

## 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

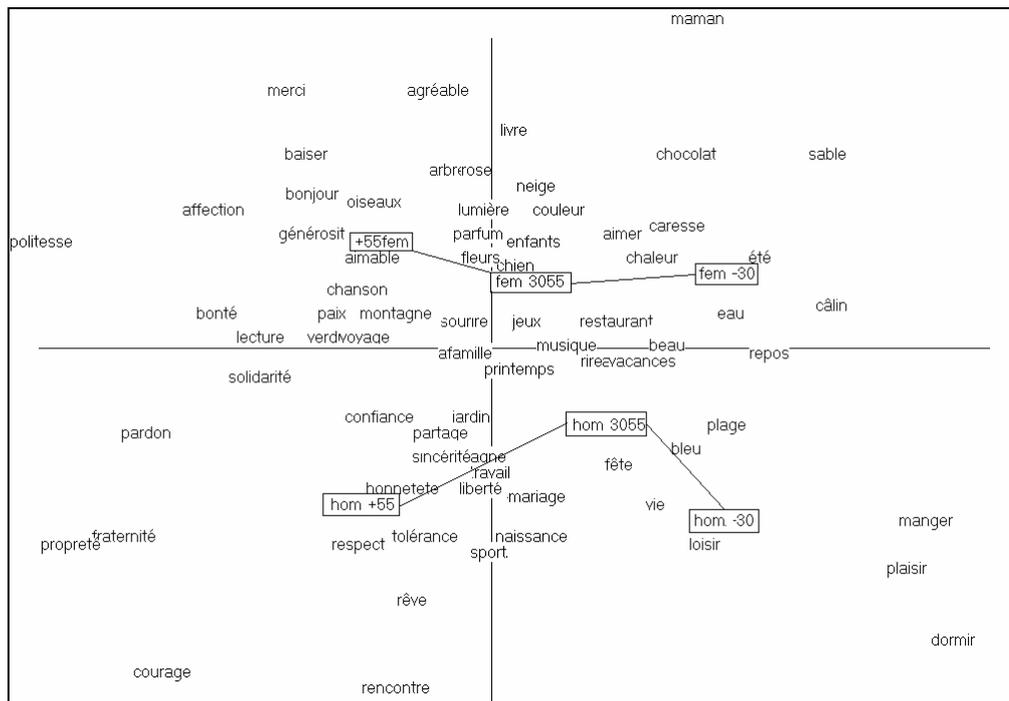
La représentation obtenue confirme la complémentarité, on pourrait presque dire l'additivité des effets de ces deux variables de base.

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).**

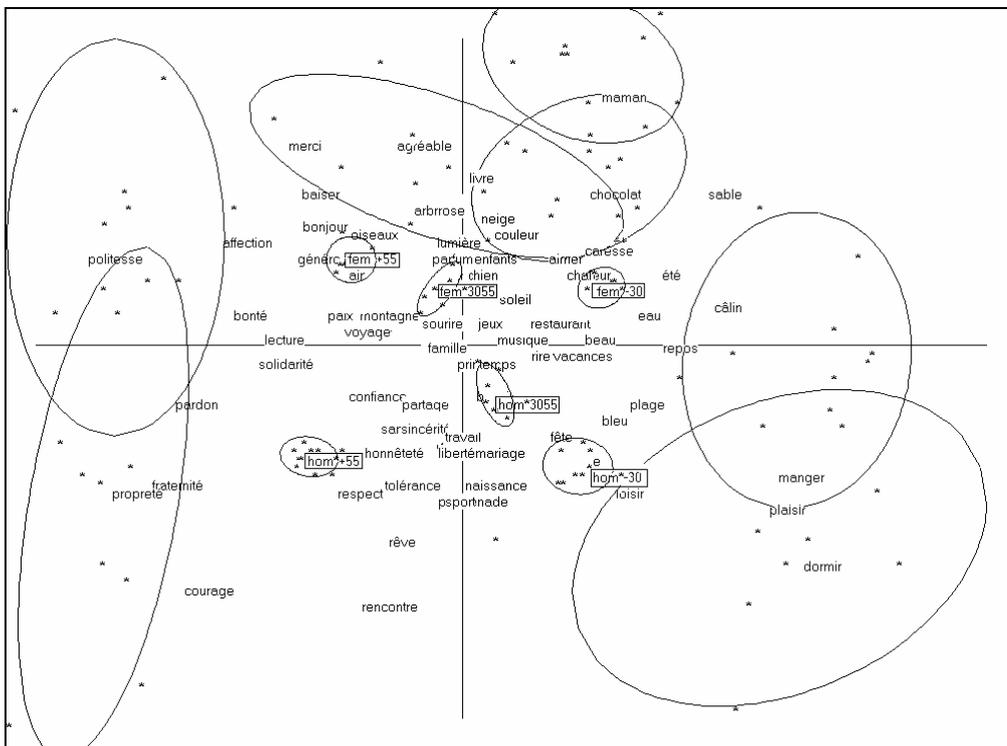
Finalement, cette figure est une synthèse des résultats lisibles sur les tableaux 4.5, 4.6 et 4.7. Elle décrit bien, dans le cadre d'un continuum nuancé, les oppositions entre sexes pour une classe d'âge donnée, et l'étendue des variations lexicales internes à chaque sexe, en fonction de l'âge<sup>1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

*bootstrap* (déjà utilisées au chapitre deux, dans un cadre différent) qui nous les fournissent<sup>1</sup>.

La figure 4.4 montre les ellipses de confiance des points-catégories (petites ellipses dans la partie centrale du graphique) : il est clair que le *pattern* observé est stable, malgré la taille modérée de l'échantillon. Comme cela est toujours le cas pour ce type de tables lexicales, les ellipses de confiance relatives aux mots sont beaucoup plus grandes (les ellipses sélectionnées à titre d'exemple concernent les mots *politesse*, *propreté*, *livre*, *maman*, *chocolat*, *manger*, *dormir*).



**Figure 4.4 : Ellipses de confiance des catégories et de mots**

Mais la grande taille de ces ellipses ne modifie pas foncièrement l'interprétation des proximités observées. En bas à droite du graphique, le mot *dormir* reste caractéristique des hommes jeunes, quelle que soit sa place dans son ellipse de confiance, de même, à gauche, les mots *propreté* et *politesse* restent caractéristiques des personnes âgées, en haut, les mots *chocolat*, *livre* et *maman* caractéristiques des femmes.

En conclusion de cette section 4 consacrée aux rapprochements entre les mots cités spontanément et certaines caractéristiques de base des

1. Sur les aspects techniques de la méthode, cf. annexe A1.4.

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.