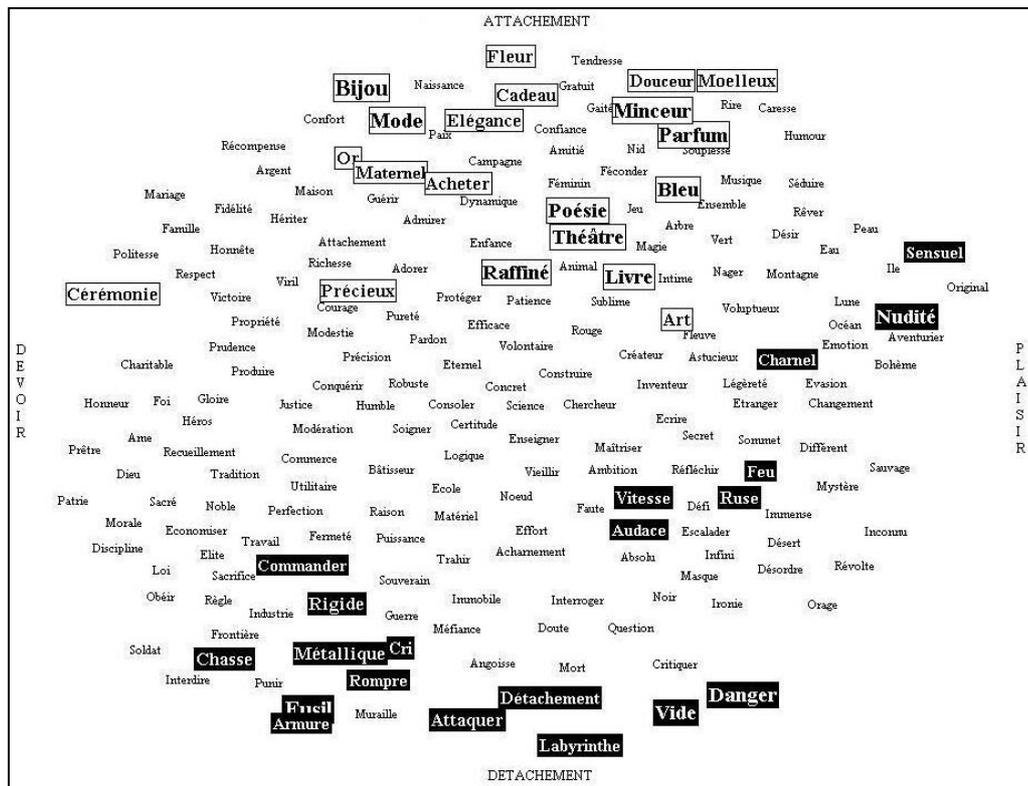


Figure 7.1 Mots caractéristiques des hommes et des femmes dans un plan sémiométrique schématisé [2, 3]

La figure 7.1 est établie à partir d’une schématisation du plan sémiométrique [2,3] qui servira de trame pour la plupart des exemples. Les mots sur fond noir sont les mots que les hommes sur-notent par rapport aux femmes (et que celles-ci, donc, sous-notent), les mots

encadrés sur fond blanc, ceux qu'ils sous-notent, et, donc, qu'elles sur-notent<sup>1</sup>. Comme on peut le constater, le résultat est assez caricatural.

Aux hommes, la violence, le danger, le détachement et la sexualité ; aux femmes, la douceur, l'élégance, la poésie, l'art et la mode. A noter que le mot le plus caractéristique des hommes est le mot *Fusil* (en partie dissimulé par le mot *Armure*)<sup>2</sup> et le plus caractéristique des femmes, *Bijou*, qui sont tous deux généralement considérés comme des symboles sexuels de leur genre respectif. Rappelons, pour nuancer cette interprétation, que le même plan principal (croisant les axes 2 et 3) est aussi obtenu lors d'analyses conduites séparément pour les hommes et pour les femmes, phénomène à propos duquel nous avons parlé de *fractalité*. Il existe des hommes qui surnotent *Bijou* et des femmes qui surnotent *Fusil*. Lorsque les deux sexes sont analysés simultanément, les deux distributions des hommes et des femmes sur le troisième axe « Attachement / Détachement » empiètent largement, mais sont très légèrement décalées<sup>3</sup>.

## 7.2 Les mots caractéristiques de deux classes d'âge

Les deux classes d'âges choisies pour cet exemple sont les classes (moins de 45 ans ; 45 ans ou plus). La figure 7.2, dont les règles de lecture sont les mêmes que celles de la figure 7.1, met en évidence une opposition très nette entre les mots caractéristiques des deux classes d'âge. Les mots sur fond noir sont sur-notés par les moins de 45 ans.

L'axe 2 « Devoir / Plaisir » étant très lié à l'âge, là encore le résultat n'a rien pour nous surprendre. La question que soulève un tel résultat est de savoir quels sont parmi ces mots ceux qui sont le produit d'une opposition d'âge et ceux d'une opposition de génération. En effet, si *Patrie* et *Soldat* semblent rendre compte d'une évolution de la société et *Modération* et *Recueillement* d'un mûrissement de l'individu, qu'en est-il de *Politesse*, de *Morale* ou de *Travail* ? On voudrait espérer que leur

---

1. Par souci de lisibilité, seuls les vingt premiers mots sont présentés ici.. En fait il y a 46 mots sur-notés par les hommes et 73 par les femmes, au seuil de 1.96. Pareillement, et pour la même raison, dans les exemples qui suivent ne figurent que les vingt premiers mots.

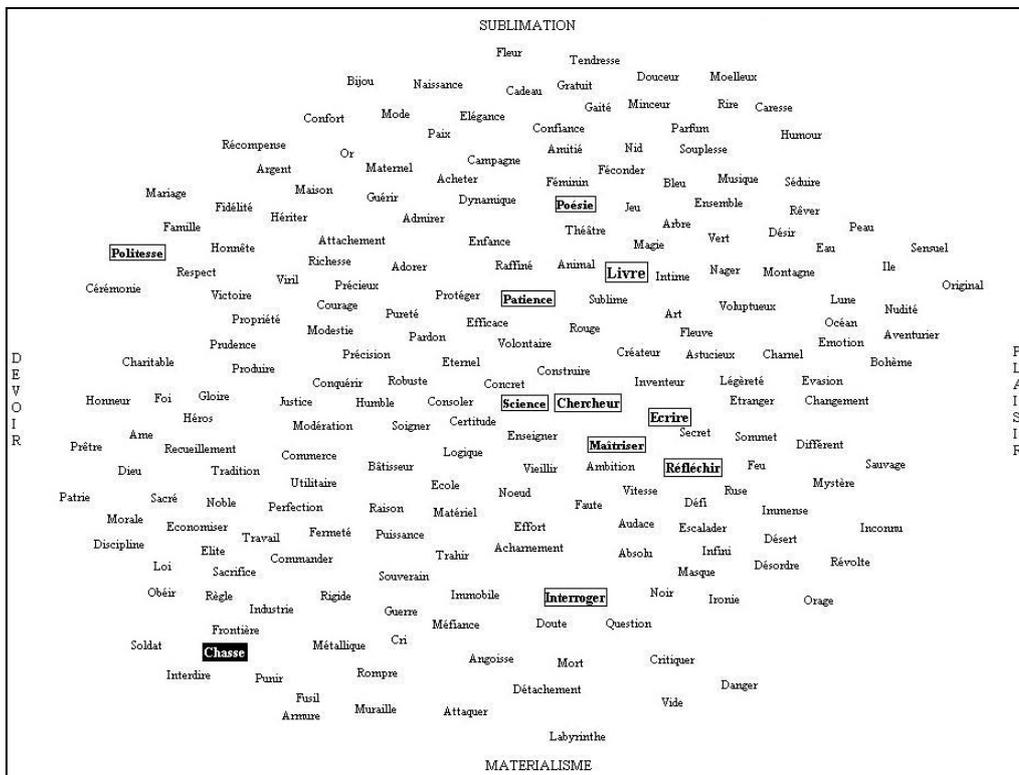
2. La taille des caractères est proportionnelle à la valeur-test du mot, autrement dit, plus les caractères d'un mot sont grands, plus celui-ci est caractéristique de la catégorie étudiée .

3. Sur ce troisième axe, les centres de gravité (points moyens) des hommes et des femmes s'opposent, avec une valeur-test de 20.0 (pour l'échantillon global de 15 686 individus).



### 7.3 Les grands amateurs de télévision

Ce résultat est issu d'une étude réalisée pour une chaîne de télévision française qui portait sur la consommation des programmes. Le groupe que nous définissons de façon conventionnelle comme « les grands amateurs TV » n'est pas constitué, comme cela est d'usage, des plus gros consommateurs, c'est à dire d'individus passant plus de cinq heures par jour devant leur écran, mais du quartile supérieur des individus regardant le plus grand nombre de programmes.



**Figure 7.3 Mots caractéristiques des « grands amateurs de télévision » dans un plan sémiométrique schématisé [2,4]**

L'axe le plus discriminant étant l'axe 4 (« Sublimation/Matérialisme »<sup>1</sup>), les résultats sont présentés cette fois dans le plan (2,4).

La figure 7.3 montre que les « grands amateurs TV » ne sur-notent qu'un seul mot, *Chasse*, et encore au seuil de 2.53 – ce qui n'est pas vraiment significatif, compte tenu des remarques précédentes sur les comparaisons multiples –, pour 12 mots sous-notés, dont neuf, *Livre*,

1. Cet axe était appelé précédemment « Esprit / Matière ».

Poésie, Chercheur, Réfléchir, Science, Ecrire, Patience, Protéger et Arbre, corrélient significativement avec le demi axe « Sublimation ». Le portrait que ce résultat esquisse des « grands amateurs » est peu flatteur, leur seule caractéristique commune étant de rejeter ce qui, d’une certaine manière, fait la grandeur de l’homme, sa faculté de maîtriser ses pulsions et de les canaliser vers des buts d’une plus grande valeur sociale.

### 7.4 Les mots caractéristiques de deux classes de jeunes (15-19 ans et les 20-24 ans)<sup>1</sup>

Dans cette application, on calcule les mots que les deux populations sur-notent à un même moment (par rapport à leur population de référence respective) et l’on fait apparaître ces mots sur le plan factoriel le plus pertinent.

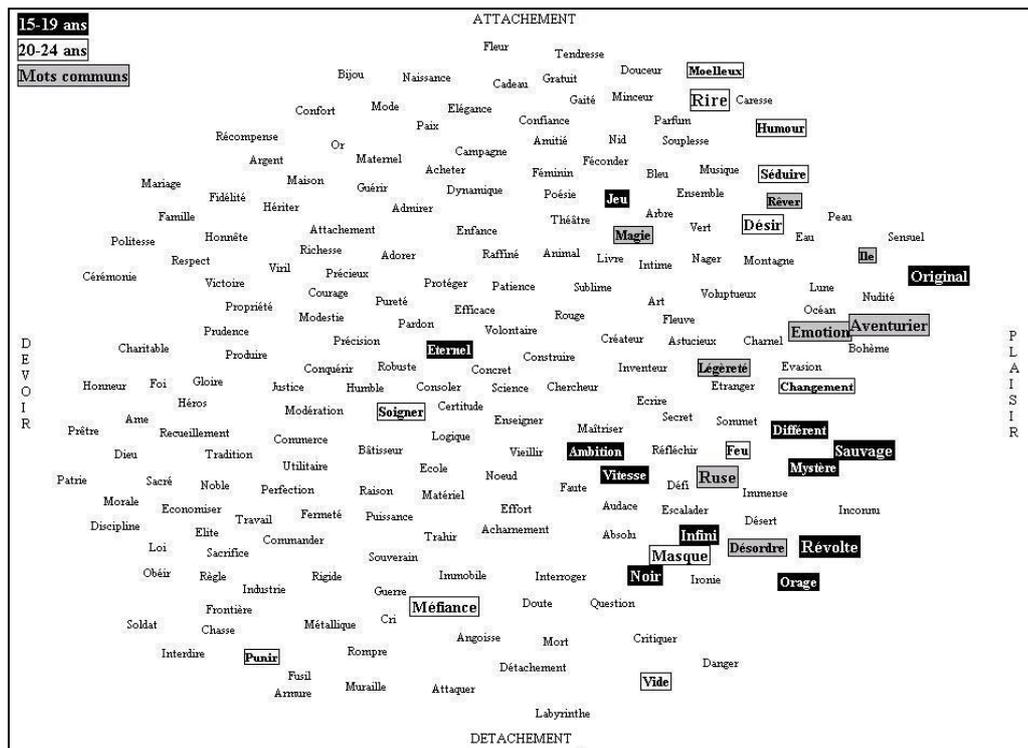


Figure 7.4 Mots caractéristiques des 15-19 ans et des 20-24 ans dans un plan sémiométrique schématisé [2,3]

1. Par souci de lisibilité, sur les cartes présentant deux populations, on ne fait apparaître que les mots sur-notés par celles-ci, distinguant trois catégories, les mots sur-notés par l’une et pas par l’autre et les mots communs.

Avec les mots *Sauvage, Mystère, Révolte, Orage, Noir*, les 15-19 ans illustrent assez bien la phrase qui ouvre l'*Aden Arabie* de Paul Nizan : « J'avais vingt ans, je ne laisserais personne dire que c'est le plus bel âge de la vie ». Mais, passé le cap fatidique des 20 ans, les jeunes semblent commencer à prendre plus de plaisir à vivre, *Moelleux, Rire, Humour*. La cause en est, peut-être, qu'ils en découvrent les voies, *Désir, Séduire*.

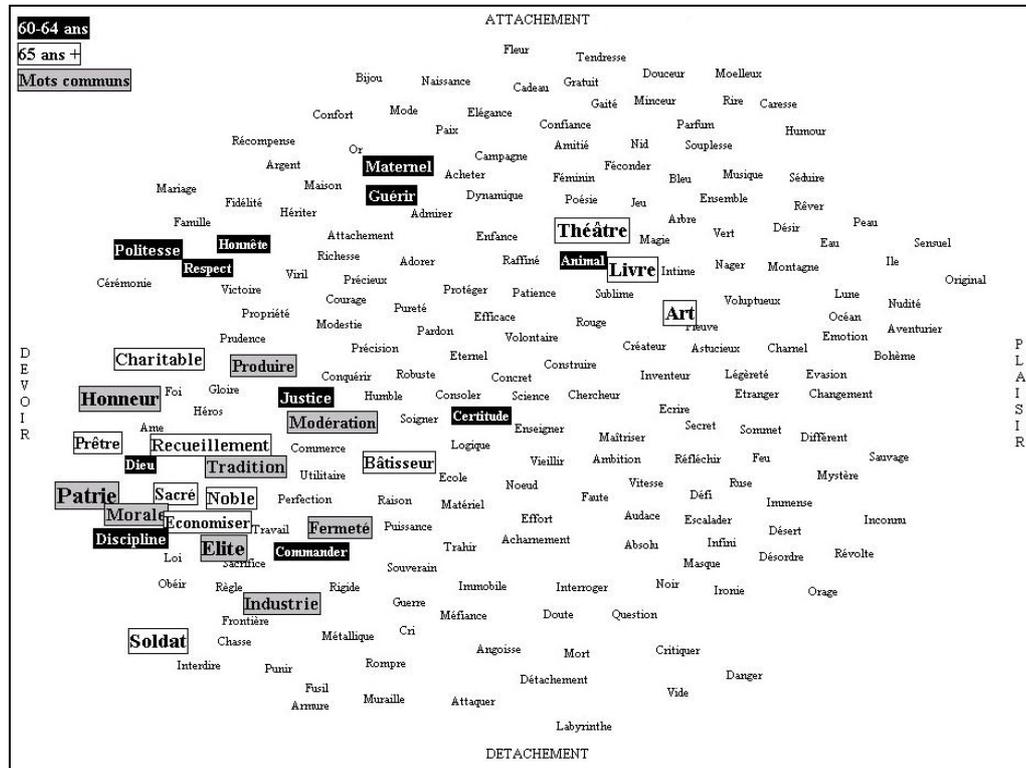
Ce résultat n'a rien de surprenant et chacun, qui se rappelle ses orageuses années de l'adolescence, en conviendra aisément. Mais que l'on puisse faire apparaître d'une manière à la fois aussi forte et évidente, ce lent glissement de l'âpre détachement, marqué par l'expression de la pulsion de destruction, de l'adolescent qui cherche encore son identité et sa place dans la société, vers une relation plus harmonieuse avec lui-même et avec les autres, illustre l'étonnant pouvoir de la statistique de révéler des phénomènes dont la présence, parce que trop diffuse, semble parfois défier l'observation.

## 7.5 Les mots caractéristiques de deux classes d'âge au delà de 60 ans

Les deux tranches d'âge (60-64 ans ; 65 ans et plus), comme on pouvait s'y attendre, investissent toutes deux massivement les valeurs de « Devoir ». Sur les vingt premiers mots les plus fortement sur-notés, quinze, pour l'une comme pour l'autre, sont des mots de « Devoir ». Sur ces quinze mots, neuf sont communs, *Morale, Patrie, Industrie, Honneur, Fermeté, Tradition, Elite, Modération* et *Produire*. Entre les six mots sur-notés par l'une et pas par l'autre, il n'existe qu'une très faible nuance – qui s'estompe, d'ailleurs, si les mots suivants sont pris en compte –, la tranche des 60-64 ans se distinguant par des valeurs plus « relationnelles », *Politesse, Respect, Honnête*, celle de 65 ans et plus, se détachant de ces contingences pour porter son regard vers un au-delà qui se profile, *Recueillement, Prêtre, Sacré*. Mais on observe des différences plus sensibles sur les axes 4 (« Sublimation / Matérialisme ») et 5 (« Idéalisation / Pragmatisme »).

Si les personnes des deux tranches d'âge investissent pareillement des valeurs de « Sublimation » et d'« Idéalisation », plutôt que de « Matérialisme » et de « Pragmatisme », les personnes de 65 ans et plus le font beaucoup plus nettement : six mots de « Sublimation », *Livre, Théâtre, Art, Charitable, Recueillement* et *Prêtre*, contre quatre, pour les 60-64 ans, *Justice, Maternel, Honnête, Dieu*, et cinq mots d'« Idéalisation »,

*Théâtre*<sup>1</sup>, *Noble*, *Recueillement*, *Prêtre* et *Sacré* contre un seul, *Dieu* – encore que le mot *Dieu* arrive en vingt-sixième position pour les 65 ans et plus, et avec une valeur de test de 3.93.



**Figure 7.5 Mots caractéristiques des 60-64 ans et des 65 ans et plus dans un plan sémiométrique schématisé [2,3]**

1. Un même mot, bien sûr, peut corrélérer avec plusieurs axes, chacune de ses « épiphanies » révélant une de ses dimensions sémantiques. Il est clair que le Théâtre de l'« Idéalisation », que connotent les mots *Cérémonie*, *Reine*, *Roi*, *Bijou*, *Château*, *Parfum*, *Masque*, *Magie*, etc., n'est pas le même que celui de la « Sublimation », que connotent les mots *Chercheur*, *Enseigner*, *Inventeur*, *Question*, *Art*, *Livre*, *Etranger*, *Paix*, *Pardon*, *Modestie*, etc. De même que le *Mariage* du « Devoir », que connotent les mots *Fidélité*, *Honnête*, *Respect*, *Politesse*, etc., n'est pas le même que celui de l'« Attachement », que connotent les mots *Confort*, *Maternel*, *Naissance*, *Confiance*, *Amitié*, *Tendresse* et *Douceur*, celui-là évoquant plutôt le mariage vu comme une institution, celui-ci, comme le cadre d'une relation d'amour entre deux êtres. Il y a pareillement une *Frontière* du « Détachement », que connotent les mots *Rigide*, *Fermer*, *Rompre*, *Muraille*, *Masque*, *Angoisse*, etc., c'est la frontière qui isole le sujet en rejetant l'autre, la mauvaise frontière, et une *Frontière* du « Devoir », que connotent les mots *Mariage*, *Propriété*, *Règle*, *Morale*, etc., c'est la frontière qui distingue et organise, créatrice et garante de l'ordre autant social que mental, la bonne frontière.

Ce glissement le long des axes 2, 4 et 5 est peut-être un effet de l'âge : après 65 ans, les pulsions commencent à s'exprimer moins fortement et il est donc plus facile de les contrôler et subséquemment, de les sublimer. Ce glissement est aussi, vraisemblablement, dû au passage, avec la retraite, d'une vie professionnelle active, qui laisse peu de place au rêve et à la réflexion, à une vie plus contemplative.

## **7.6 Hommes et femmes ayant attribué la note 7 au mot *Dieu***

Pour bien comprendre les résultats présentés sur la figure 7.6, il faut avoir présent à l'esprit la manière dont nous avons procédé : afin d'éliminer l'effet de sexe dont la prégnance aurait masqué les subtiles différences entre les notations féminines et masculines, nous avons opposé séparément les hommes et les femmes ayant attribué la note 7 au mot *Dieu*<sup>1</sup> à ceux lui ayant attribué une note inférieure. Les deux résultats ont été projetés ensuite sur une même carte.

---

1. Rappelons que les modalités proposées aux interviewés varient de - 3 (sensation très désagréable) à + 3 (sensation très agréable) mais, pour des raisons de commodité de saisie et de traitement, les notes sont ensuite recodées de 1 à 7.

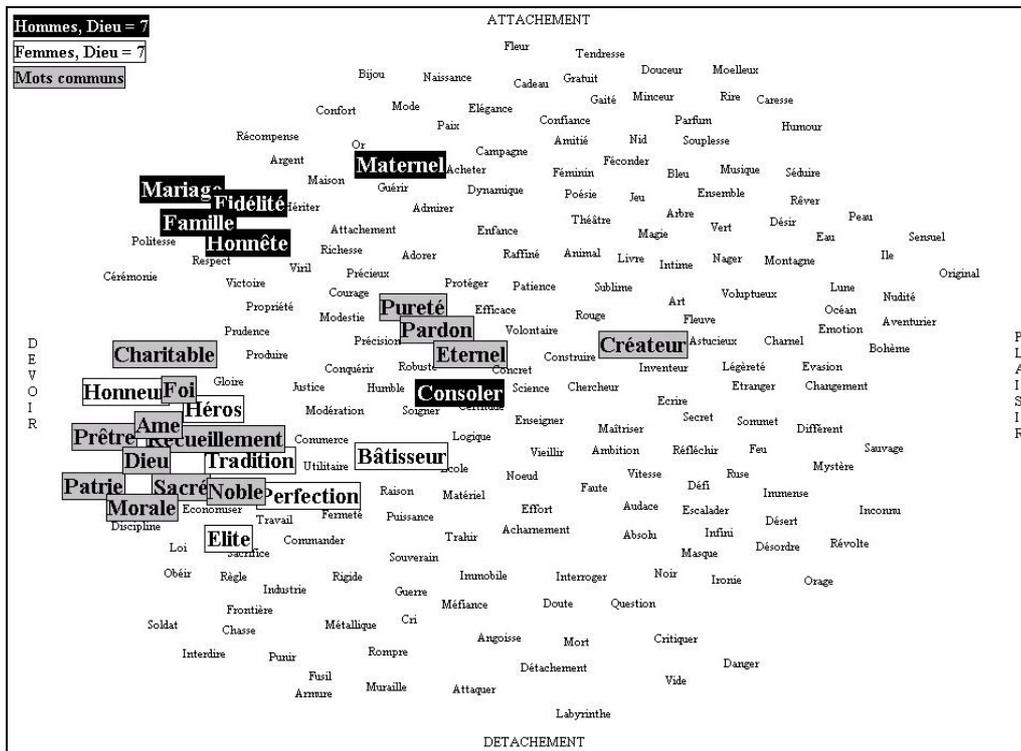


Figure 7.6. : Mots caractéristiques des hommes / femmes ayant donné la note 7 au mot *Dieu* dans un plan sémiométrique schématisé [2, 3]

Les mots sur fond noir sont ainsi les mots caractéristiques des hommes répondant que les évocations du mot *Dieu* provoquent en eux des sensations très agréables par rapport à ceux qui les trouvent moins agréables, voire très désagréables. Les mots encadrés sur fond blanc sont les mots caractéristiques des femmes pour qui le mot *Dieu* provoque des sensations très agréables, par rapport aux autres femmes. Les mots sur fond grisé sont leurs mots communs.

Le résultat tel qu’il apparaît sur la carte est si paradoxal qu’on est, dans un premier temps, tenté de penser à une quelconque erreur de codage ou de manipulation.

En effet, les mots sur-notés par les hommes sont des mots évoquant des valeurs traditionnellement réputées féminines, *Mariage*, *Fidélité*, *Famille*, *Maternel*, *Consoler*, tandis que les mots des femmes semblent

plutôt appartenir au registre des valeurs « masculines », *Honneur, Héros, Tradition, Bâtitteur, Perfection, Elite*<sup>1</sup>.

Mais, vérifications faites, il n'en est rien, et force est, donc, de nous interroger sur cet étrange résultat.

On constate d'abord que les mots qui caractérisent les hommes sont plus de l'ordre de la relation à la mère qu'à la femme d'une manière générale, qu'ils expriment avant tout une demande de protection contre l'abandon, *Fidélité* et le chagrin, *Consoler*, demande que le petit enfant, dans sa détresse, adresse à sa mère ; et que, symétriquement, les mots qui caractérisent les femmes expriment une même demande de réassurance, mais qui est, elle, adressée au père, ce « héros au regard si doux », ce *Bâtitteur* gardien d'une *Tradition* sur laquelle s'étaye le moi fragile de l'enfant en voie d'édification. Il s'agit, bien sûr, plus d'images idéales du père et de la mère, produits de fantasmes infantiles, que d'une réalité objective. Mais ne dit-on pas que le concept de Dieu est le produit d'une image idéale du père (ou de la mère) ? Alors, ce que montrerait ce résultat, c'est que, par l'effet d'un surprenant chassé-croisé, le Dieu des hommes est, statistiquement, plutôt maternel, ce serait le Dieu de miséricorde, tandis que le Dieu des femmes est plutôt paternel, ce serait le Dieu qui a dit à l'apôtre : « Pierre tu es pierre, et sur cette pierre je bâtirai mon Eglise, et les puissances du mal ne prévaudront pas contre elle »<sup>2</sup>.

Mais nous ne sommes ni des théologiens ni des psychologues, seulement des statisticiens qui s'efforcent de mesurer des phénomènes complexes avec le plus de précision et d'exactitude possible. Les interprétations présentées n'interviennent qu'à titre d'illustration. Elles voudraient attirer l'attention des spécialistes concernés, et suggérer des pistes de recherche.

---

1. Dans l'étude, portant sur 8 292 individus représentatifs de la population française, d'où ces données sont extraites, les six mots caractéristiques des hommes aimant le mot Dieu, sont tous des mots sur-notés par l'ensemble des femmes de l'échantillon, et avec des valeurs test non négligeables : *Maternel* (16.08), *Fidélité* (7.53), *Consoler* (5.86), *Famille* (4.26), *Mariage* (3.93) et *Honnête* (2.23). Les mots sur-notés par les femmes aimant le mot Dieu sont moins nettement « masculins ». Seuls deux sont significativement sur-notés par les hommes de l'échantillon : *Bâtitteur* (4.46) et *Elite* (2.36).

2. Matthieu, 16, 18.

## 7.7 Evolution des valeurs des Français entre 1996 et 1998

Les deux derniers exemples présentés sont relatifs à des comparaisons entre deux échantillons d'une même population, interrogés à deux moments différents.

Comme on l'a vu précédemment, dans les analyses synchroniques, la population de référence est (pratiquement) toujours le complément de la population étudiée. Dans les analyses diachroniques, elle est constituée par un second échantillon ayant la même structure mais interrogé antérieurement. On prendra donc garde, en observant les résultats, de ne pas confondre les deux approches. La première caractérise, à un moment donné, les valeurs d'une population donnée par rapport à une autre. Si cette autre population est le complément de la population étudiée par rapport à la population totale, ce qui est généralement le cas, on peut considérer que les mots qu'elle sur-note représentent ce que l'on peut désigner par son système de valeurs relatif<sup>1</sup>. La seconde approche caractérise une population donnée par rapport à une population homothétique observée à un moment antérieur. Dans ce cas, les mots qu'elle sur-note n'indiquent pas son système de valeurs mais l'évolution de celui-ci pendant la période séparant les deux observations. Elle est donc une mesure doublement relative.

Ce type d'analyses permet de mettre en évidence certains aspects de l'évolution à travers le temps des valeurs d'une population.

La présente application a été réalisée à partir de deux échantillons contrôlés pour leurs caractéristiques de base de 2 764 individus représentatifs de la population française. Ces échantillons sont extraits du panel postal de T.N. Sofres appelé Métascope. Si ce panel, comme tout panel, présente des biais non identifiables dus au fait que, par définition, il ne prend pas en compte les individus qui n'acceptent pas d'en faire partie, la constance des règles qui président à sa gestion en fait un terrain propice à des comparaisons temporelles. Il est *a priori* peu probable que les différences observées entre les deux échantillons puissent être uniquement des artefacts induits par des modifications de structure entre

---

1. Nous posons que l'ensemble des mots sur-notés par une population par rapport à un autre dont elle est le complément peut être considéré comme représentant son système de valeurs relatif, si celle-ci est son complément par rapport à une population représentative de l'univers dans lequel on se place. Par exemple, les mots surnotés par les lecteurs de tel titre de la presse quotidienne nationale par rapport à l'ensemble des lecteurs de cette presse représentent le système de valeurs de ceux-ci considéré du point de vue de cet ensemble. Il est clair que si l'on oppose les lecteurs de ce titre à l'ensemble des Français, le résultat sera notablement différent.

ces échantillons. Beaucoup d'indices, au contraire, nous conduisent à penser qu'elles sont le reflet d'une évolution dans la sensibilité, sinon les mentalités, des Français au cours de cette période. Les analyses réalisées dans la section 5.8 du chapitre 5 ont d'ailleurs mis en évidence la même tendance au cours de toute la période 1990 – 2002, sur un ensemble beaucoup plus étendu de six vagues d'enquêtes.

Sur la figure 7.7 les mots sur fond noir représentent les mots sur-notés par les Français de 1998 par rapport à ceux 1996.

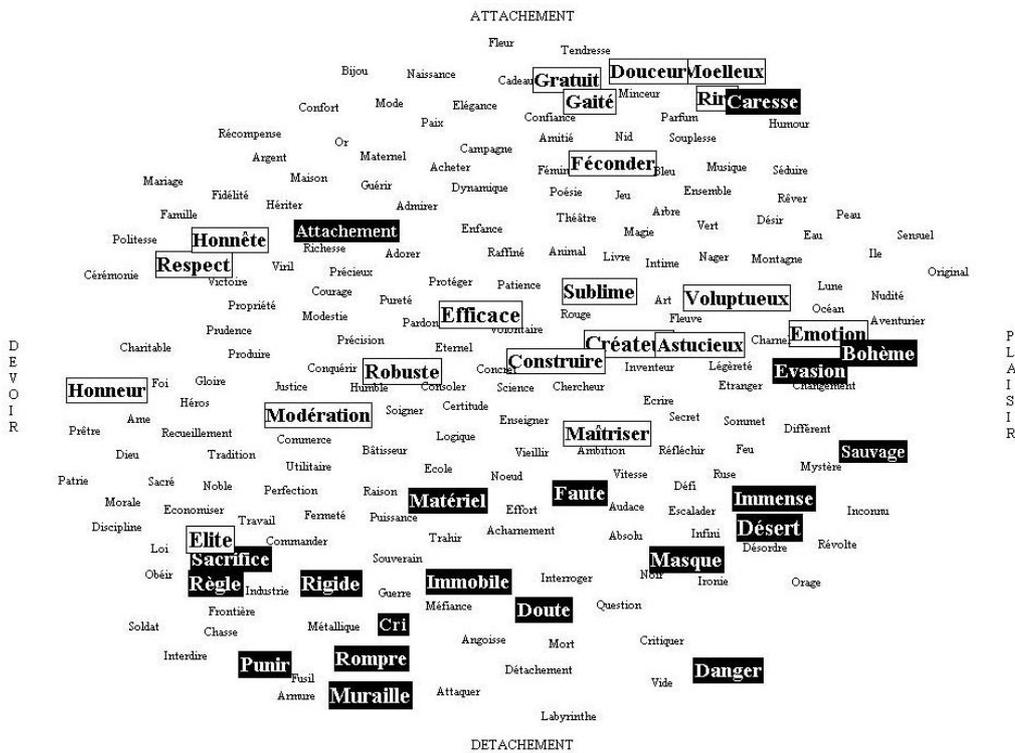


Figure 7.7 : Mots caractéristiques des années 1996 et 1998 dans un plan sémiométrique schématisé [2, 3]

Ce glissement le long de l'axe « Attachement / Détachement » caractérise, on l'a vu en section 5.8, l'ensemble de la période 1990 – 2002. Nous avons noté alors que l'attitude par rapport au questionnaire, matérialisée par la façon de noter, variait pendant la même période dans le sens d'une indifférence et d'une distance accrue vis-à-vis de l'enquête.

On pourrait interpréter cette évolution comme une crise de pessimisme ou, dans la veine de l'essai d'interprétation du chapitre 6, en

nous plaçant d'un point de vue freudien, une émergence de la pulsion de destruction dans la société française.

Cette interprétation rejoint d'ailleurs des diagnostics émanant d'horizons extrêmement divers, comme ceux d'Alain Ehrenberg ou de Tony Anatrella. De ce dernier, nous retiendrons surtout la description suivante : « Nous sommes dans le climat suicidaire d'une société qui privilégie les ruptures et les déliaisons. Il faut rompre et briser pour se donner l'impression d'exister (divorce et suicide pour s'affranchir, avortement et euthanasie par détresse ou par compassion : de curieuses motivations liées à la pulsion de mort). Il faut s'éclater pour mieux fuir et s'échapper de la vie en consommant des anesthésiants, des stimulants ou des hallucinogènes ; rester dans le primaire, le glauque (...) il faut casser le lien social, évacuer des désirs et des émotions dont on ne sait que faire et détruire toute réalité institutionnelle. »<sup>1</sup>

Le trait peut paraître outré, mais notre résultat, produit d'une mesure rigoureuse et aveugle, semble abonder dans le même sens. « Il faut rompre et briser », « casser le lien social », « rester dans le primaire, le glauque », mais n'est-ce pas ce qu'expriment les mots *Muraille* (valeur-test : 11.1), *Désert* (10.1), *Masque* (8.7), *Rigide* (8.6), *Punir* (8.3), *Rompre* (8), *Doute* (7.9), *Danger* (7.6), *Faute* (6.9), *Immobile* (6.6), *Sauvage* (6.1), *Cri* (5.6)... « Il faut (...) évacuer des désirs et des émotions dont on ne sait que faire », mais n'est-ce pas ce qu'exprime le rejet relatif des mots *Sublime* (-9.3), *Voluptueux* (-8.8), *Emotion* (-8.6), *Moelleux* (-7.7), *Douceur* (-7.2), *Rire* (-7.1), ... « et détruire toute réalité institutionnelle » : *Créateur* (-11.6), *Elite* (-10.6), *Robuste* (-9.2), *Maîtriser* (-7.8), *Construire* (-7), *Efficace* (-11), *Respect* (-8.9), *Honnête* (-7.8), *Honneur* (-7.7) ?

Il nous semble qu'il y a comme un air de famille entre le discours clinique précédent et cette litanie de mots, sélectionnés selon des critères statistiques aveugles. Les deux approches paraissent décrire un même phénomène de société, une même réalité sociale dans laquelle nous

---

1. Dans son livre « Non à la société dépressive » (Champs, Flammarion, Paris, 1995), Tony Anatrella, prêtre, psychanalyste et professeur de psychiatrie sociale, tente une explication de cette émergence de la pulsion de destruction dans les sociétés modernes, à partir du déficit d'idéal. On lira aussi avec le plus grand intérêt l'ouvrage d'Alain Ehrenberg, directeur de recherches au CNRS, « La fatigue d'être soi » (Ed. Odile Jacob, Paris, 1998) qui montre d'une manière assez convaincante comment nous avons substitué « à la vieille culpabilité bourgeoise (et à) la lutte pour s'affranchir de la loi des pères (Oedipe), la peur de ne pas être à la hauteur, le vide et l'impuissance qui en résultent (Narcisse) », c'est à dire comment nous avons substitué à la névrose « maladie de la loi », la dépression « maladie de l'insuffisance ».

sommes immergés et que nous avons spontanément peine à saisir dans sa globalité, faute d'être capables de nous en extraire pour pouvoir distinguer la part des faits et celle des fantasmes que nous projetons sur eux dans l'espoir – hélas souvent déçu ! – de neutraliser leurs effets anxiogènes.

Mais ce n'est, bien sûr, qu'une interprétation et nous laissons à chaque lecteur le soin de se faire la sienne.

Toutefois, l'étude d'évolution présentée dans la section 5.8 du chapitre 5, qui prend, elle, en compte six années d'enquêtes réparties entre 1990 et 2002, et qui porte sur plus de 15 000 interviews, confirme ce glissement vers le « Détachement » de l'ensemble de la population<sup>1</sup>. Rappelons que ce « Détachement » va de pair avec ce que l'on peut appeler *un détachement vis-à-vis des enquêtes elles-mêmes*, un désengagement, se manifestant par une moindre utilisation des notes extrêmes.

## 7.8 Conclusion

Caractériser des groupes d'individus (c'est-à-dire dans la pratique les consommateurs d'une marque ou d'un produit, les clients d'une grande entreprise de service, les lecteurs d'un quotidien ou d'un magazine...) par des mots, et positionner ces mots dans un espace dont les dimensions ont du sens, est le mode de valorisation de l'outil sémiométrique qui a été retenu dans les exemples de ce chapitre. Dans le contexte purement méthodologique du présent ouvrage, nous avons travaillé avec des groupes d'individus plus conventionnels définis par les variables socio-démographiques de base essentiellement, montrant ainsi les ouvertures et potentialités de l'outil dans les domaines des applications sociologiques ou socio-économiques.

Comme lors des chapitres précédents, les commentaires et interprétations sont surtout des invitations à commenter et interpréter. Ce sont des exemples... à propos d'exemples. Comment pourrait-il en être autrement compte tenu de la richesse et de la complexité de ce nouveau paysage statistico-sémantique ?

---

1. Notons que ce glissement concerne aussi bien les hommes que les femmes, bien que ces deux catégories de répondants occupent des positions significativement différentes le long de l'axe « Attachement / Détachement » en début comme en fin de période.

Les interprétations des travaux plus commerciaux ou industriels, s'inscrivant dans une stratégie et un cadre décisionnel bien circonscrits, seront en général plus ciblées. A l'heure de l'ergonomie, des interfaces hommes-machines, des analyses sensorielles, des médias et du multimédia, de la prise en compte généralisée de la qualité de la vie, les positionnements que permet l'outil sémiométrique apportent aux décideurs des matériaux pour la compréhension des maillons les plus énigmatiques de la chaîne production-consommation.

Nous aurons cependant atteint notre but si, à partir de ce choix restreint d'exemples, le lecteur peut avoir une idée de la richesse et de la spécificité de l'approche sémiométrique, et s'il peut maintenant mieux entrevoir les perspectives offertes par en matière de description et d'analyse.

## Conclusion

Jamais un questionnaire d'enquête n'aura été si simple dans la conception de ses questions. Jamais un questionnaire n'aura été si neutre, si primaire pourrait-on dire, pour ce qui est des réactions aux stimuli proposés : agréables ou désagréables. C'est donc une simple liste de mots qui, bien qu'elle puisse paraître quelque peu arbitraire, a permis de mettre en évidence des traits structuraux sous forme de dimensions stables.

Il aura fallu procéder à des complétions de dizaines de milliers de questionnaires pour valider ces traits structuraux dans le temps et dans l'espace et pour établir un repère socio-culturel permettant de positionner des individus et des groupes d'individus. Malgré l'épreuve périlleuse de la traduction, les mêmes traits apparaissent dans plusieurs pays différents, confirmant ainsi que les résultats ne dépendent pas de la composition exacte de la liste, et que la structure obtenue n'est ni locale, ni nationale....

Par sa transparence, sa distance vis-à-vis de l'actualité, ce repère est propice à des comparaisons intergénérationnelles, interrégionales et internationales. Mais on peut aussi positionner dans ce repère, en procédant à des enquêtes « ad hoc », des attitudes, des opinions, des produits, des supports médiatiques...

Cet outil étonnant a lui-même été objet de recherches, exposées en détail au cours des chapitres qui précèdent, et il continuera à l'être.

La seule hypothèque resterait le choix de la liste de mots. Plusieurs des travaux présentés dans cet ouvrage montrent précisément que ce choix ne constitue pas une hypothèque dirimante, et que le caractère opératoire de cette liste laisse augurer d'autres résultats et d'autres développements.

Une façon pragmatique de « vivre cette hypothèque », à défaut de la lever, est de dire que l'on observe non pas une structure universelle, mais la *trace* de ce qui pourrait être une structure universelle. Un « questionnaire total », aussi fictif et utopique que le « doctorat total » de Ionesco, permettrait d'observer cette structure. La liste de mots retenue n'en retient qu'une trace qui en est une sorte de projection, conditionnée à la fois par des critères de coût et de faisabilité. S'il est pratiquement impossible de prouver que cette « projection » est optimale vis-à-vis de ces critères, nous avons montré qu'elle n'est ni fragile, ni sans intérêt.

On ne s'étonnera pas de nous voir tirer des enseignements – et tracer des perspectives de travail et de recherche – dans des registres très divers, relevant de la statistique mathématique, de la sociologie, de la linguistique, enfin de la méthodologie d'enquête.

– *Comment passe-t-on du plaisir au sens ?*

Cette première interrogation est suscitée par la simplicité même du questionnement. Comment des questions fondées uniquement sur des sensations agréables ou désagréables peuvent-elles déboucher sur une structure sémantique locale aussi fine... ou sur une architecture globale de valeurs aussi stable ? L'émotion et la satisfaction jouent probablement un rôle important dans l'acquisition du langage comme dans tout apprentissage, et même, certainement, dans la construction progressive d'une langue. Au niveau de cet atome linguistique qu'est le mot, peu d'autres qualifications bipolaires pouvaient s'appliquer de façon exhaustive. Le caractère opératoire des réponses obtenues fut une preuve (agréable) du caractère judicieux de l'échelle choisie.

Des travaux de recherche complémentaires restent cependant à mener pour mieux analyser les mécanismes qui ont effectivement conduit de l'agréable au pertinent. Dans le cas du questionnaire ouvert utilisé au chapitre 4, pour lequel il ne s'agit plus de noter une liste fixe de mots, il est possible que le concept d'« importance » soit plus adéquat que celui d'« agrément », pour opérer une sélection et un filtrage plus sévères et

limiter ce faisant le caractère peut-être trop ludique, dans ce contexte libre, de l'exercice. Ce caractère ludique contribue, on l'a vu, à l'extrême dispersion du vocabulaire dans les réponses individuelles.

Les travaux sur les questions ouvertes devraient donc susciter de nouvelles recherches, comportant des phases d'expérimentation assez lourdes.

– ***Enseignements sur les enquêtes tendant à mesurer les évolutions... et aussi sur l'évolution des enquêtes***

Il est généralement admis que les mesures de différences entre réponses à des questions subjectives, différences dans le temps ou entre catégories, sont beaucoup plus justifiées et prudentes que des mesures absolues. Les résultats obtenus à propos de la dérive simultanée de l'instrument d'observation et de la réalité observée sont une (petite) pierre dans le jardin de ceux qui pensent qu'il est toujours licite d'observer les évolutions d'une question dont le libellé est fixe dans le temps. L'extrême homogénéité et la relative intemporalité des « questions » du questionnaire sémiométrique permet d'étudier assez finement ces phénomènes d'attitudes ainsi que les problèmes de participation et de notations. L'évolution dans le temps<sup>1</sup> illustre bien l'absence de neutralité et donc l'implication de l'instrument d'observation à la fois dans les résultats observés et dans la qualité de ces résultats<sup>2</sup>.

– ***Le modèle sémiométrique et la « double convergence »***

Parmi les apports fondamentaux des techniques statistiques connues sous le nom d'analyse des données<sup>3</sup>, nous citerons d'une part le renouveau d'intérêt apporté aux individus dans les enquêtes, ou plus

---

<sup>1</sup> Section 5.8 du chapitre 5.

<sup>2</sup> Notons que la dégradation très progressive de la participation aux enquêtes, qui concerne à la fois les refus *a priori* et la qualité des complétions est un phénomène assez largement observé dans les instituts de sondage et dans les instances de contrôles de la profession. Le questionnaire sémiométrique, et aussi, nous l'avons vu, le questionnaire ouvert du chapitre 4, peuvent contribuer à étudier certains aspects de ce phénomène, en particulier l'engagement et le degré de participation, pour les personnes ayant accepté de répondre.

<sup>3</sup> Ces techniques sont apparues dans les années 70, notamment en France, sous l'impulsion de Jean-Paul Benzécri (1973) qui préconisait une totale réécriture de la science statistique prenant en compte les nouvelles possibilités de l'informatique.

exactement à la « dimension individu »<sup>1</sup>, et, d'autre part la prise en considération de variables pouvant être très nombreuses. Corrélativement, les modèles probabilistes classiques, peu réalistes parce que beaucoup plus contraignants dans le cas *multivarié* que dans le cas d'une ou deux variables, sont moins utilisés, laissant souvent la place à des méthodes de validation empiriques à base de calculs intensifs comme les techniques de *bootstrap* auxquelles il a été largement fait appel dans cet ouvrage. A la fois les individus gagnaient en importance, et les variables se banalisaient, autrement dit, les lignes (en général : les individus) et les colonnes (en général : les variables) du tableau classique de données jouaient des rôles de plus en plus symétriques<sup>2</sup>.

D'où une potentialité de double convergence selon les lignes (individus) et/ou les colonnes (variables) qui généralise en quelque sorte la loi des grands nombres et d'autres lois limites de la statistique, fondées sur la multiplication à l'infini des seuls individus... Ces problèmes sont simplement évoqués ici, mais appartiennent à une direction de recherche importante.

Ce nouveau regard sur les tableaux de données fut, et sera, car il s'en faut de beaucoup que ces innovations aient pénétré toutes les disciplines, très enrichissant pour beaucoup d'applications en sciences humaines. La multiplication des variables change la nature des problèmes.

Les praticiens d'études de marché savent bien que quelques interviews très riches valent parfois mieux que des centaines d'interviews pauvres... ce que les cliniciens ont su de tous temps. A la limite, risquons la boutade suivante : les *Essais de Montaigne* (échantillon de taille : 1) nous apprennent beaucoup plus, sur l'homme en général et l'homme du XVIème siècle en particulier, que ne l'aurait fait une enquête de satisfaction auprès de quelques milliers de ses contemporains. Montaigne en est d'ailleurs bien conscient lorsqu'il écrit : « Chaque homme porte la forme entière de l'humaine condition », ce que Sartre exprimera de façon plus lapidaire : « Tout homme est tout l'homme »<sup>3</sup>.

---

<sup>1</sup> Les individus statistiques (ou observations) ne servaient qu'à calculer des moyennes ou des corrélations, et, détail significatif, n'étaient que rarement représentés sur les graphiques avec des identificateurs.

<sup>2</sup> D'ailleurs l'opération mathématique de compression qui est à la base des méthodes factorielles, la décomposition aux valeurs singulières, fait jouer un rôle symétrique aux lignes et aux colonnes (cf. annexe A1.2).

<sup>3</sup> Sartre J.-P. (1980) *Les Mots*, Gallimard, Paris.

Et la sémiométrie dans tout ça ? Elle constitue l'exemple emblématique de ce paradigme, le *leading case* de la double convergence<sup>1</sup>. En effet, il existe une population parente des lignes (la population française, par exemple) et une population parente des colonnes (le vocabulaire de la langue française, pour le même exemple). Peu de questionnaires se trouvent dans une situation théorique aussi favorable. Mais prendre simplement des mots au hasard dans un dictionnaire engendrerait, au centuple, les difficultés<sup>2</sup> rencontrées à propos du questionnement ouvert du chapitre 4. Parallèlement aux travaux théoriques qu'il serait intéressant de mener sur ce modèle statistique complexe, les expérimentations à venir devront aussi porter sur les problèmes de sélection et d'échantillonnage de mots.

– ***Structure globale, fractale ou individuelle ?***

Si l'on effectue des mesures sur des individus, les corrélations traduisent bien une liaison mesurée au niveau individuel. Certains notent bien à la fois les mots *Discipline* et *Patrie*, d'autres les notent moins bien, mais il y a peu d'individus qui notent bien *Discipline* et mal *Patrie*. La corrélation vient précisément de cette cohérence interne à chaque individu, cohérence que la composition du questionnaire permet de vérifier, et que les quelques légères redondances existant dans le questionnaire permettent de conforter. S'il existe une structure de corrélation, significative statistiquement, sur un échantillon de  $n$  individus pris au hasard, on doit retrouver la même structure sur un autre échantillon au hasard, ou dans un sous échantillon pris également au hasard.

Ce n'est pas à propos de ce phénomène classique que l'on a pu parler de fractalité, mais à propos du phénomène plus paradoxal suivant : des catégories semblent s'opposer sur certains axes de façon significative. En segmentant selon ces catégories, on prend un sous-échantillon délibérément non aléatoire. Et l'on s'étonne, cette fois, de retrouver la même structure à l'intérieur de ces sous-catégories. Dans le cas de

---

<sup>1</sup> L'exemple le plus pertinent permettant d'étudier l'évolution (la limite) du phénomène lorsque le nombre de lignes et le nombre de colonnes augmentent indéfiniment.

<sup>2</sup> Ouvrons le Petit Robert au hasard et pointons, aveuglément, sur des mots : « *démission, hectare, monazite, papaine, prélude, rabattre, toupie...* ». Les limites de faisabilité sont tout de suite perceptibles : il faut se restreindre aux mots connus *a priori* de tout échantillon de répondants (une petite fraction des entrées d'un dictionnaire), puis introduire une série de contraintes qui conduisent finalement à un protocole analogue (mais pas forcément identique) à celui qui a prévalu lors des premiers essais sémiométriques (*cf.* chapitre 1).

l'opposition homme-femme, la surprise vient peut-être d'une identification hâtive de certains archétypes. La liaison très significative entre les mots de l'« Attachement » et l'appartenance au sexe féminin fait penser que la différence entre sexes est seule responsable de la formation de l'axe « Attachement / Détachement ». L'existence de la même opposition à l'intérieur de chaque sexe montre donc que l'opposition entre sexes n'est en aucune façon responsable de la structure<sup>1</sup>.

En fait, ce que l'on appelle ici *fractalité* pourrait être simplement la validité « quasi-individuelle » d'une structure qui serait alors à la fois sémantique et *psycho-culturelle*. Les proximités locales observées sur les cartes sémiométriques sont influencées par les liens sémantiques qui sont universels parce que propres au vocabulaire d'une langue commune à tous les répondants. Les grandes oppositions observées sur les axes principaux proviennent de facteurs latents qui pourraient dépendre d'une organisation psychique indissociable d'un cadre culturel général (de références, de valeurs) également commun à tous les répondants, avec parfois des nuances déclinables selon certaines catégories socio-démographiques. Le lecteur aura compris que les conditionnels désignent en fait des thèmes d'approfondissement et de recherche à venir.

#### – *Pour une statistique structurale*

Ce n'est pas une activité courante de la statistique que de mettre en évidence des structures cachées, bien que cette démarche soit historiquement apparue dès le début du vingtième siècle - dans le cadre de modèles spécifiques - avec les travaux des psychologues factorialistes.

Nous nous sommes intéressés dans cet ouvrage à la notion de *fait statistique*, en considérant les outils utilisés (analyses en composantes principales, classifications, cartes de Kohonen) comme de simples instruments d'observation du *multidimensionnel* au même titre que les microscopes ou des appareils radiographiques<sup>2</sup> sont des instruments d'observation du *petit* ou de *l'opaque*. En utilisant une importante

---

<sup>1</sup> Celle-ci est stable même à l'intérieur de la plupart des catégories socio-démographiques usuelles... mais l'ordre des axes peut cependant être modifié : ainsi le second axe, fortement lié à l'âge des répondants, passe au troisième rang si l'on analyse une tranche d'âge étroite comme la tranche d'âge 40-45 ans.

<sup>2</sup> Les théories et mécanismes qui président à la réalisation d'un microscope n'ont pas ou peu de rapport avec les propriétés morphologiques des animalcules, cellules ou autres objets observés. Ils ne supposent aucune modélisation des objets observés.

panoplie de procédures de validation (méthodes de rééchantillonnage spécifiques, recodages, transformations analytiques diverses, procédures empiriques extensives), nous avons montré que l'on pouvait extraire des *structures valides* sans se contenter de contempler des *patterns suggestifs*. En fait, on commence seulement à pouvoir mettre en œuvre les outils qui permettent de donner un caractère de scientificité aux techniques d'analyse des données conçues il y a plusieurs décennies<sup>1</sup>... et c'est cette démarche de recherche de *faits statistiques* ou de *traits structuraux validés* qui a été désignée par l'expression *statistique structurale*. Les travaux présentés dans les chapitres précédents pourraient paver le début d'un chemin menant vers une telle branche disciplinaire.

---

<sup>1</sup> Si les statisticiens spécialistes ont utilisé ces instruments d'observation avec le discernement et le recul critique indispensable, beaucoup d'usages malheureux furent cependant à déplorer, en particulier à cause de l'absence de cadre conceptuel pouvant accueillir les nouvelles visualisations. L'analogie est assez frappante avec la diffusion des premiers microscopes telle que la décrit François Jacob dans *Logique du vivant* : « Quand Leeuwenhoek contemple pour la première fois une goutte d'eau au microscope, il y trouve un monde inconnu, des formes qui grouillent... Mais la pensée n'a alors que faire de tout ce monde... cette découverte permet seulement d'alimenter les conversations... ».

## ANNEXE 1

# Quelques éléments de statistique multidimensionnelle

Les méthodes d'analyse statistique exploratoire utilisées au cours des chapitres précédents visent à mettre en forme de vastes ensembles de données, à en dégager des structures et à valider ces structures. Elles relèvent de la statistique exploratoire multidimensionnelle, de l'analyse des données, ou encore du *Data Mining*, ces trois désignations étant à peu près équivalentes dans le cadre des utilisations de cet ouvrage. Nous avons utilisé à leur propos l'expression *statistique structurale* pour marquer l'importance accordée à la phase de validation des structures. Ces méthodes généralisent la statistique descriptive classique et utilisent des outils mathématiques assez intuitifs, mais plus complexes que les moyennes, variances et coefficients de corrélations empiriques de la statistique descriptive.

Sont présentés dans cette annexe les principes des techniques utilisées ou évoquées dans les chapitres précédents, l'analyse en composantes principales étant la technique d'analyse factorielle de base des applications sémiométriques. Certains développements de l'ouvrage noté [SEM 2006]<sup>1</sup> seront repris ; ils seront complétés par des travaux plus récents sur les méthodes de validation, et en particulier sur les techniques

---

1. *Statistique Exploratoire Multidimensionnelle*, [Visualisation et inférence en fouille de données], 4<sup>ème</sup> ed. L. Lebart, M. Piron, A. Morineau. Dunod, 2006.

dites de *bootstrap*, sur les cartes de Kohonen, ou sur des techniques d'analyse moins utilisées comme l'analyse logarithmique.

## A1.1 Rappel des principes des méthodes exploratoires multidimensionnelles

Les méthodes exploratoires multidimensionnelles recouvrent un grand nombre de techniques qui ont pour objectif de décrire et synthétiser l'information contenue dans de vastes tableaux de données.

### A1.1.1 Représentation géométrique et nuages de points

Au départ, les données se présentent sous forme de grands tableaux rectangulaires, notés **X**. Les lignes ( $i=1, \dots, n$ ) du tableau représentent les  $n$  individus, les sujets enquêtés par exemple, et les colonnes ( $j=1, \dots, m$ ) les  $m$  variables qui peuvent être des mesures, des caractéristiques ou encore des notes relevées sur les individus.

Afin de comprendre le principe des méthodes de statistique exploratoire multidimensionnelle, il est utile de représenter de façon géométrique l'ensemble des  $n$  individus ( $n$  lignes) et l'ensemble des  $m$  variables ( $m$  colonnes) comme deux *nuages de points*, chacun des deux ensembles étant décrit par l'autre. On définit alors, pour les deux nuages, des distances entre les points-lignes et entre points-colonnes qui traduisent les associations statistiques entre les individus (lignes) et entre les variables (colonnes).

**Tableau A1.1 :**  
**Exemple de tableau X de notes (de 1 à 7)**  
**attribuées à : m = 7 mots, par n = 12 répondants**

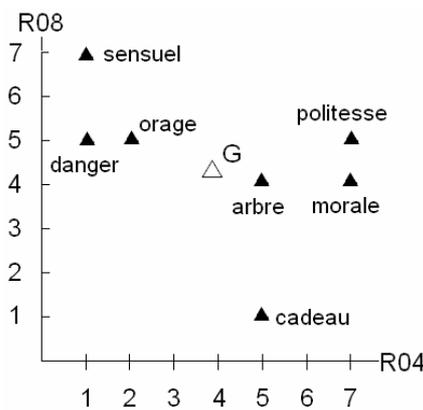
<i>mots</i> <i>Répondants</i>	arbre	cadeau	danger	morale	orage	politesse	sensuel
R01	7	4	2	2	3	1	6
R02	6	3	1	2	4	1	7
R03	4	5	3	4	3	4	3
R04	5	5	1	7	2	7	1
R05	4	5	2	7	1	6	2
R06	5	7	1	5	2	6	5
R07	4	2	1	3	5	3	6
R08	4	1	5	4	5	4	7
R09	6	6	2	4	7	5	5
R10	6	6	3	5	3	6	6
R11	7	7	6	7	7	6	7
R12	2	2	1	2	1	3	2

Dans le cas de la sémiométrie, un mot (variable) est un point dont les coordonnées sont les notes données par les  $n$  individus (répondants) : le nuage des  $m$  mots se situe dans un espace à  $n$  dimensions. De même, un individu est un point dont les coordonnées sont les notes attribuées aux  $m$  mots ; le nuage des  $n$  individus se trouve dans un espace à  $m$  dimensions.

Les figures A1.1 et A1.2 illustrent, à partir du tableau A1.1 contenant les notes attribuées à 7 mots par 12 répondants, la représentation de ces deux nuages de points intrinsèquement liés.

Le nuage des points-mots est construit dans l'espace des individus, ici à partir seulement de deux individus, R04 et R08, car deux dimensions rendent possible un graphique dans un plan (*cf.* figure A1.1).

	arbre	cadeau	danger	morale	orage	politesse	sensuel
R01	7	4	2	2	3	1	6
R02	6	3	1	2	4	1	7
R03	4	5	3	4	3	4	3
R04	5	5	1	7	2	7	1
R05	4	5	2	7	1	6	2
R06	5	7	1	5	2	6	5
R07	4	2	1	3	5	3	6
R08	4	1	5	4	5	4	7
R09	6	6	2	4	7	5	5
R10	6	6	3	5	3	6	6
R11	7	7	6	7	7	6	7
R12	2	2	1	2	1	3	2



**Figure A1.1 : Représentation du nuage des mots dans l'espace des deux répondants « R04 » et « R08 »**

De la même façon, le nuage des 12 répondants est construit dans l'espace des variables, ici à partir de deux mots, *Morale* et *Sensuel*, c'est-à-dire dans un espace de deux dimensions (cf. figure A1.2).

Pour chacun des nuages est représenté le *point moyen* appelé aussi *centre de gravité*. Il s'agit de G pour le centre de gravité des notes attribuées par les répondants (cf. figure A1.1) et de G' pour celui des répondants ayant notés les deux mots retenus.

	arbre	cadeau	danger	morale	orage	politesse	sensuel
R01	7	4	2	2	3	1	6
R02	6	3	1	2	4	1	7
R03	4	5	3	4	3	4	3
R04	5	5	1	7	2	7	1
R05	4	5	2	7	1	6	2
R06	5	7	1	5	2	6	5
R07	4	2	1	3	5	3	6
R08	4	1	5	4	5	4	7
R09	6	6	2	4	7	5	5
R10	6	6	3	5	3	6	6
R11	7	7	6	7	7	6	7
R12	2	2	1	2	1	3	2

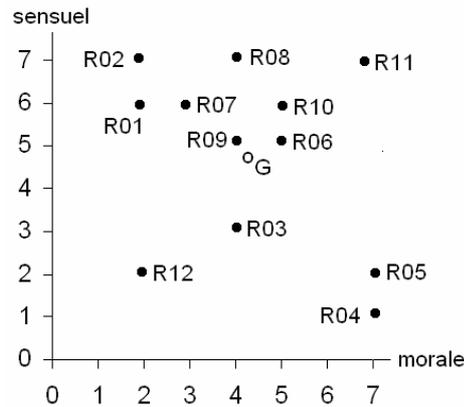


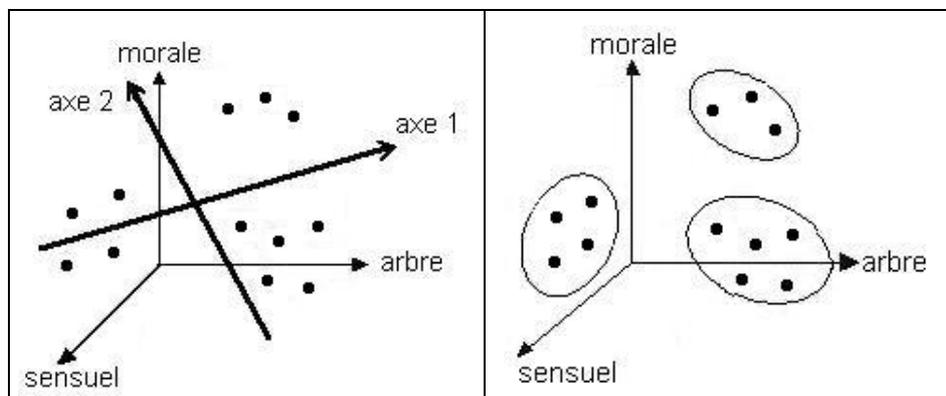
Figure A1.2 : Représentation du nuage des répondants dans l'espace des mots « *Sensuel* » et « *Morale* »

### A1.1.2 Principe et méthodes d'analyse

S'il est toujours possible de calculer des distances entre les lignes et des distances entre les colonnes d'un tableau **X**, il n'est pas possible de les visualiser de façon immédiate (les représentations géométriques associées impliquant en général des espaces à plus de deux ou trois

dimensions) : il est nécessaire de procéder à des transformations et des approximations pour en obtenir une représentation plane.

Les tableaux de distances associés à ces représentations géométriques (simples dans leur principe, mais complexes en raison du grand nombre de dimensions des espaces concernés) peuvent être décrits par les deux grandes familles de méthodes que sont les méthodes factorielles et la classification. La première consiste à rechercher les directions principales selon lesquelles les points s'écartent le plus du point moyen. La seconde consiste à rechercher des groupes ou classes d'individus qui soient les plus homogènes possibles (figure A1.3).



*Méthode factorielle*  
(recherche des directions principales)      *Méthode de classification*  
(recherche de groupes homogènes)

**Figure A1.3 : Deux grandes familles de méthodes**

Ces méthodes impliquent souvent de la même manière les individus (lignes) et les variables (colonnes). La confrontation des espaces d'individus et de variables enrichit les interprétations.

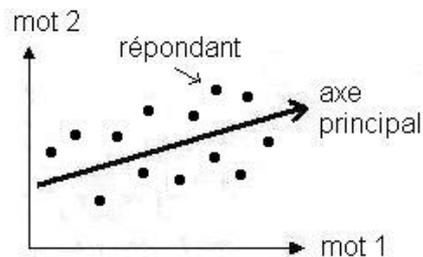
## A1.2 Les méthodes factorielles : aspects techniques

Les méthodes factorielles<sup>1</sup> permettent de gérer simultanément des quantités importantes de données et leur système de corrélations et, par une technique réalisant une sorte de *compression*, d'en dégager la structure interne, notamment sous forme de graphique-plans.

1. Elles comprennent dans la littérature statistique française des trente dernières années toutes les techniques de représentation utilisant des « axes principaux »: analyse en composantes principales, analyse des correspondances simples et multiples, analyse factorielle dite classique (en anglais : *factor analysis*) ou analyse en facteurs communs et spécifiques.

### - Recherche des sous-espaces factoriels

L'objectif est de rechercher des sous-espaces de dimensions réduites (entre trois et dix, par exemple) qui ajustent au mieux le nuage de points-individus et celui des points-variables, de façon à ce que les proximités mesurées dans ces sous-espaces reflètent autant que possible les proximités réelles. On obtient ainsi un espace de représentation, l'espace factoriel, défini par les axes principaux d'inertie et l'on représente les points du nuage dans ce système d'axes (*cf.* figure A1.4). Ces axes réalisent les meilleurs ajustements de l'ensemble des points selon le critère classique des moindres carrés, qui consiste à rendre minimale la somme des carrés des écarts entre les points et les axes.



**Figure A1.4 : Ajustement du nuage des points-individus dans l'espace des mots**

Le premier de ces axes correspond à la droite d'allongement maximum du nuage, le second maximise le même critère en étant assujéti à être orthogonal au premier, et ainsi de suite pour les axes suivants qui sont tous orthogonaux entre eux. Cette orthogonalité traduit l'indépendance (en fait, la non-corrélation) des axes.

$\mathbf{X}$  désigne le tableau de données ayant subi des transformations préliminaires (variables centrées réduites, par exemple),  $\mathbf{X}'$  son transposé.

Soit  $\mathbf{u}_1$  le vecteur unitaire qui caractérise le premier axe.  $\mathbf{u}_1$  est alors le vecteur propre de la matrice  $\mathbf{X}'\mathbf{X}$  correspondant à la plus grande valeur propre  $\lambda_1$  [*cf.* SEM 2006].

Plus généralement, le sous-espace à  $q$  dimensions qui ajuste au mieux (au sens des moindres carrés) le nuage est engendré par les  $q$  premiers vecteurs propres de la matrice  $\mathbf{X}'\mathbf{X}$  correspondant aux  $q$  plus grandes valeurs propres.

La procédure d'ajustement est exactement la même pour les deux nuages. On démontre alors qu'il existe des relations simples liant les axes

calculés dans les deux espaces, celui des individus et celui des variables (*relations de transition*).

Le vecteur des coordonnées des points sur chacun des axes, appelé *facteur*, est une combinaison linéaire des variables initiales. On dénote par  $\psi_\alpha$  et  $\phi_\alpha$  les facteurs correspondant à l'axe  $\alpha$  respectivement dans l'espace noté  $\mathbb{R}^m$  (espace dont les  $n$  points ont pour coordonnées sont les  $m$  mots) et dans l'espace noté  $\mathbb{R}^n$  (espace dont les  $m$  points ont pour coordonnées sont les  $n$  individus).

Les deux nuages de points, celui des mots et celui des répondants, sont intrinsèquement liés et révèlent exactement les mêmes structures : dans un cas, les facteurs décrivent les corrélations entre les mots, dans l'autre les associations entre les répondants.

Les plans factoriels de visualisation utilisés tout au long de cet ouvrage correspondent chacun à un couple de facteurs.

Le plan sémiométrique le plus utilisé est le plan  $(\phi_2, \phi_3)$ .

Les éléments (mots ou individus) qui participent au calcul des axes sont les *éléments actifs*. On introduit aussi dans l'analyse des *éléments supplémentaires* (ou *illustratifs*) qui ne participent pas à la formation des axes, mais qui sont projetés *a posteriori* dans les plans factoriels et peuvent aider à leur interprétation (cf. section A1.2.4).

#### - *Techniques de base et méthodes dérivées*

La nature des informations, leur codage dans le tableau de données, les spécificités du domaine d'application vont introduire des variantes au sein des méthodes factorielles.

Celles qui sont utilisées ici ne sont en fait que des dérivées de deux techniques fondamentales, l'analyse en composantes principales et l'analyse factorielle des correspondances.

L'analyse en composantes principales s'applique à un tableau de mesures numériques et sera utilisée, dans le cadre de la sémiométrie, pour traiter un tableau de notes.

Les exemples d'analyse de données textuelles présentée au chapitre 4 reposent sur l'analyse factorielle des correspondances appliquée aux tableaux de contingence lexicaux.

### A1.3 L'Analyse en Composantes Principales : aspects techniques

L'Analyse en Composantes Principales (Hotelling, 1933) s'applique à des variables à valeurs numériques (des mensurations, des taux, des mots etc.) représentées sous forme d'un tableau rectangulaire de mesures  $\mathbf{R}$  de terme général  $r_{ij}$  dont les colonnes sont les variables et les lignes représentent les individus sur lesquels ces variables sont mesurées. En sémiométrie, les variables sont donc les mots; les lignes les répondants et les valeurs numériques, les notes.

#### A1.3.1 Interprétations géométriques

Les représentations géométriques entre les lignes d'une part et entre les colonnes d'autre part du tableau de données permettent de visualiser les proximités respectivement entre les individus et entre les variables (cf. figures A1.1 et A1.2 ci-dessus).

Dans  $\mathbb{R}^m$ , deux points-individus sont très voisins si, dans l'ensemble, leurs  $m$  coordonnées sont très proches. Les deux répondants concernés sont alors caractérisés par des valeurs presque égales pour chaque variable. La distance utilisée est la distance euclidienne usuelle.

Dans  $\mathbb{R}^n$ , si les valeurs prises par deux variables particulières sont très voisines pour tous les répondants, ces variables seront représentées par deux points très proches dans cet espace. Cela peut vouloir dire que ces variables mesurent une même chose ou encore qu'elles sont liées par une relation particulière.

Mais les unités de mesure des variables peuvent être très différentes et rendre alors nécessaire des transformations du tableau de données.

#### A1.3.2 Problème d'échelle de mesure et transformation des données

On veut que la distance entre deux individus soit indépendante des unités des variables pour que chaque variable joue un rôle identique. Pour cela, on attribue à chaque variable  $j$  la même dispersion en divisant chacune de ses valeurs par leur écart-type  $s_j$  avec  $s_j^2 = \frac{1}{n} \sum_{i=1}^n (r_{ij} - \bar{r}_j)^2$ .

Par ailleurs on s'intéresse à la manière dont les individus s'écartent de la moyenne. On place alors le point moyen au centre de gravité du nuage

des individus. Les coordonnées du point moyen sont les valeurs moyennes des variables notées  $\bar{r}_j = \frac{1}{n} \sum_{i=1}^n r_{ij}$ . Prendre ce point comme origine revient à soustraire pour chaque variable sa moyenne  $\bar{r}_j$ .

On corrige ainsi les échelles en transformant le tableau de données  $\mathbf{R}$  en un nouveau tableau  $\mathbf{X}$  de la façon suivante :

$$x_{ij} = \frac{r_{ij} - \bar{r}_j}{s_j \sqrt{n}}$$

Les variables ainsi réduites et centrées ont toutes une variance,  $s^2(x_j)$ , égale à 1 et une moyenne,  $\bar{x}_j$ , nulle et deviennent comparables. D'autres transformations préalables sont possibles (cf. section 2.5 du chapitre 2).

### A1.3.3 Analyse du nuage des $n$ répondants

La transformation des données amène à effectuer une translation de l'origine au centre de gravité de ce nuage et à changer (dans le cas de l'analyse dite normée) les échelles sur les différents axes.

Pour réaliser l'analyse du nuage des points-répondants dans  $\mathbb{R}^m$ , la matrice  $\mathbf{X}'\mathbf{X}$  à diagonaliser dans cet espace, est la matrice des corrélations (dont la figure A1.4 fournit un exemple) qui a pour terme général :

$$c_{jj'} = \sum_{i=1}^n x_{ij} x_{ij'} = \frac{1}{n} \sum_i \frac{(r_{ij} - \bar{r}_j)(r_{ij'} - \bar{r}_{j'})}{s_j s_{j'}}$$

$c_{jj'}$  est le coefficient de corrélation entre les variables  $j$  et  $j'$ .

Les coordonnées des  $n$  points-individus sur l'axe factoriel  $\mathbf{u}_\alpha$  sont les  $n$  composantes du vecteur  $\boldsymbol{\psi}_\alpha = \mathbf{X}\mathbf{u}_\alpha$ .

La figure A1.4-a illustre la représentation du nuage des répondants pour le tableau de 12 répondants ayant noté 7 mots (tableau déjà présenté en section A1.1) dans le plan principal (2, 3)<sup>1</sup>. Les répondants R01 et R02 ont donné, de la même façon, des notes très contrastées et ont donné des notes élevées à *Arbre* et *Sensuel* et des notes faibles à *Morale* et *Politesse* ; ils sont par conséquent proches dans le plan et se différencient

1. Le plan (2, 3) a été considéré comme le plan sémiométrique principal compte tenu du caractère particulier du premier axe (axe dit *de taille*, cf. chapitre 5).

des répondants R05 et R04 qui se sont exprimés de façon inverse sur les mots. Le répondant R08 se distingue en ayant très bien noté *Danger* sans pour autant bien noter les autres mots, alors que R11 a bien noté tous les mots.

### A1.3.4 Analyse du nuage des variables (mots)

Les coordonnées factorielles  $\varphi_{\alpha j}$  des points-variables sur l'axe  $\alpha$  sont les composantes de  $\mathbf{u}_\alpha \sqrt{\lambda_\alpha}$  et l'on a :

$$\varphi_{\alpha j} = \text{cor}(j, \Psi_\alpha)$$

La coordonnée  $\varphi_{\alpha j}$  d'un point-variable  $j$  sur un axe  $\alpha$  n'est autre que le *coefficient de corrélation* de cette variable avec le facteur  $\Psi_\alpha$  (combinaison linéaire des variables initiales) considéré lui-même comme une variable artificielle dont les coordonnées sont constituées par les  $n$  projections des individus sur cet axe.

Les axes factoriels étant orthogonaux deux à deux, on obtient ainsi une série de variables artificielles non corrélées entre elles, appelées *composantes principales*<sup>1</sup>, qui synthétisent les corrélations de l'ensemble des variables initiales.

Sur la figure A1.4-b, comme sur la matrice de corrélations correspondante, *Politesse* et *Morale* sont très corrélés et dans une moindre mesure *Orage* et *Sensuel*. On retrouve bien les comportements des répondants où R01 et R02 vont dans la direction des bons noteurs d'*Arbre* et de *Sensuel* et des mauvais noteurs de *Morale* et *Politesse* à l'inverse de R04 et R05.

Les variables fortement corrélées avec un axe vont contribuer à la définition de cet axe<sup>2</sup>. Cette corrélation se lit directement sur le graphique puisqu'il s'agit de la coordonnée du point-variable  $j$  sur l'axe  $\alpha$ .

---

1. L'analyse en composantes principales ne traduit que des liaisons linéaires entre les variables. Un coefficient de corrélation faible entre deux variables signifie donc que celles-ci sont indépendantes linéairement, alors qu'il peut exister une relation non linéaire.

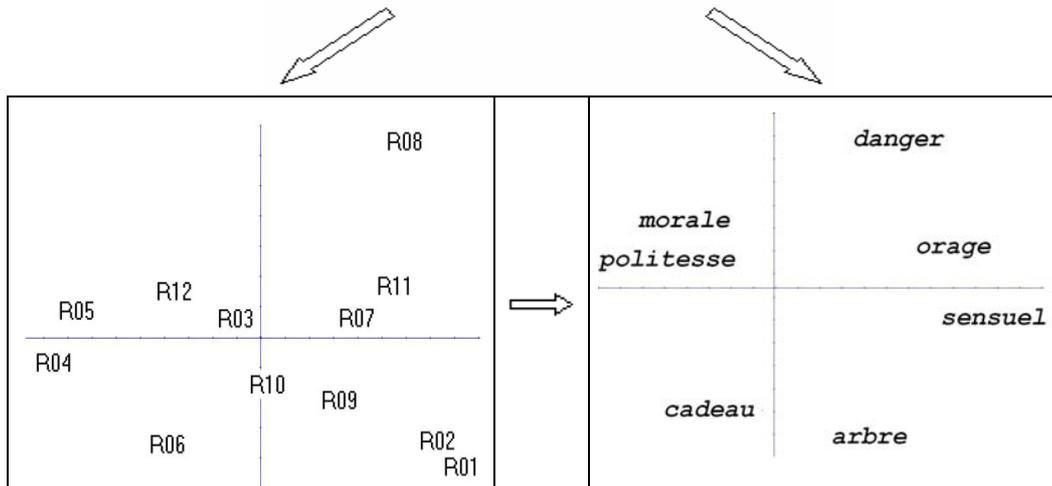
2. L'exemple n'est bien évidemment pas suffisamment représentatif pour que le plan puisse être interprété. Il a juste vocation à rapprocher le tableau de données des résultats.

**Tableau de notes (1 à 7) données à 7 mots par 12 répondants**

	arbre	cadeau	danger	morale	orage	politesse	sensuel
<b>R01</b>	7	4	2	2	3	1	6
<b>R02</b>	6	3	1	2	4	1	7
<b>R03</b>	4	5	3	4	3	4	3
<b>R04</b>	5	5	1	7	2	7	1
<b>R05</b>	4	5	2	7	1	6	2
<b>R06</b>	5	7	1	5	2	6	5
<b>R07</b>	4	2	1	3	5	3	6
<b>R08</b>	4	1	5	4	5	4	7
<b>R09</b>	6	6	2	4	7	5	5
<b>R10</b>	6	6	3	5	3	6	6
<b>R11</b>	7	7	6	7	7	6	7
<b>R12</b>	2	2	1	2	1	3	2

**Matrice des corrélations**

!	arbr	cade	dang	mora	orag	poli	sens
arbr !	1.00						
cade !	.55	1.00					
dang !	.29	.14	1.00				
mora !	.16	.62	.36	1.00			
orag !	.51	.09	.54	-.01	1.00		
poli !	.00	.63	.23	.91	-.05	1.00	
sens !	.56	-.08	.45	-.30	.68	-.37	1.00



**Fig. A1.4-a: Représentation des répondants dans le plan factoriel (2,3)**

**Fig. A1.4-b: Représentation des mots dans le plan factoriel (2,3)**

**Figure A1.4 : Analyse en composantes principales sur le tableau de notes de 7 mots par 12 répondants**

On s'intéresse surtout aux variables présentant les plus fortes coordonnées et l'on interprétera les composantes principales en fonction des regroupements de certaines de ces variables et de l'opposition avec les autres.

On notera alors que tous les points-variables sont sur une sphère de rayon 1 centrée à l'origine des axes<sup>1</sup>. Les plans d'ajustement couperont la sphère suivant de grands cercles (de rayon 1), les *cercles de corrélations*, à l'intérieur desquels sont positionnés les points-variables. Dans cet ouvrage, les cercles ne sont pas tracés dans les plans factoriels représentant les mots pour une meilleure lisibilité des libellés (le cadrage des plans factoriels aurait en effet entraîné une forte réduction d'échelle).

## A1.4 L'Analyse des correspondances

L'analyse des correspondances<sup>2</sup> s'applique en premier lieu à une table de contingence  $\mathbf{K}$ , appelé aussi tableau croisé, à  $n$  lignes et  $p$  colonnes, qui ventile une population selon deux variables qualitatives à  $n$  et  $p$  modalités. Les lignes et les colonnes jouent donc des rôles similaires.

Dans le chapitre 4, l'analyse est appliquée à un tableau croisant les 1 191 répondants spontanément aux questions ouvertes avec les 592 mots cités au moins 4 fois comme étant agréables. Une deuxième analyse porte sur 158 mots apparaissant plus de 25 fois.

### - Notations

Soit  $k = \sum_{ij} k_{ij}$  la somme de tous les éléments  $k_{ij}$  de la table de contingence  $\mathbf{K}$ . Dans le cas du tableau des questions ouvertes,  $k$  représente le nombre de fois que les 592 mots ont été cités spontanément.

On note  $f_{ij} = k_{ij}/k$  les fréquences relatives avec  $\sum_i \sum_j f_{ij} = 1$ .

---

1. L'analyse du nuage des points-variables dans  $\mathbb{R}^n$  ne se fait pas par rapport au centre de gravité du nuage, contrairement à celui des points-individus mais par rapport à l'origine. La distance d'une variable  $j$  à l'origine  $O$  s'exprime par :  $d^2(O, j) = \sum_{i=1}^n x_{ij}^2 = 1$

2. Présentée et étudiée de façon systématique comme une technique souple d'analyse exploratoire de données multidimensionnelles par J.-P. Benzécri (1973), l'analyse des correspondances s'est trouvée depuis d'autres précurseurs, en particulier C. Hayashi (1956), et a donné lieu à des travaux dispersés et indépendants les uns des autres.

On note :  $f_{i.} = \sum_j f_{ij}$ ,  $f_{.j} = \sum_i f_{ij}$ , les fréquences marginales relatives.

La table de contingence  $\mathbf{K}$  est transformé en un tableau de profils-lignes  $f_{ij}/f_{i.}$  et un tableau de profils-colonnes  $f_{ij}/f_{.j}$ .

Le point  $i$  de  $\mathbb{R}^m$  a pour coordonnées :  $f_{ij}/f_{i.}$  pour tout  $j \leq m$ .

De même, le point  $j$  de  $\mathbb{R}^n$  a pour coordonnées :  $f_{ij}/f_{.j}$  pour tout  $i \leq n$ .

Notons une différence importante entre l'analyse des correspondances et l'analyse en composantes principales : les transformations opérées sur le tableau dans les deux espaces sont identiques (car les ensembles mis en correspondance jouent des rôles analogues).

#### – Distance du Chi-deux et équivalence distributionnelle

Les distances entre deux points-lignes  $i$  et  $i'$  d'une part et entre deux points-colonnes d'autre part sont données par les équations suivantes :

$$d^2(i, i') = \sum_{j=1}^m \frac{1}{f_{.j}} \left( \frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \quad d^2(j, j') = \sum_{i=1}^n \frac{1}{f_{i.}} \left( \frac{f_{ij}}{f_{.j}} - \frac{f_{i'j'}}{f_{.j'}} \right)^2$$

La distance du  $\chi^2$  offre l'avantage de vérifier le principe d'équivalence distributionnelle. Ce principe assure la robustesse des résultats de l'analyse des correspondances vis-à-vis de l'arbitraire du découpage en modalités des variables nominales. Il s'exprime de la façon suivante : si deux lignes (resp. colonnes) du tableau de contingence ont même profil (sont proportionnelles) alors leur agrégation n'affecte pas la distance entre les colonnes (resp. lignes). On obtient alors un nouveau point-ligne (resp. point-colonne) de profil identique et affecté de la somme des fréquences des deux points-lignes (resp. points-colonnes).

Cette propriété est importante car elle garantit une certaine invariance des résultats vis-à-vis de la nomenclature choisie pour la construction des modalités d'une variable qualitative.

## A1.5 L'analyse logarithmique

L'analyse logarithmique, proposée par J.-B. Kazmierczak (1985), réalise la propriété de l'équivalence distributionnelle de l'analyse des correspondances sur des tableaux qui ne sont pas obligatoirement des tables de contingence. J.-B. Kazmierczak reprend et généralise le principe de Yule qui stipule que l'on ne change pas la distance entre

deux lignes ni la distance entre deux colonnes d'un tableau en remplaçant les lignes et les colonnes de ce tableau par d'autres lignes et colonnes qui leur sont proportionnelles (il s'agit en fait d'une généralisation du principe d'équivalence distributionnelle).

L'analyse logarithmique consiste à prendre les logarithmes des (après addition éventuelle d'une constante en cas de données négatives), puis, après les avoir centrées à la fois en ligne et en colonne, à les soumettre à une analyse en composantes principales non normée, qui coïncide ici avec une *décomposition aux valeurs singulières* [SEM, 2002].

Ainsi, si  $\mathbf{R}$  est un tableau de données  $(n, m)$  et si  $\mathbf{A}$  et  $\mathbf{B}$  sont deux matrices diagonales respectivement de dimensions  $(n, n)$  et  $(p, p)$  à éléments diagonaux positifs, la matrice  $\mathbf{ARB}$  donne lieu à la même analyse logarithmique que la matrice  $\mathbf{R}$ . Cette propriété d'invariance forte a eu pour effet de supprimer le premier axe sémiométrique sans altérer la suite des axes (section 5.7 du chapitre 5).

## A1.6 L'analyse factorielle en facteurs communs et spécifiques

L'analyse factorielle en facteurs communs et spécifiques (*factor analysis*) est probablement le modèle linéaire de variables latentes le plus ancien<sup>1</sup>. Ces modèles ont été essentiellement développés principalement par les psychologues et psychométriciens. Les développements auxquels ils donnent lieu sont complexes et diversifiés. On pourra consulter sur ce point les ouvrages classiques de Harman (1967), Mulaik (1972)<sup>2</sup>.

Mentionnons également les travaux d'Anderson et Rubin (1956) et de Lawley et Maxwell (1963) qui ont placé l'analyse factorielle en facteurs communs et spécifiques dans un cadre inférentiel classique.

### – *Le modèle de l'analyse factorielle*

Ce modèle se propose de reconstituer, à partir d'un petit nombre  $q$  de facteurs, les corrélations existant entre  $m$  variables observées. On suppose l'existence d'un modèle *a priori* :

---

3. A l'origine des principes de la méthode se trouvent Spearman (1904) (analyse monofactorielle), puis Garnett (1919) et Thurstone (1947) (analyse multifactorielle).

2. En économétrie, on distingue habituellement les modèles fonctionnels, ou à effet fixes (comme la régression multiple et le modèle linéaire dans son ensemble), et les modèles structurels ou à effet aléatoire (modèles de variables latentes).

$$\mathbf{x}_i = \mathbf{\Gamma} \mathbf{f}_i + \mathbf{e}_i$$

$(m,1) \quad (m,q)(q,1) \quad (p,1)$

Dans cette écriture  $\mathbf{x}_i$  représente le  $i$ -ème vecteur observé des  $m$  variables ;  $\mathbf{\Gamma}$  est un tableau  $(m, q)$  de coefficients inconnus (avec  $q < m$ ) ;  $\mathbf{f}_i$  est la  $i$ -ème valeur du vecteur aléatoire et non observable de  $q$  facteurs communs ; et  $\mathbf{e}_i$  la  $i$ -ème valeur du vecteur non observable de résidus, lesquels représentent l'effet combiné de facteurs spécifiques et d'une perturbation aléatoire.

On désigne par  $\mathbf{X}$  le tableau  $(n,p)$  dont la  $i$ -ème ligne représente l'observation  $i$ . De même  $\mathbf{F}$  désigne le tableau  $(n,q)$  non observable dont la  $i$ -ème ligne est  $\mathbf{f}_i'$  et  $\mathbf{E}$  le tableau  $(n,p)$  non observable dont la  $i$ -ème ligne est  $\mathbf{e}_i'$ . Le modèle liant l'ensemble des observations aux facteurs hypothétiques s'écrit :

$$\mathbf{X} = \mathbf{F} \mathbf{\Gamma}' + \mathbf{E}$$

$(n,m) \quad (n,q)(q,m) \quad (n,m)$

Dans cette écriture, seul  $\mathbf{X}$  est observable, et le modèle est par conséquent indéterminé.

L'identification de ce modèle et l'estimation des paramètres posent des problèmes complexes. Une cascade d'hypothèses *a priori* supplémentaires permet cette identification. L'application dans le cadre de cet ouvrage concerne la section 2.6 du chapitre 2.

## A1.7 Méthodes de classification hiérarchique

Les techniques de classification automatique<sup>1</sup> sont destinées à produire des groupements d'objets ou d'individus décrits par un certain nombre de variables ou de caractères. Les circonstances d'utilisation sont sensiblement les mêmes que celles des méthodes d'analyse factorielle descriptive présentées aux sections précédentes. Dans le chapitre 3, la classification est réalisée sur l'ensemble des 210 mots à partir des coordonnées de ces mots sur les axes principaux.

Il existe plusieurs familles d'algorithmes de classification : les *algorithmes hiérarchiques* qui fournissent une hiérarchie de partitions des objets et les algorithmes conduisant directement à des *partitions* comme les méthodes d'agrégation autour de centres mobiles. Les principes communs aux diverses techniques de classification ascendante hiérarchique sont simples. Il s'agit de créer, à chaque étape de

---

1. La classification est une branche de l'analyse des données qui constitue une étape fondamentale dans beaucoup de disciplines scientifiques. Elle a donné lieu à des publications nombreuses et diversifiées dont : Sokal et Sneath (1963) et Benzécri (1973).

l'algorithme, une partition obtenue en agrégeant deux à deux les éléments les plus proches.

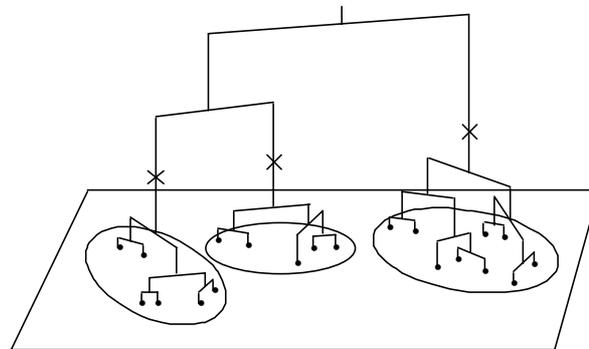
– *L'algorithme de la Classification Hiérarchique*

L'algorithme de base de la classification ascendante hiérarchique produit une hiérarchie en partant de la partition dans laquelle chaque élément à classer constitue une classe, pour aboutir à la partition formée d'une seule classe réunissant tous les éléments.

Pour  $n$  éléments à classer, il est composé de  $n$  étapes. A la première étape, il y a donc  $n$  éléments à classer. On construit la matrice de distances entre les  $n$  éléments et l'on cherche les deux plus proches, que l'on agrège en un nouvel élément.

On construit une nouvelle matrice des distances qui résultent de l'agrégation, en calculant les distances entre le nouvel élément et les éléments restants. On se trouve dans les mêmes conditions qu'à l'étape 1, mais avec seulement  $(n-1)$  éléments à classer.

On cherche de nouveau les deux éléments les plus proches, que l'on agrège. On réitère le processus jusqu'à n'avoir plus qu'un seul élément regroupant tous les objets et qui constitue la dernière partition.



**Figure A1.6: Dendrogramme ou arbre hiérarchique**

L'algorithme ne fournit pas une partition en  $q$  classes d'un ensemble de  $n$  objets mais une *hiérarchie de partitions*, se présentant sous la forme d'*arbres* appelés également *dendrogrammes* et contenant  $n - 1$  partitions (cf. figure A1.6). L'intérêt de ces arbres est qu'ils peuvent donner une idée du nombre de classes existant effectivement dans la population. Chaque coupure d'un dendrogramme fournit une partition.

Dans le tableau 3.1 du chapitre 3, on a retenu trois coupures du dendrogramme en 36, 24 et 12 classes. La coupure la moins fine en 12

classes figure en première colonne, celle en 24 classes en seconde colonne et la coupure la plus fine en 36 classes en troisième colonne. Elle est suivie de la liste des mots qui composent ces 36 classes.

## A1.8 Les cartes auto-organisées de Kohonen

L'objectif des cartes auto-organisées de Kohonen<sup>1</sup> est de classer un ensemble d'observations de façon à conserver la topologie initiale de l'espace dans lesquelles elles sont décrites. Comme les réseaux de neurones auxquelles elles sont rattachées, ces cartes obtiennent de bonnes performances pour la reconnaissance de formes. Elles ont été utilisées aux chapitres 3 (section 3.1) et 4 (section 4.2).

### - Le principe

*Les cartes de Kohonen* cherchent à représenter dans un espace à deux (parfois trois) dimensions les lignes ou les colonnes d'un tableau en respectant la notion de voisinage dans l'espace des éléments à classer. Tout comme l'analyse en composantes principales, il est utile d'imaginer au départ l'ensemble des données (les mots) comme un nuage de points dans un espace de grande dimension (celui des individus ou répondants).

Le principe est de considérer une carte comme une grille rectangulaire (parfois hexagonale) aux mailles déformables, laquelle, une fois dépliée épouse au mieux les formes du nuage de points. Les nœuds de la grille sont les *neurones* de la carte. Chaque point du nuage est projeté sur le nœud dont il est le plus proche. De fait, chaque point, décrit initialement dans un espace multidimensionnel est représenté à la fin par deux coordonnées donnant la position du *neurone* sur la carte : l'espace est réduit. L'ensemble des points affectés à un même *neurone* sont proches dans l'espace initial. Ils décrivent et regroupent des individus semblables.

On définit *a priori* une notion de voisinage entre classes et les observations voisines dans l'espace des variables de dimension  $q$  appartiennent après classement à la même classe ou à des classes voisines. Ces voisinages peuvent être choisis de diverses manières mais en général on les suppose directement contigus sur la grille rectangulaire (ce qui représente alors 8 voisins pour un *neurone*).

---

1. Introduites en 1981 par Teuvo Kohonen, elles font partie des méthodes dites *neurales* (cf. Kohonen, 1989). Elles donnent lieu à plusieurs applications relevant par exemple de l'analyse de textes, les diagnostics médicaux et industriels, les contrôles de processus, la robotique.

### - L'algorithme

L'algorithme d'apprentissage pour classer  $m$  points est itératif<sup>1</sup>. L'initialisation consiste à associer à chaque classe  $k$  un centre provisoire  $C_k$  à  $q$  composantes choisi de manière aléatoire dans l'espace à  $q$  dimensions contenant les  $m$  mots à classer. A chaque étape on choisit un mot  $i$  au hasard que l'on compare à tous les centres provisoires et l'on affecte le mot au centre  $C_{k_0}$  le plus proche au sens d'une distance donnée *a priori*. On rapproche alors du mot  $i$  le centre  $C_{k_0}$  et les centres voisins sur la carte ce qui s'exprime à l'étape  $t$  par :

$$C_k(t+1) = C_k(t) + \varepsilon(i(t+1) - C_k(t))$$

où  $i(t+1)$  est le mot présenté à l'étape  $t+1$ ,  $\varepsilon$  un paramètre d'adaptation positif et inférieur à 1. Cette expression n'intervient que pour le centre  $C_{k_0}$  et ses voisins.

Cet algorithme est analogue à celui des centres mobiles [SEM 2006], mais dans ce dernier cas, il n'existe pas de notion de voisinage entre classes et on ne modifie à chaque étape que la position du centre  $C_{k_0}$ .

L'auto-organisation de la carte de Kohonen est la conséquence de la notion de voisinage. Comme l'algorithme des centres mobiles, cet algorithme est très adapté aux applications où les données sont importantes et où il n'est pas utile de les stocker.

## A1.9 Outils de validation

Tout au long du présent ouvrage ont été utilisées les notions de *valeur-test* et de *variable supplémentaire*.

Les *valeurs-test* (section A1.9.1) sont un outil d'inférence statistique élémentaire, mais polyvalent et très utile, surtout si l'utilisateur est averti des problèmes de *comparaisons multiples* qui ne manquent pas d'intervenir (section A1.9.2).

La technique des variables supplémentaires (section A1.9.3) est un outil fondamental de valorisation des méthodes factorielles, qui permet une validation *externe* des résultats, à la fois épreuve de cohérence et enrichissement des interprétations.

---

1. On se réfère dans la présentation de l'algorithme au cours de P.Letremy et M.Cottrell (SAMOS-MATISSE, Université Paris I). Voir aussi Thiria *et al.* (1997).

Les deux autres outils de validation utilisés dans cet ouvrage sont les *intervalles de confiance d'Anderson* et les procédures de rééchantillonnage *bootstrap*.

Les *intervalles de confiance d'Anderson* (section A1.9.4) sont utilisés dans la section 2.2 du chapitre 2 pour valider des valeurs propres.

Les procédures de rééchantillonnage *bootstrap* (section A1.9.5) sont utilisées aux sections 2.3 et 2.4 du chapitre 2 pour mettre en évidence la stabilité des structures sémiométriques, et, dans la section 4.4 du chapitre 4, pour valider une analyse textuelle.

### A1.9.1 Qu'est-ce qu'une valeur-test ?

La valeur-test est un critère qui permet d'apprécier rapidement si une modalité d'une variable nominale (*i.e.* : une catégorie de répondants) a une position *significative* sur un axe. Pour cela, on teste l'hypothèse selon laquelle un groupe d'individus, correspondant à une modalité donnée d'une variable nominale supplémentaire (comme la modalité *profession libérale, cadre supérieur* pour la variable nominale *catégorie socio-professionnelle*, par exemple), peut être considéré comme tiré au hasard, sans remise, dans la population.

Dans le cas d'un véritable tirage au hasard, le centre de gravité du sous-nuage représentant le groupe (*i.e.* : la modalité) s'éloigne peu du centre de gravité du nuage global correspondant à tout l'échantillon.

On convertit alors la coordonnée de cette modalité sur l'axe en une *valeur-test* qui est, sous cette hypothèse, la réalisation d'une variable normale centrée réduite. Autrement dit, dans l'hypothèse selon laquelle une modalité a une composition *aléatoire*, la valeur-test correspondante a 95% de chances d'être comprise dans l'intervalle  $[-1.96, +1.96]$ .

On considère alors comme occupant une *position significative* les modalités dont les valeurs-test sont supérieures à 2 (pour 1.96) en valeur absolue, ce qui correspond approximativement au seuil usuel de probabilité de 5%.

Souvent les valeurs-test sont largement supérieures à ce seuil. On les utilise alors pour trier les modalités, des plus significatives au moins significatives. La valeur-test systématise la notion de *t-value* souvent utilisée dans la littérature anglo-saxonne.

Supposons qu'une modalité  $j$  concerne  $n_j$  individus. Si ces  $n_j$  individus sont tirés au hasard (c'est ce qu'on appelle l'hypothèse nulle  $H_0$ ) parmi les  $n$  individus analysés (tirage supposé sans remise), la moyenne de  $n_j$  coordonnées tirées au hasard dans l'ensemble fini des  $n$  valeurs  $\psi_{\alpha i}$  (coordonnée du répondant  $i$  sur l'axe  $\alpha$ ) est une variable aléatoire  $X_{\alpha j}$ :  $X_{\alpha j} = \frac{1}{n_j} \sum_{i \in I(j)} \psi_{\alpha i}$  avec pour espérance  $E(X_{\alpha j})=0$  et pour variance<sup>1</sup>  $Var_{H_0}(X_{\alpha j}) = \frac{n - n_j}{n - 1} \frac{\lambda_\alpha}{n_j}$

Dans la formule donnant  $X_{\alpha j}$ ,  $I(j)$  est le sous-ensemble des répondants caractérisés par la modalité  $j$  de la variable nominale.

La coordonnée  $\varphi_{\alpha j}$  de la modalité  $j$  est proportionnelle à la variable aléatoire  $X_{\alpha j}$  et s'écrit ainsi :  $\varphi_{\alpha j} = \frac{1}{\sqrt{\lambda_\alpha}} X_{\alpha j}$

On a donc  $E(\varphi_{\alpha j}) = 0$  et  $Var_{H_0}(\varphi_{\alpha j}) = \frac{n - n_j}{n - 1} \frac{1}{n_j}$

La quantité  $t_{\alpha j}$ :  $t_{\alpha j} = \sqrt{n_j \frac{n - 1}{n - n_j}} \varphi_{\alpha j}$  mesure en *nombre d'écart-types* la distance entre la modalité  $j$ , c'est-à-dire le quasi-barycentre des  $n_j$  individus, et l'origine, sur l'axe factoriel  $\alpha$ . On appelle cette quantité « *valeur-test* ». D'après le théorème de la limite centrale (*central limit theorem*), sa distribution tend vers une loi de Laplace-Gauss centrée réduite.

On doit noter que les valeurs-test n'ont de sens que pour les modalités supplémentaires (cf. section suivante), ou des modalités actives ayant des contributions absolues faibles, c'est-à-dire se comportant en fait comme des modalités supplémentaires<sup>2</sup>.

Lorsque l'on dispose d'un nombre important de modalités supplémentaires, les valeurs-test permettent de repérer rapidement les modalités utiles à l'interprétation d'un axe ou d'un plan factoriel.

---

1. Il s'agit de la formule classique donnant la variance d'une moyenne lors d'un tirage sans remise de  $n_j$  objets parmi  $n$ , en fonction de la variance totale  $\lambda_\alpha$ , qui est aussi, dans le cas des coordonnées factorielles, la valeur propre correspondant à l'axe  $\alpha$ .

2. Les coordonnées sur un axe des individus correspondant à une modalité active ne peuvent être considérées comme tirée au hasard, puisque cette modalité aura contribué à construire l'axe.

### A1.9.2 Le problème des comparaisons multiples

Le calcul simultané de plusieurs valeurs-test ou de plusieurs seuils de probabilités se heurte à l'écueil des *comparaisons multiples*, bien connu des statisticiens ; cf. O'Neill et Wetherill (1971), Saville (1990), Westfall et Young (1993), Westfall *et al.* (1999), Hsu (1996).

Supposons que l'on projette 100 modalités *supplémentaires* (cf. section suivante A1.9.3) qui soient vraiment tirées au hasard. Les valeurs-test attachées à ces modalités sont alors toutes des réalisations de variables aléatoires normales centrées réduites indépendantes.

Dans ces conditions, *en moyenne*, sur 100 valeurs-test calculées, cinq seront en dehors de l'intervalle  $[-1.96, +1.96]$  et seront, en apparence seulement, significatives. Le seuil de 5% n'a de sens en fait que pour un seul test, et non pour des tests multiples.

On résout en pratique cette difficulté en choisissant un seuil plus sévère<sup>1</sup>. Le seuil le plus sévère et pessimiste que l'on puisse imaginer est le « seuil de *Bonferroni* » (on divise le seuil initial par le nombre de tests : dans le cas de 210 tests :  $0.05 / 210 = 2.4 \cdot 10^{-4}$ ). La valeur-test unilatérale correspondante est de 3.49. Cette valeur nous fournit un garde-fou prudent à l'excès<sup>2</sup>.

Comme cela a été signalé dans le corps du texte (cf., par exemple, les notes de la section 7.2 du chapitre 7), l'interdépendance des mots ne permet pas d'appliquer aveuglément les résultats concernant les comparaisons multiples. Que conclure, en effet, lorsque plusieurs mots de sens voisins ont simultanément des valeurs-test de l'ordre de 1.96 ? Celles-ci ne sont pas significatives une par une en retenant le seuil de *Bonferroni*, mais elles se confirment et se valident mutuellement.

*Une solution pragmatique (cas multidimensionnel) : le bootstrap.*

La technique de validation par *bootstrap* dont il sera question plus loin dans cette annexe apporte une contribution intéressante au difficile problème des comparaisons multiples, car les répliques d'échantillons permettent de prendre en compte simultanément toutes les variables, et donc l'interdépendance des variables.

Il s'agit d'un test global, et non plus de tests séparés pour chaque variable. Une illustration en est donnée par la figure 4.4 du chapitre 4 qui

---

1. Les valeurs-test permettent surtout de *classer* les modalités supplémentaires par ordre d'intérêt décroissant, ce qui constitue une aide précieuse à l'interprétation des facteurs.

2. Cf., par exemple, Hochberg (1988), Perneger (1998).

représente les zones de confiance simultanées des mots, dont certains apparaissent comme significativement distincts. Dans ce cas, les tests ne sont pas réalisés isolément ni en série, mais simultanément.

### A1.9.3 Utilité des éléments supplémentaires

L'analyse factorielle permet de trouver des sous-espaces de représentation des proximités entre points-individus ou entre points-variables. Elle s'appuie pour cela sur des éléments (individus ou variables) dits *actifs*.

Il est possible d'introduire en supplémentaire d'autres points (ou éléments) que l'on ne souhaite pas faire intervenir dans la composition et définition des axes mais dont on veut connaître les positions dans les espaces factoriels<sup>1</sup>. On projette alors ces points après la construction des axes factoriels dans ce nouveau repère. Cette projection se fait de façon très simple en utilisant les formules dites *de transition*, que ce soit en analyse en composantes principales ou en analyse des correspondances.

C'est le cas lorsque l'on veut positionner les mots, variables numérique, dans l'espace des variables de contrôle (*cf.* section 5.4). On calcule, *a posteriori*, leurs coordonnées sur les axes factoriels.

C'est également le cas lorsque l'on souhaite caractériser les axes sémiométriques<sup>2</sup> par les critères socio-démographiques (variables nominales) de la population enquêtée (*cf.* section 1.4).

Ces critères définissent en fait des groupes d'individus et sont considérés soit comme des modalités de variables nominales, soit comme des individus, mis en éléments supplémentaires.

Ce sont les centres de gravité de ces groupes qui sont positionnés dans l'espace des variables. La valeur-test permet d'en apprécier la significativité sur l'axe.

Cette procédure pourrait être utilisée comme méthode alternative pour comparer des sous-populations dans le chapitre 7.

---

1. On peut citer trois raisons qui peuvent susciter la mise en supplémentaire d'un point : 1) enrichir l'interprétation des axes par des variables (de nature ou de thématique différente de celle des éléments actifs) n'ayant pas participé à leur construction ; 2) adopter une optique de prévision en projetant les variables supplémentaires dans l'espace des individus. Celles-ci seront « expliquées » par les variables actives ; 3) faire ressortir l'essentiel d'une structure masquée par l'existence d'un point actif, de faible masse, mais très excentré qui pourrait déformer le nuage.

2. Ces axes, rappelons-le, sont définis par les variables actives que sont les mots.

### A1.9.4 Intervalles de confiance d'Anderson

Anderson (1963) a calculé les lois limites des valeurs propres d'une analyse en composantes principales sans nécessairement supposer que les valeurs théoriques correspondantes sont distinctes.

L'ampleur de l'intervalle donne une indication sur la stabilité de la valeur propre vis-à-vis des fluctuations dues à l'échantillonnage supposé laplacien. L'empiétement des intervalles de deux valeurs propres consécutives suggérera donc l'égalité de ces valeurs propres. Les axes correspondants sont alors définis à une rotation près. Ainsi l'utilisateur pourra éviter d'interpréter un axe instable selon ce critère.

Si les valeurs propres théoriques  $\lambda_\alpha$  de  $\Sigma$  sont distinctes, les valeurs propres  $\hat{\lambda}_\alpha$  de la matrice des covariances empiriques  $S$  suivent asymptotiquement des lois normales d'espérance  $\lambda_\alpha$  et de variance  $2\lambda_\alpha^2/(n-1)$  où  $n$  est la taille de l'échantillon.

On en déduit les intervalles de confiance approchés au seuil 95% :

$$\lambda_\alpha \in \left[ \hat{\lambda}_\alpha \left(1 - 1.96\sqrt{2/(n-1)}\right) ; \hat{\lambda}_\alpha \left(1 + 1.96\sqrt{2/(n-1)}\right) \right]$$

Les intervalles de confiance d'Anderson concernent en fait aussi bien les valeurs propres des matrices des covariances que des matrices de corrélations. Les simulations entreprises montrent que les intervalles de confiance obtenus sont en général « prudent » : le pourcentage de couverture de la vraie valeur est le plus souvent supérieur au seuil de confiance annoncé.

Dans tous les cas, la nature asymptotique des résultats et l'hypothèse sous-jacente de normalité<sup>1</sup> font considérer les résultats comme indicatifs.

### A1.9.5 Les techniques de *bootstrap*

Face aux résultats d'une analyse factorielle, certaines questions sur la validité des axes obtenus se posent naturellement : Existe-t-il des critères pour tester la stabilité d'une structure et la valider ? Quelle est la part de l'échantillonnage des individus mais aussi, notion plus difficile, celle du choix ou de la sélection des variables ?

Nous avons vu au chapitre 2 que pour tenter de répondre partiellement à ces questions, on avait recours aux méthodes empiriques de validation. Elles consistent à perturber le tableau initial par des ajouts ou retraits

---

1. Muirhead (1982) a montré que l'hypothèse d'existence des quatre premiers moments pour la loi théorique de l'échantillon suffisait pour valider ces intervalles.

d'éléments du tableau, individus ou variables (poids, codage, etc.). L'hypothèse est la suivante : si les perturbations effectuées sur les échantillons n'affectent pas les configurations observées dans les sous-espaces, celles-ci sont supposées stables et la structure mise en évidence est alors « significative ».

Les méthodes de rééchantillonnage se proposent de systématiser cette démarche<sup>1</sup>. Celle du *bootstrap*, non paramétrique, est bien adaptée au problème de la validité des formes observées dans un plan factoriel ; elle calcule, à partir de simulations, des zones de confiance pour les positions des points-lignes et des points-colonnes.

#### – Principe du *bootstrap*

La technique du *bootstrap*, introduite par Efron (1979), consiste à simuler  $s$  ( $s$  est généralement supérieur à 30) échantillons de même taille  $n$  que l'échantillon initial. Ils sont obtenus par tirage au hasard *avec remise* parmi les  $n$  individus observés au départ, ceux-ci ayant tous la même probabilité  $1/n$  d'être choisis. Certains individus apparaîtront plusieurs fois et auront de ce fait un poids élevé (2, 3,...) alors que d'autres seront absents (poids nul).

Cette méthode est employée pour analyser la variabilité de paramètres statistiques simples en produisant des intervalles de confiance de ces paramètres. Elle peut aussi être appliquée à de nombreux problèmes pour lesquels on ne peut pas estimer analytiquement la variabilité d'un paramètre. Ceci est le cas pour les caractéristiques des méthodes multidimensionnelles où les hypothèses de multinormalité sont rarement vérifiées. L'analyse en composantes principales est un domaine d'application qui a donné à un grand nombre de travaux utilisant les méthodes de rééchantillonnage de *bootstrap*.

Prenons l'exemple de l'estimation du coefficient de corrélation  $r$  entre deux variables ou entre une variable et un facteur. Le principe consiste à calculer le coefficient de corrélation pour chaque échantillon répliqué (pour lequel on effectue un tirage avec remise des *couples* d'observations). On établit alors la distribution des fréquences du coefficient de corrélation (représentée par l'histogramme des  $s$  valeurs du coefficient  $r$  correspondant aux  $s$  réplifications). Puis on calcule à partir de

---

1. Ce sont des méthodes de calculs intensifs qui reposent sur des techniques de simulations d'échantillons à partir d'un seul échantillon. Rendues possibles par la puissance de calcul des ordinateurs, ces techniques se substituent dans certains cas aux procédures plus classiques qui reposent sur des hypothèses contraignantes. Elles sont les seules procédures possibles lorsque la complexité analytique du problème ne permet pas d'inférence classique.

l'histogramme la probabilité pour que le coefficient de corrélation d'un échantillon soit compris dans différentes fourchettes de valeurs définissant ainsi les intervalles de confiance. On obtient une estimation de la précision de la valeur de  $r$  obtenue sur l'échantillon de base sans faire l'hypothèse d'une distribution normale des données. Les bornes de l'intervalle de confiance peuvent être estimées directement par les quantiles de la distribution simulée.

Pour estimer les coordonnées factorielles issus d'une analyse en composantes principales, le principe est le même que pour le coefficient de corrélation ; on effectue sur chaque échantillon simulé, une analyse en composantes principales puis on établit une distribution de fréquences pour chacune des composantes<sup>1</sup>.

La méthode de *bootstrap* donne dans la plupart des cas une bonne image de la précision statistique de l'estimation sur un échantillon. Les recherches théoriques menées par Efron, en particulier, montrent que, pour de nombreux paramètres statistiques, l'intervalle de confiance correspondant à la distribution simulée par *bootstrap* et celui correspondant à la distribution réelle sont généralement de même amplitude.

#### – Mise en œuvre et calcul des zones de confiance

Il existe plusieurs procédures pour tester, par la méthode de bootstrap, la stabilité des coordonnées factorielles. Gifi (1981), Meulman (1982), Greenacre (1984) ont réalisé des premiers travaux dans le contexte de l'analyse des correspondances simples ou multiples. Dans le cas de l'analyse en composantes principales, Diaconis et Efron (1983), Holmes (1989), Stauffer et al. (1985), Daudin et al. (1988) ont posé le problème du choix du nombre d'axes pertinent et ont proposé des intervalles de confiance pour les points du sous-espace défini par les principaux axes. Les paramètres correspondant sont calculés à partir des échantillons répliqués et supposent des contraintes qui dépendent de ces échantillons.

Pour pallier ces difficultés, il faut se référer à un espace factoriel commun. Plusieurs variantes sont possibles.

Nous nous sommes basés dans le chapitre 2 sur deux techniques appelées ici le *bootstrap total* et le *bootstrap partiel*.

Le *bootstrap total* consiste à réaliser autant d'analyses en composantes principales qu'il y a de répliqués, moyennant une série

---

1. On trouvera des compléments sur l'intérêt et les limites de cette méthode dans les travaux de Diaconis et Efron (1983) et de Young (1994).

de transformations afin de retrouver des axes homologues au cours des diagonalisations successives des  $s$  matrices de corrélation répliquées  $\mathbf{C}_k$  ( $\mathbf{C}_k$  correspond à la  $k$ -ème réplique). Ces transformations sont des changements de signe des axes, rotations ou permutations d'axes. Cette méthode, proposée par Milan et Whittaker (1995) est en défaut s'il existe des valeurs propres très voisines.

Dans le bootstrap partiel, proposé par Greenacre (1984) dans le cas de l'analyse des correspondances, il n'est pas nécessaire de calculer les valeurs et vecteurs propres pour l'ensemble des simulations : les axes principaux calculés sur les données originales non perturbées, jouent un rôle privilégié (la matrice des corrélations initiale  $\mathbf{C}$  est en effet l'espérance mathématique des matrices perturbées  $\mathbf{C}_k$ ).

Le *bootstrap partiel* se fonde sur la projection en tant qu'*éléments supplémentaires* des points répliqués sur les sous-espaces de référence fournis par les axes principaux de la matrice de corrélation  $\mathbf{C}=\mathbf{X}'\mathbf{X}$ , provenant de l'échantillon initial, donnés par :

$$\mathbf{u}_q = \frac{1}{\sqrt{\lambda_q}} \mathbf{X}'\mathbf{v}_q$$

où  $\mathbf{u}_q$ ,  $\mathbf{v}_q$  sont respectivement les  $q$ -èmes vecteurs propres de  $\mathbf{X}'\mathbf{X}$  et  $\mathbf{X}\mathbf{X}'$  et  $\lambda_q$  la valeur propre associée.

La projection<sup>1</sup> de la  $k$ -ème réplique des  $m$  variables (mots) est donnée par le vecteur  $\mathbf{u}_q(k)$  de  $\mathbb{R}^m$  tel que :

$$\mathbf{u}_q(k) = \frac{1}{\sqrt{\lambda_q}} \mathbf{X}'\mathbf{D}_k\mathbf{v}_q$$

et  $\mathbf{D}_k$  désigne la matrice diagonale  $(n, n)$  des *poids bootstrap* associée à la  $k$ -ème réplique<sup>2</sup>.

Dans le cas du bootstrap partiel, les analyses des matrices  $\mathbf{C}_k$  ne sont en aucun cas nécessaires puisque les vecteurs propres sont obtenus à partir de l'analyse en composantes principales de la matrice  $\mathbf{C}$ .

---

1. La projection des répliques Bootstrap, dans le contexte de l'analyse en composantes principales, consiste à utiliser le fait que la coordonnée d'une variable sur un axe factoriel n'est autre que son coefficient de corrélation avec la variable « coordonnées des individus sur l'axe ». On calcule donc les répliques de ce coefficient, ce qui revient à répondre, pour chaque réplique, les individus avec les *poids Bootstrap* qui caractérisent un tirage sans remise. On obtient, comme sous-produit, des répliques de la variance sur l'axe, qui sont évidemment distinctes de ce que seraient des répliques des valeurs propres.

2. Cf. Chateau et Lebart (1996).

La variabilité bootstrap s'observe donc mieux sur le repère fixe initial, qui est d'ailleurs le moins mauvais, étant le seul à n'avoir pas été perturbé. Cette technique, éprouvée empiriquement, répond parfaitement aux préoccupations des utilisateurs dans le cas de l'analyse en composantes principales.

– *Bootstrap sur l'ensemble des variables*

Classiquement les répliques sont obtenues par des tirages avec remise dans l'ensemble des  $n$  individus. Pour tester la stabilité des structures vis-à-vis de l'ensemble des mots, nous proposons de répliquer cet ensemble par la méthode du *bootstrap total*.

Nous supposons ainsi implicitement que l'ensemble des mots du questionnaire constitue un échantillon de  $m$  mots extrait aléatoirement de l'ensemble des mots « sémiométrisables » de la langue française.

Nous cherchons à perturber cet échantillon de mots selon les mêmes principes que le *bootstrap* opéré sur les individus.

Pour cela, on appelle  $\mathbf{B}_k$  la matrice diagonale  $(m, m)$  dont les éléments diagonaux sont les poids des mots de la  $k$ -ème réplique Bootstrap  $(1, 0, 2, 0, \dots)$ . La matrice  $\mathbf{X}$  d'ordre  $(n, n)$  initiale étant supposée centrée, la matrice à diagonaliser est la matrice  $\mathbf{T}_k$  qui vaut :

$$\mathbf{T}_k = \mathbf{X}\mathbf{B}_k\mathbf{X}' = \mathbf{X}\mathbf{B}_k^{1/2}\mathbf{B}_k^{1/2}\mathbf{X}'$$

On obtient donc :

$$\mathbf{X}\mathbf{B}_k\mathbf{X}'\mathbf{v}_q(k) = \lambda_q\mathbf{v}_q(k)$$

en multipliant chaque terme par  $\mathbf{B}_k^{1/2}\mathbf{X}'$  on a :

$$\mathbf{B}_k^{1/2}\mathbf{X}'\mathbf{X}\mathbf{B}_k^{1/2}\mathbf{B}_k^{1/2}\mathbf{X}'\mathbf{v}_q(k) = \lambda_q\mathbf{B}_k^{1/2}\mathbf{X}'\mathbf{v}_q(k)$$

et en posant  $\mathbf{u}_q(k) = \mathbf{B}_k^{1/2}\mathbf{X}'\mathbf{v}_q(k)$  alors :

$$\mathbf{B}_k^{1/2}\mathbf{X}'\mathbf{X}\mathbf{B}_k^{1/2}\mathbf{u}_q(k) = \lambda_q\mathbf{u}_q(k)$$

$\mathbf{T}_k = \mathbf{X}\mathbf{B}_k^{1/2}\mathbf{B}_k^{1/2}\mathbf{X}'$  a les mêmes valeurs propres non nulles que la matrice  $\mathbf{T}_k^* = \mathbf{B}_k^{1/2}\mathbf{X}'\mathbf{X}\mathbf{B}_k^{1/2}$ . On diagonalisera la matrice  $\mathbf{T}_k^*$  de dimension  $(m, m)$

En pratique, on remplace les *poids bootstrap* nuls par des poids infinitésimaux, de façon à ce que les variables absentes d'une réplique apparaissent quand même avec le statut de variable supplémentaire.

Cette épreuve de validation est évidemment très sévère. On montre en effet que le tirage sans remise suscite approximativement, en moyenne, l'abandon d'un tiers des éléments (ici, des mots !) à chaque réplique.

La figure 2.10 de la section 2.4 du chapitre 2 illustre les effets de ces fortes perturbations sur la position des points variables.

Tous les calculs des chapitres 1 à 5 (analyses factorielles diverses, classifications, cartes de Kohonen, zones de confiances *bootstrap*), ont été réalisés à l'aide du logiciel académique *DtmVic* spécialisé dans la *fouille de données numériques et textuelles*.

## ANNEXE 2

### Quelques éléments et résultats d'analyse

#### A2.1 dictionnaire multilingue des mots

FRANÇAIS	ANGLAIS	ALLEMAND	ESPAGNOL	ITALIEN
<b>l'absolu</b>	absolute	absolut	el absoluto	l'assoluto
<b>l'acharnement</b>	persistence	hartnaeckig	el empeño	l'accanimento
<b>acheter</b>	to buy	kaufen	comprar	comprare
<b>admirer</b>	to admire	bewundern	admirar	ammirare
<b>adorer</b>	to love	anbeten	adorar	adorare
<b>l'ambition</b>	ambition	der Ehrgeiz	la ambición	l'ambizione
<b>l'âme</b>	soul	die Seele	el alma	l'anima
<b>l'amitié</b>	friendship	die Freundschaft	la amistad	l'amicizia
<b>l'angoisse</b>	anguish	die Angst	la angustia	l'angoscia
<b>un animal</b>	animal	ein Tier	un animal	un animale
<b>un arbre</b>	tree	ein Baum	un árbol	un albero
<b>l'argent</b>	silver	das Geld	el dinero	il denaro
<b>une armure</b>	armour	die Ruestung	una armadura	un'armatura
<b>l'art</b>	art	die Kunst	el arte	l'arte
<b>astucieux</b>	cunning	schlau	sagaz	astuto
<b>l'attachement</b>	attachment	die Zuneigung	el cariño	l'attaccamento
<b>attaquer</b>	to attack	angreifen	atacar	aggredire
<b>l'audace</b>	audacity	die Kuehnheit	la audacia	l'audacia
<b>aventurier</b>	adventurer	ein Abenteuer	un aventurero	avventuriero

FRANÇAIS	ANGLAIS	ALLEMAND	ESPAGNOL	ITALIEN
<b>un bâtisseur</b>	a builder	ein Erbauer	un constructor	un costruttore
<b>un bijou</b>	a jewel	ein Schmuckstueck	una joya	un gioiello
<b>bleu</b>	blue	blau	azul	blu
<b>bohème</b>	wander	ein Lebenskuenstler	bohemio	anticonformista
<b>un cadeau</b>	a present	ein Geschenk	un regalo	un regalo
<b>la campagne</b>	countryside	auf dem Land	el campo	la campagna
<b>une caresse</b>	a caress	eine Liebkosung	una caricia	una carezza
<b>une cérémonie</b>	ceremony	eine Zeremonie	una cerimonia	una cerimonia
<b>une certitude</b>	certainty	eine Gewissheit	une certeza	una certezza
<b>le changement</b>	change	die Veraenderung	el cambio	il cambiamento
<b>charitable</b>	charitable	barmherzig	caritativo	caritatevole
<b>charnel</b>	carnal	das sexuelle	carnal	carnale
<b>la chasse</b>	the hunt	die Jagd	la caza	la caccia
<b>un chercheur</b>	researcher	ein Forscher	un investigador	un ricercatore
<b>commander</b>	to order	befehlen	mandar	comandare
<b>le commerce</b>	trade	der Handel	el comercio	il commercio
<b>concret</b>	solid	konkret	concreto	concreto
<b>la confiance</b>	confidence	das Vertrauen	la confianza	la fiducia
<b>le confort</b>	comfort	die Bequemlichkeit	el confort	la comodita'
<b>conquérir</b>	to conquer	erobern	conquistar	conquistare
<b>consoler</b>	to console	trosten	consolar	consolare
<b>construire</b>	to build	bauen	construir	costruire
<b>le courage</b>	brave	der Mut	el coraje	il coraggio
<b>un créateur</b>	creator	ein Schoepfer	un creador	un creatore
<b>un cri</b>	a shout	ein Schrei	un grito	un grido
<b>critiquer</b>	to criticise	kritisieren	criticar	criticare
<b>le danger</b>	danger	die Gefahr	el peligro	il pericolo
<b>un défi</b>	a challenge	eine Herausforderung	un desafío	una sfida
<b>le desert</b>	desert	die Wueste	el desierto	il deserto
<b>le désir</b>	desire	das Verlangen	el deseo	il desiderio
<b>le désordre</b>	disorder	die Unordnung	el desorden	il disordine
<b>le détachement</b>	detachment	die Gleichgueltigkeit	el despego	il distacco
<b>Dieu</b>	God	Gott	Díos	Dio
<b>différent</b>	different	verschieden	diferente	diverso
<b>la discipline</b>	discipline	die Disziplin	la disciplina	la disciplina
<b>la douceur</b>	softness	die Sanftheit	la suavidad	la dolcezza
<b>le doute</b>	doubt	der Zweifel	la duda	il dubbio
<b>dynamique</b>	dynamic	dynamisch	dinámico	dinamico
<b>l'eau</b>	water	das Wasser	el agua	l'acqua
<b>l'école</b>	school	die Schule	la escuela	la scuola
<b>économiser</b>	economise	sparen	ahorrar	risparmiare
<b>écrire</b>	to write	schreiben	escribir	scrivere
<b>efficace</b>	effective	tuechtig	eficaz	efficace
<b>un effort</b>	effort	eine Anstrengung	un esfuerzo	uno sforzo
<b>l'élégance</b>	elegance	die Eleganz	la elegancia	l'eleganza

FRANÇAIS	ANGLAIS	ALLEMAND	ESPAGNOL	ITALIEN
<b>l'élite</b>	elite	die Elite	la élite	l'elite
<b>une émotion</b>	emotion	eine Gefuehlsbewegung	una emoción	un'emozione
<b>l'enfance</b>	childhood	die Kindheit	la infancia	l'infanzia
<b>enseigner</b>	to teach	unterrichten	educar	insegnare
<b>ensemble</b>	together	miteinander	juntos	insieme
<b>escalader</b>	to climb	hochklettern	escalar	arrampicarsi
<b>éternel</b>	eternal	ewig	eterno	eterno
<b>un étranger</b>	a stranger	ein Fremder	un extranjero	uno straniero
<b>une évasion</b>	escape	eine Flucht	una evasión	un'evasione
<b>la famille</b>	family	die Familie	la familia	la famiglia
<b>une faute</b>	a fault	ein Fehler	una falta	una colpa
<b>féconder</b>	to fertilise	befruchten	fecundar	fecondare
<b>féminin</b>	feminine	weiblich	femenino	femminile
<b>la fermeté</b>	firmness	die Standhaftigkeit	la firmeza	la fermezza
<b>le feu</b>	fire	das Feuer	el fuego	il fuoco
<b>la fidélité</b>	faith	die Treue	la fidelidad	la fedelta'
<b>une fleur</b>	flower	eine Blume	una flor	un fiore
<b>un fleuve</b>	river	ein Fluss	un río	un fiume
<b>la foi</b>	a belief	der Glaube	la fé	la fede
<b>une frontière</b>	border	eine Grenze	una frontera	una frontiera
<b>un fusil</b>	gun	ein Gewehr	un fusil	un fucile
<b>la gaieté</b>	liveliness	die Froehlichkeit	la alegría	l'allegria
<b>la gloire</b>	glory	der ruhm	la gloria	la gloria
<b>gratuit</b>	free	unentgeltlich	gratuito	gratuito
<b>guérir</b>	to recover	heilen	curar	guarire
<b>la guerre</b>	war	der Krieg	la guerra	la guerra
<b>hériter</b>	to inherit	erben	heredar	ereditare
<b>un héros</b>	hero	ein Held	un héroe	un eroe
<b>honnêté</b>	honest	ehrlich	honesto	onesto
<b>l'honneur</b>	honour	die Ehre	el honor	l'onore
<b>humble</b>	humble	demuetig	humilde	umile
<b>l'humour</b>	humour	der Humor	el humor	l'umorismo
<b>une île</b>	an island	eine Insel	una isla	un'isola
<b>immense</b>	huge	unermesslich	immenso	immenso
<b>immobile</b>	immobile	unbeweglich	immovil	immobile
<b>un inconnu</b>	an unknown	ein Unbekannter	un desconocido	uno sconosciuto
<b>l'industrie</b>	industry	die Industrie	la industria	l'industria
<b>l'infini</b>	infinity	die Endlosigkeit	el infinito	l'infinito
<b>interdire</b>	to forbid	verbieten	prohibir	vietare
<b>interroger</b>	to interrogate	befragen	interrogar	interrogare
<b>intime</b>	intimate	intim	intimo	intimo
<b>un inventeur</b>	inventor	ein Erfinder	un inventor	un inventore
<b>l'ironie</b>	irony	die Ironie	la ironía	l'ironia
<b>un jeu</b>	a game	ein Spiel	un juego	un gioco
<b>la justice</b>	justice	die Justiz	la justicia	la giustizia

FRANÇAIS	ANGLAIS	ALLEMAND	ESPAGNOL	ITALIEN
<b>un labyrinthe</b>	maze	ein Labyrinth	un laberinto	un labirinto
<b>la légèreté</b>	lightness	die Leichtigkeit	la ligereza	la leggerezza
<b>un livre</b>	book	ein Buch	un libro	un libro
<b>la logique</b>	logic	die Logik	la lógica	la logica
<b>la loi</b>	the law	das Gesetz	la ley	la legge
<b>la lune</b>	moon	der Mond	la luna	la luna
<b>la magie</b>	magic	die Magie	la magia	la magia
<b>une maison</b>	house	ein Haus	una casa	una casa
<b>maîtriser</b>	to master	beherrschen	dominar	dominarer
<b>un mariage</b>	a wedding	eine Heirat	una boda	un matrimonio
<b>un masque</b>	mask	eine Maske	una máscara	una maschera
<b>matériel</b>	material	materiell	material	materiale
<b>maternel</b>	maternal	muetterlich	maternal	materno
<b>la méfiance</b>	mistrust	das Misstrauen	la desconfianza	la diffidenza
<b>métallique</b>	metallic	metallisch	metálico	metallico
<b>la minceur</b>	thinness	die Schlankeit	la esbeltez	la snellezza
<b>la mode</b>	fashion	die Mode	la moda	la moda
<b>la modération</b>	moderation	die Maessigung	la moderación	la moderazione
<b>la modestie</b>	modesty	die Bescheidenheit	la modestia	la modestia
<b>moelleux</b>	soft	anschmiegsam	blando	morbido
<b>une montagne</b>	mountain	ein Berg	una montaña	una montagna
<b>la morale</b>	morals	die Moral	la moral	la morale
<b>la mort</b>	death	der Tod	la muerte	la morte
<b>une muraille</b>	wall	eine Mauer	una muralla	una muraglia
<b>la musique</b>	music	die Musik	la música	la musica
<b>un mystère</b>	mystery	ein Mysterium	un misterio	un mistero
<b>nager</b>	to swim	schwimmen	nadar	nuotare
<b>la naissance</b>	birth	die Geburt	el nacimiento	la nascita
<b>un nid</b>	nest	ein Nest	un nido	un nido
<b>noble</b>	noble	edel	noble	nobile
<b>un nœud</b>	a knot	ein Knoten	un nudo	un nodo
<b>noir</b>	black	schwarz	negro	nero
<b>la nudité</b>	nudity	die Nacktheit	la desnudez	la nudita'
<b>obéir</b>	to obey	gehörchen	obedecer	obbedire
<b>l'océan</b>	ocean	der Ozean	el océano	l'oceano
<b>l'or</b>	gold	das Gold	el oro	l'oro
<b>un orage</b>	a storm	ein Gewitter	una tormenta	un temporale
<b>original</b>	original	eigentuemlich	original	originale
<b>la paix</b>	peace	der Friede	la paz	la pace
<b>le pardon</b>	forgiveness	die Verzeihung	el perdón	il perdono
<b>un parfum</b>	perfume	ein Parfuem	un perfume	un profumo
<b>la patience</b>	patience	die Geduld	la paciencia	la pazienza
<b>la patrie</b>	homeland	das Vaterland	la patria	la patria
<b>la peau</b>	skin	die Haut	la piel	la pelle
<b>la perfection</b>	perfection	die Vollkommenheit	la perfección	la perfezione

FRANÇAIS	ANGLAIS	ALLEMAND	ESPAGNOL	ITALIEN
la poésie	poetry	die Poesie	la poesía	la poesia
la politesse	politeness	die Hoeflichkeit	la cortesía	la cortesia
précieux	precious	wertvoll	valioso	prezioso
la précision	accuracy	die Praezision	la precisión	la precisione
un prêtre	priest	ein Priester	un sacerdote	un prete
produire	to produce	produzieren	producir	produrre
la propriété	property	der Besitz	la propiedad	la proprieta'
protéger	to protect	beschuetzen	proteger	proteggere
la prudence	prudence	die Vorsicht	la prudencia	la prudenza
la puissance	the power	die Macht	la potencia	la potenza
punir	to punish	strafen	castigar	punire
la pureté	purity	die Reinheit	la pureza	la purezza
une question	a question	eine Frage	una pregunta	una domanda
raffiné	refined	verfeinert	refinado	raffinato
la raison	the reason	die Vernunft	la razón	la ragione
une récompense	a reward	eine Belohnung	una recompensa	una ricompensa
le recueillement	meditation	die Besinnung	el recogimiento	il raccoglimento
réfléchir	to think	nachdenken	reflexionar	riflettere
une règle	a rule	eine Regel	una regla	una regola
le respect	respect	der Respekt	el respeto	il rispetto
rêver	to dream	trauemen	soñar	sognare
la révolte	rebellion	der Aufstand	la rebelión	la rivolta
la richesse	wealth	der Reichtum	la riqueza	la ricchezza
rigide	rigid	starr	rígido	rigido
rire	to laugh	lachen	reír	ridere
robuste	robust	robust	robusto	robusto
rompre	to break	abbrechen	romper	rompere
rouge	red	rot	rojo	rosso
la ruse	craftiness	die List	la astucia	la furbizia
sacré	sacred	heilig	sagrado	sacro
un sacrifice	sacrifice	ein Opfer	un sacrificio	un sacrificio
sauvage	wild	wild	salvaje	selvaggio
la science	science	die Wissenschaft	la ciencia	la scienza
un secret	a secret	ein Geheimnis	un secreto	un segreto
séduire	to seduce	verfuhren	seducir	sedurre
sensuel	sensual	sinnlich	sensual	sensuale
soigner	to care	pflagen	cuidar	curare
un soldat	soldier	ein Soldat	un soldado	un soldato
un sommet	peak	ein Gipfel	una cumbre	una vettat
la souplesse	flexibility	die Geschmeidigkeit	la flexibilidad	la flessibilita
souverain	monarch	souveraen	soberano	sovrano
sublime	sublime	herrlich	sublime	sublime
la tendresse	tenderness	die Zaertlichkeit	la temura	la tenerezza
le théâtre	theatre	das Theater	el teatro	il teatro
la tradition	tradition	die Tradition	la tradición	la tradizione

FRANÇAIS	ANGLAIS	ALLEMAND	ESPAGNOL	ITALIEN
trahir	to betray	verraten	traicionar	tradire
le travail	work	die Arbeit	el trabajo	il lavoro
utilitaire	practical	praktisch	utilitario	utilitario
vert	green	gruen	verde	verde
la victoire	victory	der Sieg	la victoria	la vittoria
le vide	space	die Leere	el vacío	il vuoto
vieillir	to age	alt werden	envejecer	invecchiare
viril	virile	maennlich	viril	virile
la vitesse	speed	die Geschwindigkeit	la velocidad	la velocita'
volontaire	a volunteer	eigenwillig	voluntario	volontario
Voluptueux	voluptuous	lustvoll	voluptuoso	voluttuoso

## A2.2 Moyenne des notes

<u>Mots</u>	<u>moyenne</u>	<u>Mots</u>	<u>moyenne</u>	<u>Mots</u>	<u>moyenne</u>
GUERRE	1.26	IMMOBILE	3.65	AVENTURIER	4.59
TRAHIR	1.42	MASQUE	3.65	DEFI	4.60
MORT	1.77	SACRIFICE	3.66	NUDITE	4.64
ANGOISSE	1.85	NOEUD	3.71	COMMERCE	4.67
FUSIL	2.26	RUSE	3.74	INFINI	4.71
DESORDRE	2.30	ACHARNEMENT	3.75	NOBLE	4.74
ROMPRE	2.44	FEU	3.78	LOI	4.75
CHASSE	2.49	DESERT	3.88	FERMETE	4.79
DANGER	2.58	INCONNU	3.92	FOI	4.80
VIDE	2.58	SOUVERAIN	4.02	RECUEILLEM.	4.80
ATTAQUER	2.61	SAUVAGE	4.08	PRUDENCE	4.86
VIEILLIR	2.63	OBEIR	4.15	MODERATION	4.88
FAUTE	2.87	ETRANGER	4.21	UTILITAIRE	4.89
PUNIR	2.90	ELITE	4.28	DIEU	4.90
CRITIQUER	2.97	INTERROGER	4.30	SACRE	4.90
ARMURE	3.02	VITESSE	4.30	IMMENSE	4.93
INTERDIRE	3.04	QUESTION	4.31	<u>Moyenne</u>	<u>4.95</u>
CRI	3.09	MYSTERE	4.32	DISCIPLINE	4.98
DOUTE	3.09	MATERIEL	4.39	SECRET	4.98
LABYRINTHE	3.13	COMMANDER	4.41	EMOTION	4.99
MEFIANCE	3.14	DIFFERENT	4.41	MODE	4.99
MURAILLE	3.19	REGLE	4.45	CEREMONIE	5.01
REVOLTE	3.27	ESCALADER	4.47	EFFORT	5.01
IRONIE	3.29	ABSOLU	4.50	GLOIRE	5.01
NOIR	3.40	BOHEME	4.50	ORIGINAL	5.05
ORAGE	3.46	INDUSTRIE	4.51	VIRIL	5.05
RIGIDE	3.49	ROUGE	4.51	AME	5.06
SOLDAT	3.59	LEGERETE	4.55	CONQUERIR	5.06
DETACHEM.	3.61	MAGIE	4.55	EVASION	5.06
FRONTIERE	3.62	CHANGEMENT	4.56	CONCRET	5.07
METALLIQUE	3.63	PRETRE	4.56		

<u>Mots</u>	<u>moyenne</u>	<u>Mots</u>	<u>moyenne</u>	<u>Mots</u>	<u>moyenne</u>
PRODUIRE	5.07	REFLECHIR	5.48	SEDUIRE	5.86
ECONOMISER	5.08	PUISSANCE	5.53	FEMININ	5.89
LUNE	5.10	VOLONTAIRE	5.53	INTIME	5.89
HEROS	5.11	FECONDER	5.54	ARBRE	5.92
CHARNEL	5.13	BLEU	5.55	EAU	5.92
CERTITUDE	5.14	ART	5.56	VICTOIRE	5.93
HUMBLE	5.14	BIJOU	5.56	PROTEGER	5.95
SOMMET	5.14	GRATUIT	5.56	RECOMPENSE	5.96
MORALE	5.15	OR	5.56	PURETE	5.97
AUDACE	5.17	PERFECTION	5.56	MARIAGE	5.98
FLEUVE	5.17	SUBLIME	5.56	ATTACHEMENT	6.00
AMBITION	5.18	RICHESSSE	5.57	DESIR	6.00
CREATEUR	5.20	CHARITABLE	5.58	DYNAMIQUE	6.06
PATRIE	5.20	POESIE	5.61	RESPECT	6.11
MINCEUR	5.21	SOUPLESE	5.61	CONFORT	6.12
BATISSEUR	5.23	ANIMAL	5.62	ENFANCE	6.13
ECOLE	5.24	CONSOLER	5.63	REVER	6.13
VERT	5.25	MOELLEUX	5.64	MATERNEL	6.15
HERITER	5.26	PARDON	5.64	MUSIQUE	6.17
VOLUPTUEUX	5.26	ARGENT	5.65	POLITESSE	6.24
MAITRISER	5.27	LOGIQUE	5.67	CADEAU	6.25
PEAU	5.27	RAFFINE	5.67	HUMOUR	6.27
THEATRE	5.27	MONTAGNE	5.68	COURAGE	6.28
MODESTIE	5.28	SCIENCE	5.68	CAMPAGNE	6.30
ECRIRE	5.30	EFFICACE	5.69	MAISON	6.32
TRADITION	5.30	OCEAN	5.69	FLEUR	6.36
ETERNEL	5.31	PATIENCE	5.69	DOUCEUR	6.37
JEU	5.33	PRECISION	5.69	HONNETE	6.38
ROBUSTE	5.33	NID	5.70	CONFIANCE	6.42
JUSTICE	5.37	ADMIRER	5.71	GAIETE	6.46
RAISON	5.38	ENSEIGNER	5.72	NAISSANCE	6.46
NAGER	5.39	CONSTRUIRE	5.73	FIDELITE	6.48
CHERCHEUR	5.41	SENSUEL	5.74	GUERIR	6.48
PRECIEUX	5.42	ADORER	5.75	FAMILLE	6.49
ACHETER	5.43	ENSEMBLE	5.76	CARESS	6.51
ILE	5.44	HONNEUR	5.76	RIRE	6.58
INVENTEUR	5.46	ELEGANCE	5.81	AMITIE	6.59
TRAVAIL	5.46	SOIGNER	5.82	TENDRESSE	6.67
ASTUCIEUX	5.48	LIVRE	5.83	PAIX	6.72
PROPRIETE	5.48	PARFUM	5.85		

## A2.3 Corrélations entre mots et axes

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
absolu	0.27	0.08	-0.26	0.01	0.21	-0.17
acharnement	0.17	-0.04	-0.26	-0.14	-0.02	-0.07
acheter	0.42	0.06	0.14	-0.23	-0.07	0.02
admirer	0.51	0.05	0.08	-0.13	0.14	0.03
adorer	0.4	0.13	0.14	-0.23	0.19	0.17
ambition	0.41	-0.06	-0.03	-0.37	-0.07	-0.01
âme	0.36	-0.05	-0.11	0.22	0.45	-0.19
antité	0.45	0.12	0.12	0.16	-0.07	0.11
angoisse	-0.18	-0.12	-0.44	-0.1	0.06	0.21
animal	0.30	0.17	0.03	0.05	-0.03	0.13
arbre	0.41	0.21	-0.05	0.29	-0.1	-0.09
argent	0.34	-0.14	0.21	-0.48	-0.08	0.05
armure	0.02	-0.17	-0.32	-0.25	0.07	-0.11
art	0.36	0.23	-0.06	0.32	0.09	-0.26
astucieux	0.45	0.16	-0.09	-0.05	-0.23	-0.17
attachement	0.44	-0.06	0.07	0.12	-0.01	0.12
attaquer	-0.07	-0.05	-0.41	-0.31	-0.06	-0.11
audace	0.36	0.15	-0.25	-0.06	-0.17	-0.25
aventurier	0.13	0.44	-0.23	-0.18	-0.02	0.04
bâtisseur	0.48	-0.07	-0.13	0.17	-0.16	-0.22
bijou	0.41	0.06	0.29	-0.29	0.22	0.10
bien	0.33	0.26	0.04	0.04	0.12	0.02
bohème	0.07	0.40	-0.19	0.11	0.14	0.05
cadeau	0.46	0.19	0.28	-0.17	0.08	0.1
campagne	0.36	0.08	0.08	0.14	-0.05	0.17
caresse	0.33	0.27	0.16	-0.06	0.03	0.24
cérémonie	0.42	-0.17	0.08	-0.07	0.30	0.19
certitude	0.49	-0.10	-0.02	-0.09	-0.05	-0.14
changement	0.17	0.23	-0.26	-0.01	-0.04	0.01
charitable	0.51	-0.13	0.01	0.25	0.06	-0.02
charnel	0.26	0.39	-0.06	-0.21	-0.08	-0.06
chasse	0.02	-0.26	-0.19	-0.24	0.03	-0.06
chercheur	0.45	-0.03	-0.16	0.21	-0.13	-0.14
commander	0.31	-0.24	-0.17	-0.35	-0.09	-0.18

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
commerce	0.4	-0.17	-0.04	-0.23	-0.07	0.04
concret	0.43	-0.01	-0.13	0.07	-0.26	-0.2
confiance	0.53	0.04	0.09	0.19	-0.15	0.05
confort	0.5	-0.02	0.26	-0.21	-0.1	0.04
conquérir	0.41	-0.06	-0.02	-0.4	-0.01	-0.12
consoler	0.42	-0.01	-0.02	0.2	0.04	0.16
construire	0.51	0.01	-0.04	0.08	-0.21	-0.05
courage	0.58	-0.06	0.03	0.04	-0.11	0.01
créateur	0.46	-0.01	-0.07	0.15	0.15	-0.31
cri	-0.04	-0.1	-0.43	-0.05	0.04	0.21
critiquer	-0.07	0.07	-0.43	-0.02	-0.09	-0.03
danger	-0.16	0.13	-0.54	-0.1	-0.01	0.10
défi	0.24	0.18	-0.27	-0.21	-0.08	-0.06
désert	-0.01	0.24	-0.40	0.03	0.13	0.09
désir	0.31	0.39	0.03	-0.27	-0.06	0.13
désordre	-0.21	0.24	-0.32	-0.05	0.06	0.04
détachement	-0.02	-0.03	-0.39	0.07	0.11	-0.07
dieu	0.33	-0.3	0.01	0.19	0.55	-0.25
différent	0.11	0.26	-0.31	0.11	-0.03	-0.02
discipline	0.38	-0.49	-0.16	0.01	-0.07	0.17
douceur	0.52	0.24	0.17	0.06	-0.02	0.13
doute	-0.08	-0.05	-0.41	-0.01	0.09	0.3
dynamique	0.55	0.11	0.02	0.01	-0.19	0.01
eau	0.3	0.22	-0.07	0.29	0.01	0.07
école	0.36	-0.24	-0.14	0.29	-0.1	0.08
économiser	0.37	-0.37	-0.07	-0.1	-0.08	0.15
écrire	0.35	0.12	-0.14	0.26	0.06	-0.03
efficacité	0.55	-0.06	-0.06	-0.01	-0.3	-0.2
effort	0.35	-0.14	-0.35	0.16	-0.17	0.06
élégance	0.53	0.08	0.16	-0.21	0.06	-0.07
élite	0.37	-0.3	-0.1	-0.25	0.06	-0.24
émotion	0.16	0.38	-0.18	0.04	0.07	0.09
enfance	0.41	0.07	0.08	0.15	0.05	0.23
enseigner	0.4	-0.05	-0.12	0.27	-0.11	0.07
ensemble	0.39	0.25	0.01	0.09	0.02	0.04
escalader	0.21	0.2	-0.25	0.01	0.03	-0.07

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
éternel	0.36	0.03	-0.02	-0.09	0.34	0.01
étranger	0.15	0.22	-0.27	0.25	0.06	-0.01
évasion	0.17	0.35	-0.13	-0.01	0.05	-0.1
famille	0.42	-0.18	0.2	0.12	0.04	0.26
faute	-0.16	-0.01	-0.38	-0.03	0.08	0.3
féconder	0.44	0.12	0.07	-0.01	-0.02	0.16
fémmin	0.38	0.23	0.01	-0.12	-0.08	0.03
fermé	0.37	-0.32	-0.28	0.04	-0.15	-0.09
feu	-0.01	0.27	-0.38	0.01	0.03	-0.13
fidélité	0.44	-0.15	0.19	0.11	0.04	0.2
fleur	0.48	0.12	0.18	0.23	0.06	0.12
fleuve	0.34	0.25	-0.16	0.19	0.02	-0.06
foi	0.38	-0.31	0.01	0.19	0.52	-0.23
frontière	0.15	-0.26	-0.24	-0.17	0.1	0.12
fusil	-0.05	-0.22	-0.32	-0.34	-0.02	-0.11
gaieté	0.5	0.2	0.2	0.09	-0.1	0.14
gloire	0.43	-0.16	0.05	-0.44	0.16	-0.06
gratuit	0.34	0.14	0.17	-0.24	-0.01	0.06
guérir	0.49	-0.01	0.16	0.12	-0.13	0.07
guerre	-0.19	-0.04	-0.3	-0.22	0.04	-0.1
hériter	0.37	-0.17	0.18	-0.34	0.02	-0.09
héros	0.48	-0.21	0.01	-0.18	0.19	-0.12
honnête	0.53	-0.13	0.11	0.15	-0.16	0.09
honnêur	0.5	-0.3	0.02	-0.1	0.06	-0.09
humble	0.38	-0.01	-0.09	0.15	-0.01	0.1
humour	0.29	0.34	-0.03	0.02	-0.11	0.06
île	0.2	0.43	-0.07	-0.08	0.13	0.09
immense	0.32	0.21	-0.23	-0.08	0.14	-0.05
immoblie	0.02	-0.11	-0.22	-0.07	0.16	0.21
inconnu	0.09	0.19	-0.34	0.08	0.08	0.01
industrie	0.37	-0.36	-0.16	-0.03	-0.14	-0.01
infini	0.26	0.18	-0.28	0.03	0.24	-0.05
interdire	0.05	-0.28	-0.34	0.04	0.05	0.24
interroger	0.28	-0.08	-0.34	0.06	-0.06	0.07
inlime	0.42	0.24	0.04	-0.05	-0.03	0.08
inventeur	0.47	0.05	-0.18	0.12	-0.16	-0.19

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
ironie	-0.07	0.19	-0.38	-0.09	-0.04	-0.06
jeu	0.31	0.22	-0.01	-0.1	-0.01	0.02
justice	0.3	-0.12	-0.12	0.18	0.01	-0.05
labyrinthe	-0.07	0.04	-0.43	-0.11	0.11	0.17
légèreté	0.1	0.42	-0.11	-0.15	0.18	0.11
livre	0.3	0.16	-0.04	0.36	0.05	-0.1
logique	0.43	-0.13	-0.13	0.05	-0.27	-0.05
loi	0.35	-0.33	-0.21	0.07	-0.02	0.08
lune	0.28	0.35	-0.16	0.1	0.19	0.03
magie	0.18	0.32	-0.08	-0.21	0.19	0.01
maison	0.47	-0.05	0.22	-0.04	-0.06	0.15
maitriser	0.44	0.03	-0.16	-0.05	-0.24	-0.22
mariage	0.38	-0.18	0.21	-0.05	0.2	0.25
masque	0.04	0.09	-0.25	-0.19	0.16	0.23
matériel	0.34	-0.2	-0.14	-0.18	-0.19	0.06
maternel	0.49	0.02	0.16	0.19	0.08	0.22
méfiance	0.01	-0.1	-0.35	-0.2	0.00	0.31
métallique	0.11	-0.25	-0.27	-0.17	-0.06	0.05
mineur	0.36	0.11	0.13	-0.18	0.06	0.04
mode	0.38	0.05	0.17	-0.28	0.17	0.06
modération	0.4	-0.17	-0.11	0.22	-0.13	-0.05
modestie	0.41	-0.02	-0.07	0.2	-0.09	0.02
moelleux	0.38	0.34	0.13	-0.07	0.07	-0.04
montagne	0.31	0.25	-0.09	0.1	0.01	0.06
morale	0.4	-0.45	-0.11	0.15	-0.01	0.04
mort	-0.12	-0.06	-0.44	0.1	0.02	0.01
muraille	-0.02	-0.22	-0.36	-0.18	0.12	0.12
musique	0.35	0.3	0.03	0.16	0.09	-0.08
mystère	0.11	0.31	-0.32	-0.05	0.16	-0.06
nager	0.26	0.25	-0.03	-0.03	-0.01	0.04
naissance	0.33	0.01	0.15	0.09	0.1	0.35
nid	0.47	0.14	0.08	0.23	0.01	-0.02
noble	0.38	-0.11	-0.09	0.01	0.22	-0.29
noûd	0.06	-0.07	-0.33	-0.13	0.09	0.2
noir	-0.09	0.19	-0.32	-0.09	0.01	0.08
nudité	0.07	0.42	-0.15	-0.18	-0.14	-0.05

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
obéir	0.31	-0.47	-0.2	0.02	0.04	0.23
océan	0.26	0.41	-0.13	0.01	0.14	0.06
or	0.37	0.07	0.23	-0.5	0.14	0.03
orage	-0.1	0.34	-0.44	0.02	-0.01	-0.04
original	0.26	0.43	-0.17	-0.01	-0.04	-0.14
paix	0.39	0.01	0.16	0.23	-0.04	0.11
pardon	0.41	-0.05	0.02	0.27	0.13	0.05
parfum	0.37	0.27	0.19	-0.11	0.13	0.06
patience	0.41	-0.01	-0.08	0.21	-0.04	0.15
patrie	0.44	-0.45	-0.01	0.01	0.1	-0.08
peau	0.39	0.29	-0.03	-0.03	-0.01	0.1
perfection	0.38	-0.22	-0.04	-0.17	-0.01	-0.1
poésie	0.38	0.2	-0.04	0.31	0.21	-0.06
politesse	0.51	-0.24	0.12	0.08	-0.1	0.19
précieux	0.49	0.02	0.1	-0.26	0.12	-0.04
précision	0.52	-0.12	-0.1	-0.01	-0.3	-0.11
prêtre	0.37	-0.35	-0.02	0.18	0.44	-0.17
produire	0.5	-0.21	-0.1	-0.13	-0.21	-0.12
propriété	0.45	-0.23	0.11	-0.31	-0.1	-0.02
protéger	0.54	0.02	0.03	0.11	-0.06	0.15
prudence	0.46	-0.21	-0.02	0.04	-0.1	0.23
puissance	0.39	-0.11	-0.06	-0.53	0.03	-0.11
punir	0.01	-0.28	-0.43	-0.02	-0.04	0.2
pureté	0.52	-0.01	0.09	0.09	0.16	0.07
question	0.24	0.04	-0.34	0.12	-0.06	0.06
raffine	0.46	-0.03	0.04	-0.06	0.04	-0.11
raison	0.49	-0.16	-0.14	0.09	-0.15	-0.1
récompense	0.54	-0.04	0.24	-0.23	-0.01	0.02
recueillement	0.39	-0.25	-0.09	0.26	0.4	-0.18
réfléchir	0.42	0.01	-0.29	0.23	-0.21	-0.06
régler	0.37	-0.34	-0.23	0.08	-0.07	0.09
respect	0.52	-0.13	0.06	0.14	-0.1	0.12
rêver	0.3	0.47	0.07	-0.04	0.15	0.11
révolte	-0.11	0.32	-0.41	0.05	-0.13	-0.03
richesse	0.37	-0.05	0.21	-0.55	0.03	-0.07
rigide	0.15	-0.28	-0.25	-0.25	-0.01	0.26

	Axe 1	Axe 2	Axe 3	Axe 4	Axe 5	Axe 6
rire	0.42	0.28	0.15	0.04	-0.1	0.15
robuste	0.53	-0.1	-0.04	-0.02	-0.22	-0.22
rompre	-0.15	-0.19	-0.44	-0.01	0.01	0.07
rouge	0.17	0.09	-0.14	-0.01	0.03	-0.02
ruse	0.01	0.19	-0.2	-0.35	-0.11	0.14
sacre	0.45	-0.23	-0.02	0.08	0.42	-0.2
sacrifice	0.19	-0.31	-0.3	0.06	0.07	0.17
sauvage	-0.01	0.42	-0.33	-0.06	0.03	0.02
science	0.42	-0.06	-0.15	0.18	-0.14	-0.12
secret	0.32	0.09	-0.18	-0.1	0.15	0.14
séduire	0.37	0.35	0.11	-0.33	-0.03	0.1
sensuel	0.28	0.51	0.01	-0.25	-0.11	0.02
soigner	0.42	0.01	-0.09	0.18	-0.06	0.17
soldat	0.21	-0.42	-0.17	-0.15	0.09	-0.05
sommet	0.33	0.25	-0.22	0.08	0.02	-0.13
souper	0.47	0.22	0.03	0.01	-0.07	-0.06
souverain	0.32	-0.23	-0.13	-0.21	0.24	-0.16
sublime	0.42	0.27	0.03	-0.13	0.15	-0.15
tendresse	0.45	0.14	0.17	0.13	0.02	0.26
théâtre	0.32	0.14	-0.05	0.3	0.15	-0.14
tradition	0.42	-0.26	-0.05	0.04	0.11	0.06
trahir	-0.28	-0.04	-0.25	-0.2	0.12	0.06
travail	0.37	-0.31	-0.15	0.16	-0.12	0.08
utilitaire	0.46	-0.21	-0.04	-0.01	-0.22	-0.05
vert	0.29	0.25	-0.03	0.1	0.02	-0.02
victoire	0.44	-0.14	0.1	-0.31	0.04	-0.04
vide	-0.15	-0.02	-0.43	-0.13	0.05	0.2
vieillard	-0.04	-0.08	-0.36	0.19	0.1	0.12
viril	0.4	-0.06	0.04	-0.23	-0.06	-0.1
vitesse	0.08	0.2	-0.19	-0.38	-0.04	-0.1
volontaire	0.45	0.06	-0.15	0.01	-0.21	0.02
voluptueux	0.28	0.42	-0.01	-0.16	-0.04	-0.2

## Glossaire

**Algorithme** – ensemble des règles opératoires propres à un calcul.

**Analyse factorielle** – famille de méthodes statistiques d'analyse multidimensionnelle, s'appliquant à des tableaux de nombres, qui visent à calculer un nombre réduit de "facteurs" résumant approximativement l'ensemble des informations contenues dans le tableau de départ (annexe A1.2).

**Analyse en facteurs communs et spécifiques** – cette analyse correspond au modèle classique d'analyse factorielle. Elle décrit un ensemble de variables par une combinaison linéaire de facteurs communs sous-jacents et d'une variable (facteur spécifique) synthétisant la part spécifique des variables d'origine (annexe A1.6 ; chapitre 2, section 2.6).

**Analyse en composantes principales** – méthodes d'analyse factorielle\* s'appliquant aux tableaux de mesures (annexe A1.3 ; chapitre 1, section 1.3).

**Analyse des correspondances** – méthode d'analyse factorielle s'appliquant à l'étude de tableaux à double entrée composés de nombres positifs. L'analyse des correspondances est caractérisée par l'emploi d'une distance (ou métrique) particulière dite distance du chi-2 (ou  $\chi^2$ ) (annexe A1.4 ; chapitre 4, sections 4.3 et 4.4).

- Analyse logarithmique** – méthode qui consiste à transformer les données en logarithmes (après addition éventuelle d'une constante en cas de données négatives), puis, après les avoir centrées en ligne et en colonne, à les soumettre à une analyse en composantes principales\* non normée (annexe A1.5 ; chapitre 5, section 5.7).
- Bootstrap** – la technique du *bootstrap* consiste à simuler  $s$  ( $s$  généralement supérieur à 30) échantillons de même taille  $n$  que l'échantillon initial. Ils sont obtenus par tirage au hasard avec remise parmi les  $n$  individus observés au départ, ceux-ci ayant tous la même probabilité  $1/n$  d'être choisis. Cette méthode est employée pour analyser la variabilité de paramètres statistiques simples en produisant des intervalles de confiance\* de ces paramètres (annexe A1.9.5 ; chapitre 2, sections 2.3 et 2.4 ; chapitre 4, section 4.4).
- Bootstrap partiel** – variante du *bootstrap*\* dans le cas des analyses factorielles qui consiste à projeter les colonnes des tableaux répliqués (mots) comme des éléments supplémentaires\* sur les axes de l'analyse de référence, c'est-à-dire l'analyse de l'échantillon initial non perturbé (annexe A1.9.5 ; chapitre 2, section 2.3).
- Bootstrap total** – variante du *bootstrap* dans le cas des analyses factorielles qui consiste à refaire des analyses en composantes principales\* complètes sur chaque échantillon répliqué (annexe A1.9.5 ; chapitre 2, section 2.3).
- Bootstrap sur variables** – le *bootstrap* est ici réalisé non pas sur les individus, mais sur les variables, ce qui permet d'éprouver les structures observables au niveau des individus (annexe A1.9.5 ; chapitre 2, section 2.4).
- Carte auto-organisée de Kohonen** – méthode de classification\* qui consiste à représenter dans un espace à deux (parfois trois) dimensions un grand nombre de données en respectant la notion de voisinage de l'espace des éléments à classer (annexe A1.8 ; chapitre 3, section 3.1 ; chapitre 4 section 4.2).
- Classification** – technique statistique permettant de regrouper en classes homogènes des individus ou observations entre lesquels a été définie une distance.
- Classification hiérarchique** – technique particulière de classification\* produisant, par agglomération progressive, des classes ayant la

propriété d'être, pour deux quelconques d'entre-elles, soit disjointes, soit incluses l'une dans l'autre (annexe A1.7 ; chapitre 3, section 3.1).

**Coefficient de corrélation** – indice exprimant dans quelle mesure deux variables numériques varient de façon concomitante. Cet indice varie de  $-1$  à  $+1$ . Il est positif lorsque les valeurs élevées (resp. faibles) d'une variable tendent à être associées aux valeurs élevées (resp. faibles) de l'autre variable. Il est négatif lorsque les valeurs élevées d'une variable tendent à être associées aux valeurs faibles de l'autre variable (annexe A1.3 et *passim*).

**Comparaison multiple** – un problème de *comparaison multiple* se pose lorsqu'on réitère un test statistique conçu, dans son principe, pour n'être réalisé qu'une seule fois (annexe A1.9.2).

**Corpus** – ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique.

**Dendrogramme** – représentation graphique d'un arbre de classification hiérarchique\*, mettant en évidence l'inclusion progressive des classes (chapitre 3, section 3.1.1).

**Distance du chi-2** – distance entre profils\* de fréquences utilisée en analyse des correspondances\* et dans certains algorithmes\* de classification\* (annexe A1.4).

**Éléments actifs** – ensemble des éléments servant de base au calcul des axes factoriels, des valeurs propres\* relatives à ces axes et des coordonnées factorielles.

**Éléments supplémentaires (ou illustratifs)** – ensemble des éléments ne participant pas aux calculs des axes factoriels, pour lesquels on calcule *a posteriori* des coordonnées factorielles (annexe A1.9.3).

**Facteur** – variables artificielles construites par les techniques d'analyse factorielle permettant de résumer (de décrire brièvement) les éléments actifs\* initiaux (variables actives et individus actifs).

**Facteur *taille*** – Un tel facteur\* apparaît lorsque toutes les variables sont corrélées positivement entre elles. Cette caractéristique apparaît le plus souvent sur le premier axe de l'analyse en composantes principales\*, que l'on appelle alors "facteur *taille*" (chapitre 5).

**Intervalle de confiance** – Il permet d'évaluer la précision d'un estimateur et s'interprète comme une marge d'erreur liée au phénomène de la fluctuation d'échantillonnage.

- Pourcentages d'inertie (ou de variance)** – quantités proportionnelles aux valeurs propres\*, dont la somme est égale à 100. Notées  $\tau_\alpha$ . Dans le cas d'une analyse en composantes principales, les pourcentages d'inertie ne peuvent être systématiquement interprétés en termes de « pourcentages d'information ». Ils peuvent parfois être très faibles et cependant très significatifs statistiquement (cas de données fortement bruitées).
- Profil** – (d'une ligne ou d'une colonne d'un tableau de contingence\*) vecteur constitué par le rapport des effectifs composant une ligne (resp. colonne) à la somme des mêmes effectifs.
- Question fermée** – question dont les seules réponses possibles sont proposées explicitement à la personne interrogée.
- Question ouverte** – question posée sans grille de réponse préétablie, dont la réponse peut être numérique (ex: *Quel est votre âge ?*), ou textuelle (exemple, après certaines questions fermées : *Pourquoi ?*) (Chapitre 4).
- Tableau de contingence** – synonyme de tableau de fréquences ou de tableau croisé : tableau dont les lignes et les colonnes représentent respectivement les modalités de deux questions (ou deux variables nominales), et dont le terme général représente le nombre d'individus correspondant à chaque couple de modalités (annexe A1.4).
- Valeurs propres** – quantités permettant, lors d'une analyse factorielle\*, de juger de l'importance des facteurs\* successifs de la décomposition factorielle. La valeur propre notée  $\lambda_\alpha$  mesure la dispersion (variance) des éléments sur l'axe  $\alpha$  (annexe A1.3).
- Valeur-test** – quantité permettant d'apprécier la signification de la position d'un élément supplémentaire\* (ou illustratif) sur une axe factoriel. Brièvement, si une valeur-test dépasse 2 en valeur absolue, la position de l'élément correspondant a peu de chance d'être due au hasard (annexe A1.9.1, et *passim*) ; mais attention aux comparaisons multiples\* !

## Bibliographie

- Agoramétrie, *Rapports de l'association Agoramétrie*, 20 rue Rosenwald, 75015, Paris
- Anatrella T. (1995), *Non à la société dépressive*, Champs, Flammarion, Paris.
- Anderson T. W., Rubin H. (1956), Statistical inference in factor analysis, *Proc. of the 3rd Berkeley Symp. on Math. Statist.*, **5**, p 111-150.
- Anderson T. W. (1963), Asymptotic theory for principal component analysis, *Ann. Math. Statist.*, **34**, p 22-148.
- Baayen R. H. (2001), *Word Frequency Distributions*, Kluwer Academic Publishers, Dordrecht.
- Barthes R. (1957), *Mythologies*, Editions du Seuil, Paris.
- Benzécri J.-P. (1982), *Histoire et préhistoire de l'analyse des données*, Dunod, Paris.
- Benzécri J.-P. (1989), Essai d'analyse des notes attribuées par un ensemble de sujets aux mots d'une liste, *Cahiers de l'Analyse des Données*, Vol XIV, 1, p 73-98.
- Benzécri J.-P. (1973), *L'Analyse des Données, Tome 1 : La Taxinomie, Tome 2 : L'Analyse des Correspondances*, Dunod, Paris.
- Bréchon P. (2000), *Les valeurs des Français. Evolutions de 1980 à 2000*, Armand Colin, Paris.

- Chateau F., Lebart L. (1996), Assessing sample variability in the visualization techniques related to principal component analysis: bootstrap and alternative simulation methods, *in* : *COMPSTAT96*, A. Prats (ed), Physica Verlag, Heidelberg, p 205-210.
- Cottrell M., Rousset P. (1997), The Kohonen Algorithm: a powerful tool for analysing and representing multidimensional qualitative and quantitative data, *in*: *Biological and Artificial Computation : From Neuroscience to Technology*, J. Mira, R. Moreno-Diaz, J. Cabestany, (eds), Springer, p 861-871.
- Daudin J.-J., Duby C., Trécourt P. (1988), Stability of principal components studied by the bootstrap method, *Statistics*, 19, p 241-258.
- Diaconis P., Efron B. (1983), Computer intensive methods in statistics, *Scientific American*, 248, p 116-130.
- Efron B. (1979), Bootstraps methods : another look at the Jackknife, *Ann. Statist.*, 7, p 1-26.
- Efron B. (1982), The Jackknife, the Bootstrap and other resampling plans, *SIAM*, 1982, p 116-130.
- Efron B., Thisted R. (1976), Estimating the number of unseen species : how many words did Shakespeare know ?, *Biometrika*, 63, p 435-437.
- Ehrenberg A. (1998), *La fatigue d'être soi*. Odile Jacob, Paris.
- Escoufier Y. (1970), Echantillonnage dans une population de variables aléatoires réelles, *Publications de l'Institut de Statistique de l'Université de Paris*, Vol XIX, fasc. 4, p 1-47.
- Fabre J., Morlat G., Pagès J-P, Stemmelen B. (1981), Les Structures de l'Opinion Publique, *Le Progrès Technique*, n° 22-24, ANRT, Paris.
- Ferrand L., Alario F.-X. (1998), Normes d'associations verbales pour 366 noms d'objets concrets, *L'année psychologique*, 98, p 659-709.
- Freud S. (1978), *Abrégé de psychanalyse*, PUF, Paris.
- Freud S. (1985), *Trois Essais sur la théorie sexuelle*, Gallimard, Paris.
- Garnett J.-C. (1919), General ability, cleverness and purpose, *British J. of Psych.*, 9, p 345-366.
- Gifi A. (1981), *Non Linear Multivariate Analysis*, Department of Data theory, University of Leiden, (updated version: 1990, same title, J. Wiley, Chichester).
- Greenacre M. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press, London.
- Harman H.H. (1967), *Modern Factor Analysis*, Chicago University Press, Chicago.
- Hayashi C. (1956), Theory and examples of quantification, (II), *Proc. of the Institute of Statistical Mathematics*, 4, (2), p 19-30.
- Hochberg, Y. (1988), A sharper Bonferroni procedure for multiple tests of significance, *Biometrika*, 75, p 800-803

- Holmes S. (1989), Using the bootstrap and the RV coefficient in the multivariate context, in : *Data Analysis, Learning Symbolic and Numeric Knowledge*, E. Diday (ed.), Nova Science, New York, p 119-132.
- Hotelling H. (1933), Analysis of a complex of statistical variables into principal components, *J. Educ. Psy.* 24, p 417-441, p 498-520.
- Hsu, J. C. (1996), *Multiple Comparisons: Theory and Methods*, Chapman & Hall, London.
- Jacob F. (1970), *La logique du vivant*, Gallimard, Paris.
- Kazmierczak J.-B. (1985), Analyse logarithmique : deux exemples d'application, *Revue de Statist. Appl.*, 33, (1), p 13-24.
- Kleiweg P. (1996), *Een inleidende cursus met practica voor de studie Alfa-Informatica*, Master's thesis, Rijksuniversiteit Groningen, 1996.
- Kohonen T. (1989), *Self-Organization and Associative Memory*, Springer-Verlag, Berlin.
- Laplanche J., Pontalis J.-B. (1994), *Vocabulaire de la psychanalyse*, PUF, Paris.
- Lawley D. N., Maxwell A. E. (1963), *Factor Analysis as a Statistical Method*, Methuen, London.
- Lebart L. (1986), Qui pense quoi en France ? Evolution et structure des opinions en France de 1978 à 1984, *Consommation, Revue de Socio-Economie*, Dunod, p 3-22.
- Lebart L., Houzel Y. (1980), Le système d'enquête sur les conditions de vie et aspirations des Français, *Consommation, Revue de Socio-Economie*, Dunod, p 3-25.
- Lebart L., Morineau A., Warwick K. (1984), *Multivariate Descriptive Statistical Analysis*, J. Wiley, New York.
- Lebart L., Morineau M., Piron M. (2002), *Statistique exploratoire multidimensionnelle*, Dunod, Paris.
- Lebart L., Salem A. (1994), *Statistique textuelle*. Dunod, Paris.
- Lebart, L., Salem, A., Berry, L. (1998), *Exploring Textual Data*, Kluwer Academic Publishers, Dordrecht.
- Meulman J. (1982), *Homogeneity Analysis of Incomplete Data*, DSWO Press, Leiden.
- Milan L., Whittaker J. (1995), Application of the parametric bootstrap to models that incorporate a singular value decomposition, *Appl. Statist.* 44, 1, p 31-49.
- Muirhead R. J. (1982), *Aspects of Multivariate Statistical Theory*, J. Wiley, New York.
- Mulaik S. A. (1972), *The Foundation of Factor Analysis*, McGraw Hill, New York.
- Muller C. (1977), *Principes et méthodes de statistique lexicale*, Hachette, Paris.
- Muller C. (1979), Peut-on estimer l'étendue d'un lexique ? In : *Langue Française et Linguistique Quantitative*, p 399-425, Slatkine, Genève.

- O'Neill, R., and G. B. Wetherill. (1971), The present state of multiple comparison methods (with discussion), *Journal of the Royal Statistical Society, Series B*, 33, p 218-250.
- Osgood C. E. (1965), Cross-Cultural Comparability in Attitude Measurement via Multilingual Semantic Differentials, in: Steiner I., Fishbein M. (Eds): *Current Studies in Social Psychology*, Holt, Rinehart and Winston, New York.
- Osgood C. E., Suci G. J., Tannenbaum P. H. (1957), *The Measurement of Meaning*, University of Illinois, Urbana.
- Perneger T.,V. (1998), What is wrong with Bonferroni adjustments, *British Medical Journal*, 136, p 1236-1238
- Riffault H. (1994), *Les valeurs des Français*, PUF, Paris.
- Saville, D. J. (1990), Multiple comparison procedures: The practical solution, *American Statistician*, 44, p 174-180.
- Sokal R. R., Sneath P. H. A. (1963), *Principles of Numerical Taxonomy*, Freeman and co., San-Francisco.
- Spearman C. (1904), General intelligence, objectively determined and measured, *Amer. Journal of Psychology*, 15, p 201-293.
- Stauffer D. F., Garton E. O., Steinhorst R. K. (1985), A comparison of principal component from real and random data, *Ecology*, 66, p 1693-1698.
- Steiner J.-F., Auliard, O. (1992), La sémiométrie: un outil de validation des réponses, In : *La Qualité de l'Information dans les Enquêtes / Quality of Information in Sample Surveys*, ASU, (L. Lebart ed.), Dunod, Paris, p 241-274.
- Stoetzel J. (1983), *Les valeurs du temps présent : une enquête européenne*, PUF, Paris.
- Thiria S., Lechevallier Y., Gascuel O., Canu S. (1997), *Statistique et méthodes neuronales*, Dunod, Paris.
- Thurstone L. L. (1947), *Multiple Factor Analysis*, The University of Chicago Press, Chicago.
- Valette-Florence P. (1994), *Les styles de vie. Bilan critique et perspectives*, Nathan, Paris.
- Westfall, P. H., Young S. S. (1993), *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, J. Wiley, New York.
- Westfall P. H., Tobias R. D., Rom D., Wolfinger R. D., Hochberg Y. (1999), Multiple Comparisons and Multiple Tests Using the SAS System, *SAS Institute*.
- Young G. A. (1994), Bootstrap: more than a stab in the dark, *Statistical Science*, 9, p 382-418.

Ludovic Lebart  
Marie Piron  
Jean-Francois Steiner

# LA SÉMIOMÉTRIE

## Essai de statistique structurale

Au-delà de leur signification, les mots, par les souvenirs qu'ils mobilisent, ont le pouvoir de provoquer en nous des sensations agréables ou désagréables.

De cette observation simple est née une méthode, la Sémiométrie, largement utilisée en marketing et dans les études psychosociologiques. Mais, grâce à la puissance des outils statistiques actuels, son pouvoir d'investigation va, bien au-delà de ces applications pratiques, jusqu'aux confins de la psychanalyse et de la linguistique.

Il semblait indispensable qu'un ouvrage fasse le point sur les principes de cette méthode, les travaux réalisés et les applications potentielles. Le lecteur peut maintenant découvrir, au fil des chapitres, l'étendu du travail d'expérimentation, la sévérité des épreuves de validation, la profondeur et la finesse des résultats obtenus, enfin les promesses de cet outil transdisciplinaire.

Plusieurs niveaux de lecture sont possibles selon les connaissances mathématiques et statistiques du lecteur ; les développements plus techniques sont en effet regroupés dans une annexe.

Ce livre s'adresse aux spécialistes du marketing et de la communication, aux socio-économistes, aux statisticiens, aux psychosociologues, aux linguistes. Il intéresse un large public, allant des sociétés d'études et des instituts de sondage aux étudiants, professeurs et chercheurs des universités ou des grandes écoles.

LUDOVIC LEBART

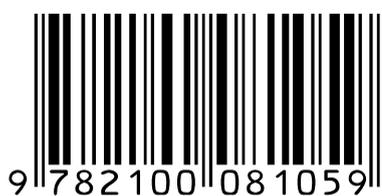
est directeur de recherches au CNRS et professeur à l'ENST (École nationale supérieure des télécommunications)

MARIE PIRON

est chargée de recherches à l'IRD (Institut de recherche pour le développement)

JEAN-FRANÇOIS STEINER

est écrivain.



ISBN 2 10 008105 5

<http://www.dunod.com>

